"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcala, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"
"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcalá, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"
"This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcalá, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)"

Editors:
Prof. Nikos Mastorakis, Technical University of Sofia, BULGARIA / Hellenic Naval Academy, GREECE
Prof. Metin Demiralp, Istanbul Technical University, TURKEY
Prof. Valeri M. Mladenov, Technical University of Sofia, BULGARIA
Prof. Azami Zaharim, Universiti Kebangsaan, MALAYSIA

Editor-in-Chief:
Prof. Nikos Mastorakis, Technical University of Sofia, BULGARIA / Hellenic Naval Academy, GREECE

Editorial Board:
Prof. Siavash H. Sohrab, Northwestern University, Evanston, Illinois, USA
Prof. Mikhail Wasfy, University of Central Florida, USA
Prof. Slobodan I. Babic, Ecole Polytechnique de Montreal, CANADA
Prof. Yiming Li, National Chiao Tung University, Hsinchu 300, TAIWAN
Prof. Petr Y. Ekel, Pontif. Cathol. Univ. of Minas Gerais, 30535-610, Belo Horizonte, BRAZIL
Prof. Tuan Pham, James Cook University, Townsville, QLD 4811, AUSTRALIA
Prof. M. Isabel García-Planas, Univ. Polit. de Catalunya, C.Mineria 1, 08038, Barcelona, SPAIN
Prof. Demetrios Kazakos, Texas Southern University, 3100 Cleburne Str., Houston, Texas 77004, USA
Prof. Olga Martin, University Politehnica of Bucharest, ROMANIA
Prof. José Mª Zamanillo Sáinz de la Maza, University of Cantabria, SPAIN
Prof. Alexander Zemliak, Puebla Aut. Univ, Av. San Claudio s/n, Ciudad, Puebla, Pue, 72570, MEXICO
Prof. Jose Carlos Quadrado, Electrical Engineering Department, ISEL, Lisbon, PORTUGAL
Prof. Pablo Luis López Espí, Universidad de Alcalá, SPAIN
Preface

This book contains important selected papers from WSEAS Conferences organized in the University of Cambridge (UK), University of Harvard (USA), Massachusetts Institute of Technology -- MIT (USA), China Jiliang University (China), Beijing Jiaotong University (China), Univ. Politecnica of Bucharest (Romania), Technical University of Sofia (Bulgaria), Tianjin University of Technology and Education (China), Zhejiang Univ. of Technology (China), University of Alcalá, Madrid (Spain), University of Cantabria, Santander (Spain), Budapest Tech (Hungary), University of La Laguna (Spain), Politehnica University of Timisoara (Romania)

The contents of this Book will be sent to the WSEAS collaborating indices after the conference: www.worldses.org/indexes

The contents of this Book are also published in CD-ROM. Both will be sent to the WSEAS collaborating indices after the conference: www.worldses.org/indexes

The friendliness and openness of the WSEAS conferences, adds to their ability to grow by constantly attracting young researchers. The WSEAS Conferences attract a large number of well-established and leading researchers in various areas of Science and Engineering as you can see from http://www.wseas.org/reports. Your feedback encourages the society to go ahead as you can see in http://www.worldses.org/feedback.htm

Expanded and enhanced versions of papers published in this conference proceedings are also going to be considered for possible publication in one of the WSEAS journals that participate in the major International Scientific Indices (Elsevier, Scopus, EI, ACM, Compendex, INSPEC, CSA .... see: www.worldses.org/indexes) these papers must be of high-quality (break-through work) and a new round of a very strict review will follow. (No additional fee will be required for the publication of the extended version in a journal). WSEAS has also collaboration with several other international publishers and all these excellent papers of this volume could be further improved, could be extended and could be enhanced for possible additional evaluation in one of the editions of these international publishers.

Finally, we cordially thank all the people of WSEAS for their efforts to maintain the high scientific level of conferences, proceedings and journals.
# Table of Contents

**Preface**  
Petronela Savin  

1. **Proposal for a Cross-Linguistic and Cross-Cultural Research Project on Idioms of Food in Europe**  
Petronela Savin  

2. **Relationship in between House Price and Annual Management Fund in High-Rise Residential in Kuala Lumpur, Malaysia**  
N. Mohd-Tawil, A. Ramly, Md N. Daud, A. I. Che-Ani, A. Zaharim  

3. **Enhancing Employability Skills through Industrial Training Programme**  
Y. M. Yusoff, M. Z. Omar, A. Zaharim, A. Mohamed, N. Muhamad, R. Mustapha  

4. **Competencies Achieved through Industrial Training Programme**  

5. **Use of Rasch Analysis to Measure Students Performance in Engineering Education**  
M. Saidfudin, A. A. Azrilah, N. A. Rodzo’an, M. Z. Omar, A. Zaharim, H. Basri  

6. **Easier Learning Outcomes Analysis using Rasch Model in Engineering Education Research**  
M. Saidfudin, A. A. Azrilah, N. A. Rodzo’an, M. Z. Omar, A. Zaharim, H. Basri  

7. **Exploring User’s Perception toward Automated Checkout Trolley in Developing Countries**  
Koh Peik See, Tan Shiang-Yen, Nasuha Lee Abdullah, Rosnah Idrus  

8. **Application of WiFi-Based Indoor Positioning System in Handheld Directory System**  
Lee Pieh Wen, Chan Wee Nee, Khor Meng Chun, Tan Shiang-Yen, Rosnah Idrus  

9. **Mixed Number and Clifford Algebra**  
Md. Shah Alam, Sabar Bauk  

10. **The Effect of Environment on Mine**  
Soheila Khoshnevis Yazdi, Bahram Skakouri  

11. **Consumers’ Complaining Behavior. An Exploratory Research**  
Muntean Andreea Cipriana, Stremtan Filimon, Gavrila Paven Ionela  

12. **Ability Training through Partnered Education**  
Catherine W. Cook, Frank P. Decaro, Nicole E. Decaro  

Zamalia Mahmud, Mohd Saidfudin Masodi, Ida Rosmini Othman, Jamilah Othman
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Issues and Challenges in NSDI Implementation</td>
<td>86</td>
</tr>
<tr>
<td>Noor Habibah Arshad, Fuziah Abu Hanifah</td>
<td></td>
</tr>
<tr>
<td>GEE-Smoothing Spline Approach on Semiparametric Models for Correlated Nominal Data: Estimation</td>
<td>92</td>
</tr>
<tr>
<td>and Simulation Study</td>
<td></td>
</tr>
<tr>
<td>Noor Akma Ibrahim, Suliadi Arema</td>
<td></td>
</tr>
<tr>
<td>Portable Multi-Channel Environmental Data Measurement System for Monitoring Global Warming</td>
<td>102</td>
</tr>
<tr>
<td>M. Amir Abas, A. Khusairy Azim, M. Hilmi Fadzil, M. Dahlui, A. M. Arshad</td>
<td></td>
</tr>
<tr>
<td>Parametric Optimization in EDM of Ti-6Al-4V using Copper Tungsten Electrode and Positive Polarity</td>
<td>107</td>
</tr>
<tr>
<td>A Statistical Approach</td>
<td></td>
</tr>
<tr>
<td>M. M. Rahman, Md. Ashikur Rahman Khan, K. Kadirkama, M. A. Maleque, Rosli A. Bakar</td>
<td></td>
</tr>
<tr>
<td>A Wide Tuning Range Voltage-Controlled Oscillator with Active Inductors for Bluetooth Applications</td>
<td>114</td>
</tr>
<tr>
<td>Jenn-Tzey Yang, Shao-Kang Hsieh, Ping-Jung Tsai</td>
<td></td>
</tr>
<tr>
<td>Non-Atomic set Multifunctions</td>
<td>118</td>
</tr>
<tr>
<td>Alina Gavrilut</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Diffused set Multifunctions</td>
<td>124</td>
</tr>
<tr>
<td>Alina Gavrilut, Anca Croitoru, Nikos E. Mastorakis</td>
<td></td>
</tr>
<tr>
<td>The Rademacher Complexity Model over Acoustic Features for Improving Robustness in Hypernasal</td>
<td>130</td>
</tr>
<tr>
<td>Speech Detection</td>
<td></td>
</tr>
<tr>
<td>E. Delgado, F. A. Sepulveda, S. Rothlisberger, C. G. Castellanos</td>
<td></td>
</tr>
<tr>
<td>Intensive Quenching of Tools in Water Salt Solutions</td>
<td>136</td>
</tr>
<tr>
<td>N. I. Kobasko, A. A. Moskalenko, E. A. Mazurenko, A. M. Medvedev</td>
<td></td>
</tr>
<tr>
<td>Why Database for Cooling Capacity of Various Quenchants should be Developed?</td>
<td>142</td>
</tr>
<tr>
<td>Nikolai Kobasko</td>
<td></td>
</tr>
<tr>
<td>New Approach in Manufacturing of High Quality Superstrengthened Steels</td>
<td>148</td>
</tr>
<tr>
<td>Nikolai Kobasko</td>
<td></td>
</tr>
<tr>
<td>MFM and XPS Analysis of Low Cost Black Nickel Solar Absorber</td>
<td>154</td>
</tr>
<tr>
<td>Ricardo Luiz Perez Teixeira, Renata Antoun Simao, Bruno Coelho, Armando Coelho Oliveira</td>
<td></td>
</tr>
<tr>
<td>Constructive Principles of the Italian System used to Build the Bastioned Fortification of Oradea</td>
<td>159</td>
</tr>
<tr>
<td>Fortress</td>
<td></td>
</tr>
<tr>
<td>Valentin Capotescu</td>
<td></td>
</tr>
<tr>
<td>A Robust Control by Extended Static Output Feedback for Discrete-Time Uncertain Linear Systems</td>
<td>165</td>
</tr>
<tr>
<td>Addison Rios-Bolivar, Flor Narciso</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Optimization of Limit Characteristics in Opened Queueing Networks</td>
<td>171</td>
</tr>
<tr>
<td>Gurami Tsitsiashvili</td>
<td></td>
</tr>
<tr>
<td>Design of Control Systems based on Vector Error</td>
<td>176</td>
</tr>
<tr>
<td>Oleg N. Agamalov</td>
<td></td>
</tr>
<tr>
<td>The Possible Causes of Tax Evasion in Romania</td>
<td>185</td>
</tr>
<tr>
<td>Comaniciu Carmen</td>
<td></td>
</tr>
<tr>
<td>Swarm Intelligence for Mixed Pixel Resolution</td>
<td>190</td>
</tr>
<tr>
<td>V. K. Panchal, Nitish Gupta</td>
<td></td>
</tr>
<tr>
<td>Modulation Transfer Function Compensation through a Modified Wiener Filter for Spatial Image Quality Improvement</td>
<td>195</td>
</tr>
<tr>
<td>Jessica Wong Soo Mee</td>
<td></td>
</tr>
<tr>
<td>Sorting Networks on FPGA</td>
<td>201</td>
</tr>
<tr>
<td>Devi Prasad, Mohamad Yusri Mohamad Yusof, Smruti Santosh Palai, Ahmad Hafez Nawi</td>
<td></td>
</tr>
<tr>
<td>Prediction of Wear Mechanisms in High Speed Steel Hobs using Artificial Neural Network</td>
<td>204</td>
</tr>
<tr>
<td>M. Jalali Azizpour, H. Mohammadi Majd</td>
<td></td>
</tr>
<tr>
<td>Charging for Reactive Power Employing Generation Distribution</td>
<td>209</td>
</tr>
<tr>
<td>Francisco David Moya Chaves</td>
<td></td>
</tr>
<tr>
<td>Modeling and Prediction of Ship Squat using ANFIS and GMDH-type Neural Network</td>
<td>215</td>
</tr>
<tr>
<td>K. Salmalian, M. Soleimani</td>
<td></td>
</tr>
<tr>
<td>Modelling on Pedestrian Accidents in Malaysia</td>
<td>224</td>
</tr>
<tr>
<td>Noorizam Daud, Haslinda Ab. Malek</td>
<td></td>
</tr>
<tr>
<td>Deep and Surface Learning of Elementary Calculus Concepts in a Blended Learning Course</td>
<td>228</td>
</tr>
<tr>
<td>R. Haripersad</td>
<td></td>
</tr>
<tr>
<td>Towards a General Legal Framework of an International Objective Environmental Liability?</td>
<td>237</td>
</tr>
<tr>
<td>Cristian Mares</td>
<td></td>
</tr>
<tr>
<td>Measuring and Improvement the Productivity by PMCI Model in Fuzzy Environment (Case Study)</td>
<td>243</td>
</tr>
<tr>
<td>Ghassem Haj Ebrahimi, M. Ali Sobhanallah, A. Mirzazadeh</td>
<td></td>
</tr>
<tr>
<td>Study of Soil Erosion and Sediment Transport in Small Watersheds</td>
<td>249</td>
</tr>
<tr>
<td>E. B. M. Beilicci, R. F. Beilicci, C. A. Girbaciu</td>
<td></td>
</tr>
<tr>
<td>Indoor Air Quality in Apartment Buildings of Estonia</td>
<td>257</td>
</tr>
<tr>
<td>Alo Mikola, Teet-Andrus Koiv</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Simplified Design Approach of Rectangular Spiral Antenna for UHF RFID Tag</td>
<td>262</td>
</tr>
<tr>
<td>Makroum El Mostafa, Rifi Mounir, Latrach Mohamed, Benbassou Ali</td>
<td></td>
</tr>
<tr>
<td>Maximal Allocated Benefit and Minimal Allocated Cost for Interval Data</td>
<td>269</td>
</tr>
<tr>
<td>Sahand Daneshvar, Mozhgan Mansouri Kaleibar</td>
<td></td>
</tr>
<tr>
<td>An Investigation of the Protection Performances for Current-Limiting Fuses using Stainless Steel Wire</td>
<td>276</td>
</tr>
<tr>
<td>Narong Mungkung, Somchai Arunrungrusmi, Toshifumi Yuji</td>
<td></td>
</tr>
<tr>
<td>Assessment and Forecast Stability of Slopes in the Carpathian and Sub-Carpathian Area of DamboviNa and Prahova. Classification and Risk Factors</td>
<td>284</td>
</tr>
<tr>
<td>Alexandru Istrate, Madalina Frinculeasa, Gica Pehoiu</td>
<td></td>
</tr>
<tr>
<td>Numerical Results for the Blood Micro-Circulation Modelled by a Porous Media</td>
<td>290</td>
</tr>
<tr>
<td>Yvon Maday, Noura Morcos, Toni Sayah</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>297</td>
</tr>
</tbody>
</table>
Proposal for a Cross-linguistic and Cross-cultural Research Project on Idioms of Food in Europe

PETRONELA SAVIN
Faculty of Letters
“Vasile Alecsandri” University of Bacău
Bacău, 157, Mărăşeşti Boulevard, 600115
“Alexandru Ioan Cuza” University of Iaşi
Iaşi, 11, Carol I Boulevard, 700506
ROMANIA
gabureanupetronela@yahoo.com

Abstract: - Terms like “cross-linguistic” or “cross-cultural” are often used in phraseology research but the main question as to what idioms actually constitute the so-called “phraseological uniformity of Europe” has never been systematically studied. The present paper describes our first steps of a research project which aims at creating a collection of European idioms containing those units structures regarding the human being's nourishment and feeding, which exceeds the perspective of a simple compilation of phrasemes that are separately treated and alphabetically ordered, but organised as organic elements associated to a coherent system.

Key-Words: - European idioms, onomasiological field, human nourishment, systemic organization

1 Introduction

The present article concerns idioms of food that are found across various languages in Europe. The idioms are seen as the central category of phraseology, distinguished from other categories. Unlike proverbs, sayings and quotations which differ from idioms by always making up an entire sentence, an idiom is a structure in which the meaning of the entire word formation does not correspond with the sum of the meaning of the individual words.

Phraseology researchers appear to have some vague knowledge about a cross-linguistic approach of idioms, but the set of problems connected with this perspective has never been methodically studied. There is no tradition of idioms studies that could be comparable to the rich tradition of international cooperation in the field of proverb research, represented by Abquilina [1], Kuusi [2], Paczolay [3], [4]. All these authors present extensive collections of proverbs that are common in many languages, but which are not primarily set in linguistics but in a wide-ranging ethnological-folkloristic and cultural framework. However, some researchers interested in making collections of the idioms in Europe and beyond have especially noticed common aspects. Matti Kuusi, in his important book “Regen bei Sonnenschein. Zur Weltgeschichte einer Redensart” (“Rain in the sunshine. Toward the world history of a saying”), published in 1957, succeeded in recording more than 3,000 variants of the saying mentioned in the title, taken from hundreds of languages and dialects spoken in various countries and continents, which represents an impressively extensive research material without modern media of today [5]. Menac, in 1987, also presents an inventory of several dozens of common European idioms, extracted from six languages of diverse genetic relationships, namely two Slavonic (Croatian, Russian), two Germanic (German, English) and two Romance (French and Italian) languages [6]. Korhonen, in 1991, realizes a comparative study of idioms which covers nine European languages: German, Finnish, French, Italian, English, Swedish, Russian, Hungarian and Estonian [7]. Other attempts to analyze idioms across several languages fall far behind the above-mentioned studies.

2 Problem Formulation

The cross-linguistic similarities of idioms have never been systematically studied. Regarding an onomasiological approach of units structures,
research are much less convincing. Several “cross-cultural” studies start from the traditional grouping of idioms into “thematic groups” (e.g. idioms with body parts, animal or garment constituents), mistakenly referred to as “onomasiological” approach as well.

3 Problem Solution

In response to the current situation of onomasiological cross-linguistic studies of idioms, our project’s main objective is the elaboration of an European dictionary containing those idioms regarding the human being's nourishment and feeding, a dictionary which exceeds the optical of a simple collection of phrasemes that are separately treated and alphabetically ordered, but organised as organic elements associated to a coherent system. This dictionary is aimed at taking profits from current linguistic space, providing an answer to Eugeniu Coşeriu’s challenge regarding the study of language from the perspective of culture’s universality and having in mind the various demands of linguistic research which, as compared with other subjects, entails “the most numerous connections with the man’s way of being and with all the human activities in general” [8].

This task requires substantial empirical preliminary work, which cannot be carried out by one single researcher. An international, multi-language cooperation of linguists will be necessary to organize the corpus of idioms in as many European languages as possible.

3.1 Argument for an European phraseological dictionary of field of human nourishment

According to Dubois [9] and Rey [10], the dictionaries can be compiled by authors who are not linguists, but this does not mean that the dictionary does not assume a particular perspective on language, even if lexicographers are not aware of it. Béjoint argues that the most important currents of theoretical linguistics had echoes in practical lexicography, even if this happened after a certain period of time [11].

Obviously, lexicography, as an independent discipline, benefits from the evolution of linguistic research, though the focus in current lexicographic work has shifted to the structure and functions of dictionaries.

A method and a linguistic discipline with valuable results in the current research is onomasiology. Designed by German linguists as a discipline studying the names of concepts in a certain language, onomasiology combines with the fields method to delineate lexical class (paradigm) within which they can rigorously analyze the semantic relationships. These methods contribute to the delimitation of our corpus dictionary, including the European idioms founded on images of human nutrition. A dictionary of an onomasiological field which is exclusively made up of a phraseological corpus it’s a <cultural dictionary>, in the sense that A. Ray and S. Delesalle give to this phrase, a dictionary to answer socio-cultural questions [12]. This approach also subsumes the ethnolinguistics, the European version of what Anglo-Saxons called linguistic anthropology. In Eugen Coşeriu’s terms, this linguistic discipline is aiming at “the study of language variety and variation in close contact with civilization and culture of a community” [13] while for A. Duranti “is an interdisciplinary field dedicated to the study of language as a cultural resource and speaking as a cultural practice” [14].

In the modern lexicographical theory of functions, cf. Bergenholtz-Tarp [15], it is considered that dictionaries have to meet the communicative needs and the cognitive needs of the users. By satisfying the communicative needs it is meant the help in solving a problem with text production, reception, translation, while by satisfying the cognitive needs it is meant the help in achieving systematic knowledge about a linguistic phenomenon or a non-linguistic phenomenon or context. An European phraseological dictionary of an onomasiological field obviously responds to the cognitive needs. It can be an instrument of great utility not only for linguists, but also for the anthropologists considering that language, at the level of idioms, preserves the deepest structures of the folk mentality.

3.2 The methodology and expected results of the project

In the former part of the dictionary the corpus of phrasemes based on images of feeding will be structured on onomasiological nuclei of the type NOURISHMENT/FOOD, BEVERAGES, CHARACTERISTICS, ACTIVITIES, ORGANS, INSTRUMENTS, PROCESSES regarding feeding, which are in turn divided into subcategories reaching the level of the food image which is phraseologically exploited. At the level of each image we will identify the signifying directions of phrasemes.
The choice of human nutrition as the topic for study will be the occasion to reveal some essential mechanisms of the reflection of history which is not strictly related to the description of events in the language, because in order to illustrate the linguistic specificity of an idiom we must study the representation forms of the quotidian structures in the language, feeding being one of them, together with housing and clothes-making [16].

Concerning the project submitted here, the methodology is specific to a lexicographic approach, which entails the following stages:

1. Inquiring into the works on general and European linguistics regarding the study of phraseology, as well as into the lexicographic sources of the European languages;
2. Selecting the phraseological corpus of the onomasiological field of nourishment in the European languages and organising it into phraseological nuclei;
3. Identifying the structure of each phraseological nucleus;

In this stage each term – food image with its phraseological usage – will be extracted from the lexicographic source. We will compare the structures from different sources and we will write down the meaning and register differences, if the case. We will directly organise the extracted material on the abovementioned phraseological nuclei: NOURISHMENT/FOOD, BEVERAGES, CHARACTERISTICS, ACTIVITIES, ORGANS, INSTRUMENTS, PROCESSES regarding human feeding.

4. Elaborating each dictionary article;

At the level of each section we will register the food images which present phraseological illustration. Each food image, for instance EAT, incorporated into the section ACTIVITIES regarding human feeding, is investigated in a separate chapter where all its phraseological units organised on signifying directions will be presented. For example:

EAT „to take into the body by the mouth for digestion or absorption”
1. To overwhelm or defeat thoroughly
Eng. eat somebody alive
Eng. eat somebody for breakfast [17]
Fr. il veut le manger tout crû [18]
It. mangiarsi vivo (qcn.)
It. mangiarsi in insalata (qcn.)
It. mangiarsi in un bucone (qcn.) [19]
Rom. a mânea pe cineva de viu
Rom. a mânea fript pe cineva [20]
Rus. съесть живьем кого-л. [21]

2. To desire/ want someone physically someone physically
Fr. dévorer des yeux (qcn.)
It. mangiare con gli occhi / sguardo (qcn.)
Prt. comer com os olhos [22]
Rom. a mânea (pe cineva) din ochi (cu ochii)
Rus. пожирать глазами
Blg. поглядя с очи [23].

The above example holds that feeding phraseology enables the discovery of mechanisms of interaction between man and environment. Negative relationship with otherness is best expressed through violent act of ingestion. But from the digestive womb to the sexual womb it is only one step, confirming the connection between greed and sexuality [24].

These homologous idioms in different languages demonstrate the existence of phraseological universals, in fact, the general form of representing the world in direct correlation with human experience. In this context, the study of European idioms regarding the human being's nourishment and feeding is a form of highlighting the deep structures of collective mentality.

5. Elaborating the contents, the title indexes for food images, the bibliography;

The usage of this dictionary is meant to be an accessible and productive one. By continuously citing the bibliographical resources we aim at guiding the reader to the essential works in the domain of European lexicography. The elaboration of the contents and the appendix with title indexes
we intend to overcome the common difficulty of the reader when turning to dictionaries whose material is alphabetically organised.

6. Formulating conclusions regarding the specific character of the dictionary, some guidelines for usage, the abbreviations list employed in the dictionary;

Having in view the fact that this book is set to organise the phraseological material, and not to record it alphabetically, we need to present the general parameters for conceiving the dictionary, as well as some guidelines for its usage.

7. Publishing the Phraseological European Dictionary: The Onomasiological Field of Human Nourishment at one of the internationally acknowledged publishing houses.

4 Conclusion

As a condition of existence, food is one of the most important chapters in human development. Feeding is defined by what is generally human nature, characterized by a spontaneous automatism. This belongs to nature. And, also, it is subject to restrictive rules, being relative and individual. By this, feeding is a part of culture. Both dimensions of the food act, relate to nature, as source needed to revitalization, and cultural dimension, as way of knowledge and understanding life, are valued by language. Idioms are the language section where the food pictures show the true relationship between culture and nature in the popular mentality. Our project tries, in the process of meaning of idioms, to highlight food, as a cultural sign, in the specific elements of European culture and of human culture generally.

The outcome of this project, an European phraseological dictionary of the onomasiological field of human nourishment, represents the first attempt of this kind. This dictionary is not organised as a simple collection of phraseologisms that are separately treated and are alphabetically ordered, but as organic elements associated to a coherent system. The former part of the dictionary entails the structuring of the phraseological material on onomasiological nuclei, reaching the level of a food image corresponding to a series of phraseological valorisations, organised on signifying directions.

The volume ensuing this project, European Phraseological Dictionary: The Onomasiological Field of Human Nourishment, will constitute a philological resource-tool which will promote the scientific activities in the domain of ethno-linguistic research. The analysis of idioms could transgress the level of language as a system, as it focuses on their importance at the level of speech, but also at the level of discourse, as a method of identifying language acts as performance acts. Studies in semantics, stylistics and even pragmatics centred on phraseology based on human nourishment at the level of discourse, could prove their productivity in the space of contemporary linguistics.

Considering the established objectives and the results which are expected, the project provides answers to some requirements related to the strategies of developing and improving European science, in order to increase the visibility and prominence of cultural research at an international level.

References:
Acknowledgements:
Financing of this article was supported in part by the project CNCSIS PD-582/2010 and in part by project POSDRU/89/1.5/S/49944.
Abstract: Living in Kuala Lumpur, city centre of Malaysia, bring matters of choosing dwelling as shelter. The escalating of land promotes vertical residential building. Sharing lifestyle in this type of dwell, needs a pull amount of money to manage and maintain the building ensuring of good condition and also prevent from decreasing value. Therefore, it is very important to have enough fund to bring all facilities in order. Previous study on residential schemes in Malaysia found that 50% of housing complex obtain less than 50% payment of service charge, whereas no complex gets collection above 80%. In the other words, 100% collect less than 80%. In this study, research was done to analyse the relationship in between house price and annual management fund. It was collect among 150 schemes of high-rise in Kuala Lumpur. It also limited to the non low-cost unit with the price RM250,000 and below. Study shows that both house price and annual management fund is linear function with gradient of 0.11 or we determine the relationship is 11% among this two element.

Key-Words: Management corporation, High-rise housing management, High-rise residential, regression coefficient

1 Introduction
Housing is a basic need for all and increased population and income will lead to the increase in housing demand. The increase of working group indirectly will increase the demand for houses since they are the productive group and have the purchasing power. The concept of high-rise or strata living in this country is not new. It started in 1958 with the two high-rise complexes of Pangsa Sulaiman Court and Pangsa Jalan Loke Yew (Leong, 1980). In mid-sixties, high rise construction was emphasised to reinstate squatters and accommodate urbanisation (Alinah, 2004). However, due to the increasing urbanisation and the scarcity of land the development of skycrapers and high-rise residential complexes has offered the best solution. Nevertheless, the demand for house differs between the early needs and the demand from house buyers now. The trend now is more towards the quality living. The house price also increases parallel with the demand for the house (Tiun, 2006).

Conversely, living in this kind of scheme means sharing a lot of things. Conditionally, this is because buyers are now paying more attention to value added features such as security, privacy, covered parking space, swimming pool, landscaped garden and many others. Subsequently, to have a good living in high-rise properties, a good management is needed to ensure everything goes in the right track. And to have a good management, good finance is most critical (Adi, 2007). Unfortunately, the experience on managing high rise residential properties in the country is very new and inconsistent (Tiun, 2006). While fund collection is not in the good order, a management agent and residence have their own reason in facing this problem (Jamila, 1994). Management declares that they cannot manage and maintain the property as it should be since they do
not have enough capital resources while the residents complain that they refuse to pay because the maintenance done was commensurate with the fee. The problems appear when residents cannot understand the sharing concept. Some of them are declining because they do not use all the facilities (Jamila, 1994). Although extensive effort are being undertaken to collect fees, it seems that collection is a critical issue (Tiu, 2003). Norngainy (2005) found that just about 50% of complexes could only manage to collect less than 50% maintenance fees.

This paper discusses and suggests a way of solving the problem. It will focus on the fund or deposit concept.

2 Literature Review

To most individuals housing represents the largest single investment in life. People are beginning to realise that effective facilities management is very important to ensure that everything goes well. Facilities management is an activity with the wide range of businesses; it is not an easy task. It involves proper planned activities to achieve goals and objectives. In the other words, property management plays a big role in sustaining the property value and maintaining high return on their investment (Tiu, 2006). The focus of property management and facilities management is often associated with maintenance job and collecting rent (Singh, 1969). But the continuous growth trend of high-rise residential buildings indicates that there is a need for an effective ownership and property management system to have a quality living experience among high-rise residents in this country.

Under such residential schemes, a multiplicity of owners have their interests over the same parcel of land. Each residential unit is owned by different owners. The main characteristics of this form of property are individual ownership of a unit, shared ownership of common property and membership of a body corporate responsible for the management of development (Christudason, 2004). This makes managing such complex is the difficult task. Therefore, a management body is needed to manage the building together with the public area (common property) such as parking area, lighting and other common property in the buildings that are shared together.

A residential high-rise unit is different from other residential property types with regard to its management responsibilities. Unlike the case for single-owner dwellings where responsibilities for managing and maintaining the property lie exclusively with the owner, a high-rise residential complex needs to be organised and maintained by a Management Corporation and, because of that, gives rise to the issue of facilities management (Linariza and Ashok, 2003). Adding to the difficulty, owners come from different backgrounds and ethnicity to contribute to the need for proper and systematic residential complex management. This is to ensure that the management serves the interests of the majority of owners and at the same time does not neglect the interests of the minority (Liaia, 1998).

The Strata Title Act was introduced in 1985 as to support the National Land Code. This Act is to ensure that all high-rise residential issues can be addressed effectively. An important provision in the Act is the requirement for the establishment of Management Corporation to manage housing complex. To ensure that the building functions properly and complies with the law, residents need to contribute to management fees. All building management tasks are costly, especially the maintenance parts. The high service charges contribute to discontentment among residents who often fail to see the necessity for contributions and flow of the money. The Management Corporation, on the other hand, has responsibility to keep service charges to more realistic levels and find alternatives in reducing maintenance cost.

3 Issues Regarding Service Charge in High Rise Residential

House ownership scheme will determine housing management activities. Ownership is classifiable into two categories, i.e. single ownership, and multiple ownerships or one building with different owners. This research focuses on multiple-ownership housing developed by government and private sector. According to Tiun, (2003), high rise housing management is governed by two main legislations, the Housing Development (Control and Licensing) Act 1966 and the Strata Title Act 1985 (West Malaysia) (1985 Act).

According to Tan and Teo (1990), there are three parts to housing management philosophy: the maintenance management (tero-technology), rental and pledge maintenance, and community development management. This aspect was enlightened by Priemus et al (1999) as in Gruis and Nieboer (2004) that housing management comprises of four categories: technical management (such as maintenance and refurbishment), social management, financial management (such as treasury and rental policy) and ownership management (such as renting, buying and selling). While Priemus et al (1999) look at all aspects, the present research concentrates on financial and management aspects only.
Each building needs to be managed and maintained continuously as to cope with normal wear and tear. Lack of attention to maintenance will contribute to heightened wear and tear (Ines, F. C. & Jorge, D. B., 2002). Tiun (2003) argues that housing management plays an important role especially in maintenance activities to keep building in good condition. Moreover, it is important to make sure all facilities are managed wisely in order that they are safe for habitation (Ahmad, 2003). This aspect is the responsibility of the management corporation as mentioned in Section 43 of Act 318.

The management needs financial resources to provide the services to residential property. Towards this, each resident is required to pay service charge with regard to the services given. This charge includes payments for insurance, refurbishment, cleaning, maintenance, lighting for public area and other costs to comply with law, rule and ordinance (Ismail, O., 1993). Service charge is defined as money collected by developer, agent or management body from the unit owners for maintaining and managing all common areas in the residential complex. This charge is also known as a fee which, in addition, all residents are obliged to pay (Government of Malaysia, 2003). For units with strata title, the amount charged by Management Corporation is based on the share contribution of each unit. Nonetheless, some owners are reluctant to pay because they consider the charges to be too high (Jamila, 1994).

It is typical with any management activity, finance is the most critical element in effective management of high-rise housing. Without adequate fund, proper management of high-rise housing will be affected. Fee collection and arrears present the two greatest challenges for any management body (such as developer, local authority, management corporation and residence organisations) undertaking the management of strata housing scheme (Liass, 1998; Roerup, 1998; Jamila, 1994; Mohd Razali, 2001; Sapian, 2003; Tiun, 2003; PKNS, 2004; Eddy, 2004). These problems are more frequent among low and medium cost housing complexes. In this regard, Ahmad, et al (2005) found financial factor as the most important contributor to the effectiveness of facilities management in multiple-ownership management. Therefore, serious consideration needs to be given to this factor if management problem is to be resolved.

Since service charge and more specifically maintenance charge and sinking fund, are a form of investment expenditure from owners’ point of view, it is paramount to pay attention to the value returned from the investment. Viewed in this manner, owners obviously would expect to get the best services out of the fees they pay (Liass, 1998). For this reason, Management Corporation needs to ensure that the quality of services they provide is at par with the fees charged. Further, the management corporation needs to negotiate with owners and seek fee resolutions that owners are comfortable to accept.

Although the Strata Title Act was introduced in 1985, owners’ awareness of the importance of service charge payment is still very poor. This is evident from the substantial arrears that have accumulated over the years. PKNS as a housing management body, for example, claimed that their service charge arrears have risen to RM8.1 million by March 2003. In the case of DBKL (Kuala Lumpur City Hall), its 2004 budget summary showed housing accounts as in deficit by RM38 million by the 31 December 2004. Tiun, (2003) claims that there are certain high-rise housing complexes where the arrears are in excess of 60%. When the delivery of services is affected as a result of non-payment, its effect is felt among those who pay the service charge as well. This tends to encourage the latter to take the same action of not paying. Consequently, the Management Corporation has to shoulder the burden of non-payment, to the extent that they may have to draw on their own financial resources or sinking fund to cover the management expenses.

Studies show that owners give various reasons for their reluctance to pay service charges. Their reasons range from the fact that they do not fully utilise the facilities (Roerup, 1998), to that the charge for service quality does not match the fees charged (Tiun, 2003). Jamila, (1999) contended that owners regret paying the management fee. In the case of the gymnasium usage, for example, the initial fee was announced as maintenance fee and the monthly fee as gymnasium management fee. Residents find this burdensome and eventually refuse to declare ownership of the facilities. According to Teo (1993), although Management Corporation is legally permitted to prosecute in the court of law under Sections 52(2), 53A, 53(2) and 55A of Strata Title Act, the prosecution is rarely resorted to for impracticality reason (Roerup, 1998; Tiun, 2003), such a step might affect overall complex occupants’ pride (Sapian, 2003). Also, although developer is allowed to resort to the disconnection of electricity or water supplies, again such an action is rarely taken (Roerup, 1998; KPKT, 1999; JPN, 2001). When most occupants default on their payment of service charges, the fund runs out. This impinging on the scheduled execution of most activities and affects the effective management of the complex (Tiun, 2003).
The above scenario portrays the importance of bringing management costs to realistic levels. Owners of high-rise residential complexes are increasingly aware of this, which creates the need for the introduction of some form of an index or benchmark on costs of high-rise housing management. This is so that occupants are better informed on whether or not they have paid reasonable cost, and on whether or not the services they receive are deserving of what they paid.

4 The Model
Gruijs and Nieboer (2004), quoting from Priemus et al. (1999), contend that there are four main areas in housing management: the technical management, financial management, social management and ownership management. The current study focuses on financial management, and the samples are drawn from the Management Corporation of strata title schemes where first annual general meeting (FAGM) has been held. Such a management body will continue in its function until it is terminated. In Korea there is a housing rental system called as Jonsei deposit. This system has been developed hundred years ago and only practised in Korea. In this system, the tenant will pay the landlord a large lump-sum amount of money of about 25%-70% of the house price. The lump-sum deposit will be deposit in any bank to get the interest as the payment for the house rent. after completing the tenure contract, the deposit will be return back to the tenure (Yeong-Hee Jang, 2001).

The flow of the system can be done to manage the fund deposit. It needs to generate income from the investment to fulfil the management fund and must also have a yearly surplus that must be deposited back to the deposit fund. This is to ensure the fund will supply enough money for the whole life of the building. Starting from the early stage of proposed building by the developer with facilities we can calculate how much money they need to manage the building for one year. Then it will be determined for all units. With the fund deposit model, a deposit can be determined for each house. These will be include in the house price. After that the fund deposit body will invest the money and make sure it will goes as what it should. After a year, the management corporation can apply from the fund. This formula will consider the return and the increase of the management fund needs. We will also assume that the building will remain for infinity. This is to ensure the deposit have enough money as long as the building exists.

This can be shown from diagram below where the lump sum amount of service charges for a management fund can over come the criticality in collecting those funds. Tiun (2003) and Norngainy (2005) find out the critical indebtedness amongst the unit owner. Although various methods used to collect that charges, the management still face difficulty such as facing with black mail and vandalism of their property (Tiun, 2003). If this concept can be accepted, therefore this challenge can be managed.

![Diagram](https://example.com/figure1.png)

**Figure 1: Conceptual flow chart**

**Lumpsum Deposit Management Fund**

Figure 1 shows the process of integrating the lump sum amount of service charge as management fund deposit. When the developers proposed the building, they are also proposing the facilities for residential complex. The house price offered consists of residential location and type of facility provided. In this stage the amount of the management fund per annum must be determine. Through this the amount of the deposit will be derive by the formula which consider the fund itself is perpetuity and sustained as the building itself. With this, we hope that the challenge in collecting management fund will no longer be a gap to both parties.
3.1 Fundamental at the model
In managing fund, the management financial principle must be followed. In developing this model, we assume that the service charge already known. This model is based on annuity concept. Annuity is constitutes a stream of over a period of time. Since, we assume the management fund has an annual increment, it is not suitable to use normal annuity. So, researcher used continuous annuity which is actuarial science formulation to get the cycle of deposit fund. This function is well known in insurance and pension scheme. Continuous annuity will give us present value. In this study, present value is the deposit amount that needs to be incurred in the house price.

\[ A = \int_{0}^{T} f(t) e^{-rt} \, dt \]  \hspace{1cm} (1)

where

- \( A \) = present value
- \( f(t) \) = annuity payment per year function
- \( r \) = return
- \( t \) = time
- \( T \) = payment time period

(Haussler, 1996)

In this paper we want to know what is the relationship in between the home price and the service change per year. The analysis is done by using Regression Analysis.

4.0 Analysis
Ratio of House Price to Service Charge, \( \alpha \)

Survey was conducted for 150 high-rise residential in Klang Valley’s surrounding. The chart 1 below shows the relation between house price and service charge yearly.

5.0 Conclusion
The challenge in collecting fee from the residents of high-rise residential has been seen as a critical issue in our country. Although many legislations have been introduced, it seems that there is a gap in persistance. Both parties, the management and the residents, are pointing at each other. In this paper, the study shows that if a fund can generate income for the management fund yearly, both parties will get the benefit. The residents will only pay the fee up to the period of the repayment loan (not for the whole of their live) whereas the management will not suffer depletion of the management fund.

References:
Relationship in between House Price and Annual Management Fund in High-Rise Residential in Kuala Lumpur, Malaysia

Enhancing Employability Skills through Industrial Training Programme

Y.M. YUSOFF, M.Z. OMAR, A. ZAHARIM, A. MOHAMED, N. MUHAMAD, R. MUSTAPHA

Dept. Of Eng. Sciences and Mathematics, College of Engineering, Universiti Tenaga Nasional, Putrajaya, Selangor, MALAYSIA

Centre for Engineering Education Research, Faculty of Engineering and Built Environment Universiti Kebangsaan Malaysia, Bangi, Selangor, MALAYSIA

Centre for Engineering Education Research, Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, MALAYSIA

yuzainee@uniten.edu.my, zaidi@eng.ukm.my, azami@eng.ukm.my, azah@eng.ukm.my, hamidi@eng.ukm.my, ramlee@pkrisc.cc.ukm.my

Abstract: Employers nowadays consider a graduate’s academic achievement alone is not sufficient for hiring the engineering graduates. They sometimes find achievement other than academic such as employability skills to be important in the recruitment process. Hence, the purpose of this study is to examine the employers’ perceptions on the performance of graduates and interns as an effort from Faculty of Engineering and Building Environment of Universiti Kebangsaan Malaysia (UKM) to meet the job market demands. Two different sets of questionnaires were used to collect data. First, data collected on the employers’ perceptions on skills possessed by new engineers in their workforce from a sample of 302 employers of engineering graduates in Klang Valley. Second, data collected on assessment of the performance of engineering industrial training students from 305 employers of interns. The study reveals that attributes exposed in industrial training programme fulfil the attributes required by employers. It shows that the industrial training programme is an essential component of engineering curricula. The programme develops undergraduates’ employability skills through activities that gave the interns an exposure to industry practices. The industrial training also strengthens employability skills of engineering students through relevant work experience. The similar trend on percentage of agreement by employers also indicates that the attributes attained by UKM’s students in industrial training meet the criteria required by employers.

Key-Words: Graduates, interns, employers, employability skills, performance.

1 Introduction

The importance of employability skills is now greater than ever in engineering sectors. Most engineering industries are now using highly developed equipment, systems, and systematic processes, requiring highly trained, integrated and multi-skilled engineers [1]. As a result, employers place main concern on employability skills in hiring new engineers causing a significant increase on demand for skilled engineering graduates in Malaysia [2-7]. Yet, previous studies indicate that the supply of entry level engineers in Malaysia does not meet current demands, and future expectations in engineering sectors [2; 3; 4]. One of the key factors is the graduates are not ready to enter the workforce [2-7]. They lack employability skills needed by employers [2-5], competencies or capabilities of personal skills [6] and not equipped with the relevant skills [7]. Therefore, employers’ perceptions and expectations have its influential in determining the employability skills needed. Several universities have conducted a study to obtain their perceptions and expectations in various disciplines. The best universities require engineering students to obtain excellence academic achievements, however, common queries still on: “who they will be; what they will do; where they will do it; why they will do it …” [1].

Malaysian employers feel it is a beneficial for a new engineer to have acquired certain employability skills than just studied certain courses [7]. Although professional education, skills and training expected to be exposed by employers, rather than by universities, employers still concerned about soft skills own by graduates, in addition to, academic achievement when they recruit new engineers.

Engineering education in Malaysia has long been considered as a process of refining characters and
skills for employment in the workplace. With the tremendous changing in the employment system, the tertiary education must take more responsibility in developing their students’ personal qualities in terms of employability, and implement appropriate programs to meet the changing needs in industry. One approach to help increase the employability skills are through an industrial training or internship programme for engineering undergraduates’ students.

2 Engineering Graduates and Industrial Training Students

Unstable economic environment and globalisation influenced the recruitment practice, which also affect to higher education system in Malaysia. This can be seen in the advertisement that stated clearly the criteria required for engineering post in Malaysia nowadays. Malaysian engineering graduates have strong basic engineering knowledge and sufficient technical competency [2; 5; 9], but employers complaint on the entry level job applicants still lacking of generic skills [5] such as communication skills, decision making and interpersonal skills [2; 5; 10]. Employers need an engineer with a solid theoretical background and equipped with essential soft skills and employability skills [2]. Robinson (2000) defined employability skills as “those basic skills necessary for getting, keeping, and doing well on a job.” Engineering employability skills are highly related to nontechnical skills or abilities. These nontechnical skills not only help graduates to succeed in getting employed but also doing well in the workplace [8]. Obviously, engineering graduates should have employability skills to apply and practice the knowledge in the workplace.

Most engineering companies today expect higher education providers (HEP) prepare for “work-ready” graduates, who have sufficient skills and abilities to work immediately [2; 3; 7]. Industrial training is one of approaches to help students to prepare for “work-ready” graduates. It is part of the learning process to apply theories learned in college and to develop technical skills and employability skills. Industrial training programme helps students to discover different roles and activities that give them an opportunity to experience different challenges in a real workplace. The activities provide exposure to realities of the working environment and an opportunity to gain ‘hands-on’ working experience in related industry. The industrial training programme is part of curriculum in Malaysian engineering programme. Duration of the industrial training is between 2 to 6 months, subject to the requirements of the degree programme. Even though the educational background is significant to employers, it is also necessary to have real-world experience in the area before entering the workforce. Therefore, industrial training programme is one way of providing experience by which these skills can be attained.

Upon completing the training, the students have to submit a technical report as part of the assessment of their performance in industrial training. The assessment also includes the input by the employers on the soft skills and knowledge as showed by the students during the training. This input allows the faculty to assess the students’ abilities as well as to ascertain whether they have met the objectives of the industrial training set by accreditation requirement.

Accreditation standards set by the Engineering Accreditation Council (EAC) require a minimum continuous period of two months of industrial training for each student. The industrial training supposes to provide the interns an exposure to professional practice in an engineering-practice environment [12]. It is essential for the interns to familiarise themselves with common engineering processes and practices and to expose themselves to a wide variety of processes required at a level appropriate to them. Therefore, the interns of an engineering programme are expected to attain programme outcomes (PO) specified for industrial training programme. The Faculty of Engineering and Built Environment, UKM has set seven POs (of 12 overall Pos) listed in Table 1 to be achieved by the internship programme.

These programme outcomes for industrial training describe the expectation of knowledge and abilities should be attained after completing the programme. Therefore, it is necessary to have feedback from employers on the performance of interns, to see whether the objectives of industrial training programme have been achieved. This feedback also will provide information on how the outcomes meet the employers’ requirement. Parts of an effort to achieve this, a study to obtain information on the perceptions of employers regarding to employability skills through interns and graduates need to be done. Hence, this study seeks employers’ perceptions on Malaysian engineering graduates attached to them and performance of Universiti Kebangsaan Malaysia (UKM) industrial training students.
Table 1: The relevant Programme Outcomes for UKM Industrial Training Programme

<table>
<thead>
<tr>
<th>Programme Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO1 Ability to acquire and apply knowledge of science and engineering fundamentals.</td>
</tr>
<tr>
<td>PO2 Ability to communicate effectively, not only with engineers but also with the community at large.</td>
</tr>
<tr>
<td>PO6 Ability to function effectively as an individual and in a group with the capacity to be a leader or manager as well as an effective team member.</td>
</tr>
<tr>
<td>PO7 Understand the social, cultural, global and environmental responsibilities of an engineer, and the need for sustainable development.</td>
</tr>
<tr>
<td>PO9 Ability to design and conduct experiments, as well as analysing and interpreting data.</td>
</tr>
<tr>
<td>PO10 Ability to function in multi-disciplinary groups.</td>
</tr>
<tr>
<td>PO11 Having the knowledge of contemporary issues.</td>
</tr>
</tbody>
</table>


3 Methodology

3.1 Survey on Engineering Graduates

The selection of the respondents employed a random sampling method. The questionnaires developed to gather information on the performances of engineering graduates on adopted fifteen employability skills. These 15 employability skills listed in Table 2, required by Malaysian employers is quite similar to the attributes required by the accrediting body, Engineering Accreditation Councils (EAC) of Malaysia. The questionnaires sought to collect data on the perception of the employers towards performance of new engineers in their organisation using a five point Likert-type scale. The employers responded to level of satisfactory on knowledge, skills and experience owned by engineering graduates in their work place. The responses “1” indicated “Not Satisfactory at all” and “5” indicated “Most Satisfactory”. The researchers had targeted about 400 employers with sample limited to engineering employers in a limited geographical area of Malaysia. Data collected through face-to-face interview, online survey and snow-ball sampling using a set of questionnaires. In the end, 337 engineering employers responded out of 500 questionnaires given out and only 302 usable responses were analysed. For interpretations and conclusions of study finding, the data analysed using the software package for social sciences (SPSS 16.0).

Table 2: List of Employability Skills

<table>
<thead>
<tr>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Have adequate background knowledge</td>
</tr>
<tr>
<td>B. Ability to apply knowledge</td>
</tr>
<tr>
<td>C. Ability to function effectively in group</td>
</tr>
<tr>
<td>D. Capability to function as a leader in group</td>
</tr>
<tr>
<td>E. Ability to carry out leader/ manager's instruction</td>
</tr>
<tr>
<td>F. Behave professionally and practice good ethics</td>
</tr>
<tr>
<td>G. Appreciate social and cultural responsibilities</td>
</tr>
<tr>
<td>H. Awareness on environmental responsibility</td>
</tr>
<tr>
<td>I. Recognize the needs of lifelong learning</td>
</tr>
<tr>
<td>J. Ability to extract information</td>
</tr>
<tr>
<td>K. Ability to practice listening skills and giving feedback</td>
</tr>
<tr>
<td>L. Ability to communicate in public or with community at large</td>
</tr>
<tr>
<td>M. Ability to express ideas verbally</td>
</tr>
<tr>
<td>N. Ability to make decision</td>
</tr>
<tr>
<td>O. Ability to work independently</td>
</tr>
</tbody>
</table>

3.2 Survey on Internship Students

The research study was an empirical nature, with a set of questionnaires handed to employers of interns at the end of two months of the industrial training. The internship programme is part of the curriculum where third year students undergo a 20-week (from the 12th April to the 3rd August 2009) training after successfully completed six semesters of studies. A set of questionnaires were delivered to a total of 305 employers to assess the interns under their charge, of which 295 usable responses were obtained, resulting in a response rate of 97 percent. The questionnaires use a five-point Likert-type scale to the attributes/skills measured. In the assessment, fifteen employability skills listed with scale “1” indicates “extremely weak” and “5” indicates “excellent”. A brief description of each of the attribute/skill is included in the questionnaire to avoid misunderstanding.
4 Findings and Discussions

The responses on scale 4 represent “satisfactory” and 5 represent “most satisfactory” take into account as “well-performed” skills by graduates. Meanwhile, the responses on scale 4 represent “good” and 5 represent “excellent” take into account as “well-performed” skills by interns. The statistical results demonstrate that the employer’s assessment on performance of the skills owned by engineering graduates and engineering interns are satisfactory. Figure 1 illustrates the percentage of agreement on well-performed employability skills as identified by employers on engineering graduates and interns.

The three highest percentage of employers agreed that the well-performed skills by graduates were the ability to function effectively in group (93%), ability to carry out leader/manager's instruction (92%) and ability to express ideas verbally (92%). The last three skills that are not substantially performed by graduates are the capability to function as a leader in group (74%), have adequate background knowledge (72%) and ability to communicate in public or with community at large (71%). The employer’s perception on employability skills apparently shows that the employers recognised all the skills performed by graduates as being, at least “satisfactory”. They were quite satisfied on skills performed by graduates but still emphasis on the lacking of ability to communicate in public or with community at large among graduates. This study also agreed with some other previous work in that employers quite disappointed with various generic skills of engineering graduates such as communication skills, leadership skills and decision making [2; 3; 7; 11]. However, surprisingly, this study shows that only 72% of employers agreed that graduates have adequate background knowledge and capability to function as a leader in group. The employers satisfied with the performance of the UKM students particularly in an ability to carry out leader/manager's instruction (96%); appreciate social and cultural responsibilities (95%); and Ability to function effectively in the group and work independently (92%). The other skills also assessed as well-performed skills by interns are: Recognize the needs of lifelong learning, Ability to practice listening skills and giving feedback, Awareness on environmental responsibility, Ability to extract information, Ability to apply knowledge, Act professionally and practice ethical ethics, and Ability to express ideas verbally. The skills that employers agreed not well-performed by interns are ability to communicate in public or with community at large (77%); ability to make decision (75%); and capability to function as a leader in group (67%). This situation is perhaps due to the interns lack of experience. Therefore, the students need to involve actively in various activities to gain experiences, in order to enhance their communication skills, leadership skills and the level of confident to make a decision.

The analysis of data on both graduates and interns shows they have a similar pattern on the level of satisfactory of all skill assessed in the survey. Figure 1 shows that the trend of the graph is similar except for the attribute of “appreciate social and cultural responsibilities”. The interns were evaluated as 95% by their employers/supervisor as well-performed whereas the graduates only scored a mere 76% on this attribute. The level of satisfactory on skills performed suggested that all of the skills rated as satisfactory, with lowest mean equal to 3.9. This does not imply interns and graduates are skilful enough to enter the workforce. The assessment done on graduates and interns has shown that both are lack of employability skills particularly in decision making, leadership and communication skills. The
students need to equip themselves with these critical skills to better prepare themselves before entering the market. The universities also perhaps need to provide more opportunities for the students to gain more industrial experience, and likewise, the industries to be more open and supportive with such programmes.

![Percentage of Employers Satisfied With the Performance of Engineering Graduates and Industrial Training Students](image)

**Fig. 1: Percentage of Employers Satisfied With the Performance of Engineering Graduates and Industrial Training Students**

5 Conclusion

The survey invited views on the performance in fifteen different employability attributes/skills based on previous studies and accreditation requirement. The employers’ percentage of agreement on well-performed skills shown by graduates and interns have similar trend, although they were from different cohort and were evaluated in different situations and job scope. This shows that industrial training is an essential component of engineering curricula to prepare for “work-ready” graduates for industries. An exposure to industry practice is vital for students to acquire employability skills through industrial experience. The similar trend on percentage of agreement by employers also indicates that the attributes attained by students in industrial training have met with the criteria as required by employers.

The programme outcomes as set by the faculty for the internship programme seemed to have been achieved as demonstrated by the outcomes achieved by the interns. Some of the “well-performed” interns have even received conditional job offers by the employers (i.e. by joining them upon the completion of their studies). Thus, it can be said that a successful industrial training tenure can become a powerful tool to increase a student’s employability capability.

Acknowledgment

The authors wish to thank the Committee for Industrial Training, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, for sharing with us the data of their internship programme.

References:

Enhancing Employability Skills through Industrial Training Programme

2


Competencies Achieved Through Industrial Training Programme

M. JAMIL1, M.Z. OMAR2, S.A. OSMAN3, N.T. KOFLI4, M.N.A. RAHMAN5, K. MAT6
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
43600, UKM Bangi, Selangor
MALAYSIA
lin@eng.ukm.my, 2zaidi@eng.ukm.my, 3saminah@eng.ukm.my, 4maverick@eng.ukm.my,
5mnizam@eng.ukm.my, 6kamarul@eng.ukm.my
web: http://www.ukm.my

Abstract: - This paper discusses the assessment on performance criteria obtained by the trainees during industrial training. A total of 303 trainees from four departments under Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, have successfully carried out the program with placements in various organisations. The training has been carried out for the duration of 20 weeks, from the 12th April to the 28th August 2009. In this study, 96% employers have given their valuable feedbacks on the performance of 290 trainees. From the employers’ perception, trainees have obtained high performance rating in many criteria including ability to carry out instructions (96%), social and multi-racial awareness (95%) and professionalism and work ethics (94%). However, the employers are mostly dissatisfied with trainees’ leadership ability which recorded the lowest score with percentage of 67%.

Key-Words: - employers’ perception, industrial training, performance criteria

1 Introduction
In the era of globalization, Malaysia needs to build up intellectual capital and transform workforce into skilled manpower in order to succeed in this newly competitive world. Based on evident from developed nations, skilled manpower is the driving force in many successful countries. Furthermore, future national development of our country is to bring forth challenges that require manpower to be innovative and equipped with knowledge and training with related skills. Industrial training program refers to experience in the real working environment that is relevant to professional development prior to graduation and an aid to prospective employment. It is an essential element in the development process of professional skills and work ethics required to become an engineer. The program is an extremely valuable component of professional courses in university education [1].

At Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (a.k.a. National University of Malaysia), the industrial training program is part of the curriculum where third year students undergo 20 weeks of industrial training after successfully completed six semesters of studies. Students from four departments involved in industrial training program which (training period from the 12th April to the 28th August 2009) are from Department of Civil and Structural Engineering (JKAS), Department of Electrical, Electronic and System Engineering (JKEES), Department of Mechanical and Materials Engineering (JKMB) and Department of Chemical and Process Engineering (JKKP). Students undergoing their industrial training are supervised by industrial training coordinator from each departments as well as industrial supervisors. Students also have to prepare a written report of their industrial training experience, which has to be submitted to their respective departments while the supervisor is required to evaluate the performance of the students by filling up the assessment sheet. The evaluation is very important to measure the competencies of trainees with respect to some performance criteria from the employers’ perspective. The feedbacks and evaluation from employers are very useful to further improved the industrial training program and equip students with desired criteria needed by employers upon graduation. Thus, this study was carried out to investigate employers’ perception and consequently measure the performance of students.
2 Assessment Methodology
This study involved activities including distributing assessment sheet to participating organization, collecting data and analyzing findings. This study utilized both quantitative and qualitative methods. The quantitative method will be used to determine the performance of students that highly related to the quality of human capital produce by institutions. On the other hand, the qualitative method will be used to capture the various comments, feedbacks and suggestion by industry towards improving the quality of industrial training program. The assessment of trainees by their employers was carried out using Likert scales. It is a psychometric scale commonly use as the familiar five-point bipolar response. The assessment is justified on educational grounds and soft skill. The assessment will reflect the performance of the students in industrial context, as judged by the employers who are closely concerned with the progress.

3 Results and Discussion
Data was collected through assessment sheet which contained both the profiles of the trainees and performance inventory. Results of the assessment are discussed and divided accordingly as follow.

3.1 Profiles of Respondents
Sample of this study is the students from four departments as shown in Fig. 1 which involved 59 student from JKAS (19%), 67 students from JKEES (22%), 80 students from JKKP (26%) and 97 students from JKMB (32%). A total of 303 students underwent the industrial training in various organisations of different industrial backgrounds in Malaysia. The primary data for this research was gathered through assessment sheets filled up by supervisor or relevant personnel in the organization participating in the industrial training program. The assessment sheet contains items that capture general information and items that probe on the expectations and perceptions of employers towards the trainees. A number of 290 employers (as shown in Fig. 2) have given their valuable feedbacks which involved 51 employers of students from JKAS, 67 employers (JKEES), 77 employers (JKKP) and 95 employers (JKMB). These numbers of employers represents 96% feedbacks out of the total number of organizations providing placement for industrial training for the year of 2009. Thus, it shows that many organisations have shown a great interest in assisting and further improved the industrial training program and curriculum development.

Fig. 1: Student Profiles Based On Department

3.2 Performance Criteria
The performance criteria measured by employers are shown in Table 1. Each criterion is coded as P1 to P20 respectively. Twenty criteria were evaluated based on Likert scale. The scale ranges from a group of five categories from least to most, asking employers to indicate the level or degree of certain criteria for each trainee. When responding to the assessment item as shown in Table 1, employers will specify their level of agreement to a statement for every performance criteria listed. The assessment sheet is designed to meet the objective of the program and simultaneously can be used for evaluation of trainees’ performance in industry. The scale indicator ranging from 1-5 represents Very
Weak (1), Weak (2), Average (3), Good (4) and Excellent (5). To simplify this paper, rating given to scales 1 and 2 are grouped together as ‘weak’, those for scales 4 and 5 are grouped as ‘good’ and scale 3 is considered as ‘average’. For this paper, only result for ‘good’ are shown.

The evaluation by the employers is given in Fig. 3. It shows the percentage of students evaluated as “good” by the employers. From the result, generally the students have performed well during the industrial training and employers also quite satisfied with the performance possessed by the students. As shown in Fig 3, the highest score was given by the employers to criteria $P_{16}$ (ability to carry out instructions). The result of 96% of students obtained high marks may indicates that majority of trainees are good and fast learners. The figures also indicate that the trainees were able to understand project objectives and translate them into detailed tasks, expanding, and implementing. However the lower score is recorded by criteria $P_{17}$ with only 67% of students able to function as a leader.

### Table 1: Performance Criteria Measured By Employers

<table>
<thead>
<tr>
<th>Code</th>
<th>Performance Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_1</td>
<td>Independent</td>
</tr>
<tr>
<td>P_2</td>
<td>Ability to make decision</td>
</tr>
<tr>
<td>P_3</td>
<td>Ability to express ideas (written)</td>
</tr>
<tr>
<td>P_4</td>
<td>Ability to express ideas (verbal)</td>
</tr>
<tr>
<td>P_5</td>
<td>Non-Verbal skill</td>
</tr>
<tr>
<td>P_6</td>
<td>Ability to communicate with public</td>
</tr>
<tr>
<td>P_7</td>
<td>Negotiation skill</td>
</tr>
<tr>
<td>P_8</td>
<td>Listening skill</td>
</tr>
<tr>
<td>P_9</td>
<td>Ability to interact</td>
</tr>
<tr>
<td>P_10</td>
<td>Ability to extract information</td>
</tr>
<tr>
<td>P_11</td>
<td>Lifelong learning</td>
</tr>
<tr>
<td>P_12</td>
<td>Disciplined and motivated</td>
</tr>
<tr>
<td>P_13</td>
<td>Environmental awareness</td>
</tr>
<tr>
<td>P_14</td>
<td>Social and multi-racial awareness</td>
</tr>
<tr>
<td>P_15</td>
<td>Professionalism and work ethics</td>
</tr>
<tr>
<td>P_16</td>
<td>Ability to carry out instructions</td>
</tr>
<tr>
<td>P_17</td>
<td>Ability to function as a leader</td>
</tr>
<tr>
<td>P_18</td>
<td>Ability to function as team player</td>
</tr>
<tr>
<td>P_19</td>
<td>Ability to apply knowledge</td>
</tr>
<tr>
<td>P_20</td>
<td>Adequate background knowledge</td>
</tr>
</tbody>
</table>

**Fig. 3: Students’ Performance during Industrial Training**

### 3.3 Achievement Level

Overall, the 20 measured criteria have been categorized into three different achievement level of performance indicated as high, medium and low as shown in Fig. 4, Fig. 5 and Fig 6 respectively. High achievement level comprises performance criteria with rating percentage of students more than 90%. Another achievement levels are medium (80%-89%) and low (below 80%). Eight criteria are considered as high achievement level with high rating more than 90% as shown in Fig. 4. The eight criteria include $P_1$ (Independent) $P_8$ (Listening skill), $P_{11}$ (Lifelong learning), $P_{12}$ (Disciplined and motivated), $P_{13}$ (Social and multi-racial awareness), $P_{15}$ (Professionalism and work ethics), $P_{16}$ (Ability to carry out instructions) and $P_{18}$ (Ability to function as team player). Also of some relief is that many aforementioned criteria fall into this group are related to positive attitude and soft skills. These data may imply that the engineering education of the
The faculty has been successfully providing enough emphasis on preparing quality graduates required by engineering profession and job market.

Furthermore, other seven criteria also obtained high approval rating from employers. Fig. 5 shows recorded percentage of more than 80% which include criteria related to personal soft skills such as ability to express ideas in written and verbal (P3 and P4), negotiation skill (P7), ability to interact (P9) and ability to extract information (P10). Thus, the data prove that engineering student are equipped with soft skills which play a vital role for professional success and help them to excel in the workplace. Competency on soft skills also has become an important criteria required by employers (2). The employers also satisfied with the criteria of environmental awareness (P13) which recorded percentage of rating 89%. This data may indicate that sustainable development and awareness towards green environment being practiced in university level have become an added value to students. Furthermore in some engineering courses offered by faculty, the understanding and importance of sustainability and cost-effectiveness in design and development of engineering solutions is emphasized. The criteria of ability to apply knowledge (P19) also recorded satisfying result with percentage of 86%. Ability to apply knowledge in problem solving is one of the key aspects of innovation practice in engineering company as discussed by Baxter et al. (2009)[3].

Fig. 4: High Achievement Level of Performance Criteria

As mentioned earlier, the employers are most dissatisfied with the criteria P17 with only 67% student shows the ability to function as a leader, scoring the lowest percentage (see Fig. 6). The figure also shows the low achievement level of performance for criteria related to ability to make decision (P2), non-verbal skill (P5) and ability to communicate with public (P6) with percentage of 75%, 77% and 77% respectively. The data shows that these four criteria are related to leadership characteristic that have to be strengthened among the students. However, since this only involved four criteria, the result therefore is not much significantly worrisome since all other criteria have received high approval rating by employers. All these criteria are among the soft skills desired by employers in order to work in multidisciplinary team which expose trainees to the real world in industrial sectors. To further improved the program and hence to further benefit the student, engineering education has to seriously evolve from providing students solely with technical skills to providing them with courses that provide students with the non-technical or soft skill.
4 Conclusion
In conclusion, this study exhibit employers’ assessment on students’ performance during industrial training. The performances showed by the students are satisfactory and fulfil the organisations’ needs with almost all criteria as evaluated by the employers have scored around 80% or more “good” rating. The results also reveal the importance of positive attitudes as well as soft skill characteristics in the real work environment. The employers also have forwarded some suggestions to the university to further strengthen the cooperation between the university and industry, such as through technical visits and organising industrial talk by the representatives from industries. The feedbacks also indicate that industrial training is viewed by the employers or organization as essential and would benefit both university and the industry.

References:
Use of Rasch Analysis to measure Students Performance in Engineering Education

SAIDFUDIN, M 1, AZRILAH, A.A 2, RODZO’AN, N. A 3, OMAR, M.Z 4, ZAHARIM, A 5 AND BASRI, H 6

1 Program Director, 2 P.hD, Program Coordinator, Exec.Dip. in Quality Management, UTM SPACE, 3Master (Candidate), Dept. of Mathematics, Fac. of Science, University Teknologi Malaysia, 81300 Skudai, MALAYSIA
4Assoc.Prof., 3 Professor Centre for Engineering Education Research, Fac. of Engineering and Built Environment,
6Professor, Ir; Deputy Vice Chancellor (A), Universiti Kebangsaan Malaysia, 43600 Bangi, MALAYSIA

Abstract: - The Engineering Accreditation Council of Malaysia (EAC) adopts the American Accreditation Board of Engineering and Technology 2000 (ABET) requirements which promote outcome based education (OBE) learning process. OBE calls for the evaluation of the subjects learning outcomes (LO) as specified in the Programme Specification. This good practice is implemented in the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM) teaching and learning processes. Evaluation method has been largely dependent on students’ performance carrying out tasks such as tests, quizzes or submission of assignments. Instrument construct were based on Bloom’s Taxonomy whilst the evaluation on the students performance output were assessed based on Students Observed Learning Outcomes (SOLO) Taxonomy which gives an indication on the student’s achievement of the subject expected LO. However, the measurement of the student’s achievement from the observed outcomes remain vague. This paper describes a measurement model using Rasch Analysis which can be used to measure the subject LO of an undergraduate engineering subject. An overview of the measurement model and its key concepts are presented and illustrated using the final exam paper given through subject KKKF1134 – Introduction to Engineering. Results obtained were assessed against the course LO maps for consistency and used as a guide for future improvement of the teaching method and style. The study shows that Rasch model of measurement can classify grades into learning outcomes more accurately especially in dealing with small number of sampling unit.

Key-Words:- Learning Outcomes, instructional objectives, performance assessment, Quality, continuous improvement.

1 Introduction
The assessment of student learning begins with educational values and a vehicle for educational improvement. Educational values should drive not only what we choose to assess but also how we do so. Where questions about educational mission and values are skipped over, assessment threatens to be an exercise in measuring what's easy rather than a process of improving what we really care about [1]. It’s greatest contribution comes on campuses where the quality of teaching and learning is visibly valued and worked at. On such campuses, the push to improve educational performance is a visible and primary goal of leadership; improving the quality of education is central to the institution's planning, budgeting, and personnel decisions. On such campuses, measuring learning outcomes to generate useful and meaningful information is seen as an integral part of decision making, and avidly sought. [2]

The Engineering Accreditation Council of Malaysia (EAC) adopts the American Accreditation Board of Engineering and Technology 2000 (ABET) requirements which promote outcome based education (OBE) learning process. OBE calls for the assessment of the subjects learning outcomes (LO) as specified in the programme specification. IHLs in Malaysia conducting any engineering programmes must assess the learning outcomes of its teaching and learning processes as a prerequisite to obtain EAC accreditation hence measurement.
2 Overview of Measurement Principles

Measurement has been grossly misunderstood and overlooked in many circumstances especially in the field of social science. Many researchers in social science are frustrated when existing instruments are not well tailored to the task, since they then cannot expect sensitive, accurate, or valid findings [3]. However, modern measurement method as practiced using item response theory with a focus on Rasch measurement model provides the social sciences with the kind of measurement that characterizes measurement in the natural sciences; i.e. the field of metrology.

The fundamentals of measurement must comprised of the instrument to be used for purpose which has specific unit of an agreed standard amount. An instrument must have the correct construct of linear scale which can be zero set and duly calibrated. A valid instrument can then be replicated for use independent of the subject hence measurement taken thereof is therefore a reliable data for meaningful analysis and examination to generate useful information [4]. This information is of utmost importance to be the prime ingredient in a particular decision making.

3 Measurement Method

Responses from the students in an examination, test or quizzes is normally marked against a marking scheme comprising keywords; where when there is a match then the student would be given a mark or otherwise. This is truly counting the number of correct answers which is then added up to give a total raw score. The raw score is only giving a ranking order which is known as Bernoulli random variable; in our case a smart student or otherwise. This is truly counting the number of correct answers which is then added up to give a total raw score. The raw score is only giving a ranking order which is known as Bernoulli random variable; in our case a smart student or otherwise.

Rasch Measurement Model is expressed as the ratio of an event being successful as;

$$P(\theta) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$  

Equ.(1)

where;

- $e = \text{base of natural logarithm or Euler’s number; 2.7183}$
- $\beta_n = \text{person’s ability}$
- $\delta_i = \text{item or task difficulty}$

Rasch exponential expression is a function of Logistic Regression which resulted in a Sigmoidal ogive and can be transformed into simpler operation by reducing the indices by logarithm:

$$\ln[P(\theta)] = \ln \left[ \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \right]$$  

Eq. 2

Now $\ln[P(\theta)]$; as the probability of a successful event, $x=1$ is reduced to the expression in equation 6 and can be construed simply as the difference of person ability; $\beta_n$ and the item difficulty; $\delta_i$, which can be represented as;

$$\ln[P(\theta)] = \beta_n - \delta_i ; \quad \text{Equ.(3)}$$

The very reason why the need to transformed it to logit is primarily to obtain a linear interval scale. It can be readily shown mathematically that a series of numbers irrespective of based used is not equally spaced but distant apart exponentially as the number gets bigger while a log series maintain their equal separation; thus equal interval [6]. This equal separation is mathematically shown in Table 1. The difference between $\log_{10}5$ and $\log_{10}2$ is constant and remain of equal distant between $\log_{10}50$ and $\log_{10}500$.
\[ \log_{10} 20 \] which similarly hold true for \[ \log_e \]; thus the theorem can now be universally applied.

<table>
<thead>
<tr>
<th>Numerical series</th>
<th>( \log_{10} )</th>
<th>( \log_e )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.301</td>
<td>0.694</td>
</tr>
<tr>
<td>5</td>
<td>0.699</td>
<td>1.609</td>
</tr>
<tr>
<td>10</td>
<td>1.000</td>
<td>2.303</td>
</tr>
<tr>
<td>20</td>
<td>1.302</td>
<td>2.997</td>
</tr>
<tr>
<td>50</td>
<td>1.699</td>
<td>3.912</td>
</tr>
<tr>
<td>100</td>
<td>2.000</td>
<td>4.606</td>
</tr>
</tbody>
</table>

This enable us to construct a log-odd ruler of probability an event taking place with the odd-of success as shown in Figure. 2 with unit termed as logit, derived from the term ‘log-odd unit’; as unit of measurement of ability akin to meter to measure length or kilogram to weight.

In order to achieve an equal interval scale, we can introduce logarithm of the odd probabilistic value. Maintaining the same odd probabilistic ruler as in Figure 2, starting with 0.01 to 100, we can create an equal interval separation between the log odds units on the line, hence the measurement ruler with the logit unit [8]. This can be verified by computing the value of \( \log_{10} 0.01 \) \((10^{-2})\) equals to \(-2.0\); value of \( \log_{10} 0.1 \) equals to \(-1\); value of \( \log_{10} 1 \) equals to 0 and so forth. Figure 3 shows the newly established logit ruler as a linear scale with equal interval separation. It is just like looking at a thermometer with ‘0’, as water being ice and 100 as boiling point whilst the negative extreme end as \(-273^\circ\text{C}\), the point where all atoms of any element come to a standstill.

Before delving any further, it is best to look at the analysis Summary Statistics as in Table 2. The prime information we are looking for in this table is the overall students’ LO ability reflected by the Person Measure \( \mu_{\text{PERSON}} \). SD=0.48 shows that the students is very much within target though we noted that poor students are LM171, LM242 with
PI159 measured -1.77 logits being lowest whilst the best students are PC134, PC129, LC186 and LI152 measured at +1.62 logits being topmost.

Both student’s ability measurements can give us some indication where are the students are on the probability scale; -ve means they are to the left of the ‘thermometer’, +ve means they are located to the right of the scale. Now we can sense and have a better appreciation if the students are in trouble or not since now their performance is duly measured on sound metrology principles thus generasibility.

The PIDM Map is the heart of Rasch analysis [9]. The vertical dashed line represents the ideal lessytoybest continuum of quality. Items and students now share the same linear measurement units known as logits. On the left hand side of the dashed line, the items are aligned from too easy to too hard, starting from the bottom. The distribution of student positions is on the right side of the vertical dashed line in increasing order of ability; the best naturally being at the top and the poorest student is at the bottom of the rung. Letter “M” denotes the student and item mean, “S” is one standard deviation away from the mean and “T” marks two standard deviations away from the mean.

In Rasch Model, since we are interested in the person’s ability for a given task, it is most prudent to zero set the scale where the item mean is zero when the ability is deemed 50:50 being the tipping point.

There are 127 Males against 116 Females students in this study. Male students shows a mean of, \( \mu_M = -0.02 \) logits where they were found to performed slightly better than their Female counterparts with a lower mean of, \( \mu_F = 0.04 \) logits. Generally, the students separation, \( G = 0.66 \) is such a small value that indicates that there is not enough differentiation among students ability to separate them into distinct performance level or strata.

Table 3 shows the pattern of responses. The structure calibration; ‘s’ is assessed to confirm the rating classification used is true where s-value being the separation between each structure category label;

- e.g; \( s_{2,3} = -0.88 - 0.27 = -0.61; < 1.4, \text{ Not OK.} \)
- \( s_{3,4} = 0.57 - (-0.88) = 1.45; < 1.4, \text{ OK.} \)

The separation shall be in the range where 1.4 < s < 5. It is noted in Figure 5 that the difference for each category are irregular where the difference between category 2, 3 and 4, 5 are all less than 1.4. It can be seen that classification 2 is well submerged and 4 just below 3and 5. Therefore, the ability classification A,5 > 90; B, 4 > 80; C, 3 > 70; D, 2 > 60 and Fail,1 < 60 is not reflective of this cohort person separation. Hence, we need to re-classify the rating and, in Rasch this is termed as collapsing. On the other hand if s > 5, then the category cluster need to be split instead. Subsequent to the re-scoring, if it is found that the SD is larger, then the new re-scoring will taken as the better measurement. Otherwise, it shall be retained.

This pattern of dichotomous response gave rise to concern of students lacking partial knowledge. In engineering, prudent engineers must possess some partial knowledge particularly in design concept or engineering philosophies. A mechanical mindset alone does not suffice and they need a more rigorous program to change their mindset to be ‘ingenious’.
Rasch has a unique ability in recognizing the students development based on the students responses. Table 4 shows the Person Measure Fit Order. This table gives an indication the validity of the person responses whether it fits the model; i.e. the fundamental of Rasch Model.

Table 4: Person Fit Order

<table>
<thead>
<tr>
<th>ENTRY</th>
<th>TOTAL</th>
<th>COUNT</th>
<th>MEASURE</th>
<th>MODEL</th>
<th>INFIT</th>
<th>OUTFIT</th>
<th>FIT-MEASURE</th>
</tr>
</thead>
<tbody>
<tr>
<td>93</td>
<td>24</td>
<td>9</td>
<td>-0.64</td>
<td>-1.02</td>
<td>1.70</td>
<td>4.76</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>20</td>
<td>9</td>
<td>-0.54</td>
<td>-2.02</td>
<td>4.14</td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>222</td>
<td>28</td>
<td>9</td>
<td>-0.15</td>
<td>1.42</td>
<td>2.14</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>204</td>
<td>27</td>
<td>9</td>
<td>-2.02</td>
<td>-1.75</td>
<td>1.72</td>
<td>3.28</td>
<td></td>
</tr>
<tr>
<td>208</td>
<td>24</td>
<td>9</td>
<td>-0.54</td>
<td>-1.24</td>
<td>1.32</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>32</td>
<td>9</td>
<td>-0.22</td>
<td>-2.52</td>
<td>2.42</td>
<td>0.29</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>27</td>
<td>9</td>
<td>-0.71</td>
<td>-3.29</td>
<td>2.53</td>
<td>0.51</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>25</td>
<td>9</td>
<td>-0.25</td>
<td>-2.60</td>
<td>2.93</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>31</td>
<td>9</td>
<td>-0.21</td>
<td>-1.60</td>
<td>2.13</td>
<td>0.60</td>
<td></td>
</tr>
<tr>
<td>207</td>
<td>25</td>
<td>9</td>
<td>-0.52</td>
<td>-1.24</td>
<td>1.32</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>187</td>
<td>25</td>
<td>9</td>
<td>-0.77</td>
<td>-2.76</td>
<td>2.16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>78</td>
<td>25</td>
<td>9</td>
<td>-0.22</td>
<td>-1.24</td>
<td>0.83</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>40</td>
<td>9</td>
<td>1.02</td>
<td>4.51</td>
<td>1.32</td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>55</td>
<td>9</td>
<td>-0.52</td>
<td>-1.24</td>
<td>1.32</td>
<td>0.64</td>
<td></td>
</tr>
<tr>
<td>189</td>
<td>25</td>
<td>9</td>
<td>-0.20</td>
<td>-1.24</td>
<td>0.83</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>27</td>
<td>9</td>
<td>-0.27</td>
<td>-1.24</td>
<td>0.83</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>18</td>
<td>9</td>
<td>-0.77</td>
<td>-2.76</td>
<td>2.16</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td>243</td>
<td>19</td>
<td>9</td>
<td>-1.40</td>
<td>-6.40</td>
<td>1.64</td>
<td>-0.00</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>26</td>
<td>9</td>
<td>-0.20</td>
<td>-1.24</td>
<td>0.83</td>
<td>0.66</td>
<td></td>
</tr>
</tbody>
</table>

Student PM097 attempted all the 9 domains but obtained only 24 out of possible 45 since each domain has a maximum rating 5. Coded demographically as PM; “Perempuan Melayu”, her ability is measured as -0.64 logit which means she is on the lower side of the scale.

Rasch examine item or person fit by looking at two types of fit values known as infit and outfit. Rasch typically examine ‘outfit’ which is less threatening to measurement and easier to manage. Hence, we look at “outfit MNSQ” where the mean square (MNSQ) outfit for the students is expected to be near 1.0. Acceptable MNSQ outfit shall be between 0.5 and 1.5.

Table 4 gives the Person misfit responses; the topmost being worst where the data provided are outfit to the model thus multi-dimensionality. PM097 shows MNSQ Outfit=7.00 which means a far item is not correctly assessed. Closer examination of the scagolam pattern response in Table 5 shows that item 8, 7 and 5 are somehow not well assessed and need to be reviewed. Those rating 1 could have been more appropriately be higher value. Similarly, item 4 should have been a 1 instead of 4. Rasch would ask the researcher to identify the reasoned argument ‘why’ does this happen.

However, the point measure correlation can give more interesting pattern of responses. Though the acceptable value is in the range of 0.38 to 0.85, perhaps we shall start to worry as it approaches near zero or make it worst negative value. It simply means the respondent is behaving the opposite way. Let us have a look at Table 5 Student LI153. He appeared to have scored a high 40 out of 45 from 9 items, but with a high MNSQ Outfit= +2.73 and a very low point measure correlation= -0.17, which is near zero. Analysing his pattern of response in Table
5 Scalogram confirms that he is peculiar. Take note he fai
red poorly on the easy items, on the left but
scored well on the difficult items on the right.
Despite his high score, Rasch identified his pattern
of responses is different that warrants justification
thus response validity.

Table 5: Scalogram of Unexpected Responses

<table>
<thead>
<tr>
<th>Person</th>
<th>Item</th>
<th>EASY</th>
<th>TOUGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>222</td>
<td>3</td>
<td>-50</td>
<td>+503</td>
</tr>
<tr>
<td>36</td>
<td>4</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>79</td>
<td>5</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>180</td>
<td>6</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>15</td>
<td>7</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>135</td>
<td>9</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>200</td>
<td>10</td>
<td>+05</td>
<td>-505</td>
</tr>
<tr>
<td>187</td>
<td>11</td>
<td>+05</td>
<td>-505</td>
</tr>
</tbody>
</table>

These outcomes do not meet Rasch model expected outcomes. This major finding raise some conclusions, for example, the student underestimated the easiest items hence careless errors. Conversely, for the difficult items, suspects probably have special interest or knowledge on the topic and/or comfort answering statement-based question. On the other hand, it makes sense that the student may simply guess the answers for the questions. Rasch has this particular predictive properties embedded in the model to make it a very reliable validation model.

5 Conclusion
Rasch Model provides a sound platform of measurement equivalent to natural science which matches the SI Unit measurement criteria where it behaves as an instrument of measurement with a defined unit and therefore replicable. It is also quantifiable since it’s linear. Rasch Model has made it very useful with its predictive feature to overcome missing data [3].

The logit ruler has been developed with purpose to measure ability; in this case students learning ability of specific learning outcomes. It can define the students profile and most important we are now able to validate a question construct on line. It is a noble innovation where the ability ‘ruler’ can transform ordinal data into measurable scale. It’s graphical output is great which gives better clarity for quick and easy decision making.

The measurement conducted reveals the true degree of cognitive learning abilities of the Engineering undergraduates based on Blooms Taxonomy [10]. Previously, lack of such measurement in Engineering Education has made the necessary corrective actions in the form of skills development, education and competency training difficult to formulate. This major problem faced by Engineering Education Administrators in an IHL to design the necessary curriculum to mitigate the going concern is therefore resolved. Rasch has all the capabilities to rigorously analyse examination results more accurately thus making evaluation clearer to read and easier to understand [11].

Acknowledgements
The authors wish to acknowledge the financial support received from the Centre for Engineering Education Research, University Kebangsaan Malaysia as research grant in the effort of improving the quality of teaching and learning in engineering education.

References:
Use of Rasch Analysis to Measure Students Performance in Engineering Education


Easier Learning Outcomes Analysis using Rasch Model in Engineering Education Research

SAIDFUDIN, M \(^1\), AZRILAH, A.A \(^2\), RODZO’AN, N. A \(^3\), OMAR, M.Z \(^4\), ZAHARIM, A \(^5\) AND BASRI, H \(^6\)

\(^1\) Program Director, \(^2\) P.hD, Program Coordinator, Exec.Dip. in Quality Management, UTM SPACE,
\(^3\) Master (Candidate), Dept. of Mathematics, Fac. of Science, University Teknologi Malaysia, 81300 Skudai, MALAYSIA
\(^4\) Assoc.Prof., \(^5\) Professor Centre for Engineering Education Research, Fac. of Engineering and Built Environment,
\(^6\) Professor, Ir; Deputy Vice Chancellor (A), Universiti Kebangsaan Malaysia, 43600 Bangi, MALAYSIA

Abstract: - It has been a predicament of mankind to comfortably live in the wrong since the famous flat-earthers theory and subsequent Ptolemic believe that the world is the centre of the universe. Despite such widely accepted highly scholastic wrong Greek philosophies, the world does not collapse. Similarly in Engineering Education (EE), we have been doing the traditional academic achievement reporting the classical way believing the number of A’s as a measurement of achievement and progress. Such practise is only an inference arising an observation made like the expansion of mercury due to heat energy obtained from the surrounding. However, we need to define the quantum of one degrees centigrade; 1\(^{\circ}\)C before a scale can be meaningful and become of good use. An overview of the measurement model and its key concepts are presented and its application illustrated using the final exam paper given through KKKF1134 – Introduction to Engineering. The students performance output were assessed based on Students Observed Learning Outcomes (SOLO) Taxonomy which gives an indication on the student achievement of the subject expected LO i.e. Students’ Profiling. The study shows that Rasch model of measurement can classify grades into learning outcomes more accurately especially in dealing with small number of sampling unit.

Keywords: - Learning Outcomes, instructional objectives, performance assessment, Quality, continuous improvement.

1 Introduction

Learning is a complex process. It entails not only what students know but what they can do with what they know; it involves not only knowledge and abilities but values, attitudes, and habits of mind that affect both academic success and performance beyond the classroom. Assessment should reflect these understandings by employing a diverse array of methods, including those that call for actual performance, using them over time so as to reveal change, growth, and increasing degrees of integration. Such an approach aims for a more complete and accurate picture of learning, and therefore firmer bases for improving our students' educational experience [1].

A good assessment recognizes the value of information for the process of improvement. Assessment approaches should produce evidence that relevant parties will find credible, suggestive, and applicable to decisions that need to be made.The point of assessment is not to gather data and return "results"; it is a process that starts with the questions of decision-maker that involves them in the data gathering and subsequent analysis;
1. How do you assure the correct instrument is used for purpose ? and subsequently;
2. What is the correct method of such data analysis ?

It is of utmost importance on the onset this fundamentals of measurement must be correct. Analysis must be based on valid data and duly interpreted to generate a reliable report with meaningful information for prudent decision making towards continuous improvement of teaching and learning. In an earlier paper, it was shown how academic reporting using Rasch Analysis proved to be more meaningful and make students classification hence better management to improve their achievement in meeting the targeted learning outcomes [2].
2 Overview of Data Types
Fundamentally there are two types of data; quantitative and qualitative type [3]. It was generally perceived as countable and non-countable. Total marks of a student obtained in an exam gave a rank order but the distant between the next student ability having lower or higher marks is never the same. In reality, crudely speaking we are only counting the number of correct answers. However, it has been grossly misunderstood and treated like a quantitative data which is somehow blatantly added and subtracted and even multiplied or divided.

Modern measurement method as practiced using item response theory with a focus on Rasch measurement model now provides the social sciences with the kind of measurement that characterizes measurement in the natural sciences i.e. the field of metrology [4]. The fundamentals of measurement calls for an instrument to be used for purpose to have specific unit of an agreed standard amount [5]. An instrument must have the correct construct of linear scale which can be zero set and duly calibrated. A valid instrument can then be replicated for use independent of the subject hence measurement taken thereof is therefore a reliable data for meaningful analysis and examination to generate useful information. This information is of utmost importance to be the prime ingredient in a particular decision making.

3 Measurement Method
Responses from the students in an examination, test or quizzes is normally marked against a marking scheme comprising keywords; where when there is a match then the student would be given a mark or otherwise. This is the traditional ‘park and mark system’. In theory, at this stage truly the assessors is only counting the number of correct answers which is then added up to give a total raw score. The raw score only give a ranking order which is deemed an ordinal scale that is continuum in nature [6]. It is not linear and do not have equal intervals which contradicts the nature of data fit for the due statistical analysis. It does not meet the fundamentals of sufficient statistics for evaluation [7].

Rasch focuses on constructing the measurement instrument with accuracy rather than fitting the data to suit a measurement model with of errors. By focusing on the reproducibility of the latent trait measurement; in this case the students’ LO instead of forcing the expected generation of the same raw score, i.e. the common expectation on repeatability of results being a reliable test, hence the concept of reliability takes its rightful place in supporting validity rather than being in contentions. Therefore; measuring LO ability in an appropriate way is vital to ensure valid quality information can be generated for meaningful use; by absorbing the error and representing a more accurate prediction based on Rasch probabilistic model [8].

An attempt of a student to answer a question can be seen as a chance of him being able to get the correct answer or successfully accomplishing a given task. Now, for a given normal score of 7/10 which is normally read as 70%; there is need of a paradigm shift to read it as the odds of success being 70:30; thus a ratio data. A mark of 6/10 shall now be seen as odd of success 60:40 and, so on. After all percentage is statistically recognized only a data summary; which is somehow largely confused as a unit of measurement.

This enable us to construct a log-odd ruler of probability an event taking place with the odd-of success as shown in Figure. 1 with unit termed as logit, derived from the term ‘log-odd unit’; as unit of measurement of ability akin to meter to measure length or kilogram to weight.

In order to achieve an equal interval scale, we can introduce logarithm of the odd probabilistic value. Maintaining the same odd probabilistic ruler as in Figure 1, starting with 0.01 to 100, we can create an equal interval separation between the log odds units on the line, hence the measurement ruler with the logit unit. This can be verified by computing the value of \( \log_{10} 0.01 \) \((10^{-2})\) equals to -2.0; value of \( \log_{10} 0.1 \) equals to -1; value of \( \log_{10} 1 \) equals to 0 and so forth. Figure 2 shows the newly established logit ruler as a linear scale with equal interval separation. It is just like looking at a thermometer with ‘0’, as water being ice and 100 as boiling point whilst the negative extreme end as \(-273^\circ C\), the point where all atoms of any element come to a standstill.
Thus, we now have a valid construct of an instrument to measure the students ability for each defined LO.

4 Results and Discussion
The test was administered on 1st year Engineering and Architecture students from the Faculty of Engineering and Built Environment, University Kebangsaan Malaysia (UKM) for the course code KKKF1134 – Introduction to Engineering. The result from the tests were assessed based on SOLO Taxonomy [9] and ran in Winsteps v 3.6.8, a Rasch analysis software; to obtain the logit values. Figure 3 shows the Person-Item Distribution Map (PIDM) where the persons; i.e., the Students is on the left whilst the items; the learning topics were plotted on the right side of the logit ruler as in Figure 3. By virtue of the same ruler with the same scale; then the correlation of the person, $\beta_n$ and item, $\delta_i$ can now be established. In Rasch, the probability of success can be estimated for the maximum likelihood of an event as;

$$P(\theta) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}}$$  \hspace{1cm} \text{Equ.}(1)$$

where;

- $\theta$ = base of natural logarithm or Euler’s number; 2.7183
- $\beta_n$ = person’s ability
- $\delta_i$ = item or task difficulty

The PIDM Map as in Figure 3 is the heart of Rasch analysis. On the right hand side of the dashed line, the items are aligned from easy to difficult, starting from the bottom. The distribution of student positions is on the left side of the vertical dashed line in increasing order of ability; the best naturally being at the top and the poorest student is at the bottom of the rung. In Rasch Model, since we are interested in the person’s ability for a given task, it is most prudent to zero set the scale where the item mean is zero when the ability is deemed 50:50 being the tipping point. Rasch analysis tabulates the item’s location in a very clear graphical presentation which is easy to read and easier to understand. Each item can be coded with attributes of Bloom’s Taxonomy that is assessed affecting the students learning process [10]. This will enable in depth analysis of their study pattern to be evaluated meaningfully.

Before delving any further, it is best to look at the analysis Summary Statistics as in Table 1. The prime information we are looking for in this table is the validity of this assessment. The value of Cronbach-\( \alpha \) = 0.33 is disturbingly low which is well below the acceptable level 0.6 and, in normal statistical analysis this test evaluation would have been disregarded. However, Rasch analysis offer a better evaluation where it shows the two components of the test; the Person and the instrument, i.e. item reliability. Rasch found the Person Reliability rather low at 0.31 and a very high Item Reliability of 0.99. This conclude that the students need further scrutiny and yet we can proceed with the analysis as the instrument has a very high reliability in measuring what is supposedly to be measured. This is where Rasch has the major strength as the better model is making measurement [11].

<table>
<thead>
<tr>
<th>Table 1: Summary Statistics</th>
</tr>
</thead>
</table>

Fig. 3: Person-Item Distribution Map: Item Location
The Summary of 8 measured items gave a measurement of Maximum item = +1.82 logit and minimum item = -1.71 logit. One item is identified to be classified as minimum extreme score. Close study revealed in Table 2 - Item Measure shows the item to be JKKP = -7.42 logit.

### Table 2: Item Measure

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>JKKP</td>
<td>-7.42</td>
</tr>
</tbody>
</table>

Item measures gave the indication on the level of difficulty the students encountered in attempting a given task. Now we can sense and have a better appreciation if the students have trouble or not since now their performance is duly measured on sound metrology principles hence JKAS is the most difficult task whilst JKEES is the easiest. JKKP point measure correlation = 0.00 with extreme measure, a match of 100% means the item cannot discriminate between a good and a poor student.

Generally, the item separation, G=11.67 is a big value which indicates that there is a very good differentiation of item difficulty to separate the students into distinct difficulty levels. So, if sample separation is 2, then strata are (4*2+1)/3 = 3, means;

Separation= 2: The test is able to statistically distinguish between high and low performers.

Strata= 3: The test is able to statistically distinguish between very high, middle and very low performers.

Thus, a student separation G=11.67 was computed into the strata formula which yielded a distinct 15.89 strata. This indicate there is a large separation between a very easy question and a very difficult question. This call for a review of the assessment done to close the gap.

Rasch has a unique ability in recognizing the students development based on the students responses. Table 3 shows the Item Misfit Order. This table gives an indication the validity of the person responses whether it fits the model or not. The topmost being worst where the data provided are outfit to the model thus multi-dimensionality.

### Table 3: Item Misfit Order

<table>
<thead>
<tr>
<th>Item</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>JKKP</td>
<td>-7.42</td>
</tr>
</tbody>
</table>

Rasch examine item or person fit by looking at two types of fit values known as infit and outfit. Rasch typically examine ‘outfit’ which is less threatening to measurement and easier to manage. Hence, we look at “outfit MNSQ” where the mean square (MNSQ) outfit for the item is expected to be near 1.0. Acceptable MNSQ outfit shall be between 0.5 and 1.5.

Closer examination of the scalogram pattern response in Table 4 shows that Person 158 onwards to the end are somehow not well assessed and need review. Those rating 1 could have been more appropriately be higher value.

### Table 4: Scalogram of Item Misfit Response String

Further analysis of the expected value is shown in Table 5 – Item Most unexpected Response Prediction. The item Z-STD = 3.80 is beyond the upper limit +2.00. It can be generalized that the item has been under rated and Rasch would ask the researcher to identify the reasoned argument ‘why’ does this happen. One possible conclusion is that these cohort could have been careless in attempting
their works which lead to such a grossly under rated works.

Table 5: Item Most Unexpected Response Prediction

<table>
<thead>
<tr>
<th>Item</th>
<th>Observation</th>
<th>Expected</th>
<th>Item</th>
<th>Person</th>
<th>Item</th>
<th>Observation</th>
<th>Expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.29</td>
<td>1.61</td>
<td>1.70</td>
<td>-2.21</td>
<td>1</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>2</td>
<td>1.45</td>
<td>2.00</td>
<td>2.00</td>
<td>-1.69</td>
<td>2</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>3</td>
<td>1.28</td>
<td>1.78</td>
<td>1.70</td>
<td>-1.77</td>
<td>3</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>4</td>
<td>1.35</td>
<td>1.05</td>
<td>2.04</td>
<td>-1.74</td>
<td>4</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>5</td>
<td>1.35</td>
<td>1.05</td>
<td>2.04</td>
<td>-1.74</td>
<td>5</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>6</td>
<td>1.25</td>
<td>1.62</td>
<td>2.04</td>
<td>-1.74</td>
<td>6</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>7</td>
<td>1.35</td>
<td>2.05</td>
<td>2.04</td>
<td>-1.74</td>
<td>7</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>8</td>
<td>1.35</td>
<td>2.05</td>
<td>2.04</td>
<td>-1.74</td>
<td>8</td>
<td>1.70</td>
<td>1.61</td>
</tr>
<tr>
<td>9</td>
<td>1.60</td>
<td>1.95</td>
<td>2.04</td>
<td>-1.74</td>
<td>9</td>
<td>1.70</td>
<td>1.61</td>
</tr>
</tbody>
</table>

Table 5 shows a very interesting finding where the most difficult item; Item 4 -JKAS was found to be the reversed where it is observed to be generally over rated with quite low point measure correlation of 0.28. Conversely, for this difficult item suspects could probably have special interest or knowledge on the topic. On the other hand, they could have a very kind hearted assessor who is gave away marks rather easily. Rasch has this particular predictive properties embedded in the model to make it a very reliable validation model.

5 Conclusion
Rasch Model provides a sound platform of measurement equivalent to natural science which matches the SI Unit measurement criteria where it behaves as an instrument of measurement with a defined unit and therefore replicable. It is also quantifiable since it’s linear. Rasch Model has made it very useful with its predictive feature to overcome missing data [12].

The logit ruler has been developed with purpose to measure ability; in this case students learning ability of specific learning outcomes. It can define the students profile and most important we are now able to validate a question construct on line. It is a noble innovation where the ability ‘ruler’ can transform ordinal data into measurable scale. It’s graphical output is great which gives better clarity for quick and easy decision making [13].

The measurement conducted reveals the true degree of cognitive learning abilities of the Engineering undergraduates[14]. Previously, lack of such measurement in Engineering Education has made the necessary corrective actions in the form of skills development, education and competency training difficult to formulate. This major problem faced by Engineering Education Administrators in an IHL to design the necessary curriculum to mitigate the going concern is therefore resolved. Rasch has all the capabilities to rigorously analyse examination results more accurately thus making evaluation clearer to read and easier to understand. This method of reporting was found to consistent with research done in other countries treating ordinal data the correctly by application of Rasch Analysis to obtain a more meaningful information of the item validity hence prudent LO measurement [15].

Acknowledgements
The authors wish to acknowledge the financial support received from the Centre for Engineering Education Research, University Kebangsaan Malaysia as research grant in the effort of improving the quality of teaching and learning in engineering education.

References:
Easier Learning Outcomes Analysis using Rasch Model in Engineering Education Research


Exploring User’s Perception toward Automated Checkout Trolley in Developing Countries

KOH PEIK SEE¹, TAN SHIANG-YEN¹, NASUHA LEE ABDULLAH¹, ROSNAH IDRUS²,³
¹School of Computer Sciences, ²National Advanced IPv6 Center, ³CEMACS
Universiti Sains Malaysia
MALAYSIA
sherrykoh85@hotmail.com, tsey09_com036@student.usm, nasuha@cs.usm.my, rosnah@cs.usm.my

Abstract: - Retailers nowadays have relied on technology-based self-service options such as the self-checkout system to provide better customer value through pleasant shopping experience. Recently, automated checkout shopping trolley technologies have taken a further step than the previous introduced self-checkout system in terms of greater convenience, ease-of-use and greater efficiency. As this technology is considerable new in the market of developing countries, hence it is vital to explore the perception of consumers toward the use of automated checkout trolley. The purpose of this paper is to understand the perceptions and expectations of both enterprise users and end users toward the new checkout system. Findings of this study indicate that both consumers and retailers from developing countries possess positive attitude toward the automated checkout trolley system, and are enthusiastic to try this emerging technology. The main concerns of the users are the secureness of payment and the variability of payment methods, which forms their core criteria in determining whether to adopt the system. The feedbacks from them are valuable in providing better automated checkout system and facilitate adoption rate of the technology.

Key-Words: - Technology-based self-service (TBSS), Automated checkout, Technology management

1 Introduction

Technology-based self-services (TBSS) have gradually repositioning as one of the important mechanisms to deliver superior customer value, especially in retailing sector [1]. Mass retailers such as hypermarkets and department stores are the pioneers and main adopters of self-service technologies (SST). In this emerging trend of technologically-oriented service, customers utilize self-service technologies in order to provide service for themselves with or without help from a retail contact employee [2]. Studies postulate that self-service technologies create customer value by helping customer checkout more quickly, providing them a simple convenience and usually leading to higher perceived service quality [1, 3, 4]. One of the most of common self-service technologies in retailing sector is the self-checkout system.

 Provision of the technology-based self-service options such as self-checkout system entails advantages to both the enterprise users (retailers) and end users (shoppers). Benefits from the enterprise users mainly derive from cost efficiency resulted from reduced staff requirements, indirectly reduce personal training costs and other variable costs which proportionate to the number of staffs employed [5]. On the other hand, shoppers are benefited from the reduced checkout time, better service, perceived privacy and anonymity [4]. Despite the advantages offered, researches indicate that adoption of self-service technologies not necessary contribute to better customer value [6, 7]. Instead, customer dissatisfaction may occur due to frustration in using the SST. In fact, in order to reap the time efficiency benefit of using the self-service checkout system, shoppers required to able to use the system reasonably competent. An inexperienced customer can cause even more delays than an inexperienced cashier on a conventional register, and older customers may expect the attendant to assist them directly with scanning items, preventing the attendant from dealing with other customers who actually require intervention [7].

Recently, the introduction of automated checkout trolley indicates the potentiality to alleviate the shortfalls of existing self-checkout system. The differences between two, are the self-checkout system commonly exists as a stationary kiosk which located in close proximity to cashier-driven cash register in a retail setting, and there is often a head cashier nearby who oversees user interaction with the self-checkout system. Whereas the automated checkout trolley, by its name, it is a trolley features
a built-in computer which allow users or shoppers to scan the items that being put into the trolley and to check out the trolley of items by simply passing through the sensor device.

As the automated checkout trolley is a variation of self-service technology, it may inherit the common problem of SST that the end user’s adoption rate is lower than enterprise user’s adoption rate [1]. In order words, enterprise users (retailers) is comparatively more aggressive in adopting the technology, but the end users (shoppers) is less willing to adopt the technology, which eventually lead to the failure of the self-service technology given the fact the successfullness of the technology will ultimately depend on the acceptance of end users. Hence, this implies that perceptions and expectations of both end user and enterprise user toward the automated checkout trolley are vital inputs to the development and enhancement of the automated checkout trolley. In addition, the main streams of technology-based self-service researches have focused on the user behaviors and assessment of acceptance of the technology through service quality measurement. Scarce research has studied the system requirements from practical level which able to provide sufficient details of system requirements for the use of practitioners. Moreover, main stream of TBSS studies usually are done in Western context, application of TBSS in Asian context are less well understood. Their perceptions and expectation constitute the requirements of the technology to become success, especially in Asian contingent.

Therefore, the purpose of this paper is to assess the perceptions and expectation of both end users and enterprise users toward automated checkout trolley system. The understanding of the user requirements will able to provide better solution to the existing model and enhance the adoption rate by end users.

2 Automated Checkout Trolley
2.1 System Overview
Automated checkout shopping trolley (ACT) system presented in this paper refers to the whole system working together to make the “trolley” or “cart” capable to conduct automatic checkout. In other words, the under discuss trolley system not only refers to the tangible “trolley”, but also the database system, networks, technologies used to identify sale items, sensors and the software as well.

In general, the most tangible module of automated checkout trolley system is trolley itself that attached with a built-in computer and sensor, which connected with the database system. In this environment, sale items are often tagged with automatic identifier such as Radio Frequency Identification (RFID) tag. The trolley automatically scans the items which the items are put into the trolley. The ultimate goal of the ACT system lies on its capability to check out the trolley of items at once, by pushing the trolley through the sensors. At the same time, the total amount of item purchased will be deducted from the user’s credit card or cash card.

The automated checkout trolley system puts service and checkout in the consumer’s hands, reducing reliance on the point-of-sale for customer service and freeing store personnel to provide customer service in the aisle.

2.2 Developed Conceptual Prototype of ACT
In order to enable evaluation by users, the research team has produced a conceptual prototype of the automated checkout trolley system, named as S-Cart with details of its feature and functionalities. The proposed S-Cart has a mounted computer and sensor device on near the end of trolley which facing the user. User can interact with the system through the touch-sensitive monitor or navigational buttons. Users can access to various information using the interactive display. The following list shows the type of information provided by the S-Cart.

- Promotion and advertisement information
- Product details (e.g. contents, ingredients, tips)
- Check location of particular merchandise
- Check price of item
- Sum up prices of all items in the trolley
- Live chat function with supporting staffs

In order to use the S-Cart, users need to insert their credit card or cash card (a member card issued by the retailers, which users can reload the card with money at the customer service counter) into the card slot. The mounted computer on the trolley is connected to the retailer’s database system and credit card verification system through WiFi 802.11 network. Data of Credit card and cash card are sent over the network to be verified. Other information aforementioned is transmitted to the trolley using the WiFi network as well.

When the merchandises are put into the cart, the RFID sensor will update the list of items added into the cart and display the updated total price of items. The S-Cart adopts passive RFID tags as the automatic identification technique to identify merchandises in the trolley. The main reason of
adopting passive RFID tag is the cost efficiency and energy efficiency provided by the method.

After the checkout process, customers are required to push the S-Cart back to the trolley collection area, in order to remove their credit card or cash card from the S-Cart. The S-Cart will produce alert message and sound to remind customer to retrieve their card from the S-Cart.

3 Methodology
The purpose of this study is to explore the perceptions and expectation of Asian users toward the S-Cart system. Both enterprise users (retailers) and end users (consumers) are included in the study. In the first part, this study adopts questionnaire survey to identify the existing service problems faced by consumers prior to S-Cart adoption and consumer’s acceptances toward the S-Cart. Second part of the study involves the enterprise users (employees of retailers) in in-depth interviewer, in order to understand their perceptions toward the idea of S-Cart.

3.1 Questionnaire Survey for End-Users
In this part, the unit of analysis was the individual shopper. Questionnaires are distributed to individual shopper in hypermarkets. The research team collected back 86 sets usable questionnaire. The questionnaire can be divided into three session, which first session examine the perceived service quality of hypermarkets, the second session examine the acceptance of users toward the S-Cart and the payment method in S-Cart, whereas third session contained open-ended questions which attempt to solicit subjective perception of user toward to idea of S-Cart and the features they would like to S-Cart to have. Both sessions one and two are in the form of five point Likert Scales, which coded from strongly agree with value of 5 and strongly disagree with value of 1. The purpose of first session of the survey is to identify the service problems in the hypermarkets, thus new feature of S-Cart can be proposed to confront the problems. It attempts to examine the perceived service quality, by the factors from “The Service Quality Model” and “Evaluation of Customer Satisfaction”. Table 1 show the factors and aspect which the factor attempt to measure. The details of items used to measure each aspect aforementioned are as shown in Table 2. It is an administrated survey so that researchers able to explain the idea of S-Cart to respondents in better pictures, and to solicit further response whenever it is necessary.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Aspects</th>
<th>Service Quality Model</th>
<th>Customer Satisfaction Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comfortable</td>
<td>Storehouse environment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Consistency of price tag</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Service Quality</td>
<td>Service Quality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Ease of Access</td>
<td>Efficiency and Timeliness</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Accountability</td>
<td>Service Quality</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2 Items used in Questionnaire

Questions

Storehouse Environment
- Store atmosphere and décor are appealing.
- A good selection of products was presents.
- Merchandise displays are attractive.
- Advertised merchandise was in stock.
- The store layout makes it easy for customers to find what they need.

Consistency of Price Tag
- Price tags are consistent with the cashier's scan.
- Prices are known by cashier in order to avoid arguments.

Service Quality
- Promoters are in full support and helpful.
- Sincerity in solving customer’s problem
- Employees in this store are able to handle customer complaints directly and immediately.

Efficiency and Timeliness
- Payment process is fast and efficient (< 5 Minutes)
- Sufficient counters are available
- Staffs able to handle unexpected case effectively and avoid customer waiting

3.2 In-Depth Interview with Enterprise Users
Three interviewees of a selected case company, who holding different positions (ranged from front end manager to customer service staff) are interviewed. The result of interview is recorded and transcripted. Research team first informs the interviewee regarding the result of survey – consumer’s perception toward the S-Cart and the perceived problem with existing service. The purpose of the interviews is to understand the perceptions of enterprise users toward to idea of S-Cart, and their
concerns or expectations on the S-Cart system. The general questions asked are:

a. What is your comment on the S-Cart?
b. What is your concern regarding the S-Cart?
c. What functionality or features which you would like to have on the S-Cart?
d. Do you think the S-Cart system is applicable in the hypermarkets?

4 Result

Customer satisfaction is the vital contributor to competitive advantage. Thus, a new creation or development of system to be used by customer must be able to meet their requirements and help them to solve existing problem. Otherwise the system is developed only for the sake of development. The results of this study are organized into two parts, which first part is the result of questionnaire survey whereas the second part is the result of in-depth interviews with the representatives of the hypermarkets.

4.1 Result of Questionnaire Survey by Shoppers

The 86 samples constitute of 49 female shoppers and 37 male shoppers, which as illustrated in Fig. 2.

Table 3. Descriptive Statistics

<table>
<thead>
<tr>
<th>Factors</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storehouse environment</td>
<td>1.83</td>
</tr>
<tr>
<td>Price Tag Accuracy</td>
<td>2.70</td>
</tr>
<tr>
<td>Service Quality</td>
<td>2.81</td>
</tr>
<tr>
<td>Efficiency and Timeliness</td>
<td>3.10</td>
</tr>
</tbody>
</table>

The finding indicates that consumers or shoppers are satisfied with the storehouse environment of the hypermarkets, but dissatisfied with the other three aspects which are price tag accuracy, service quality, and efficiency and timeliness, with increasing dissatisfactory from tag accuracy to efficiency timeliness. These results indicate that the hypermarkets can improve customer satisfaction by solving the problems of efficiency and timeliness, service quality, and price tag accuracy.

In addition, the finding of survey indicates that approximately 70 percent of respondents agreed that they overspent when shopping in the hypermarkets, as shown in Fig. 3.

This implies that behavior of overspending is prevalent. Hence a mechanism which is able to help shoppers to control their spending is much appreciated. In other words, the budget control and alert feature proposed in the S-Cart system is a practical and appropriate system requirement.

Session two of the survey examine the acceptance of shoppers toward the use of S-Cart and the payment methods used in the S-Cart, which the credit card and cash card methods. As illustrated in Fig. 4, about 88 percent of respondents are willing to use S-Cart or automated checkout trolley system when shopping in hypermarkets. This finding indicates that majority of the shoppers in hypermarkets are willing to embrace the S-Cart or Automated Checkout Trolley (ACT) system, which also implies of the practicality of implementing this system in the hypermarkets of developing countries.

On the other hand, 59 percent of the respondents not accept the credit card and cash card as payment methods used in the S-Cart, as shown in Fig. 5.
With the complementary qualitative response acquired, shoppers responded that they feel unsecure and worry to use credit card or cash card as the payment method used in S-Cart. The use of credit card as payment method is much concerned by respondents as the risks and amount of money involved are considerable high. This finding implies that security is single most important system requirement of S-Cart system. Mechanisms which allow to users to verify the transaction, password encryption of card information, and secured transmission of transaction information should be implemented in order to increase the acceptance level of the payment method. In addition, alternative payment methods must be provided, which able to cater for the needs of customer who not own credit card and avoid customer frustration by forcing customer to use certain payment methods.

Session three of the survey consists of 2 open-ended questions to solicit subjective comments of respondents toward the S-Cart and the features that they would like to S-Cart to have. The advantages of open-ended question are that it allows respondents to provide creative, unconstrained feedbacks about the S-Cart. The following are the comments from respondents.

- Respondents are impressed with the convenience and efficiency which realizable by the S-Cart
- Multi language supports should be provided
- Respondents wish to have item-location associated floor directory function to help them navigate effectively
- Respondents are pleasant with the budget control and alert functions.
- Respondents suggest use cash as alternative payment method.
- Respondents concern about the security of the system since credit card or cash card is used, and amount is deducted automatically from the account. The risk of forgetting the card in S-Cart is perceived as high.
- Password verification should be used to initial the card access, and also to verify the transactions.
- The robustness of S-Cart is concerned as well as customers often put their child inside the cart during shopping.
- Respondents request for adding “Touch and Go” card as one the payment method. “Touch and Go” card is a multi-purpose cash card in Malaysia that can be used in public transportations and road tolls.

Based on the qualitative feedbacks from respondents, it indicates that most shoppers responded that they are excited about the idea and enthusiastic to try the S-Cart, provided if it is use-to-use, convenience, and secure. As reconfirmed by finding from session 2 of the survey, system security is the main concern of shoppers, which form the vital system requirements of the S-Cart system in order to be succeeded. The second most prevalent comments from the respondents are regarding the payment methods, they request that the various payment methods such as cash should be available to make the S-Cart usable by wider range of users.

4.2 Result of Interviews with Enterprise Users

Enterprise users of the S-Cart system are referred to the staffs of the retailers or hypermarkets. Their perceptions and expectations toward the S-Cart system are paramount as they potentially influence the decision of whether adopting the automated checkout trolley system or not. As aforementioned, staffs ranged from frontline manager to customer service staffs are interviewed.

Overall, the interviewers possess of positive attitude toward the ideas of S-Cart. They believe the use of S-Cart will significantly improve the checkout process, reduce personnel-related expenditures, improve the customer service, and increase customer satisfaction. At the same time, they have valuable comments as well as concerns on the S-Cart system, which are the main concerns that sought by this paper. The following are the comments or concerns of the interviewees.
• Cash payment must be available
• How the system handle those merchandises which cannot be scanned
• How can the system handle items (such as vegetables and grocery) that need to be weighted?
• How the system handle merchandises which must be paid in internal counter (such as medicines and cosmetics).
• Is the system able to prevent fraud in transaction, especially for credit payment?
• Will the S-Cart system prevent shoplifting techniques
• We are concerning about the robustness of the trolley as vandalisms are quite pervasive.
• The cost of the trolley is high, and we cannot afford the trolley to be stolen.
• Since the S-Cart use RFID tags on the merchandise, it will increase our costs.

The feedbacks from the representatives of the hypermarkets help this study to derive at the system requirements of the S-Cart system from the perspective of enterprise users. Most of the concerns are related to the system security and operations. To the retailers, the security and integrity of the S-Cart to verify credit card or transaction are most important as the compromises of these features are disastrous to them.

Furthermore, the operability of the S-Cart in handling routine operations or practices of the hypermarkets is vital as well. For instance, the system must be capable to handle items that paid in internal counter, to avoid redundant checkout and to avoid vulnerable to shoplifting. Additionally, RFID tags may not appropriate for all the merchandises (e.g. when the RFID price tag is more expensive than the price of merchandise or merchandise that need to be weighted). Thus, the S-Cart must able to use other tagging method or identification techniques to scan the items. Interviewees suggested alternative payments methods, especially cash are, as similar to the comments by the shoppers.

5. S-Cart solve Retailers’ Problem

Through the questionnaire and interviews, the following are the common problems identified in the hypermarkets which agreed by both customers and the retailers. This section presents how the adoption of S-Cart able to solve the various service-related problems. Solving these problems implies improved service quality, customer value, and eventually leads to customer satisfaction. Table 4 shows the solution provided by S-Cart to service-related problems.

<table>
<thead>
<tr>
<th>Problems</th>
<th>Solution offered by S-Cart System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inconsistency of price tags</td>
<td>All prices are directly retrieved from central database. No longer need to physically change the price tags, retailers can change the price and other information from the database system.</td>
</tr>
<tr>
<td>Limited price checker device</td>
<td>S-Cart itself has price checker function. Users just need to drop an item into trolley and it would display price, as well as other product information such as expire date, content, ingredients, and tips.</td>
</tr>
<tr>
<td>Trolley stolen</td>
<td>S-Cart as a solution to prevent shopping cart lost, as users need to push the cart back to its dedicated storage area in regain their credit card in the S-Cart.</td>
</tr>
<tr>
<td>Long Queuing Time</td>
<td>S-Cart is an automated checkout system that significantly fastens the checkout process. It helps to reduce the loads of cashier counters, while maintaining the service level.</td>
</tr>
</tbody>
</table>

Additionally, as the S-Cart utilizes RFID tag, it make possible for the hypermarkets to create a vast database of customer information [8]. Later the retailers can extract business intelligence from the knowledge base, which help the retailers delivery product and service based on customers’ needs, preferences, or past transaction [9].

6. Discussion

The findings from this study indicate that respondents possess of positive attitude toward the S-Cart system and most of them are enthusiastic to try the automated checkout trolley system. On the retailers’ side, they are willing to adopt this automated checkout system, given its advantage to reduce operational and personnel costs. These results suggest that the shoppers and hypermarkets in developing countries have considerable high level of acceptance toward this evolutionary way of shopping. In other words, the rate of adoption and successfulness of the automated checkout system are likely to be high in developing countries.

The most common concerns of the both enterprise and end users are the security of payments and the method of payments to be used for the automated checkout trolley system. The developers of such automated checkout system must ensure that the transaction made using the system is highly secured and users are allowed to verify the transaction. The transaction security forms the core functionality of the system which must be fulfilled in order for the users to accept the system.

Moreover, the findings indicate that the payment method used for the automated checkout trolley should convenience and allow cash as one of the payment method because the credit holders in developing countries are not prevalent as in...
developed countries, and cash payment may provide great convenience and the feeling of secured. Besides, forcing customers to use certain payment methods are tend to result in frustration [7].

Enterprise users (retailers or hypermarkets) perceived the compatibility of the automated checkout system with their existing operations or activities as important criteria in the evaluation of adopting such systems. In fact, hypermarkets tend to have their own unique business process. Hence, the automated checkout trolley system must flexible enough to enable the system to customized or tailored to existing business process or practices. For instance, the system must be able to handle items which must paid in internal counter, rather require the hypermarket to change their process. On the other hand, a standardized system is more likely to be perceived as disruptive and risky since the adopter need to change their existing structure or process in order to use the system, and thus the likelihood of adoption would be lower.

The automated checkout trolley system is a complementary mechanism to the conventional checkout line, but not a total replacement of it. Consumers should be given the freedom to choose among the automated checkout trolley or conventional checkout line. Solely dependent on the automated checkout trolley system is not practical as well, given that certain customer would prefer to feel of customer service provided by real human [7, 10]. Additionally, social pressure might make it impractical; because of fully adoption of the automated system potentially impair the employment opportunities.

6. Conclusion
The findings of this study provide information about the perceptions of population in developing countries toward the automated checkout trolley, which is useful for the providers and suppliers of automated checkout trolley. The findings provide insights for these practitioners in defining the system requirements, from the users’ point of views. Additionally, this study forms the foundation for more sophisticated or specific research which wishes to investigate in the domain of automated checkout trolley system.

This study is subjected to certain limitation as well. Firstly, the sample used in the study is the consumers from Malaysia, which the results might not be able to fully generalize to represent all the developing countries. Secondly, the assessment of users in this study is based on the conceptual prototype of the system, which might be varied from the assessment using fully developed system.

Future works can assess the perceptions of users based on the fully developed system, in order to represent their real experiences and perceptions of actual use of the automated checkout trolley system. Furthermore, future works can apply theories or models to derive at fortified hypotheses and statistical testing on the hypotheses, in order to enhance the result of questionnaire survey.

Acknowledgment
Appreciations go to School of Computer Sciences, Universiti Sains Malaysia, and Incentive APEX in supporting the conference fee.

References:
Application of WiFi-based Indoor Positioning System in Handheld Directory System

LEE PIEH WEN, CHAN WEE NEE, KHOR MENG CHUN, TAN SHIANG-YEN, ROSNAH IDRUS
School of Computer Sciences
Universiti Sains Malaysia
MALAYSIA
chimera.nav.tech@gmail.com, tsy09_com036@student.usm, rosnah@cs.usm.my

Abstract: - The purpose of this paper is to introduce a handheld indoor directory system which built on WiFi-based positioning techniques. The whole system consists of three modules, namely mobile phone module, Kiosks module, and Website and Database module. The mobile phone module is the main frontend module which installed on users’ mobile devices, whereas the Kiosks and website are to provide supports and maintenances to the main module. User can download maps and application from the Kiosks or website. On the other hand, administrator uses the website to manage the database. The proposed system is aimed to provide a new ways of providing indoor floor directory, which offers capabilities to retrieve customizable information, to navigate interactively, to enable location-awareness computing and most importantly the portability of the directory system. Additionally, the proposed system is designated to cater for the needs of visually challenged persons by incorporating screen reader and improved accessibility.

Key-Words: - WiFi Positioning, WLAN positioning, Navigation system, Mobile Application, Visually impaired

1 Introduction

Navigation has always been an issue, especially when a person is in an unfamiliar area. Directories and maps are the vital facilities which assist visitors to reach their desired destination or to have a feel of the building layout. A common form of directories is in the form of a physical board, placed in the strategic positions which highly visible to the visitors.

It is usual that vast numbers of information and labels are shown on the directory boards. However, humans have the cognitive limitations to process huge amount of information. Visitors often need to spend long time to filter out irrelevant labels and information, in order to focus on their point of interests. They might also subject to disruption due to lost trace in the congested directory. At the same time, the physical directory board is subject to the limitation of maximum user at a single point of time. Additionally, the main inefficiency of the conventional directory board is its immobility. Users need to memorize the place of interests, route to reach the place, and it is not uncommon that users need to memorize multiple places and routes.

Approximately 314 million people are visually impaired worldwide, and about 87 percent of the world’s visually impaired live in developing countries [1]. Despite of the high distribution of visually impaired persons in developing countries, the indoor facilities to assist the group is much less sophisticated compared to developed countries. The Visually challenged persons often encounter tougher issues when they newly arrive at a building. Moreover, they are almost impossible to use the conventional directory, as it does not provide the accessibility for those visually challenged. For instance, the fonts used in the directory are often small and enriched with different colors and a style in order to make the directory has more attractive appearance [2].

In relevant to aforementioned issues, an interactive and customizable directory which has the portability to be carried around would be much appreciated. Furthermore, the implementation cost of the new directory must be considerable low, in order to allow it being widely adopted and beneficial to larger groups of user. Therefore, the purpose of this paper is to introduce a handheld directory system, which allows users to use their mobile device as an interactive and customizable indoor directory. The proposed floor directory system is built on WiFi-based positioning techniques to provide the capability of real time positioning, interactive navigation, and the provision of location-triggered information. Furthermore, a kiosk with touch screen acts as a part of the whole system which not only to allows users to download maps of the area through Bluetooth, but to cater for the needs of users whose mobile device does not support WiFi connection. Additionally, the proposed floor directory system caters for the needs of visually challenged persons...
by incorporating dedicated features and functionalities such as screen reader and legibility of the interfaces.

Next section of this paper presents and discusses various positioning techniques. Then, the discussion narrowed down to WiFi-based positioning and various wireless positioning techniques, which form the foundation of this research. Section 3 presents the proposed directory system in terms of its overall architecture and system design. Section 4 presents the implementation of the proposed system.

2 Literature Review

2.1 WiFi-based Positioning System

In order to remedy the shortfall of GPS technology, many researchers have proposed alternate indoor navigation systems, e.g. ultrasonic-based system, pseudo-satellite technology, Bluetooth, RFID-based system, wireless signal positioning, TV signal positioning, and IP address positioning [3]. However, most of these positioning systems rely on proprietary infrastructure and often entail expensive deployment [4]. As a kind of the alternate navigation system developed, WLAN-based positioning system has received much attention recently. The scope of this paper is focus on the WLAN-based positioning system, which also refers as WiFi-based positioning system.

Compared to these other positioning systems, wireless positioning system (WPS) that based on Wireless Local Area Network (WLAN) infrastructure has certain advantages, including availability and readiness of infrastructure, low deployment cost, easy to implement, and signal stability. Currently, given that many buildings are equipped with WLAN access points (shopping malls, universities, office buildings, airports, etc.), WLAN based positioning system has become practical to use these access points (usually in the form of wireless routers) to determine user location in these indoor environments [4]. In other words, the WLAN based positioning system utilizes the existing communication equipment, thus significantly reduce the implementation efforts and costs. Major part of the positioning system can be implemented in software.

Moveover, provided that most of the mobile device such as cell phones are enabled with wireless radio communication network interfaces (such as Wi-Fi), protocols which to provide location estimation based on the received signal strength indication (RSSI) of wireless access points are becoming more-and-more popular, precise, and sophisticated.

Additionally, WLAN is also known as a stable system due to its robust Radio Frequency (RF) signal propagation compared with the other systems. For instance, the performance of video or IR-based positioning systems is negatively affected by line-of-sight (LOS) obstructions or light conditions, such as fluorescent lighting or direct sunlight [5].

2.2 Wireless Positioning Techniques

There are numbers of techniques have been developed by researchers to determine or estimate the position of the mobile device by different properties of signal from the WLAN Access Points (APs). Bose and Heng summarized various WLAN-based positioning techniques into Cell Identity (Cell-ID), Time of Arrival (TOA), Time Difference of Arrival (TDOA), Angle of Arrival (AOA), and signal strength based method [6].

Cell identity (Cell-ID) makes use of the radio coverage of an identified cell to indicate the location of a mobile device. It does not require complex operation such as time synchronization and multiple Access Points. The main shortcoming of this technique is its accuracy since usually the coverage of a cell is wide and due to its simplicity. Moreover, the presence of high rise buildings and many stationary points in an urban setting make this method inaccurate due to multi-path propagation and signal reflection [7].

In Angle of Arrival (AOA), the position of the mobile device is determined by the direction of incoming signals from other transmitters whose locations are known. Triangulation technique is used to compute the location of the measured mobile device. However, a special antenna array is required on the AP and be capable of mounting them under static conditions [5].

Time of Arrival (TOA) measures a distance using the travel time of a radio signal from a transmitter to a receiver. Once the distances from a mobile device to three stationary devices are estimated, the position of the mobile device with respect to the stationary devices can easily be determined using the intersecting circles of trilateration. Its application requires very accurate and tight time synchronization of the transmitter and receiver, which is difficult to achieve for close ranges [8, 9].

As a remedy to the shortfall of TOA, Time Difference of Arrival (TDOA) was developed, which utilizes the time difference between receiver and two or more receivers [6]. Thus, TOA require time synchronization between transmitters and receivers, whereas TDOA only requires synchronization between receivers.
Signal Strength based technique uses the signal attenuation property of the radio wave – Received Signal Strength Indication (RSSI) to measure the distance from a receiver to transmitter using the distance-to-signal-strength relationship. One common approach of RSSI-based system is fingerprint approach, which entails two phases: a training phase and a tracking phase. In the training phase, the received signal strength information is filtered, interpolated, and eventually stored in a database as sample points. In the tracking phase, the position is determined by comparison with the received signal strength sample points stored in the database [10]. The accuracy of this system is a function of the sample points' sampling space, an estimation method and the structure of the database. However, such a method requires the time consuming survey procedure.

3. Handheld Indoor Directory System

3.1 System Overview
The proposed interactive floor directory system, named “Guide Phone”, is a solution to provide floor directory and indoor navigation services, by utilizing mobile phone itself as both an interactive directory panel and as a signal processing device for navigational purpose. In other words, the proposed solution intended to change the conventional way of providing indoor directory, which usually in the form of physical directory board located inside a building. The purpose of Guide Phone is to provide interactive, portable, customizable ways to navigate within an indoor environment. The following subsections present the features of the proposed “Guide Phone”.

3.1.1 Interactivity
Guide Phone is coined as interactive directory system, given that it capable of providing richer information contents, rather than just static information as in the conventional directory board. Additionally, users can request only the information which is in their interests. In other words, information is tailored and customized according to personal needs. For instance, users can search a specific location by name, level, or other characteristics. Richer information such as products, contacts, and direction can be provided to users. This implies significant time savings, enhanced user satisfaction and improved efficiency.

3.1.2 Location-Awareness Information
As the Guide Phone use WLAN-based positioning technique, it able to provide location-awareness information to users. As the most fundamental function, users can track their current location in the map. In advance, customized information can be sent to user when they are in certain positions, to provide further description or relevant information regarding the current surroundings. Specifically, shopping mall or commercial building can capture the benefit of this functionality to customize their advertisements and to increase the flow of crowd into their premises.

3.1.3 Not Required to Connect to Access Points
Apart from that, the positioning technique does not require the mobile device to connect to a wireless access point (AP) such as wireless router, but only need to detect the signal transmitted from the access points. This feature is valuable given the reason that it is common that most of the APs are encrypted with passwords, thus make the positioning system which require connection to the APs infeasible. Nevertheless, Guide Phone only needs to detect the strength of the wireless signal in order to estimate the location of the users.

3.1.4 Flexible and Self-Contained
Flexibility represents the capability of Guide Phone to use different maps with minimal efforts by downloading the map of particular building from the Kiosk provided or the management website. In other words, the capability of the Guide Phone is expandable to cover any new building or place. The Guide Phone is a self-contained system when it is in action. The sensor, signal processing application, database are all included in the mobile phone. Thus, the phone application can operate independently without request any service from server.

3.2 Overall Architecture
The whole “Guide Phone” system consists of three main modules, namely mobile phone module, Kiosk module, and Website and Database module, as illustrated in Fig. 1.

![Fig. 1 Main Modules of Guide Phone System](image_url)
3.2.1 Mobile Phone Module
Mobile Phone module is the application which installed on the users’ mobile device. The purpose, input, operation, and output of the module are indicated as in Table 1.

<table>
<thead>
<tr>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>To enable searching of lots/rooms by user</td>
</tr>
<tr>
<td>To determine current location of user</td>
</tr>
<tr>
<td>To provide sufficient information of a lot/room.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>User’s search request would be needed in order to perform the search functions.</td>
</tr>
<tr>
<td>3 MAC address, latitudes, longitudes and signal strengths will be needed to calculate the current location of the user using triangulation.</td>
</tr>
<tr>
<td>User’s selection of the lot/room will determine the result of the full details of the lot/room.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Operations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database and query function</td>
</tr>
<tr>
<td>A mobile database is stored internally in the mobile phone and contains all the data similar to the database of the web and the kiosk, which contain the maps and layout data. Users define the criteria used to search or filter the destination of interests.</td>
</tr>
<tr>
<td>Wi-Fi detection function</td>
</tr>
<tr>
<td>To detect the available access points for triangulation purposes. It also includes functions to determine the access points with the strongest signal strength.</td>
</tr>
<tr>
<td>Triangulation function</td>
</tr>
<tr>
<td>Consists of functions and algorithms to carry out triangulation based on the coordinates given in order to calculate the current location of the user based on the strength of wireless signals.</td>
</tr>
<tr>
<td>Location mapping function</td>
</tr>
<tr>
<td>Location mapping consists of functions which enables the exact location of the user to be displayed in the form of an image so that the user can navigate around the area. Basically it displays the results of triangulation.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of search results.</td>
</tr>
<tr>
<td>The user’s current location on a map.</td>
</tr>
<tr>
<td>Details on a lot/room selected by the user.</td>
</tr>
</tbody>
</table>

3.2.2 Kiosk Module
As part of the whole Guide Phone system, touch sensitive kiosks are provided in the building. Users can download the Guide Phone application and maps of the building from the Kiosk through Bluetooth or USB connection. Moreover, the Kiosks act as complementary facilities to cater for the needs of users whose mobile device does not has WiFi adapter. It provides the identical functions available to Mobile Phone module, with exception of the positioning tracking system. This is due to the reason that as the Kiosks are fixed in certain location, it does not make sense to provide the real-time positioning function.

3.2.3 Website and Database Module
The website and database can be described as the backend portion of the Guide Phone system. The main purpose of this module is to allow administrators to manage the maps and other data such as location of wireless points, details of rooms or premises, and so on. However, users can access to the frontend of this module to perform the exact same functions of Kiosk module, which are to download map, mobile module application, and search functions.

3.3 Hardware and Software Architecture
Fig. 2 indicates the overall hardware and software architecture of the Guide Phone system.

![Fig. 2 Overall System Architecture](image)
location and with this information, triangulation for a user’s current location can be determined. The access points can also function as a gateway for users to access the local server to get the Guide Phone application as well as the database and maps of that specific area.

There are two ways for the end user to download the Guide Phone application. One of it would be through direct download whereby the user can connect their phone to the Internet and then download the application to their mobile phones. Another method would be to access the website on the central server using a computer, download the application and transfer the application to the phone. After downloading the application, the user can now use the Guide Phone application.

3.4 Process Flow for End User’s Operations
Firstly, a mobile user must turn on his mobile phone’s Wi-Fi connectivity. After it is turned on, the user will start the Guide Phone application. If the application is not yet available in the mobile phone, the user can download the application and the mobile database containing a location’s data from the website. After the application is loaded, the user selects the database of his current location. If the database is not available, he can download it. Once the database has been loaded onto the application, the user can now access the functions of the application. The user can choose to search for a lot/room or search for his current location.

If the user wants to search for a lot/room, the user can search either by category, name or level. Once the search is complete, the list of search results will be displayed in the user’s mobile phone. The user can then select the exact lot/room he was looking for or search again. If the user selects a lot/room, the lot/room details would then be displayed along with the lot/room’s picture. After getting the information needed, the user can opt to search for his current location, conduct a search again or end the application.

If the user selects to search for his current location, the application will then detect the access points available in the area and extract 3 access point’s MAC address. If the access points cannot be detected, the application will refresh until a few access points are detected. These 3 access points have the strongest signal strength. The MAC address will then be stored into the phone and the mobile database will be accessed to find the coordinates of the access points based on their MAC address. With the coordinates, the current location of the user will be calculated using triangulation. The results will then be displayed to the user. The user can then opt to end the application or continue to browse around the application. The overall system use case diagram is shown in the appendix session, which provides the overall view of the system functionalities.

4 System Implementation
The proposed system has been implemented in a building located in a university. The following shows the user interfaces of the system main modules.

4.1 Interface of Mobile Device Module
The application is developed to run on Windows Mobile, using C# language. Fig. 3 shows one of the main interfaces which users can see the description of the building, and a menu of available actions is provided at the bottom. Fig. 4 illustrates the interface of search by category. Relevant results will be shown according to the category pre-assigned to the rooms.

![Fig. 3 Main Menu](image1) ![Fig. 4 Search by Category](image2)

The option “search by level” implies the function will query the database to retrieve the room which on the level or floor that match the criteria selected by users. With the “search by name”, it provide the flexibility for user to customize the search by typing in the keywords that related to the room. The search current location shows the map with the real-time position of users, which shown in Fig. 5. By clicking on the area in the map, user can view the details of the particular selected area, as illustrated in Fig. 6.
4.2 Interfaces of Kiosks and Website
The Kiosks and website interfaces which accessible to the users are similar in terms of functionality. Fig. 7 shows a sample of the user interface on Kiosks. The functions available are identical to the functions offered in mobile phone module, with the exception of search by current location.

A user can view the directory for the location. Upon hover over any lots, the area will be highlighted and the name for the lot will appear. User can also click on the lot to view further details. To make the kiosk accessible by even the blind users, a screen reader (NVDA, an open source screen reader for Windows) is used. Alternate texts are used for images to allow blind users to visualize the items which appear on the kiosk screen.

5. Conclusion
Although the implementation involves the most fundamental functionalities, but it is resulted in sufficient data and information for the purpose of assessing the practicality and technical feasibility of the Guide Phone system.

The proposed Guide Phone as a handheld directory system is promising innovation. Provided the pervasiveness of WiFi access points in current buildings such as shopping malls, office and administrative buildings, and tourism spots, the proposed Guide Phone system can be implemented with low cost by utilizing the existing infrastructures. In addition, Smartphones with WiFi adapter have penetrated the market in developing countries than ever before. WiFi or WLAN adapters have become necessities of modern mobile devices which not limited to mobile phone but also other devices such as tablet PC, netbook, and iPod. These imply that the proposed system can be implemented not only in Window Mobile platform, but can be used in other platform in the future.

Nevertheless, one of the limitations of the proposed directory is that the triangulation algorithm used to estimate the position of mobile device requires at least three WiFi access points in order to successfully function. Moreover, more sophisticated functions such as location-specific message, voice guided direction to destination, and inter-users connectivity can be introduced into the Guide Phone system in future.

References:


---

**APPENDIX – OVERALL USE CASE**

**Mobile Phone**

- **End User**
  - Search Lot/Room by Category
  - Search Lot/Room by Name
  - Search Lot/Room by Level
  - Search Current Location
  - View Lot/Room details
  - View About
  - Choose location map and database

**Kiosk**

- **End User**
  - Search Lot/Room by Category
  - Search Lot/Name by Name
  - Search Lot/Room by Level
  - View Current Location
  - View Lot/Room Details

**Website and Database**

- **End User**
  - Browse Web
  - Download Map
  - Download Application
  - Search Lot/Room by Category
  - Search Lot/Room by Name
  - Search Lot/Room by Level
  - Update Database
  - View Database
  - Access Database
  - Delete Database
  - Administrator

**End User**

**End User**

**End User**

**End User**

**End User**
Mixed Number and Clifford Algebra

Md. Shah Alam$^{1,2}$ and Sabar Bauk$^1$

$^1$Physics Section, School of Distance Education, Universiti Sains Malaysia, 11800 Penang, Malaysia
$^2$Department of Physics, Shahjalal University of Science and Technology, Sylhet, Bangladesh

E-mail: shahalam032003@yahoo.com

Abstract: - Scalar and vector quantities are well known. Here we presented another type of quantity which is the sum of a scalar and a vector which we called mixed number. We have presented the Geometric product of Clifford algebra and compared it with the mixed product.

Key-Words: Mixed number, Geometric product, Mixed product

1. INTRODUCTION

In Clifford algebra, a new virtual unit $j$ has been introduced. $j^2 = 1$, $j \neq 1$, $j^* = -j$ and be named as hyperbolic virtual unit [1, 2]. We also known that in Clifford algebra, taking an equation [3-6]

$$A \cdot B = A.B + A \wedge B \quad \ldots \ldots (1)$$

Where $A$ and $B$ are two vectors, $A.B$ is a dot product $A \wedge B$ i.e. $A$ wedge $B$ which is different from the usual cross-product in the sense that it has magnitude $AB\sin \theta$ and shares its skew property $A \wedge B = -B \wedge A$, but it is not a scalar or a vector: it is directed area, or bivector, oriented in the plane containing $A$ and $B$. This product (equation 1) is called geometric product.

In the second section we introduce another quantity which we called mixed number. In the third section we compare the mixed number algebra with Clifford algebra.

2. MIXED NUMBER

Mixed number [7-11] $\alpha$ is the sum of a scalar $x$ and a vector $A$ like quaternion [12-14] i.e. $\alpha = x + A$

The product of two mixed numbers is defined as $[15, 16]$

$$\alpha \beta = (x + A)(y + B) = xy + A.B + xB + yA + iA \times B \quad \ldots \ldots (2)$$

The product of Mixed numbers is associative, that is if $\alpha, \beta$ and $\gamma$ are three Mixed numbers defined by $\alpha = x + A, \beta = y + B$ and $\gamma = z + C$ we can write

$$(\alpha \beta) \gamma = \alpha (\beta \gamma) \quad \ldots \ldots (3)$$

Basis Mixed number $i, j$ and $k$ are defined in terms of mixed number $\alpha$ by

$$\alpha = x + A = (x + A_1 i + A_2 j + A_3 k)\text{ where } x, A_1, A_2, \text{ and } A_3 \text{ are scalars and } i, j \text{ and } k \text{ are basis Mixed numbers with the properties that }$$

$$ii = jj = kk = 1 \text{ i.e. } i^2 = 1, j^2 = 1, k^2 = 1 \text{ and }$$

$$ij = i k, \quad jk = i \quad ki = i j, \quad ji = -i k, \quad kj = -i, \quad ik = -ij$$

where $i = \sqrt{-1}$.

Rule of multiplication of mixed numbers has been shown in the table – 1.

<table>
<thead>
<tr>
<th></th>
<th>$1$</th>
<th>$i$</th>
<th>$j$</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1$</td>
<td>$1$</td>
<td>$i$</td>
<td>$j$</td>
<td>$k$</td>
</tr>
<tr>
<td>$i$</td>
<td>$i$</td>
<td>$1$</td>
<td>$ik$</td>
<td>$-ij$</td>
</tr>
<tr>
<td>$j$</td>
<td>$j$</td>
<td>$-ik$</td>
<td>$1$</td>
<td>$ii$</td>
</tr>
<tr>
<td>$k$</td>
<td>$k$</td>
<td>$ij$</td>
<td>$-ii$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Table-1: Multiplication table for Mixed numbers.

Taking $x = y = 0$ we get from equation (2)

$$A \otimes B = A.B + iA \times B \quad \ldots \ldots (4)$$

This product is called mixed product and the symbol $\otimes$ is chosen for it.

2.1 Consistency of mixed product with Pauli matrix algebra.

It can be shown that [17]

$$(\sigma.A)(\sigma.B) = A.B + i\sigma.(A \times B) \quad \ldots \ldots (5)$$

where $A$ and $B$ are two vectors and $\sigma$ is the Pauli matrix. If we consider $\sigma$ as an unit matrix the equation (5) will be same as equation (4).
2.2 Consistency of mixed product with Dirac equation.

Dirac equation \((E - \beta \cdot P - \alpha \cdot m)\psi = 0\) can be operated by the Dirac operator \((t - \alpha \cdot V - \beta n)\) then we get
\((t - \alpha \cdot V - \beta n) \{(E - \alpha \cdot P - \beta m)\psi\} = 0 \quad \ldots(6)\)

For mass-less particles i.e. for \(m = n = 0\) we get
\((t - \alpha \cdot V)(E - \alpha \cdot P)\psi = [\{tE + V \cdot P + i\sigma \cdot (V \times P)\} + \{-i(\sigma \cdot P + E\sigma \cdot V)\}]\psi = 0 \quad \ldots(7)\)

where \(\psi\) is the wave function.

Putting \(t = 0\) and \(E = 0\) in the equation (7) we get
\((\alpha \cdot V)(\alpha \cdot P)\psi = \{V \cdot P + i\sigma \cdot (V \times P)\}\psi = 0\)

or, \((\alpha \cdot V)(\alpha \cdot P) = \{V \cdot P + i\sigma \cdot (V \times P)\} \quad \ldots(8)\)

If we consider \(\alpha\) and \(\sigma\) are both unit matrices the equation (8) will be same form as equation (4). Therefore from equation (4) and (8) it is clear that Mixed product is consistent with Dirac equation.

2.3 Applications of mixed products in dealing with differential operators

In a region of space where there is no charge or current, Maxwell’s equation can be written

\[
\begin{align*}
(i) \quad \nabla \cdot E &= 0 \\
(ii) \quad \nabla \times E &= -(\partial B)/(\partial t) \\
(iii) \quad \nabla \cdot B &= 0 \\
(iv) \quad \nabla \times B &= \mu_0\varepsilon_0(\partial E)/(\partial t)
\end{align*}
\]

(9)

From these equations it can be written as [18]

\[
\begin{align*}
\nabla^2 E &= \mu_0\varepsilon_0(\partial^2 E)/(\partial t^2) \\
\nabla^2 B &= \mu_0\varepsilon_0(\partial^2 B)/(\partial t^2)
\end{align*}
\]

Therefore from equation (4) and (9) we can write

\[
\begin{align*}
\nabla \otimes (\nabla \otimes E) &= \nabla \cdot \{-i(\partial B)/(\partial t)\} \\
&= \nabla \cdot [i(\partial B)/(\partial t)] \nabla \otimes B \\
&= -i(\partial\partial t) \{\nabla \cdot B + i(\nabla \times B)\} \\
&= -i(\partial\partial t) \{0 + i\mu_0\varepsilon_0(\partial E)/(\partial t)\}
\end{align*}
\]

or, \(\nabla \otimes (\nabla \otimes E) = \nabla \cdot (\nabla \otimes E) = \mu_0\varepsilon_0(\partial^2 E)/(\partial t^2) \quad \ldots(12)\)

It can be shown that \(\nabla \otimes (\nabla \otimes E) = \nabla^2 E \quad \ldots(13)\)

From equation (12) and (13) we can write

\[
\nabla^2 E = \mu_0\varepsilon_0(\partial^2 E)/(\partial t^2)
\]

which is exactly the same as shown in equation (10)

Similarly using mixed product it can also be shown that

\[
\nabla^2 B = \mu_0\varepsilon_0(\partial^2 B)/(\partial t^2)
\]

Therefore mixed product can be used successfully in dealing with differential operators.

2.4 Elementary properties of mixed product

(i) Mixed product of two perpendicular vectors is equal to the imaginary of the vector product of the vectors.

(ii) Mixed product of two parallel vectors is simply the scalar product of the vectors.

(iii) It satisfies the distribution law of multiplication.

(iv) It is associative.

3. COMPARISON OF CLIFFORD AND MIXED NUMBER ALGEBRA

The algebra of 3-dimensional space, the Pauli algebra, is central to physics, and deserves further emphasis. It is an 8-dimensional linear space of multivectors, which we can written as [19]

\[
M = \alpha + a + ib + i\beta \quad \ldots(14)
\]

scalar vector bivector pseudoscalar

The space of even-grade elements of this algebra,

\[
\psi = \alpha + ib, \quad \ldots(15)
\]

is closed under multiplication and forms a representation of the quaternion algebra. Explicitly, identifying \(i, j, k\) with \(i\sigma_1, -i\sigma_2, i\sigma_3\), respectively, we have the usual quaternion relations, including the famous formula

\[
i^2 = j^2 = k^2 = ijk = -1. \quad \ldots(16)
\]

Finally it can be shown that [19]:

\[
a \times b = -i a \wedge b. \quad \ldots(17)
\]

Thus \(a \wedge b\) is taken before the multiplication by \(i\). The duality operation in three dimensions
interchanges a plane with a vector orthogonal to it (in a right-handed sense). In the mathematical literature this operation goes under the name of the 'Hodge dual'. Quantities like \( a \) or \( b \) would conventionally be called 'polar vectors', while the 'axial vectors' which result from cross-products can now be seen to be disguised versions of bivectors.

Using equation (4) and (17) we can write

\[
A \otimes B = A.B + i (-i A \wedge B)
\]

Or,

\[
A \otimes B = A.B + A \wedge B
\]

\[\text{……(18)}\]

The right hand side of equation (1) and (18) are same. Therefore, we can say that mixed product is directly consistent with the geometric product of Clifford algebra.

4. CONCLUSION

We have clearly explained the mixed number algebra. Mixed product is derived from mixed number algebra. It has observed that mixed product is directly consistent with Pauli matrix algebra, Dirac equation and Geometric product of Clifford algebra.

References:
[3] Baylis Willian E. Clifford (Geometric) Algebra with Applications to Physics, "Mathematics and Engineering", Birkh
Abstract—With all of the news reports of global warming and other environmental troubles facing society today, one might wonder if mineral science has any role to play in solving these problems. It certainly does! Traditionally, the major application of minerals and mineral science has been in the understanding of how rocks form (petrogenesis) and behave. This is essentially the study of the solid part of the earth, or petrology. Environmental mineralogy is by no means new, but it is certainly seeing an impressive rise in scope and activity. Monographs review the immense literature and diverse applications of environmental mineralogy. Recently, the Mineralogist published special “green” issues that focus on environmental mineralogy. The important role of minerals in many interesting environmental problems makes today a very exciting time for mineralogy. Environmental mineralogical topics as the effects of minerals on human health, minerals that form in surficial acidic environments, and microbe–mineral interactions.

Mineral resources are an important source of wealth for a nation but before they are harnessed, they have to pass through the stages of exploration, mining and processing. Different types of environmental damage and hazards inevitably accompany the three stages of mineral development. The negative effect on the environment of the activities involved in harnessing the minerals. An attempt will also be made to examine the possible precautions and remedies that can be applied in order to mitigate the effect of adverse environmental impact of mining activities.

Clearly, the major goal of this special issue is to highlight some of the important research that is occurring in the new, multidisciplinary field of environmental mineralogy. Another goal is to reveal to the members and officers of the mineralogical societies an indication of the range of mineralogical research that can be considered pertinent to the environmental sciences. Lastly, the society officers and the journal editors want to encourage the submission of more manuscripts on this challenging topic.

Environmental mineralogy is by no means new, but it is certainly seeing an impressive rise in scope and activity. One indication of this is the number of recent books and monographs that review the immense literature and diverse applications of environmental mineralogy. The sidebar lists numerous examples of such review volumes. Recently, both the American Mineralogist and Mineralogical Magazine published special “green” issues that focus on environmental mineralogy. The important role of minerals in many interesting environmental problems makes today a very exciting time for mineralogy.

Key-Words: environmental damage, life cycle, mineralogy, minerals, mineral Economics, mining.
1 Introduction
Mineralogy is the study of chemistry, crystal structure, and physical (including optical) properties of minerals. Specific studies within mineralogy include the processes of mineral origin and formation, classification of minerals, their geographical distribution, as well as their utilization. History of mineralogy early writing on mineralogy, especially on gemstones, comes from ancient Babylonia, the ancient Greco-Roman world, ancient and medieval China, and Sanskrit texts from ancient India and the ancient Islamic World. Systematic scientific studies of minerals and rocks developed in post-Renaissance Europe. The modern study of mineralogy was founded on the principles of crystallography and microscopic study of rock sections with the invention of the microscope in the 17th century (Needham, 1986).
Mineralogy was heavily concerned with taxonomy of the rock forming minerals; to this end, the International Mineralogical Association is an organization whose members represent mineralogists in individual countries. Its activities include managing the naming of minerals (via the Commission of New Minerals and Mineral Names), location of known minerals, etc.

2 Types of Mineralogy
Mineralogy is an interdisciplinary science, as is geology, in which the principles of physics and chemistry are applied to Earth materials. There are numerous other applications of minerals outside the realm of petrology or even geology, and the importance of minerals extends into many areas of scientific and technological pursuit including materials science, environmental science, medicine, biology, and engineering (Ramsdell, Lewis, 1963).

2.1 Physical Mineralogy
Physical mineralogy is the specific focus on physical attributes of minerals. Description of physical attributes is the simplest way to identify, classify, and categorize minerals, and they include:
- crystal structure
- crystal habit
- twinning
- cleavage
- luster
- color
- streak
- hardness
- specific gravity

2.2 Chemical Mineralogy
Chemical mineralogy focuses on the chemical composition of minerals in order to identify, classify, and categorize them, as well as a means to find beneficial uses from them. There are a few minerals which are classified as whole elements, including sulfur, copper, silver, and gold, yet the vast majority of minerals are chemical compounds, some more complex than others. In terms of major chemical divisions of minerals, most are placed within the isomorphism groups, which are based on analogous chemical position and similar crystal forms.

3 Environmental Mineralogy
International Mineralogical Association (IMA) Working Group on Environmental Mineralogy and Geochemistry definition: Environmental mineralogy [and geochemistry] is an interdisciplinary field dealing with systems at, or near, the surface of the Earth where the geosphere comes into contact with the hydrosphere, atmosphere and biosphere. This is the ‘environment’ on which plants and animals (including humans) depend for survival and which can be disrupted by human activity, particularly that associated with exploitation and utilization of Earth’s resources.
Environmental mineralogy is obviously a very diverse sub-discipline of mineralogy, and a comprehensive review of its many aspects is far beyond the scope of these columns. Two applications of environmental mineralogy (Rakovan, 2008):
1. Remediation of heavy-metal contaminants that are currently in the environment.
2. The design of solid forms of radioactive waste for stable disposal.
One significant class of environmental pollutants is heavy metals such as Pb, As, Cd, Cr, and Hg, which are toxic, even at low concentrations, and may act as carcinogens. These may be naturally present, but human activity has greatly increased the flux of biologically available forms of heavy metals in the environment.
If the metals are present in a form that cannot get into plants and animals or is nonreactive (i.e., cannot be metabolized), they are not bioavailable and pose a much lower health risk. This leads to one possible strategy for contaminant remediation. Instead of removing the metals from the environment, the idea is to change their speciation...
to a stable, non bio available form, such as being incorporated into the crystal structure of a mineral with low solubility (Bostick et al. 2000). Mineralogists and other scientists (Wajima et al. 2007) are investigating the potential of converting paper sludge into usable minerals such as zealots. Minerals as indicators of specific environment conditions (chemical and physical). The environment parameters that can be inferred from the presence of a specific mineral are: (Irwin, 1997)

- pH
- Radix conditions
- Presence of specific anions
- Drainage with landscape position and
- Temperature & climb sequences
- Transferred versus in situ regolith

Minerals can be used as indicators of specific environmental parameters on a variety of scales ranging from the micro (thin section) to macro (hand specimen) to profile (vertical differences) to landscape (hydrology, topography) to even continental scale.

3.1 pH – Acid or Alkaline with Dominant Anion
The pH is the master variable and pH of surface and ground waters (shallow and deep) have a direct effect on the precipitation and stability or persistence of secondary minerals in the regolith.

3.2 Redox Conditions
Specific minerals in the regolith are indicators of radix state of the environment under which they formed. They form and persist in a restricted Eh pH range of oxidation potential. Pyrite (black) and mono sulphides (“black ooze”) are indicators of reducing environments, where sulphur reducing bacteria catalyze oxidation of organic matter and link it to reduction of Fe3+.

3.3 Drainage or Water Activity
The rate of flow of groundwater vertically and laterally (landscape) through a profile, sediments or sedimentary rock can affect the nature of the minerals that form, and the presence of resulting minerals can be used as indicators of the rate of flow of water or in thermodynamic terms the water activity.

3.4 Landscape or Catenaries’ Position
Regolith minerals can be used as indicators of past landscape position, but interpretation is a combination of drainage and pale landscape evolution or substrate type. The iron oxides, especially hematite and goethite, having their genetic pathways controlled significantly by Water activity or saturation, provide the best indicators of drainage linked to centenary position.

3.5 Temperature & Climate
Generally, there are few reliable mineral indicators of current and past temperatures. The main indicators are the iron oxides and hydroxides. Magnetite, commonly found in globules (nodules, mottles) within soil and on the surface as ferruginous lag, is a likely indicator of forest-fires. Iron oxides (goethite, hematite) present in soil matrix and in globules when heated to temperatures above 3000°C in the presence of organic matter (top soil organics), transform to magnetite. Therefore, the presence of magnetite is used as an indicator of forest fires.

3.6 Transferred vs. In Situ Regolith
Regolith is broadly classified into that produced as a result of weathering of the basement rocks in situ) and that produced due to the transportation processes, which is essence is surficial sediments.

4 Types of Environmental Damage
“Environment” as used in this paper has three components, namely, the sum total of external conditions in which organisms exist; the organisms themselves including the floral and faunal community; and the physical surroundings such as landforms. All these three aspects, which include various entities such as air, water, land, vegetation, animals including human, landscape and geomorphologic features, historical heritage etc. are adversely affected one way or the other during the course of mineral development (Aigbedion ,2007).

4.1 Air, Land and Water Pollution
Varying degrees of pollution of air, water and land occur in the course of mineral development depending on the stage and scale of activities attained. While only minor pollution occurs during mineral exploration, more intense air and water pollution emanates from the exploitation stages, particularly if carried out on a large scale. In the oil-producing areas of the country oil spillage of differing intensity resulting from burst pipelines, tanks, tankers, drilling operations, etc. is a common phenomenon.
4.2 Damage of Vegetation
Vegetation in form of natural forest or crop plantation is usually the first casualty to suffer total or partial destruction or degradation during the exploration and exploitation of minerals in a locality. The vegetation damage is more extensive at the time of mine development and mining operations and is more expensive when crop plantation is affected.

4.3 Ecological Disturbance
Another adverse effect of mineral extraction and processing activities, which may not be immediately felt, is the disturbance of the ecosystem with possible adverse consequences on the floral and faunal community in general.

Oil spillage produces a devastating ecological disturbance in the oil-producing states as well as in areas where leakage occurs due to natural breakage of oil pipe line or illegal bunkering. The plants, animals, soils and water are affected.

4.4 Degradation of Natural Landscape
A common negative effect of mining minerals from the earth’s surface is the destruction of its natural landscape, creating open space in the ground and generating heaps of rock wastes that cannot be easily disposed off. These phenomena are amply demonstrated in several parts of Nigeria, where commercial mining or quarrying had occurred in the past or is currently taking place.

In the Younger Granite Province, especially the Jos Plateau, tin and columbite mining has resulted in the destruction in places of the scenic landscape which is replaced by unsightly large irregular holes and heaps of debris produced by the opencast method of mining (Brooks, 1974).

4.5 Geological Hazards
Mining operations normally upset the equilibrium in the geological environment, which may trigger off certain geological hazards such as landslide, subsidence, flooding, erosion and tremors together with their secondary effects.

Minor earth tremors are generated due to blasting of rocks in various quarries. Villages and settlements in the neighbourhood of the quarries have experience unpleasant earth movements when the rocks are blasted (Ajakaiye, 1985).

5. Mine Development: Exploration, Planning, Approval and Construction
Mine development consists of a sequence of activities:
- Prospecting and exploration work to locate and delineate the ore resource.
- Economic, environmental and technical feasibility assessment of the ore body.
- Planning and design of the mine layout, site infrastructure and the mining sequence;
- Obtaining relevant government permits and approvals.
- Construction and commissioning of the operation.

Most environmental impacts during this stage of the mining life cycle are typically associated with planning, exploration and construction.

5.1 Planning
Effective planning, commencing at this stage and continuing throughout the life of the mine, has a great influence on minimizing the impact of the operation. Planning is important to avoid or reduce adverse environmental impacts over the life of the mine and after closure. Planning is most effective when the entire life of mine is compassed. Defining the final objectives of mine closure from the outset allows an optimum balance between operational, rehabilitation and closure goals to be selected, thus minimizing the cost of these activities.

5.2 Exploration
Exploration activities have some environmental impacts. These are largely related to land disturbances from the clearing of vegetation, construction of camps, access roads, drilling sites and sumps for drilling fluids and fines.

Effective planning of the exploration activities reduces potential impacts by using existing infrastructure where possible, taking appropriate care during the construction of access tracks and containing any drilling fluids and fines in sumps.

5.3 Construction
Construction activities have significant potential to have adverse environmental impacts. During this phase, often a large transient workforce is employed, workforce numbers tend to peak and material and equipment movements tend to be large. Impacts are typically related to land disturbance caused by earthworks, air emissions from dust, noise from equipment and construction activities and heavy volumes of traffic on access roads. In
many cases, specialized third party companies and consultants conduct mine construction activities. Underground operations control exhaust particulates by installing particulate filters on the exhausts of mobile equipment. Retrofitting equipment with new, more efficient and leaner engines also reduces exhaust particulate and gas emissions, as well as contributing to reduced fuel consumption and maintenance costs.

5.4 Other Potential Impacts
A variety of other impacts may occur because of extraction activities, including noise and vibration due to blasting and the operation of equipment. In some places, unsightly landscapes have been improved by maintaining buffer zones, planting green belts or the constructing barrier fences. Prompt rehabilitation of disturbed areas that are most visible can reduce the visual impact and improve relations with the local community.

6 Socio-environment problems
Some socio-environmental problems are sometimes created as a result of certain peculiarity of the mineral industry. Since minerals are exhaustible and irremovable commodities, the life of a mine and, consequently, the mining activities in a place have a limited time. The stoppage of mining activities imposed by depletion of the available reserves often leads to migration of people from the mining areas to other places. This may result in the formation of “ghost towns”, which are abandoned towns and previous bubbling mining communities.

- **Radiation hazards**: Exposure to natural radiations emitted by some radioactive minerals is a major source of health hazards. The radiation intensity increases when the minerals are concentrated.
- **Precautions and remedies for environment damage**: In order to minimize the ill effects of mineral mining and processing, certain precautionary measures must be taken by both the government and the mining and processing companies. Since some damage to the environment is inevitable in the course of mineral exploitation, usually, the only option left is to apply some remedy to the damage. The remedy or compensation should depend on the type, extent and magnitude of damage, which can be permanent or redeemable in which case the damage effect fades away as the causative factor is withdraw. The environmental impact of mining and processing activities can extend for many kilo meters from the operation site.

7 Environmental Management
Environmental management is an activity that is ongoing throughout the entire mining life cycle, from initial exploration to the final closure of the operation and handing over of the site. It encompasses and influences all the different activities of the mining life cycle and is most effective when integrated with the day-today management and planning of the operation. Environmental management consists of systems, procedures, practices and technologies that vary depending on the specific characteristics of the surroundings ecosystem, the legal framework and the mining methods and beneficiation processes employed.

As appreciation of the complexity of environmental considerations has grown, a variety of management tools have been developed to assist company’s better control the effects of their operations and provide a higher level of environment protection. These tools include: environmental management systems (EMS); environmental impact assessment (EIA); environmental technology assessment (EnTA); environmental auditing; life cycle assessment (LCA); cleaner production and environmental reporting (United Nations Environment).

7.1 Environmental Management Systems
The ISO 14000 Environmental Management system and the European Union Eco-management and Audit Scheme (EMAS) standards provide widely recognized frameworks and guidelines that companies can use to develop their own environmental management systems. These are third-party certified, providing a means by which the company can demonstrate its commitment to improved environmental performance. Many phosphate rock and potash mining companies are looking at the ISO 14000 standards to guide the development of systems that are specific to their needs.

Recognition of the growing importance of environmental management systems is evident in a number of countries including many developing countries, where producers understand the competitive advantage of meeting the increasing demands of consumer markets such as the European Union and Japan. As community awareness has
grown, demands on industry have increased accordingly. In many countries, both industry and governments are becoming more accountable for the decisions and actions they take. This is partly because of the increased availability of information concerning their actions and more stakeholder involvement in the decision-making process. In future, community concerns may shift away from a focus on environmental damage at the mine site to the need to balance competing demands for limited natural resources such as fresh water and agricultural land. The mining industry, as a consumer of natural resources, will not be isolated from these pressures. Foresight and the adoption of adequate and effective solutions to arising issues will assist the industry's response.

Governments are responding to community pressure through new legislation. The European Union's Integrated Product policy is one such response. Policy implementation will lead to increasing producer responsibility for their products. It will also increase the use of market instruments to internalize the external environmental costs of products over their entire life cycle. This responsibility will concern not only companies' own activities but also those of their suppliers, transporters and service providers Mining firms with good reputations to keep and healthy financial resources are more likely (although not certain) to engage in socially responsible mining. Such firms are typically international or they are domestic firms with ties to reputable international firms. In the short run, socially responsible mining costs more and requires considerable financial resources but this is the sine qua non of long-term viability.

There remains a need to create positive demonstration effects to promote sustainable mining through tangible examples of successful, modestly sized operations that practice responsible mining. Rushing to expand the mining sector may increase the likelihood of further accidents that are almost certain to undermine the sector's growth potential. A strategy of mining expansion focused more on the quality than the quantity of mining operations will pay higher dividends over the medium to long term. The costs of environmental irresponsibility and bad social practices on the part of one mining company are borne by all companies in the region, and peer pressure should reduce the likelihood of major accidents and improve the credibility of the industry. Therefore, the private sector in developing countries should seek to institute mechanisms for collective self-evaluation and self-policing with regard to member companies' environmental performance and contributions to community development.

And lastly, for the foreseeable future, there will be a need to provide increased support for the enforcement of all mining laws and provisions by the appropriate government agencies and local communities. It should be recognized that effective enforcement and manageable levels of investment are related. When enforcement is weak, prudence requires that the vetting of prospective investors be more stringent. However, strong enforcement allows greater latitude in opening the mining sector to investors.

8 Conclusions

Any environmental problem, there is need to assess or measure the magnitude of the problem. This can be done by direct measurements such as calculating the value of economic trees removed and changes in farm produce due to pollution; land, road and property reparation cost; water treatment cost; and the costs of treating diseases directly traceable to the environmental damage. However, direct measurements of environmental damage are not always possible because the damage may be intangible, subtle or even slow to appear. Under such a situation the cost of providing an alternative environment or renewable resources elsewhere, if possible, can be considered.

A major issue concerning the remedy or compensation for environmental damage resulting from mining and processing activities is that those who bear the costs of the environmental damage are the people who live in the environment and not the producing companies. The problem requires the intervention of government through appropriate legislation that can compel the mining/processing companies to internalize the reparation or replacement costs, which are so far borne by the people who live in the environment. Safe disposal of unavoidable waste in stable and aesthetically acceptable structure must be enforced through legislation. Thus, withdrawing the area from mining would have economic impacts on the local communities and may have social and economic effects that are regional or national.

Extraction and processing of mineral resources is the backbone of the national economy in many developed and developing countries of the world. However, the great danger posed by mineral exploitation such as abandoned sites, biodiversity damage, use of hazardous use of chemicals with potential health risk to mine workers and
neighbourhood communities deserved urgent attention. Emphasis should shift from waste disposal to waste minimization through sorting; recycling of regents and water. Safe disposal of unavoidable waste in stable and aesthetically acceptable structure must be enforced through legislation.

References:


Consumers’ complaining behavior. An exploratory research

MUNTEAN ANDREEA CIPRIANA
STREMTAN FILIMON
GAVRILĂ PAVEN IONELA
Management - Marketing Department, Faculty of Science
1 Decembrie 1918 University of Alba Iulia
Nicolae Iorga Street, no. 11-13, Alba Iulia, Alba County
ROMANIA
bologandreea@yahoo.com http://www.uab.ro

Abstract: Nowadays company’s profitability and viability depend considerably by the way that this company administrates its clients’ satisfaction and especially dissatisfied experiences (Cătoiu, Bolog 2006). The dissatisfied clients will tell to other clients about their negative experience, they will change their attitude about the product or service; they will register their complaint that will affect in a negative way the company’s image and the company’s products’ perception.

The authors of this article want to contribute to the improvement of the complaining behavior of the dissatisfied clients, considering the fact that in our country researchers in this area are modest. In this purpose it was developed a pilot survey which objectives were referring to building a socio-demographical profile of the consumer that register complaining. Data gathered and analyzed from an 112 persons sample living in Alba Iulia, lead to the conclusion that the majority are recording their complaints when are dissatisfied by a product or a service, no matter their gender, they have high monthly incomes, they have a high education level or are private entrepreneurs, and they are young.

The accomplished research represents a theoretical and empirical background of a future research that will contribute to a better understanding in this area.

Key Words: consumer, complaints, complaining behavior, satisfaction, dissatisfaction, consumers’ rights

1 Introduction

The management of consumers’ complaints represents an important aspect of the marketing activity of any organization. A major difficulty that these companies might have consists in understanding the way in which their clients respond to dissatisfaction. Exploratory research developed had as purpose the identification of some demographic variables that are influencing the complaining behavior, such as age, monthly income and the socio-professional category.

The assumptions from which this research starts consist in the fact that there are demographic differences among those who file a complaint and those who don’t do this when they are dissatisfied with a product, service or commercial unit.

In Romania were conducted few studies concerning the consumers’ complaints behaviour. Accordingly, the main purpose of this study consists in a better understanding of how dissatisfied consumers react. For this reason, the pilot survey carried out for gathering information has the character of an exploratory research, without any pretension of generalizing the results.

2 Literature review

Complaining behavior occurs after the acquisition and consumption of a product or a service when the consumer is dissatisfied (Bearden & Teel 1983; Folkes, Koletsky and Graham 1987). According to the non-confirmation paradigm, developed by Oliver in 1980, a consumer will be dissatisfied when the perceived performance of the product is below his expectations formed before the purchase.

Jacoby and Jaccard (1981) define the complaining behavior as being “an action made by an individual in order to communicate a negative aspect concerning a product or a service either to the producing or distributing company or to a third organization”.

Day et al. (1981) suggests in one of their studies that the consumer’s dissatisfaction may have important
implications such as changing the attitude towards the brand or boycotting the selling points.

The complaints are a response to dissatisfaction and their attainment highly depends on the dissatisfaction’s intensity (Yi 1990). While the dissatisfaction is a mainly affective response, the complaining behaviour is a rather cognitive response. The factors influencing the complaints behaviour are closely related to the individual predisposition to complain, to the opportunity of filing a complaint, to the procedure of filing a complaint or the consumer’s lack of information.

The research conducted in the field highlighted the associations between the complaining behaviour on one hand, and the gender of the dissatisfied consumers who file complaints (Duhaime and Ash 1980), different personal and situational factors (Day 1984), the socio-professional category and the income (Jacoby and Jaccard 1981), the life style, the values and demographical characteristics (Morgansky and Buckley 1986), on the other hand.

With regard to consumer’s reaction to dissatisfaction, most research paid particular attention to the complaining behaviour as a consequence of his dissatisfaction (Andreassen 2001, Ved 1991). It is estimated that there are many types of consumer’s responses to dissatisfaction, namely: taking no action, changing the product / brand, switching the supplier, filing a complaint to the seller, producer or a third person/institution, telling other people about the product that dissatisfied him (Goodwin, Ross 1990).

The complaining behaviour, as a reaction to dissatisfaction, varies considerably from one consumer to another. Most often, consumers take no action in order to reduce or eliminate the problems encountered on the market. Day and Bodur showed that most of the dissatisfaction cases for which no action was taken occur in the case of perishable goods (49.6%), followed by durable goods (29.4%) and services (23.2%).

The results of different studies have led to the conclusion that a large number of dissatisfied consumers do not file a complaint (Davidow, Dacin 1997, Jeffery ș.a. 1995, Valenzuela ș.a. 2002). This lack of action could have a negative impact for both marketers and consumers. First, by not expressing the dissatisfaction, the consumers are deprived of obtaining a recovery of this unpleasant experience for them. Second, the limited actions of the consumers do nothing but hide the problems of the market that the company should correct. Additionally, the company sees the consumer’s complaints as an important source of information for making some decisions about creating more competitive products.

There are studies that have attempted to establish profiles of the consumers who complain the dissatisfaction and have tried to find answers to questions like: “What types of consumers are willing to file complaints?”, “What are the differences between the consumers that file complaints and those who compliment?”, etc. (Cordoș 2003, Dong-Geun, 2003, Lau ș.a. 2001).

When establishing such profiles, the authors considered especially the demographic variables such as income, occupation, education, age, marital status, social integration, mobility, etc. According to Robinson and Berl, consumers who file complaints are generally young, have a relatively higher income and were not loyal to a brand even when they were satisfied. Also, it is appreciated that the complaints are filed only by those dissatisfied consumers who have enough resources to do this and have high expectations regarding their resolution.

Regarding the relationship between the complaints behavior and the product’s characteristics, some authors (Keng 1995) concluded that it is more likely to file a complaint for products that do not demonstrate their performance, and this may have a negative impact on the company’s image. It was also demonstrated that there is a direct link between the product’s price and the complaining behavior: the higher the price for a product, the higher probability for a dissatisfied consumer to file a complaint (Keng ș.a. 1995).

In one of their studies, Gronhaug and Zaltman (1981) describe three patterns of the complaining behavior, called the resources, learning and personality patterns. The resources pattern takes into account the access to time, money and power as determinants of the complaining behavior; the learning model suggests that the complaints will be filled by experienced and better educated consumers as they know their rights better, regarding the personality model, the consumers who file complaints are more aggressive and more confident in themselves than those who do not file a complaint.

3 Problem formulation

A pilot survey was conducted in a shopping center, on a sample of 112 people, inhabitants of Alba Iulia. In the survey’s questionnaire was included a “filter” question, placed at the beginning, which sought to exclude from
the sample the ones who were not dissatisfied with the product or service: “Have you ever been dissatisfied with a product or service?” With the help of closed and mixed questions there were obtained information on the complaining behaviour of those included in the sample.

The research objectives were established as:
1. Identifying the extent to which the dissatisfied consumers have filed a complaint based on gender, income, age or socio-professional category;
2. Identifying the reason for which the dissatisfied consumers have not filed a complaint;
3. Identifying the organization to which the complaint was filed (distributor, producer or state authority);
4. Identifying the type of product/service for which they filed a complaint;
5. Identifying the repurchase intentions.

Research’s hypotheses were formulated in this shape:
1. $H_0$: statistically gender influences insignificant the measure in which the dissatisfied clients are register their complaints;
2. $H_0$: statistically monthly income influences in a significant way the measure in which the dissatisfied clients are recording their complaints.

4 Problem solution

Of the 112 questioned subjects, 6 persons (respectively 5,53%) said that they have never been dissatisfied with a purchased product or service.

Considering the dissatisfied clients, 70,80% had no ulterior action, and the difference of 29,20% recorded a complaint. It is important to underline that those consumers that recorded a complaint they are not doing this each time when they are unsatisfied, but they have done it at least one time.

As it can be observed from the table below (Table 1), among those that recorded a claim 62% are men, and among those that had no action 56% are women.

<table>
<thead>
<tr>
<th>Investigated persons’ gender</th>
<th>Complain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Women</td>
<td>38%</td>
</tr>
<tr>
<td>Men</td>
<td>62%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: realized by author based on research data

Null hypothesis “$H_0$: statistically gender influences insignificant the measure in which the dissatisfied clients are register their complaints” was verified using $\chi^2$ Test. For 1 degree of freedom and 0,05 significance level, the calculated value for $\chi^2$ (calculated = 2,623) was situated under the level of $\chi^2$ theoretic ($\chi^2$ theoretic = 3, 841), so the null hypothesis is accepted. In conclusion, statistically there are no significant differences between women and men regarding the complaining behavior.

From the data presented in the table below (Table 2) it can be observed that the ones that are not recording complaints are with modest monthly incomes (49,30% of all), while the dissatisfied consumers that are recording complaints have monthly income above the average (48,50% of all).

Null hypothesis “$H_0$: statistically monthly income influences in a significant way the measure in which the dissatisfied clients are recording their complaints” was verified using $\chi^2$ Test. The calculated value for $\chi^2$ was 16,87 and it is higher than the theoretical value of $\chi^2$ for 2 degrees of freedom and a 0,05 significance level. In conclusion the null hypothesis is not accepted, so the complaining behavior is significantly depending on the consumers’ monthly income.

Table 1. Complaining behavior depending on the investigated persons’ gender

<table>
<thead>
<tr>
<th>Income category (personal, monthly average)</th>
<th>Complain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bellow 1000 lei</td>
<td>19,3%</td>
</tr>
<tr>
<td>1001-2000 lei</td>
<td>32,2%</td>
</tr>
<tr>
<td>Above 2000 lei</td>
<td>48,5%</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>

Source: realized by author based on research data

Private entrepreneurs, high education level persons and clerks that are recording their claims when they are dissatisfied in a bigger measure than the rest of the socio-professional categories (Table 3). One possible explanation of this result could be that the persons that are included in these socio-professional categories know better their rights as consumers.
Table 3. Complaining behavior depending on the investigated subjects’ socio-professional category

<table>
<thead>
<tr>
<th>Socio-professional category</th>
<th>Complain</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Pupil</td>
<td>-</td>
<td>6,66%</td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>6,4%</td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>Worker</td>
<td>3,2%</td>
<td>30,6%</td>
<td></td>
</tr>
<tr>
<td>Clerk</td>
<td>22,5%</td>
<td>10,6%</td>
<td></td>
</tr>
<tr>
<td>High education level</td>
<td>25,8%</td>
<td>6,66%</td>
<td></td>
</tr>
<tr>
<td>Private entrepreneur</td>
<td>29,03%</td>
<td>9,33%</td>
<td></td>
</tr>
<tr>
<td>Domestic</td>
<td>3,2%</td>
<td>9,33%</td>
<td></td>
</tr>
<tr>
<td>Unemployed</td>
<td>6,4%</td>
<td>5,33%</td>
<td></td>
</tr>
<tr>
<td>Retired</td>
<td>-</td>
<td>6,66%</td>
<td></td>
</tr>
<tr>
<td>Other category</td>
<td>3,2%</td>
<td>6,66%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Source: realized by author based on research data

Considering the age, claiming behavior is specific to the people with the age bellow 40 years old (Table no. 4). The dissatisfied consumers that have above 60 years are recording complaints in a small measure.

Table 4. Complaining behavior depending on the investigated subjects’ age category

<table>
<thead>
<tr>
<th>Age category</th>
<th>Complain</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>16 – 25 years</td>
<td>24%</td>
<td>6,45%</td>
<td></td>
</tr>
<tr>
<td>26 – 40 years</td>
<td>52%</td>
<td>25,8%</td>
<td></td>
</tr>
<tr>
<td>41- 60 years</td>
<td>22,66%</td>
<td>54,8%</td>
<td></td>
</tr>
<tr>
<td>Above 60 years</td>
<td>1,33%</td>
<td>12,9%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
<td>100%</td>
<td></td>
</tr>
</tbody>
</table>

Source: realized by author based on research data

The lack of trust regarding the problem solving is the main raison for which the most of the investigated subjects are not recording complaints when they are dissatisfied. As well, the ones that are not recording complaints do not record them because of their lack of time (34,66%) or appreciate that the product represented for them a small importance (12%) (Fig. 1).

Fig. 1. Weight of the persons that are not recording complaints depending on the reason for which they are not recording complaints

Source: realized by author based on research data

Among the subjects that had recorded the majority (45,33%) had addressed to the distributor/selling point from where they purchased the product related to which they were dissatisfied and the smaller part of them (20%) directly to the producer (Fig. 2).

Fig. 2. Weight of consumers that recorded complaints depending on the recording point

Source: realized by author based on research data

From the total of 75 unsatisfied consumers that have recorded complaints, the biggest part of them registered their complains related to long term using products (respectively 50%) and for services (29%). In case of unsatisfied clients related to the perishable products, had been recorded complaints by 21% of the subjects investigated.
The dissatisfied consumers that had recorded complaints and were solved (their number representing 43% of the total subjects that recorded complaints) intend in a big measure to repeat the purchase, while the ones that were dissatisfied and did not recorded their complaints do not intend to repeat the purchase.

5 Conclusions

The exploratory research developed allowed to made some conclusions.

First of all, complaining behavior is influenced by demographical variables like gender, monthly income, age and socio-professional category. Statistically there are not significant differences between the women’s complaining behavior and the men’s complaining behavior, but statistically there are significant differences between the complaining behaviors of the subjects belonging to different monthly income category.

The subjects that are recording their complaints when they are dissatisfied by a product or a services are usually young with high education level.

In the second place, the main reasons for which part of the dissatisfied consumers do not record their complaints is related to the lack of time and to the lack of trust that their problems will be solved.

In the third place, consumers recorded their complaints first at the selling point where they purchased the products and less to a third part.

Generally the complaints were made for long term using products and services and less for perishable products;

The conclusions resulting after the exploratory research are underlying the importance of future studies, especially due to the managerial implications of complaining behavior researches. In this way, the companies which want to increase the number of dissatisfied consumers who complain should facilitate this and also should assure them that their problems will be solved. Also, these companies should focus on dissatisfied consumers who complain to identify those socio-demographic characteristics that make them to have a positive attitude regarding complaining behavior.

6 Limits

The main limit of this study consists in that of the fact that the results cannot be generalized because the inquiry developed has the characteristics of the pilot survey and the sample is statistically unrepresentative.

References:


Abstract: - An ability training partnership model was designed for high school graduates with disabilities to execute results that were incorporated into a descriptive study. There are limited resources and employment for high school graduates with disabilities. The objective of the study was to develop a descriptive analysis of an ability training model as partnered efforts between colleges, universities and organizations seeking to provide resources to high school graduates with disabilities. One of the purposes of the research was to establish a proactive position for colleges, universities, and organizations against barriers that address disabilities stereotypes and practices in employment. The nature of disabilities has been study for the many years. However, few studies have been directed towards collaboration between colleges, universities, and organizations to develop programs for ability training or on-the-job training to high school graduates with disabilities. Hiring individuals with disabilities makes good business sense. These individuals can be qualified, productive, and dependable in the workplace. This explanatory study was designed to take information from learning theories and on-the-job training programs for high school graduates with disabilities and build a picture of current and future trends involving employment training for these individuals.

Key-Words: ability training, disabilities, special needs, employment, learning theories, training programs, partnered education

1 Introduction
In 1945, diplomats met to form the United Nations. The World Health Organization (WHO) came into force April 1948, and is a specialized agency of the United Nations (UN) that coordinates authority on international public health [1]. The WHO’s constitution and mission states its objective are “the attainment by all peoples of the highest possible level of health, to combat disease, specially key infectious diseases, and to promote the general health of the people of the world” [1].

An estimated 10% of the world’s population experience some form of disability and according to WHO, about six hundred million
people live with disabilities of various types due to – chronic diseases, injuries, violence, infectious diseases, malnutrition, and other causes closely related to poverty [1]. According to WHO, disabilities is an umbrella term covering impairments, activity limitations, and participation restrictions: impairment is a problem in body function or structure; an activity limitation is a difficulty in executing a task or action; and participation restriction is a problem involvement in life situations. The International Classification of Functioning, Disability and Health, known more commonly as ICF, is WHO’s framework for measuring health and disability at both individual and population levels [1]. The ICF definition of disability is “the outcome or result of a complex relationship between an individual’s health condition and personal factors and of the external factors that represent the circumstances in which the individual lives” [1]. Importantly, the WHO organization is committed to work towards ensuring equal opportunities and promoting the rights and dignity of people with disabilities [1].

In order to provide the highest possible level of health to the disabled people of the world, WHO, in partnership with other UN agencies, are assisting by: promoting early intervention and identification of disability; supporting the integration of community-based rehabilitation services; and facilitating the inclusion and participation of people with disabilities in their societies. In addition, WHO promotes strengthening collaborative work on disability across the United Nations system and academia, private sectors and non-governmental organizations, including disabled people’s organizations [1].

2 United States Legislation
Legislation in the United States (U. S.) has always played a major role in the history of education; much of the progress in educational needs of children and youths with disabilities is attributed to laws requiring states to include students with special needs in the public education system [2]. In 1975 a federal law was passed called the Education for All Handicapped Children Act and in 1990 the law was amended to become the Individuals with Disabilities Education Act (IDEA). In 1997, the federal law was again amended to ensure that all children and youths with disability have the right to a free, appropriate public education [2]. This federal legislation specified that to receive federal funds, every school system in the United States must provide a free, appropriate education for every student regardless of any disabling condition [2]. Thus, during the 1990s, the first comprehensive civil rights laws for people with disabilities were passed in the United States.

Cornerstones of IDEA and other federal laws focused on early childhood intervention with free, appropriate public education for every child or youth between the ages of three and twenty-one regardless of the nature or severity of the disability [2] and The American with Disabilities Act (ADA) prohibited discrimination against people with disabilities in employment, in public services, in public accommodations, and in telecommunications [2, 3]. According to Buck [4], students with disabilities have more academic options resulting from the ADA of 1996; the federal law ensures the right of individuals with disabilities to nondiscriminatory treatment in all aspects of their lives with protections of civil rights in the specific areas of employment, transportation, public accommodations, state and local government, and telecommunications. The law created opportunities for people with disabilities to overcome their shortcomings and master skills in the classroom; importantly, these skills can transfer into the real world [2]. The primary intent of the federal special education laws passed in the past decades has been to require educators to focus on the needs of individual students with disabilities [2]. Thus, with changes in the education system, students with disabilities have access to all the educational benefits that fully able bodied students enjoy.

3 U. S. Students with Disabilities
Integration into the larger society for people with disabilities began in the 1960s and continues today. A key factor toward more integration of people with disabilities into
society was normalization, which is the belief that we should use means that are as culturally normative as possible to establish or maintain personal behaviors and characteristics [2]. Under this principle of normalization, the means and the ends of education for students with disabilities should be as much like those for nondisabled students as possible; other trends in integration have involved deinstitutionalization, self-determination, and inclusion [2].

According to Hallahan and Kauffman [2], early intervention and a program of education or treatment developed a better outcome for a child with disabilities. Studies by Bricker [5] in 1995 and Kaiser [6] in 2000, suggested a child’s early learning provided the foundation for later learning, so a program of early intervention would have the child go further in learning more complex skills. Early intervention was likely to provide support for child and family. In addition, early intervention helped families adjust to having a child with disabilities and to finding additional support service, such as counseling, medical assistance, or financial aid [5, 6].

Kahn, Hurth, Kasprzak, Diefendorf, Goode, and Ringwalt [7] described approaches to long-term services for student with disabilities, which included state and local infrastructure, personnel development, service providers, and community settings that impact access and quality of services. According to Kahn et al. [7] there was a need for identifying students with learning disabilities early – early collaborative intervention was designed to bring together information about the student’s strengths and needs, which allowed educators to think about the student and match support and service to that student [7]. After determining the student’s strengths, needs, and interest, individual assessments undertook to understand the student’s strengths. Developed plans for differentiated instruction are then offered with appropriate challenges and high-end learning opportunities [7]. According to Coleman and Hughes [8], the allocation of resources followed the supports and services, and as the needs of the student increased, the educational resources combined to provide greater support. Thus, the goal of educational programs for disabled children should be the fullest possible development from preschool to completion of high school.

4 Employment
What happens to students with disabilities when they leave high school? Many students with disabilities leave high school equipped with the skills and supports needed to realize their goals for adulthood; however, many students do not [9]. According to the U. S. Department of Education 2000 [10], students with disabilities who graduated with a high school diploma has remained at about 25%. Students who do not complete high school are more likely to face difficulties in adult living than are those who have a high school diploma. In addition, studies by Malian and Love [11], Wagner, Blackorby, Cameto, and Newman [12], along with Yelin and Katz [13], suggested students with disabilities who do not complete high school are more likely to have lower levels of employment and wages – and higher rates of problems with the criminal justice system.

According to Goldstein, Murray, and Edgar [14], employment earnings of high school graduates with disabilities were slight higher than those of peers without disabilities for the first four years after leaving high school (because most of their peers were attending college). However, by the fifth year the earnings of graduates without disabilities outpaced those with disabilities [14]. Thus, attending college or postsecondary training programs would increase the likelihood of obtaining employment and expand success as an adult [15].

5 Early Intervention
In order to address the importance of effective transition and preparatory programs to facilitate entry into the workplace, one must begin at the foundation of the planning process, which is that of early intervention. Howlin, Goode, Hutton and Rutter [16] posit that a child’s school district Special Education Department or committee is required by both federal and state regulations before the age of fourteen to develop a transition plan from school to life after school. This
entails meeting with the parent or guardian, the child and community agencies when applicable to discuss the skills and knowledge that the child will need as an adult. The concern however, is not whether programs are in fact in existence for a child with disabilities. Rather, the attention must be directed toward the quality of the programs during the developmental process, along with the quality of training and development provided to those responsible for implementing and monitoring of the success or failure of such programs. The implication is that individuals working directly with the child serve as the catalyst for positive, negative or neutral outcomes, consequently warranting the need for immediate address as each child continues to move throughout the levels of the educational process.

According to The Colleges with Programs for Learning Disabled Students in 2010 [17], The Americans with Disabilities Act mandates that colleges and universities provide services and/or accommodations for students with learning disabilities [2, 3, 18, 19]. Yet, the key to the success of the college program will largely be determined by the quality of early intervention found in the educational foundation provided during elementary and high school years. Consequently, the skills, knowledge, and abilities which have been acquired during this time in conjunction with external environmental factors are crucial toward the personal and professional outcomes of a disabled adult. Nuehring and Sitlington [20] conceded that the goal in educating students with disabilities must be to help them become productive members of society as adults; however, few schools have directed these students into meaningful employment opportunities that are appropriate for their strengths and achievements. Although an adult vocational service could assist a student in this area, the high schools must serve their function in preparing students to become productive citizens as well.

6 Transition
An understanding of the definition of the term “transition” must therefore occur prior to the actual program development process within a high school or institution of learning. The 1997 amendments to the Individuals with Disabilities Education Act (IDEA) [2, 3, 17, 18, 19] according to Nuehring & Sitlington [20] defined transition services as a coordinated set of activities for a student with a disability that is 1) designed within an outcome-oriented process that promotes movement from school to post-secondary education, adult services, independent living or community participation; and 2) is based on the individual need of the student, taking into account the student’s preferences and interests. This is a multi-faceted outcome oriented approach that encompasses future vocational placement, residential options, funding and community resources.

Sherman [21] conceded that parents and guardians must take responsibility with regard to being careful about drafting goals and objectives in the transition individual education plan (IEP) in as much as services must be provided through the age of twenty-one or until the goals and objectives are met. Individuals that share the same diagnosis may not necessarily share the same needs, therefore a collective effort between home and school may increase the likelihood of the development of realistic and attainable transition-based strategies. Anonymous [22] stated that a primary initiative in the field of education has been to promote the quality of scientific research that may serve as the foundation for instructional practice and that a tenet of the No Child Left Behind Act was that instructional practices needed to be research-based. Because of the range of questions that are important, the variability in the population, and the ecological and sociopolitical contexts of special education, important research questions that would lead to the improvement of educational practices for students with disabilities may require unique applications of scientific research methodology.

7 Disability
Amendments to the Americans with Disabilities Act (ADA) signed into law on September 25, 2008, clarified who is covered by the law’s civil rights protections [2, 3]. The ADA Amendments Act of 2008 revised the definition
of “disability” to more broadly encompass impairments that substantially limit a major life activity (Americans with Disabilities Act, [2, 3]. The amended language also stated that assistive devices, auxiliary aids, accommodations, medical therapies and supplies (other than eyeglasses and contact lenses) have no bearing in determining whether a disability qualifies under the law (Americans with Disabilities Act, [2, 3]. Changes additionally clarify coverage of impairments that are episodic or in remission that substantially limit a major life activity when active such as epilepsy or post-traumatic stress disorder. These amendments took effect as of January 9, 2009 [2, 3].

8 Issues in Transition Planning

The challenge presents itself in the realization that one cannot group all students with disabilities together in terms of educational programs and transition-based strategies when a child becomes of age to receive such services. For example, children in many school districts throughout the United States may be receiving 15-30 minute allocations of speech and occupational therapy per month, while the actual need may fall within the range of a minimum of 2 hours of speech therapy per week. An additional area of concern is that school districts that use the IEP as a specific funding source for individual students can be problematic because of budget constraints. Therefore, students that require higher levels of services are not receiving as such and ultimately be on the road for becoming institutionalized. Teachers have the added task of practicing a triage program while the students which illustrate progressive improvement are receiving the individual services necessary and those with less improvement may be relegated to the equivalent of a day care setting. According to Hernandez, Wadsworth, and Nietupski [23], counseling professionals who served persons with disabilities in vocational settings often measured career outcomes through the assessment of employment outcomes that occurred in the competitive labor market. The underlying assumption therefore, was that successful community employment led to increased empowerment and is associated with positive independent living outcomes for persons with disabilities [23]. However, there is still a need to further understand the actual factors that lead to placement and economic success, especially in the transition stage.

9 Research Studies

A research study conducted by Fabian, Lent, and Willis [24] reported that transition activities in which students with disabilities could explore and match their interests (i.e. job tryouts, job shadows, and internships) led to successful job entry. Beale and Holinsworth [25] reported that low job production, job dissatisfaction, and personal distress occurred when jobs did not match student interests. In essence Hernandez and et al. [23] concurred that the match between career interests and employment opportunities has been described by many career theories as a significant fact in transition planning stages. On the other hand, Beale and Holinsworth [25] cautioned that unfortunately study participants with more severe disabilities received lower wages than did peers with less severe impairment, despite achieving employment which matched their interests. Consequently, students with disabilities may benefit from exposure to diverse career and interest opportunities and not necessarily positions found in the low wage category. However, the process does not end with a matching process. Downs and Carlon [26] stated that although the “old” bureaucratic system might not have placed the person in his or her preferred setting, the “new” system must monitor the workplace and/or residential site through meaningful and continued support. As a result, Downs and Carlon [26] suggested that the individual receive support while on the job for a minimum of 90 days after initial hire and then on a standard or as-needed basis. In addition, one-support person per employer process instead of the one-support person per person with a disability would begin with the employer need rather than employ.

10 Future Transition Research

The need for continued research in the area of transition planning for students with disabilities is of paramount importance and timing is critical as the numbers continue to increase for students
awaiting an uncertain future as well as just beginning the process within the educational system nationwide. Studies must take into account students who fail to be employed because of their disability or disabilities. It is imperative that the researcher examines the system from the beginning stages when the student enters the system at a young age throughout the process of young adulthood and ultimate transition into the workplace.

An examination of the needs assessment, program development, planning and implementation must be reviewed on a continual basis. Individuals responsible for transition planning implementation vis-à-vis directly providing educational and socially based curriculum and training, must be properly trained and qualified to do so. This is a cost-effective strategy in as much as an effective transition program can make the difference between independence vs. institutionalization. It is also necessary to examine how children with disabilities are characterized in terms of educational levels. In theory, one may state that the child is graduating from high school, yet the reality may be that the student is graduating with a sixth grade education.

11 Partnership Resources

According to Carter, Trainor, Cakiroglu, Cole, Swedeen, and Owens [27], career development and early work transition personnel reported having few community partners to support and enhance these ability training experiences. Carter et al. [27] surveyed 135 Chambers of Commerce and other employer networks to examine (a) whether and how these networks have partnered previously with local high schools on youth-focused career development activities, (b) the extent to which they would consider such involvement to be feasible, and (c) the influence of disability status of youth on their responses. Carter et al. [27] stated most respondents considered a number of youth-focused support activities feasible; however, most had limited previous involvement or experiences and their views were clearly influenced by the disability status of those youths [27]. Lastly, the study by Carter et al. [27] recommended expanding the employer networks in supporting career development and early work experiences of students with disabilities. To assist with the development and implementation of ability training programs for youths with disabilities, organizations have a variety of networks and resources from which to select. Resources such as the Business Leadership Network, Easter Seals, local Achievement Centers, Vocational Rehabilitation Centers, and Workforce Development Boards have programs designed with partnerships in mind. Although there are differences in the resource entities, they all share the same mission: to provide job development, on-the-job-training, as well as follow up and support services.

There is need for partnerships between institutions of higher education and places of employment. According to Durlak, Rose and Bursuck [28], an increasing number of students with disabilities were looking to postsecondary education and training to help them achieve success in career development and eventual job placement. Past research by Bursuck, Rose, Cowens, and Yahaya [29] Rose and Bursuck [30], Bursuck and Rose [31], Benz, Doren, and Yovanoff [32], along with Cameto and Wagner [33], found an increasing number of community colleges, 4-year colleges, and universities were developing and providing services for students with learning disabilities. Bursuck and Rose [31] found some postsecondary programs were reaching out to high school students in an effort to facilitate the transition to higher education, but high schools were not necessarily actively involved in this process.

Despite the urgent messages from researchers (e.g., Ward [34]; Bursuck & Rose [31]), the teaching of self-determination skills must occur at the secondary level; evidence of the existence of such programs were minimal Durlak et al. [28]. According to Ward [34], self-determination was one component of secondary students’ readiness for adulthood. In addition, the concept of self-determination varied according to its usage Algozzine, Browder, Karvonen, Test, and Wood [35]. However, Wehmeyer [36] defined self-determination as an
individual’s freedom to exercise choice and make decisions associated to the quality of life; including education, work, and other important personal matters. Durlak et al. [28] proposed a model training program to teach self-termination skills to high school students with disabilities. The results suggest that some students can acquire, maintain, and generalize skills that focus on the self-determination skills of self-advocacy and self-awareness [28]. The results of the study by Durlak et al. [28] had implications for parents, educators, and employers. In particular, higher education was in the position of being able to teach students about the opportunities and expectations of the adult world and about ways in which acquisition of those skills will enable them to negotiate in that world [28]. Lastly, according to Gillespie [37], it was the fundamental responsibility of school personnel to make routine assessment of basic academic skills and learning abilities and disabilities. Thus, the study by Durlak et al. [27] was a small step to education school personnel on the characteristics of students with disabilities.

Schooling and partnerships between colleges and organizations involves fostering in teaching and learning. Partners must focus on the need of the students with disabilities and have authentic assessments as those that sample the actual knowledge, skills, and dispositions of students with disabilities in teaching and learning contexts. According to Whipp and Scanlan [38], teaching and learning communities should be inclusive of students across multiple dimensions of abilities or disabilities. While evidence shows that current partnerships between colleges, universities, and organizations exist as pockets of innovation, it is suggested [38] that a systemic effort across institutions and employers to improve education and life opportunities for students with disabilities. Clearly, a contemporary challenge for both institutions and employers is to become simultaneously more efficient in their use of resources (e.g., human, fiscal, material, and tools) and more ambitious in their partnership outcome aims that include the elimination of gaps (between high school and post-secondary education) in achievement for students with disabilities [38]. Consequently, these partnerships might contribute to the broader improvement in teaching and learning so that career development and work experiences are not missed by students with disabilities [38].

Are these partnerships helping youths with disabilities be qualified, productive, and dependable in the workforce? In efforts to examine these partnerships critically, it was suggested by Whipp and Scanlan [38] the use of conceptual frameworks that cross defined boundaries between colleges and organizations. Whipp and Scanlan [38] described four such frameworks: 1) a justice framework that draws from theory and on-the-job-training so students are touched by direct experience for the formation of the of the whole person; 2) an ethical care framework that pulls from social and behavioral theories so students grow academically, emotionally, morally, and physically; 3) a learning framework based on sociocultural learning theory and a professional learning community to deliver social and cultural learning situated in the contexts of everyday living and work; and 4) a vocational framework derived from business, management, and economic teachings oriented toward results and problem solving innovations. Each framework provides a valuable perspective from which to examine partnerships and reforms that are systemic and oriented toward social and educational justice for youths with disabilities [38].

12 Ability Training Partnership Model

The term model has a wide range of uses, from a physical scale model to a set of abstract ideas. Modeling is a resource utilized in the flow of decision making: decision makers have the ability to analyze the problems, identify the text techniques that can be used to resolve individual segments of the defined problem, and eventually elect or develop a model flow that will properly employ the techniques for problem resolution [39, 40]. Al-Fedaghi [41] stated that information processing models have evolved since 1949 and methods have been proposed to extract requirements from policies and regulations using formal models [39]. Geogini,
Massacci, Mylopoulos, and Zannone [42] described a framework that enables modeling actors and goals and their relationships: thus, modeling can be designed as an instrument based on information factors deemed to be important for partnership decisions and an analytical framework to understand ability training programs for youths with disabilities [39, 40, 42]. A model represents a real world situation and is employed to aid decision makers with preferred solutions to problems through the evaluation of alternatives [39, 40].

The following Ability Training Partnership Model (ATPM), Figure 1, suggests the partnership importance between colleges, universities, organizations, and students with disabilities, and can be seen as characterized by five criteria:

1) control over the conditions and context of the training, roles, experiences, and practices;

2) training conducted within the context of student’s abilities;

3) training that provides for deliberative or personalistic or critical reflection that addresses personal growth, professional relations, and the social context of the work from students with disabilities;

4) training should enable the partners to interrogate their own practices in the educational goals through their decisions about instruction and assessment within the institutional structures of the partners; and

5) the nature of the relationship between the partners is dialogic where students with disabilities and their evaluators enter into dialogue aimed towards success in ability training programs, resulting in positive independent living outcomes [43].

Efforts to create new forms of on-the-job-training programs, such as ability training through partnered education for youths with disabilities, has sought to transcend the limits of traditional programs. The unique way in which the partners are at work within the ATPM model is based on the belief that all partner action need to be geared towards the efficient social and practical fulfillment of on-the-job-training competencies involving students with disabilities [43].

**Fig. 1**

![Ability Training Partnership Model](image-url)

Source: Cook, DeCaro, and DeCaro, 2010
13 The Star Center Model
According to DeCaro and DeCaro [44], a real-world practical model was developed by The Star Center Foundation, Inc. which incorporates a holistic combination of medical, therapies, and educational support. This particular model works well with children and young adults ranging in ages from three to twenty-one years of age. The model is incorporated in a school system within the state of Florida and what makes this model extremely interesting is that it blends the disciplines of psychiatry and neurology together to gain insight into a child in their formative years. The program also provides for the children the different types of therapies such as speech, occupational, physical and applied behavior analysis. What makes this program unique is that for the first time many children will be under the care of a physician who would actually be watching their progress to see if the primary diagnosis that was given to the child was correct and progressive adjustments can be made accordingly. In addition, DeCaro and DeCaro [44], find that today a number of teachers that are graduating from schools of education are not necessarily qualified or prepared to address some of the severe mental disabilities which could be dangerous if appropriate intervention training is lacking. Teachers in our schools are being seriously injured by students who are suffering from a number of severe disorders, while behaviors continue to escalate. Unfortunately, it is after the teachers are hurt that schools take the necessary steps to see that the child is properly evaluated [44].

As the child moves toward adulthood, we find in a number of students with disabilities placed on sedatives which may not necessarily be helpful and may in fact impede the learning process. Therefore, what some school systems may unknowingly be promoting is the pre-positioning of institutionalization for the students. Alternatively, if we advance The Star Center Foundation’s model, one can try to first teach the person how to learn and to be socially aware of their surroundings with the goal of living in the community as a contributing member of society. This would increase the likelihood in later years for assisting in making the transition process into college more viable. What is currently being developed between The Star Center Foundation and Greenleaf University is a specialized on-line curriculum which would make young adults with specific disabilities more active and interactive within this curriculum design [44]. The future holds promise for young adults in achieving an academic and/or vocational degree which would help to remove the stigma of being “burdens” and rather become perceived as being viable contributors to our society. One of the benefits of this state of the art program is to identify those special gifts that these individuals may possess such as may be found in the areas of mathematics, music, the arts and technology, to name a few. The key is to be aware that individuals with disabilities learn differently from the typical student. Therefore, once we understand the individualized path of learning, we have thus succeeded [44].

14 Conclusions
The nascent field of disabilities education is growing more clearly defined within the United States. This paper has conceptualized how partnerships between higher education and the workplace could inform and promote disabilities education in communities. They could benefit from a forged partnership based on a mission of learning with professional development resources. Such partnerships can be natural, convenient, and symbiotic settings for students with disabilities [38]. According to Gamoran and Long [45], to counter the persistent evidence of inequalities in educational opportunities at all levels for many students in the United States, there were indications that some partnerships are offering a positive alternative for those who have been traditionally marginalized in schools [38]. The number of students with disabilities has increased in the United States [46]. The Office of Special Education reported in their 25th Annual Report to Congress [46] that 5.8 million students with disabilities were served in the public education system, with 49.2 percent of those students enrolled in special education programs due to specific learning disabilities [46]. Thus, in order to ensure successful
outcomes in terms of providing the opportunity for each child with a disability to attain his or her maximum potential, it is imperative that the proper resources and training be provided to those individuals working directly with the child in order to provide to the child the necessary resources and training preparation for transitioning from the high school to workplace environment. A collective effort between the parents or guardian of the child and the educational system increases the likelihood of successful outcomes for all parties. The Ability Training Partnership Model supports partnering from high school to post-secondary education and is a robust resource to improve education and life opportunities for students with disabilities.

In conclusion, the research has illustrated that matching the job to the interest of the student at various levels of cognitive or physical functioning has resulted in successful outcomes and possible new insight for future planning, along with monitoring progress during various intervals of employment. It must additionally be understood that an individual with a disability is not necessarily destined to become or remain a low-wage earner. The investment of time, effort and funds in the long-term can yield cost and social benefit for all parties. In essence, laws alone are insufficient if they are not properly upheld and programs are not designed to ensure that children with disabilities along with society will benefit as a whole. As a result, Beale and Holinsworth [25] contended that the benefits of matching between occupation and interests for young adults at various levels of functional severity of impairment, may offer new insight into the best practices for promoting the transition from high school to the workplace.

15 Recommendations

Partnership is vital to the development of ability training for students with disabilities. For this reason, partners, such as college, universities, and organizations, must work together to offer ability training for students with disabilities. Important issues include participant evaluation and selection, program orientation and standards, identifying participant needs for positive behavior support, legislative documentation and policy requirements, useful and relevant knowledge funneling to practitioners and families, and securing knowledgeable personnel for the ability training program. Most important is that partners be committed to action step communication from development, implementation, to assessments and outcomes in the ability training programs. In addition, the partners must have a common purpose of improving student self-achievement by ensuring effective instructional strategies for success in the workplace environment. Finally, partnerships can broaden, contribute, and promote collaborative ability training models for the future. Much future research remains to be done if we are to understand and facility partnerships in ability training for youths with disabilities.

References:
2009, pp. 7-23.


Overcoming Measurement Hurdles in Statistical Education Research using Rasch Measurement Model

ZAMALIA MAHMUD 1, MOHD SAIDFUDIN MASODI 2, IDA ROSMINI OTHMAN 3, JAMILAH OTHMAN 4

1 Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 48000 Shah Alam, Selangor MALAYSIA
2 Program Coordinator, Exec.Dip in QMS / QMS ISO9000 Lead Assessor, School of Professional Advancement & Continuing Education, University Teknologi Malaysia, 81300 UTM Skudai, MALAYSIA.
3 e-mail: zumal669@salam.uitm.edu.my, rosmini@tmsk.uitm.edu.my, jamilaho@tmsk.uitm.edu.my
4 e-mail: saidfudin@gmail.com

Abstract: Research involving volunteer participants in educational settings is known to be difficult, and when the study is on a small sample, it becomes even more challenging. This study aims to measure changes in postgraduate students perception on statistics over a series of statistics remedial classes. Five overlapping test forms were used so that each student will eventually experience both the typical test ambience and open book test at least once. This paper describes an attempt using Rasch measurement model to measure postgraduate students’ problems in learning statistics. Using data collected from post classes survey responses, it was evaluated on how it relates to the latent trait abilities being scrutinised and whether it corresponds to the attributes to be measured. Rasch Person-Item map and measure appears to be more rigorous and has plentiful of vigor in revealing the correlations between the postgraduate students’ performance ($\beta_n$) and item difficulty, ($\delta_i$) in classifying the students learning ability. In this study, Rasch measurement model has been able to serve as a ruler to more accurately measure students’ knowledge mastery development.

Keywords: Learning outcomes, learning ability, performance, item difficulty, Rasch person-item ma,

1 Introduction

Student learning begins with the assessment of educational values. Assessment lends itself to an educational improvement where it should help students to inculcate the value of learning and continuously help them to achieve their learning goals. However, sometimes we tend to overlook the values of educational mission. Assessment is often seen as an exercise in measuring what's easy rather than a process of improving what we really care about [1].

Proper assessment is expected to lead to improvement when it is part of a larger set of conditions that promote change. Assessment alone without emphasising changes will be meaningless. Assessment without using proper measurement method could also lead to incorrect diagnosis of student learning.

With current education reforms in Malaysia, the Ministry of Higher Education has empowered the Malaysian Qualification Agency (MQA) to ensure that the quality of education through various strategic planning and actions be made visible and successfully implemented.

Firstly, the challenge for most higher learning institutions (HLIs) is to understand the concept of quality education and how to provide quality education. Various strategic actions have been introduced by MQA to assist colleges and universities in providing quality education. The recent activity being the assessment of student learning outcome following the Outcome Based Education (OBE) principle [2]. Students from various disciplines who are currently enrolled in a statistics course are subjected to the measurement of their learning outcomes. Unfortunately, among the statistics educators much less is said about the success rate of OBE implementation as expected by the education reformist. The reason being there is no clear procedures on how these learning outcomes should be precisely measured apart from the usual pen and paper assessment. Even though certain Engineering faculties gave positive response and input on the successful implementation of OBE [3] but to many others, the measurement path remains unclear.

Hence, this study intend to discuss an alternative and precise measure of students’ learning using Rasch Measurement Model. The discussion revolves around a case study on a small sample of postgraduate students who attended a statistics remedial classes in preparation for the analysis of their research data. The
advantages and strength of the Rasch analysis is discussed in the next section.

2 Overview of Measurement Principles

Measurement has been grossly misunderstood and overlooked in many circumstances especially in the field of social science. Many researchers in social science are frustrated when existing instruments are not well tailored to the task, since they then cannot expect sensitive, accurate, or valid findings [4]. However, modern measurement method as practiced using item response theory with a focus on Rasch measurement model provides the social sciences with the kind of measurement that characterizes measurement in the natural sciences; i.e. the field of metrology.

The fundamentals of measurement must comprised of the instrument to be used for purpose which has specific unit of an agreed standard amount. An instrument must have the correct construct of linear scale which can be zero set and duly calibrated. A valid instrument can then be replicated for use independent of the subject hence measurement taken thereof is therefore a reliable data for meaningful analysis and examination to generate useful information [5]. This information is of utmost importance to be the prime ingredient in a particular decision making.

3 Measurement Method

Responses from the students in an examination, test or quizzes is normally marked against a marking scheme comprising keywords; where when there is a match then the student would be given a mark or otherwise. This is the traditional ‘park and mark system’. In theory, at this stage truly the assessors is only counting the number of correct answers which is then added up to give a total raw score. The raw score only provide a ranking order which is deemed an ordinal scale that is continuum in nature. It is not linear and do not have equal intervals which contradicts the nature of data fit for the due statistical analysis [6]. It does not meet the fundamentals of sufficient statistics for evaluation.

In Deterministic Model, these data set would normally be put on a scatter plot to establish the best regression line. However, estimate or prediction from ordinal responses on the student learning outcomes (LO) attributes are almost impossible due to the absence of equal interval on a linear scale. The normal solution in linear regression approach is to establish a line which fits the points as best as possible; which is then used to make the required predictions by interpolation as necessary as shown in Fig. 1.

\[
y = \beta_0 + \beta_1 m \quad \text{Equ. (1)}
\]

\[
y_i - \hat{y}_i = e_i \quad \text{Equ. (2)}
\]

In obtaining the line of best fit, there exist differences between the actual point; \(y_i\) and the predicted point; \(\hat{y}_i\). This difference is referred as error; \(e_i\)

By accepting the fact that there is always errors involve in the prediction model, the deterministic model of Eq. (1) renders itself less reliable. This can be resolved by transforming it into a probabilistic model by including the prediction error into the equation.

In Rasch philosophy, the data have to comply with the principles, or in other words the data have to fit the model. In Rasch point of view, there is no need to describe the data. What is required is to test whether the data allow for measurement on a linear interval scale specifically in a cumulative response process i.e. a positive response to an item stochastically implies a positive response to all items being easy or otherwise. This dichotomous responses which can take only two values, 0 and 1 indicates the failure and success of an event, respectively. Rasch Measurement Model is expressed as follows

\[
P_m(x_i | B_n, D_i) = \frac{e^{(B_n-D_i)}}{1 + e^{(B_n-D_i)}} \quad \text{Equ. (3)}
\]

The model expresses the probability of obtaining a correct answer (1 rather than 0) as a function of the size of the difference between the ability (\(B\)) of the person (\(n\)) and the difficulty (\(D\)) of the item (\(i\)). Rasch exponential expression is a function of Logit Model which resulted in a Sigmoidal ogive and can be transformed into simpler operation by reducing the indices by natural logarithm. This is represented as

\[
\ln [P(\theta)] = \beta_n - \delta_i \quad \text{Equ. (6)}
\]

This transforms it to logit for the purpose of obtaining a linear interval scale. It can be readily shown mathematically that a series of numbers irrespective of based used is not equally spaced but distant apart
exponentially as the number gets bigger while a log series maintain their equal separation; thus equal interval [7]. This equal separation is mathematically shown in Table 1. The difference between $\log_{10}5$ and $\log_{10}2$ is constant and remain of equal distant between $\log_{10}50$ and $\log_{10}20$ which similarly hold true for $\log_e$, thus the theorem can now be universally applied.

Table 1. Comparison of Numerical and Log Intervals

<table>
<thead>
<tr>
<th>Numerical series</th>
<th>$\log_{10}$</th>
<th>$\log_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.301</td>
<td>0.694</td>
</tr>
<tr>
<td>5</td>
<td>0.699</td>
<td>1.609</td>
</tr>
<tr>
<td>10</td>
<td>1.000</td>
<td>2.303</td>
</tr>
<tr>
<td>20</td>
<td>1.302</td>
<td>2.997</td>
</tr>
<tr>
<td>50</td>
<td>1.699</td>
<td>3.912</td>
</tr>
<tr>
<td>100</td>
<td>2.000</td>
<td>4.606</td>
</tr>
</tbody>
</table>

An attempt of a student to answer an item renders him some level of difficulty to choose a preferred response. Rather than summing up the frequency of response rating which is a complete disregard of allowable mathematical operation in ordinal data to obtain a so thought ‘score’ and subsequently the ‘item mean’. Rasch sees it as a chance of a person selecting a given option hence in a rating of 1 to 5, from ‘Dislike to Really Like’, he has a 20-odd chance of selecting a choice of preference. So, if he chose ‘Dislike’ the odds will be 20:80, next slight dislike; 40:60 and so forth. This is known as odds of event.

We can put this on a probability scale as in Fig. 2. For a given frequency of 5/100, an odd of success is 5:95.

<table>
<thead>
<tr>
<th>Rating</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f/\sum f$</td>
<td>5</td>
<td>15</td>
<td>45</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>Odds</td>
<td>5:95</td>
<td>15:85</td>
<td>45:55</td>
<td>25:75</td>
<td>10:90</td>
</tr>
</tbody>
</table>

This enable us to construct a logit ruler which is the log-odd of an event taking place with the odd-of success, as shown in Figure 2 with unit termed as, derived from the term ‘log-odd unit’, as unit of measurement of ability akin to meter to measure length or kilogram to weight.

In order to achieve an equal interval scale, we can introduce logarithm of the odd probabilistic value. Maintaining the same odd probabilistic ruler as in Fig. 3, starting with 0.01 to 100, we can create an equal interval separation between the log odds units on the line, hence the measurement ruler with the logit unit [8]. This can be verified by computing the value of $\log_{10} 0.01 (10^{-2})$ equals to -2.0; value of $\log_{10} 0.1$ equals to -1; value of $\log_{10} 1$ equals to 0 and so forth. Figure 3 shows the newly established logit ruler as a linear scale with equal interval separation. It is just like looking at a thermometer with ‘0’, as water being ice and 100 as boiling point whilst the negative extreme end as -273°C, the point where all atoms of any element come to a standstill.

Thus, we now have a valid construct of an instrument to measure the students ability for each defined LO. Rasch offers the opportunity to measure every respondents responses irrespective of the type of test taken either traditional or Open book style. Similarly, all the respondents whether he has taken the earlier test and subsequently take the post test and vice-versa; Rasch has the unique strength to predict the likelihood of their expected responses [9].

This predictive ability of Rasch Model enable to explain the link between the latent trait under scrutiny. This connection is described as the Maximum Likelihood Estimate (MLE) of an event based on Rasch-Guttman Scalogram henced developed [10]. Thus, Rasch analysis is the formal testing of an outcome scale; i.e. MLE against a mathematical measurement model developed by the Danish mathematician George Rasch as in Equation (6).
4 Results and Discussion

The test was administered on postgraduate students pursuing their doctoral research degree in Universiti Teknologi MARA (UiTM) attending the Research Method class. The result from the tests were analyzed using Winsteps 3.6.8, a Rasch based analysis software, to obtain the logit values. Fig. 4 shows the Person-Item Distribution Map (PIDM) where the persons; i.e. the Students is on the right whilst the items; the topics learned in statistics were plotted on the left of the logit ruler. By virtue of the same ruler with the same scale; then the correlation of the person, $\beta_n$ and item, $\delta_i$ can now be established.

The PIDM Map is the heart of Rasch analysis [6]. The vertical dashed line ( - + - - + - ), separating the Items and Person, represents the ideal less-to-best continuum of quality. Items i.e., topic learned in statistics and Person, i.e. the students, now share the same linear measurement units known as logits. On the left hand side of the dashed line, the items are aligned from too easy to too hard, starting from the bottom liken to a high jump bar; the higher the more difficult is the task. The distribution of student positions on the right side of the vertical dashed line in increasing order of ability; the best naturally being at the top and the poorest student is at the bottom. Letter “M” denotes the student and item mean, “S” is one standard deviation away from the mean and “T” marks two standard deviations away from the mean.

On the right side of the PIDM as in Fig. 4 shows the Students Location where the students were separated by the type of test taken, traditional method and the open book test to evaluate their trend of learning in Statistics. Rasch Analysis tabulates the students’ location in a very clear graphical presentation which is easy to read and easier to understand. Each student can be coded with attributes or factors that is deemed to affect their learning process. This will enable in depth analysis of their study pattern to be evaluated meaningfully.

Prior delving further, it is best to look at the analysis Summary Statistics as in Table 3. The lead information we need in this table is the overall students’ LO ability reflected by the Person Measure mean ; $\mu_{PERSON}$ equal +0.38logit ($P[\theta]=0.6411$). This gives the indication that generally the students performance and perception towards statistics under scrutiny is very much within target where the item difficulty has broader range compared to the students’ ability. Note that the poorest student ability (A0922713) measured at Mean$_{PERSON}$ equal -0.28logit being lowest whilst the best students (B0623915) is measured at Max$_{PERSON}$ equal +1.78logit ($P[\theta]=0.8557$) being best compared to the item range of difficulty where the hardest item correspond to item 3Fa_Graph, measured at a very high Max$_{ITEM}$ equal +6.36logit ($P[\theta]=0.9983$) and the easiest item is item 07EP_Employable measured at Min$_{ITEM}$ equal -2.91logit ($P[\theta]=0.0517$). This shows that there are items which is extremely difficult far beyond the students’ ability and equally many items being too easy as well.

![Figure 4. Person-Item Distribution Map: Ph.D students Learning Statistics in Research Method](image-url)
Rasch analysis provides two very important primary information. Two major component is involved in any test; the Person and the Question Paper. Cronbach-α has the limitation of providing only the overall test result but it can be at stake if we encounter a problem with the Question Construct. Rasch analysis shows both the Person Reliability and Item Reliability score at 0.88 and 0.96 thus showing the data validity for further analysis. Further to that, it gave a clearer classification of Person and Item separation instead of the traditional Confirmatory Factor Analysis (CFA) where CFA alone is able to show unidimensionality among items only. Rasch analysis involves lesser operation to validate the data for further analysis.
Generally, the students separation, G equal 2.75 is a typical value achievable in Malaysian experience that separate them into distinct performance level; Good, Mediocre and Poor students or strata. Strata can be calculated using the formula:

\[ \text{Strata} = \frac{(4 \times \text{student separation} + 1)}{3} \quad \text{Equ}(7) \]

which yielded a four(4) weak different strata. This is clearly reflected by the PIDM in Figure 4. Generally, it grouped the students into three (3) separate profiles:

Group 1: Poor students: (Male, N=0, 0%; Female, N=1, 5.26%) and she is from Group A, who does not have much problem with Central Tendency, Measure of Dispersion, Reliability Test and very familiar with SPSS applications only.

Group 2: Mediocre students: (Male, N=2, 10.52%; Female, N=7, 36.84%) 79% (7/10) from Group A, includes the two (2) males, 21% (2/9) from Group B. They found Graphical Presentation easy, with some good command of Data Tabulation but encountered split difficulty in Test of Normality and Parametric Hypothesis Testing.

Group 3: Good students: (Male, N=0, - %; Female, N=9, 40.52%) All female; 20% (2/10) from Group A and 80% (7/9) of them are from Group B. They have good understanding of all the topics taught but have difficulties in Logistic and Multiple Regression. They also have little knowledge in other statistical software, e.g. Minitab, SAS.

The PIDM and Summary Statistics was derived from Rasch Fundamental Theorem based on the scalogram as shown in Table 4. Rasch has refined Guttman Scalogram not only sorting by rank order but by the nature of responses to the items being examined. The result shows Person 16 (B0623915) as being the most able at the top of the table whilst Person 9 (A0922713) the poorest being at the bottom of the table. Rasch also has arranged the items in a way where easy items which theoretically every Person will score well on the left and the difficult items where Person encounter difficulties attempting the task thus low score. Item 113 (07EP_Employable) is the easiest item being on the left and item 31 (04BType of Study) is the hardest where none of the students managed to answer it correctly.

This give rise to a table of person misfit response as reflected in Table 5. Since the students could not respond correctly to Item 31 hence too difficult and categorized as “maximum measure”, it has been discarded from further analysis. Rasch analysis generated two (2) useful information; Outfit Mean Square (MNSQ), z-standardized score and point measure corelation (PMC). Mean Square below 0.5 or exceeding 1.5 need to be investigated further. Item 144 (3Fa_Graph) high MNSQ equal 3.53 indicates there are poor students who has made some lucky guesses or probably having a specific interest on that topic but warrants a consideration. This is further reflected by the PMC as a negative value (-0.47) where the item is measuring the reverse; smart students failed to answer the item where else the poor student somehow answered it correctly as shown in Table 4.

Rasch translate these data to be tested against the model by graphical presentation as shown in Fig. 5 and Fig. 6. If the item is very easy, graphically the item characteristic curve will show the responses on the upper end of the modeled line and vice-versa when the item is very difficult all the responses will be at the lower end of the curve. This allows us to predict the rest of the ‘possible responses’ in between.

![Figure 5. Item Characteristic Curve: Easy Item](image)
Rasch has a unique ability in recognizing the students development based on the students response pattern from the scalogram. Table 5 shows the Person Measure Fit Order. This table gives an indication of person misfit response where values (5,5,3,2,5,0) are not plotted according to the expected Rasch model.

Table 5. Misfit person response

In Rasch, we test the data if it fits the model as opposed to deterministic model where the line of best fit with permissible error is delineated based on least square method. In Table 5, Rasch Analysis revealed that Person no. 6A062_14 gave the most misfitting response string followed by A0122913. In case where items does not show any response, Rasch will estimate the likelihood of the event.

Table 6: Maximum Likelihood Estimate of Unexpected Responses
If observed responses do not fit the expected outcomes, Rasch will compute the residual difference. The larger the residual signify the further deviation from the expected outcome.

Table 6 shows the MLE for the most unexpected responses from this assessment. It was found that the largest standardized residual is 3.89 followed by 2.08 for item 129 (23EP_Irrelevant) and 1Fb_Quantity respectively all relates to Person A0122913. Rasch estimated the expected response to be 1.85 and 2.08 as compared to 5 in the actual response.

This give rise to the strength of Rasch Model where it is able to predict all the missing data as seen in Table 5. The main hurdle in this research is therefore overcomed. Though every effort is taken to ensure all the students took all the test in any of the possible scenario, it is certain that some will not attend all the classes despite the compulsion for this remedial class. Rasch predictive property enabled us to conduct this research irrespective of the students having taken the test before or after each identified session. Rasch has this particular predictive properties embedded in the model to make it a very reliable validation model; for both the person and item response string.

5 Conclusion

Rasch Model provides a sound platform of measurement equivalent to natural science which matches the SI Unit measurement criteria where it behaves as an instrument of measurement with a defined unit and therefore replicable. It is also quantifiable since it is linear. Rasch Model has made it very useful with its predictive feature to overcome missing data.

The logit ruler has been developed with the purpose of measuring ability; in this case students learning ability of specific learning outcomes. It can define the students profile and most important we are now able to validate a question construct on line. It is a noble innovation where the ability ‘ruler’ can transform ordinal data into measurable scale. It’s graphical output is great which gives better clarity for quick and easy decision making.

The measurement conducted reveals the true degree of cognitive learning abilities of statistics learners [11] based on Blooms Taxonomy [12]. Previously, lack of such measurement in educational research for assessment was difficult to formulate. This major problem faced by education researchers in an IHL to design a comprehensive assessment method is therefore resolved. Rasch has all the capabilities to rigorously analyse examination results more accurately thus making evaluation more comprehensive with better accuracy, clearer to read and easier to understand.

References


Issues and Challenges in NSDI Implementation

NOOR HABIBAH ARSHAD, FUZIAH ABU HANIFAH
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam, Selangor
MALAYSIA
habibah@tmsk.uitm.edu.my

Abstract: - The challenge to design, build, implement and maintain a National Spatial Data Infrastructure (NSDI) requires attention from various disciplines and examination of numerous issues and challenges. It is reported that numerous SDI-related activities at the national, state and local levels in many countries share similar core objectives to stimulate coordinated collection, dissemination and use of spatial data by public and private entities. NSDI is about improving data sharing and use, so citizens, communities and governments can easily make use of spatial data to solve problems. The main objective of this paper is to determine the major issues and challenges in NSDI implementation. Major key issues and challenges identified are selected based on the comparison and matching of input from the subject matter expert. The result of this study revealed that out of sixteen, three issues that have been cited by all the experts gives the most impact to the implementation of NSDI namely availability of quality digital data, lack of awareness on GIS and SDI, and institutional arrangements. The determination and elaboration of the issues and challenges would greatly influence the strategy chosen for SDI development and its effectiveness.

Key-Words: - National Spatial Data Infrastructure, Spatial Data Infrastructure, Geographic Information System, Digital Data, Institutional Arrangements, Capacity Building.

1 Introduction
The National Infrastructure for Land Information System (NaLIS) was developed by the Ministry of Land and Cooperative Development. NaLIS is the Malaysian Government effort to establish National Spatial Data Infrastructure (NSDI). In 1997, NaLIS Secretariat was established within the IT Department of the Ministry of Land and Cooperative Development. The scope and duty of the unit was to carry out the development of geospatial data infrastructure that encompasses policies, standard, technology, law as well as support that includes data preparation in agencies at Federal, State, and Local Authorities at all levels [8,9].

In 2002, NaLIS Secretariat was superseded by the Malaysian Centre for Geospatial Data Infrastructure (MaCGDI). The NSDI for Malaysia is a geospatial data infrastructure that comprises technology, policies, standards and procedures for land related agencies to cooperatively produce and share land information. NSDI, which for Malaysia is called Malaysian Geospatial Data Infrastructure (MyGDI), provides a basis for spatial data exploration, evaluation, and application for users and data providers [9].

Despite the successful establishment of NSDI in Malaysia, there are a number of issues and challenges related to NSDI development from conceptual, technical, political, institutional and financial perspectives. Therefore, the challenge of designing, building, implementing and maintaining an NSDI draws on many different disciplines and requires examination of such issues and challenges. It is also essential to understand the significance of human and societal issues, all of which contribute to the success of the Malaysian NSDI development.

2 National Spatial Data Infrastructure
Advanced spatial information and visualization technologies, including geographic information systems (GIS), remote sensing (RS), global positioning systems (GPS), image processing have enhanced the methods and tools for collecting, disseminating, sharing, integrating, and using spatial information. Access to such information as input to the planning and implementation of various projects indicates its effective use. To address the need for easy access to accurate, consistent, and up-to-date spatial information, spatial data infrastructure (SDI) is created by many countries at all levels [11].

An SDI encompasses policies, fundamental data sets, technical standards, access network (technologies), and human resources including users and data providers necessary for the effective collection, management, access, delivery, and utilisation of spatial data at different political/administrative levels [1,13]. SDI developments range from local to state/provincial, national, and international regional levels, to a global level. The design and implementation of an SDI is not only a matter of technology but also one of designing institutions, the legislative and regulatory frameworks and acquiring new types of skills [1,2]. The ultimate objectives of these initiatives, as summarised by [3], are to promote
economic development to stimulate better government, and to foster environmental sustainability. Ideally, an SDI should provide benefits to all parties.

The development of SDI is driven by the business needs and technological developments to support the rapidly expanding geospatial information industry. According to [6] and [5] the infrastructures for spatial information should be designed to ensure that spatial data are stored, made available and maintained at the most appropriate level, that it is possible to combine spatial data from different sources in a consistent way, share them between several users and applications. Policy makers need fast and easy access to different types of information in order to make sound decisions. Spatial data is often crucial to the decision making process.

NSDI is a way to overcome the barriers and allow different departments and organizations to supply the NSDI with spatial information and open standards fill in the gaps between geospatial islands of data sources [10,14]. NSDI is necessary and implemented by many countries, with objective to enable on-line access to geospatial information, avoid duplication of effort in the collection of data and to ensure the accuracy, timeliness, correctness and consistency of data to be used in planning, development and management of land-based resources. Most of these are driven by the national or federal government.

The importance of NSDI is recognized by majority of developed nations and a number of emerging nations. This is evident from surveys reported by [4] on comparative analysis of NSDI implementations in various countries such as Australia, Canada, United States, Europe, Japan, France, India and Malaysia. Furthermore, NSDI are underway in many countries. Being the way to interoperability, NSDI is valuable in many respects. NSDI is a must for the e-government and e-business. NSDI will be a major component of e-government because almost 80 percent of all the information is spatial.

Establishment of NSDI is critical to provide the foundation of a properly working system for geospatial data sharing. This study will identify the issues and challenges in the implementation of NSDI. The information about the current development in NSDI implementation, experiences and analysis covered by the local and other countries will be used to propose the suitable enhancement in the Malaysian context.

3 Research Approach

Interviews with ten SDI experts from various countries (Malaysia, Australia, German, Japan and Belgium) were done to identify the issues and challenges which are important in long running success of an SDI initiative. The interactive inquiry includes interaction with key experts in the field, which involved interviews with the NSDI experts from local and abroad. These interviews aimed at finding out the issues and challenges for a success implementation of NSDI. For the purpose of this research, some guided questions related to critical issues relevant to the promotion and dissemination of NSDI have been designed which include several aspects of NSDI development such as policies, standards, framework data, organisational and enabling technology involved. The interviews were done with the key experts who have vast experience in NSDI implementation and involve directly in development of GIS based applications and various data sharing activities.

4 Issues and Challenges

Detail elaboration of the major issues and challenges discussed during the interviews are as depicted below:

a. Availability of quality digital data

Spatial data depositories developed and managed by government agencies at all levels contain a varied set of spatial datasets. Data availability depends upon content and completeness of databases on variety of spatial themes. The current and accurate spatial data must be readily available to contribute to local, state and national development and contribute to economic growth, environmental quality and stability, and social progress. This would be best achieved by making accurate and timely spatial data readily available to support decision makers and to do so with minimum duplication of effort and at a reasonable cost. However, information needed to solve cross-jurisdictional problems is often unavailable. Users have to spend considerable amount of time to order and process the raw data to produce the data products they need in the analysis. Even if available, a spatial dataset may or may not be obtainable. Some of the problems faced by organisations were that the digital spatial data needed for good governance are limited and only available on certain conditions that are restricting its extensive use. Relevant data is often hard to find and frequently it is not in compatible forms. Furthermore, framework data does not exist for broad geographic areas and the information describing data is often non-existent.

The readiness of core framework dataset are crucial, particularly the geo-reference and topographical maps, which are used as the underlay for thematic data and provides the basis for many geospatial datasets used by public and private industry. Use of spatial information for resource management and decision-making is limited due to problems on how to combine the different data sets.

b. Institutional arrangements

Existing systems serve primarily their own clients, without concern for the needs of other potential users.
This leads to the duplication of efforts and sometimes inefficient use of resources, both financial and human. Sharing information in a fully transparent manner is not the main characteristic of the culture. In most organisations, communication is linked to hierarchy and authority. Due to the strong vertical organisation culture of government and administration, there is no real encouragement of cross sectoral communication. Each ministry or department undertakes its own mandate, trying to create its own database and information system, following its own needs and priorities. Information is handled in a strictly vertical direction, following hierarchies.

c. **Lack of awareness in GIS and SDI**

Lack of awareness of the potential of GIS among public sector institutions, non-governmental organizations as well as the private sector means, that the use of geographic information systems is still low. To raise awareness in GIS and SDI, it is necessary to demonstrate the benefits, obtain support and promote investment for implementation of the NSDI within provider and user communities. NSDI Communications Plan should be implemented to communicate progress and obtain stakeholder feedback on NSDI implementation through the NSDI workshop series.

There is an increasing number of spatial data producers who provide substantial geographic databases and products. However, means for disseminating information about these databases and products are still limited. Spatial data clearinghouses and NSDI-related communication channels are in the formative stages and not widely known or used by the majority of potential users of spatial data and products. Information such as data type (what), location (where), quality (accuracy, currency, completeness), and ownership (whose) can be obtained rather efficiently if the level of awareness among potential users is increased.

d. **Capacity building**

The capacity building concept is often used within a narrow meaning such as focusing on staff development through formal education and training program to meet the lack of qualified personnel in a project in the short term. Capacity building measures should be addressed in the wider context of developing and maintaining institutional infrastructures to meet short and medium term needs. Capacity building for NSDI and the adoption of its wider concepts and levels can assist NSDI coordinating agencies to speed up the progress in the development of NSDI initiatives.

e. **Clear SDI directives**

NSDI is national in scope, and must meet the needs of a wide range of geospatial user communities, data producers and different areas of the private sector. A clear NSDI directive is necessary to lay down general rules aimed at the establishment of the infrastructure for spatial information in the community. NSDI for a country should be build upon infrastructures for spatial information and operated by the member of the states. This Directive should apply to spatial data held by or on behalf of public authorities and to the use of spatial data by public authorities in the performance of their public tasks.

Problems voiced regarding the availability, quality, organisation, accessibility, and sharing of spatial information are common to the various levels of public authority. Solving these problems requires measures that address exchange, sharing, access and use of interoperable spatial data and spatial data services across the various levels of public authority and across different sectors. An infrastructure for spatial information in the community should therefore be established. The NSDI should assist policy-making in relation to policies and activities that may have a direct or indirect impact on the environment.

It should be possible to combine spatial data from different sources across the community and share them between several users and applications. The directive should make possible for spatial data collection at one level of public authority to be shared between other public authorities; that spatial data are made available which do not restrict their extensive use.

f. **Lack of funding on GIS-based Projects**

Many organisations are lacking of enough resources to use and analyze the spatial data. Many of current geospatial application projects require handling multi-terabytes of data. In order to conduct such projects, users have to buy expensive high-performance hardware and specialized software. In many cases, those resources are only purchased for a specific project and when the project is finished, the resources will be set idle. Because of the above problems, applying geospatial data to solve the scientific and social problems is a very expensive business and only few organisations can afford such luxury.

Efforts should be made to keep distribution costs of spatial data to a minimum. However, additional funding will be required for some agencies. Thematic data represent a key component of many geospatial data products and services. Thematic data are used in a variety of industries and sectors. Frequently, clients request specific types of thematic data to fulfil a decision-making or planning process requirement, which requires in some cases substantial amounts of funding to do research, collect and maintain the datasets. Governmental departments make up the largest user group of geospatial data in a country. These governmental clients in some cases cannot use needed data to make a planning or policy
g. **Adoptions of standards**

The major technical obstacles to data sharing reside in the lack of application of a national standard for spatial data, incompatible classification schemes and the absence of data documentation or metadata. Spatial data sources should conform to common standards that enable integration with other data, where such integration enables efficient and effective solutions for users. Sharing and use of these datasets is encouraged for data providers to make priority datasets available through the NSDI. Actions should be taken to identify priorities and support development of nationally consistent and integratable spatial datasets that meet user needs. Adoption of common classification systems, spatial referencing system, content standards and data models are promoted to facilitate geodatabase development at all levels. A minimum set of best practice data standards should be applied that facilitate integratability and develop reference implementations to support the user-driven development of integrated national geodata centre.

In many cases, the temporal and spatial coverage, resolution, origination, format and map projections are incompatible. Among the technical challenges are integrating multiple standards and specifications together and communicating complex standards. The adoption of a core set of standards focused on documented data quality and interoperability. Users are able to ascertain the quality of existing spatial data and its fitness to meet their needs. A set of NSDI data quality standards should be approved and promote further improvement of the quality of metadata records held by data providers.

h. **Access Delivery Mechanism**

More organizations and individuals are realizing the benefits of having access to others’ data. The access issues are related to the organizations and the individuals. Financial issues, which are related to data access, include cost-sharing, charges for access and services, profit, and ownership. The major issues involved in the data access environment are related to technical standards of data access and sharing. The metadata (information about the data relating to source, quality, limitations, and other aspects), data standards, data transfer procedures, and system interoperability are the key factors in the NSDI environment. Progress continues on resolving the technical issues related to data access, but the major obstacles are institutional and managerial, including policies for data access, data sharing arrangements and agreements, pricing, rights and responsibilities, restrictions on access and use, maintaining privacy and security, and control of the data-sharing environment.

The current absence of guidelines and operational evaluations in the data access area of the GIS and SDI field presents a challenge for those trying to establish data access policies and procedures for their own organizations. Actions should be taken to identify barriers to access to spatial data and develop institutional arrangements to lower the barriers to data access and use. Coordination arrangements in jurisdictions need to be strengthened to develop agreements of relationship management and consensus building for NSDI. However, in most countries, the lack of uniformity across different jurisdictions within a country often creates problems in attempts to integrate various dataset at national level. There is an increasing number of spatial data producers who provide substantial geographic databases and products. Means for disseminating information about these databases and products, however, are still limited. Therefore, the data and service providers need to advise potential users about the availability of their spatial data and services.

i. **Lack of Knowledge and skills**

Design and management of an appropriate NSDI is crucial to ultimate integration of the geoinformation into the organization. The development process must incorporate the organizational aspects of NSDI initiation, development, and operationalization. Successfully handling all the components of the NSDI development process requires skill and knowledge, including the ability to recognize early the skills and expertise required and the level present in the organization, as well as how much outside assistance must be obtained.

j. **Data Interoperability**

Data interoperability is among the core issues in the process of geospatial knowledge discovery and utilisation. Many data users spend considerable time on assembling the data and information into a ready-to-analise form, even when the analysis is very simple. Data collected by different organisations are often incompatible. The data may cover the same geographic area but use different geographic bases and standards. The need for interoperability is also strong for municipality and government applications. Due to the lack of interoperability, governments and municipalities face serious problems. These are mainly poor quality and high cost of services, and low economical revenues. Geodata should be as seamless as possible, with coordination across jurisdictions and boundaries where possible.

Most data is collected, processed, and maintained not according to national standards. In addition, users, faced with difficulties to maintain data integrity across databases, and to enable the data integration.

k. **Cultural (reluctant to change)**
All communities and societies have a culture – a system of shared meaning. Similarly, any initiative or function, including the sharing of information, also has a specific culture, which needs to be promoted to prepare the environment for developing/pursuing the specific activity. Whether that culture is weak or strong is important to both a coordinating agency and individual parties. Therefore, sharing knowledge and information requires a specific culture – a culture for sharing.

The spatial databases being built up are “stand alone systems”, using individual philosophies and technologies. Most of these implementations are technology driven and isolated implementations related to specific environmental issues. Different agencies are often supported by a different technology. Every organisation has its own data sharing policies. There are also organisations, which do not have any data sharing policies. Due to these issues, data sharing and exchange among several organizations is hard to achieve. Data sharing involves a lot of political negotiation, agreements on standards, agreement on costs sharing and agreements on maintenance. Successful GIS and NSDI implementation and adoption usually require some degree of organizational change, which can be very difficult to effect because organizations are naturally resistant to it.

This resistance has many sources, including bureaucratic operating procedures, personal resistance, and personnel habits. Making the changes to implement the SDI in the organization can be complicated. Multiple transitions may be needed to achieve the desired state.

1. **Language barriers**
Multilingual adaptation presents an important problem for configuration management. The software use must support a multiple-value variable for each of the different languages in the system. However, multilingual support issues are not usually considered part of the configuration management processes. People traditionally provide the translation capability. It is necessary to identify clearly the new or modified functionality in order to be able to coordinate the translation processes.

Multilingual support of data offered by Web applications presents several problems. For instance, catalog services should support cross language information retrieval. There are many geoinformation resources that are cataloged using only one language, but users that make their queries in one language, may be interested in resources that have been described in another language. The user is more interested in the resource (map, image or multimedia) than in the metadata describing it. Therefore, catalogs must provide users with mechanisms facilitating multilingual search without forcing cataloging organisations to describe their resources in all possible languages. Most standards for service specifications do not take into account the problems with internationalisation. The main difficulties have been in the internationalisation of legends in Web Map views because the Web Map Service standard does not support the management of names of layers in multiple languages.

m. **Sosio and political stability**
The significance of sosio and political issues is essential that NSDI practitioners need to understand as they determine and contribute to the success of NSDI developments. Developing a successful NSDI initiative depends much upon issues such as political support within the community, clarifying the business objectives which the NSDI is expected to achieve, sustaining a culture of sharing, maintaining reliable financial support and enlisting the cooperation of all members of the community. The communities concerned are expecting to get benefits from their investment in NSDI in terms of improved corporate performances and cooperation. If the success rate of NSDI initiatives is to be improved, it is clear that attention needs to be paid to understanding the community and organisational issues within which NSDI is supposed to be developed.

n. **Metadata availability**
Metadata availability is a serious issue in many NSDI developments. This creates problems to users who want to use the data. Without metadata, they do not know the detailed data contents. The creation of metadata must be given priority and should be done along the data preparation activities. Metadata adds informative and quality aspects to the data. There has been a wide range of metadata standards proposed. ISO 19115 became an international metadata standard for geographic information. In addition, most of the countries and international organisations have proposed different metadata profiles.

o. **Legal arrangements**
The development of NSDI not only comprises of technical aspects but also is supported by economic, social, organisational and legal measures. Three types of policies promote spatial data availability, each with a different purpose, access, reuse and sharing. Among the policies that hinder the availability of spatial data are those dealing with privacy, liability and intellectual property. Intellectual property rights (IPR) in particular endanger the availability of spatial data for access, reuse, and sharing and pose a considerable threat to the development of NSDI. Many agencies use their intellectual property rights on spatial data to gain additional funding for their activities.

p. **Partnership arrangements**
Responsibility for generating, maintaining, and
distributing the data is not widely shared by different levels of government and the private sector. The costs of generating, maintaining, and distributing such data should be justified in terms of public benefits; overlap and duplication efforts among participating organisations are still persists. Access arrangements should recognise confidentiality, privacy, security and intellectual property rights. Upon agreement, partners should contribute equitably to the costs of collecting and managing the data, and should be allowed to integrate the resulting information into their own databases, for their own use and for further distribution to their stakeholders. There should be an attempt to harmonise terms and conditions for use where practical.

4 Conclusion
Development of a spatial data infrastructure is a challenging task, as it requires identification and examination of a large number of issues and challenges. Interviews with expert resulted in identification of sixteen major issues and challenges. From the sixteen issues and challenges, three issues as cited by all the experts give the most impact to the implementation of NDSI namely availability of quality digital data, lack of awareness on GIS and SDI, and institutional arrangements. Since the experts are from local and international arena, most of the issues and challenges identified can be considered common to many countries. All the experts agreed that data is a national capital asset and it should be collected once and shared through NSDI at all level where the usage of spatial data can be done effectively. All the issues and challenges discussed could influence the strategy chosen for SDI development and its effectiveness. It is also a step towards the development of a fully shifting of NSDI to become an infrastructure for services.

References:
GEE-Smoothing Spline Approach on Semiparametric Models for Correlated Nominal Data: Estimation and Simulation Study

1NOOR A. IBRAHIM AND 2SULIADI
1Institute for Mathematical Research and Dept. of Mathematics
Universiti Putra Malaysia.
UPM Serdang, Selangor Darul Ehsan 43400.
MALAYSIA
2Dept. of Statistics
Bandung Islamic University
Jl. Tamansari No. 1 Bandung 40116
INDONESIA

nakma@putra.upm.edu.my, suliadi@gmail.com

Abstract: - In this paper we proposed GEE-Smoothing spline in the estimation of semiparametric models for correlated nominal data. The method can be seen as an extension of parametric generalized estimating equation to semiparametric models. The nonparametric component is estimated using smoothing spline specifically natural cubic spline. We use profile algorithm in the estimation of both parametric and nonparametric components. The properties of the estimators are evaluated using simulation studies.

Key-Words: - Generalized estimating equation, Nominal data, Properties of the estimator, Smoothing spline, Simulation study.

1 Introduction
Nominal data are common in many fields of study, such as in marketing research, biology or in social sciences. The data arise from the situation where the subject has a chance to be at several possibilities of levels. The difference between ordinal and nominal data is that in ordinal data the levels can be ordered whilst the nominal levels cannot. Hence in the modeling, ordinal data can also be modeled as nominal data. Correlated nominal data usually comes from longitudinal or area-based studies. In longitudinal study, the observations from the same subject are usually correlated whereas observations from different subject are independent. Whilst in the area-base study, the observations from the same area are correlated and from different areas are independent.

Since nominal data can be seen as a member of exponential family distribution [1], the general parametric method used in the analysis of these data is the class of generalized linear model (GLM). But GLM assumes that the observations are independent. Reference [2] proposed marginal model for correlated data from the exponential family distribution called generalized estimating equation (GEE). Attention have been made with this method considering the efficiency of the estimators (see: [3]-[6]).

Several authors extended GEE for binary data to GEE for ordinal data. The main attention is in the construction of the covariance matrix. Reference [7] proposed GEE for correlated ordinal data that can be seen as an extension of GEE given by [2] and [3]. They estimated the association parameter (correlation) using the second set of estimating equation for the association. Reference [8] also proposed marginal model for correlated ordinal data. They used global cross (odds) ratio to obtain the association parameter, where the odds ratio is estimated by other set of estimating equation. A more similar method of [7] was given by [9], but they rather estimated the covariance matrix than the correlation matrix. Another approach on GEE for correlated ordinal data was proposed by [10], where the association parameter is obtained as minimization solution of the objective function.

The method discussed above is in the class of parametric models which assume that the relation between the response and covariates can be specified in the form of known linear or non linear function. In this model, the function is known except the (finite) parameters. It is often that the relation of the response with one or more covariates is unknown, that means the relation can be any function but unknown. This leads us to the
nonparametric model. The nonparametric model is usually used in two situations: when the researcher knows that the relation does not follow any linear or non linear parametric function or, to explore the relationship between the response and the covariate.

Semiparametric model is a combination between parametric and nonparametric model. In this model some covariates are modeled parametrically and others are modeled nonparametrically. Several semiparametric methods for correlated data have been proposed. Reference [11] proposed local polynomial kernel (LPK)-GEE to analyze longitudinal data and studied the property of the estimator. They used profile-based estimating equation as proposed by [12]. They found that if independent working correlation is used, then it would produce \( \hat{\theta}_{\text{GEE}} \) consistent parameter estimate of the parametric component. If other working correlation is used, then the parameter estimate of the parametric component. If other working correlation is used, then the parameter estimate of the parametric component is \( \hat{\theta}_{\text{GEE}} \)-inconsistent, except that the nonparametric function is under smoothing. Thus in LPK-GEE, one must ignore the within subject correlation in order to obtain an efficient estimator. This means within subject observations should be assumed independent, hence the working correlation matrix must be an identity matrix. This result was definitely different to GEE given by [2], that is the estimator is consistent even if incorrect working correlation is used.

Reference [13] studied in-consistency of LPK-GEE by studying the locality of kernel regression and comparing it with P-spline and smoothing spline. Their result is that the kernel is local for both independent and non-independent working correlation. P-spline and smoothing spline are local if data are independent and non-local when data are correlated. Motivated by this result [14] proposed nonparametric regression for correlated binary data using GEE-Smoothing spline. Reference [15] also proposed semiparametric estimation using GEE-Smoothing spline for correlated binary data.

In this paper we propose semiparametric estimation for correlated nominal data using GEE-Smoothing spline. This method can be seen as an extension of GEE-Smoothing spline from binary data [15] to nominal data, or an extension of parametric GEE for ordinal data given by [7] to semiparametric estimation for nominal data. We use profile algorithm in the estimation of parametric and nonparametric components. We evaluate the properties of the estimator using simulation study.

The outline of this paper is as follows. Section 2 gives brief review of generalized estimating equation and nonparametric regression using smoothing spline. The GEE-Smoothing spline for correlated nominal data is given in Section 3. Section 4 gives the simulation studies and their results. The conclusion and discussion are given in Section 5.

2 Generalized Estimating Equation and Smoothing Spline
2.1 Generalized Estimating Equation
Generalized estimating equation was proposed by [2]. This method can be seen as an extension of quasi-likelihood by introducing working correlation into the estimating equation [16]. Suppose there are \( n \) subjects and each subject is observed \( n_i \) times, for \( i = 1, 2, ..., n \). The response for the \( i-th \) subject at \( j-th \) time is \( y_{ij} \) with respective vector of covariates \( x_{ij} \) for \( j = 1, 2, ..., n_i \). It is assumed that the responses from the same subject are correlated and from the different subjects are independent. Let the marginal distribution of \( y_{ij} \) follows exponential family distribution with probability distribution function

\[
 f(y_{ij}) = \exp\left( \frac{y_{ij} \theta_i - b(\theta_i)}{a' (\theta_i)} + c(y_{ij}, \phi) \right)
\]

where \( \theta_i \) is the canonical parameter. The first two moments of \( y_{ij} \) are

\[
 E(y_{ij}) = b'(\theta_i) = \mu_i \quad \text{and} \quad Var(y_{ij}) = b''(\theta_i)a(\phi).
\]

The relationship between \( \mu \) and covariates through a link function

\[
 g(\mu_i) = \eta_{ij} \quad \text{with} \quad \eta_{ij} = x_{ij}^T \beta,
\]

where \( \beta = (\beta_1, \beta_2, ..., \beta_p)^T \) is \( p \times 1 \) vector of regression coefficient. Let \( y_i = (y_{i1}, y_{i2}, ..., y_{in_i})^T \) be the \( n_i \times 1 \) vector of response variable with \( E(y_i) = \mu_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{in_i})^T \) and \( X_i = (x_{i1}, ..., x_{in_i})^T \) be \( n_i \times p \) matrix of covariate for the \( i-th \) subject, and \( x_{ij} = (x_{ij1}, x_{ij2}, ..., x_{ijn_i})^T \). Let \( R_i(\alpha) \) be \( n_i \times n_i \) matrix which fulfills the requirement of being correlation matrix and \( \alpha \) be \( s \times 1 \) vector which fully characterizes the \( R_i(\alpha) \). The \( R_i(\alpha) \) is called “working correlation matrix”. Defined

\[
 V_i = A_i^{1/2} R(\alpha) A_i^{1/2}
\]

where \( A_i = \text{diag}\{\hat{\sigma}^2(y_{ij})\} = \text{diag}\{\hat{\sigma}^2(y_{ij})\} \). The estimating equation for \( \beta \) is defined as
\[ \sum_{i=1}^{n} D_i^2 V_i^{-1}(y_i - \mu_i) = 0 \]  

(3)

where \( D_i = \partial \mu_i / \partial \beta \) = \((\partial \mu_i / \partial \theta_0)(\partial \theta_0 / \partial \eta_i)(\partial \eta_i / \partial \beta) = A_i A \lambda X_i \) and \( V_i \) as in (2). The estimate of \( \beta \) is obtained as solution of (3). The iteration procedure for \( \beta \) uses modified Fisher scoring algorithm and estimation of \( \alpha \) and \( \phi \) use method of moment. The iterative procedure for \( \beta \) is

\[
\hat{\beta}^{i+1} = \hat{\beta}^i - \left( \sum_{i=1}^{n} D_i (\hat{\beta}^i)^2 V_i^{-1}(\hat{\beta}^i) D_i (\hat{\beta}^i) \right)^{-1} \times \left( \sum_{i=1}^{n} D_i (\hat{\beta}^i)^2 V_i^{-1}(\hat{\beta}^i) (y_i - \mu_i(\hat{\beta}^i)) \right)
\]

(4)

with \( V_i(\hat{\beta}^i) = V_i(\hat{\beta}^i, \hat{\alpha}(\hat{\beta}^i, \hat{\phi}(\hat{\beta}^i))) \). This method treats the association parameter as nuisance. This means that the focus of the study is regression parameter. It produces low efficiency for the correlation parameter. Some more efficient methods to estimate association parameter has been proposed. For more details see [3]-[6]. Reference [2] showed that \( \beta \) is consistent even if we use incorrect working correlation, as long as the mean is correctly specified, but the most efficient estimate of \( \beta \) is obtained if the true correlation structure is used.

2.1 Smoothing Spline

Smoothing spline is one of the nonparametric regression methods. This method is focused on how to obtain the unknown function, that can be any arbitrary function but comes from a specific class of function, by imposing a roughness penalty to the objective function. Reference [17] gave a good introduction to smoothing spline focused in natural cubic spline. They applied the method to independent data: continuous data and general member of exponential family distribution. They also discussed the semiparametric model based on smoothing spline.

Suppose a model with systematic component \( y_i = f(t_i) + \varepsilon_i \) for \( i = 1, 2, ..., n \), where the function \( f \) can be any arbitrary function and be estimated by minimizing an objective function, sum squares residuals \( \text{RSS} = \sum (y_i - f(t_i))^2 \). Without any restriction, we can take \( f(t_i) = y_i \) and \( S(\lambda) = 0 \). But the estimate is a very rough function and it is only an extrapolation of the data. One way to avoid too rough \( f \) is by imposing roughness penalty to the objective function. There are several measurements of smoothness of a curve, one of them is the integrated square of the second derivative of \( f \), \( \int |f''(t)|^2 dt \). Thus now the objective function is

\[
S(\lambda) = \sum_{i=1}^{n} (y_i - f(t_i))^2 + \lambda \int |f''(t)|^2 dt ,
\]

and known as penalized sum square. The objective function has two components, (1) the goodness-of-fit of data and the roughness penalty. Reference [17] gave procedure to obtain the roughness penalty \( \int |f''(t)|^2 dt \) for the class of natural cubic spline as follows.

Suppose given \( n \) real numbers \( t_1, t_2, ..., t_n \) in the interval \([a, b]\). A function \( f \) in the interval \([a, b]\) is cubic spline if two conditions are satisfied: (i) in each interval \([a, t_1], (t_1, t_2], ..., (t_n, b)\), the function \( f \) is polynomial cubic spline, (ii) the polynomial pieces fit together at point \( t_i \) in such away \( f, f', f'' \) are continuous at each \( t_i \). Hence the function \( f \) is continuous in \([a, b]\). Cubic spline is called natural cubic spline if its second and third derivative of \( f \) is zero at \( a \) and \( b \), thus \( f \) is a linear function on two extreme interval \([a, t_1] \) and \([t_n, b] \). Suppose \( f_i = f(t_i) \), and \( f_i = f'(t_i) \) and by definition \( \gamma_1 = \gamma_n = 0 \). Let \( f \) be \( n \times 1 \) vector of \( (f_1, f_2, ..., f_n)^T \) and \( \gamma \) be the \((n-2) \times 1 \) vector of the second derivative \((\gamma_2, \gamma_3, ..., \gamma_{n-1})^T \). Vector \( f \) and \( \gamma \) completely specify the curve \( f \). These two vectors are specified by two matrices \( Q \) and \( R \) which are defined as follows.

Let \( h_i = t_{i+1} - t_i \) for \( i = 1, 2, ..., n-1 \). Let \( Q \) be the \( n \times (n-2) \) matrix with elements \( q_{ij} = \delta_{ij} \), \( i = 1, 2, ..., n \) and \( j = 2, 3, ..., n-1 \), given by \( q_{ij} = h_i \delta_{ij}, \gamma_q = -h_i \delta_{ij} - h_j \delta_{ij} \), and \( q_{ij} = h_i \delta_{ij} \). The symmetric matrix \( R \) is \((n-2) \times (n-2) \) matrix with elements \( r_{ij} \), for \( i \) and \( j \) running from 2 to \((n-1) \), given by

\[
r_{ij} = \frac{1}{2} (h_i + h_j); \quad r_{ii} = r_{ij} = \frac{1}{6} h_i ;
\]

for \( i = 2, 3, ..., n-1 \).

Matrix \( R \) and \( Q \) are numbered in non standard way. The matrix \( R \) is strictly diagonal dominant, in which \(|r_{ij}| > \sum_{j \neq i} |r_{ij}| \). Thus \( R \) is strictly positive definite, hence \( R^T \) exists. Define a matrix \( K \) by

\[
K^* = QR^{-1}Q^T.
\]

The important result is a theorem given [17] that: The vector \( f \) and \( \gamma \) specify a natural cubic spline \( f \), if and only if the condition

\[
Q^T f = R \gamma
\]

is satisfied. If the above condition is satisfied then the roughness penalty will satisfy

\[
\int |f''(t)|^2 dt = \gamma^T R \gamma = f^T K^* f.
\]

(5)
From the theorem above, the roughness penalty can be computed easily, since it is just multiplication of vector and matrix. Using (1), \( f \) can be easily estimated (see [17] for more detail).

3 GEE-Smoothing Spline for Correlated Nominal Data

Suppose there are \( n \) subjects, and each subject is observed \( n_i \) times. The response for the \( i \)-th subject and \( j \)-th measurement is \( y_{ij} \) with the respective \( p \times 1 \) vector of parametric covariates \( x_{ij} = (1, x_{ij1}, ..., x_{ijp(i)})' \), and nonparametric covariate \( t_{ij} \) for \( i = 1, 2, ..., n \) and \( j = 1, 2, ..., n_i \). It is assumed that the response has nominal scale, with levels \( 1, 2, ..., q \). Since \( y_{ij}^* \) has \( K = q + 1 \) levels, then we can define a new dummy variable, \( y_{ijr} \), for \( r = 1, 2, ..., q \). The new variables are defined by \( y_{ijr} = 1 \) if \( y_{ij}^* = r \), and \( y_{ijr} = 0 \) otherwise. The first and the second moment of \( y_{ijr} \) are \( E(y_{ijr}) = \mu_{ijr} \) and \( \text{Var}(y_{ijr}) = \mu_{ijr}(1 - \mu_{ijr}) \) respectively. It should be noted that \( y_{ijr} \) and \( y_{ijr'} \) are not independent, with \( \text{Cov}(y_{ijr}, y_{ijr'}) = \mu_{ijr}\mu_{ijr'} \) for \( r \neq r' \) ([11], [18]). The marginal expectation of \( \mu_{ijr} \) is linked to the covariates through a logit link function:

\[
\logit(\mu_{ijr}) = \log \left( \frac{\mu_{ijr}}{1 - \mu_{ijr}} \right) = \eta_{ijr}.
\]

The semiparametric model takes the form

\[
\eta_{ijr} = \beta_{ir0} + \beta_{ir1}x_{ij1} + \beta_{ir2}x_{ij2} + \cdots + \beta_{irp(i)}x_{ijp(i)} + f_i(t_{ij}), \quad (7)
\]

In model (7), we model the covariates \( x_{ij1}, x_{ij2}, ..., x_{ijp(i)} \) parametrically and the covariate \( t \) non-parametrically. We assume that \( f \) is a smooth function and \( \int [f''(t)]^2 \, dt < \infty \). In estimation it is better to write \( f(t_{ij}) \) in (7) in the form of vector and matrix, that can be constructed as follows. Suppose that all \( t_{ij} \)'s have \( s \) different values, for all \( i \) and \( j \), denoted by \( t(1) < t(2) < ... < t(s) \) and defined \( s \times 1 \) incidence vector \( N_q = (N_{ij1}, N_{ij2}, ..., N_{ijq})' \), where \( N_{iju} = 1 \) if \( t_{ij} = t(u) \) and \( N_{iju} = 0 \) otherwise, for \( u = 1, 2, ..., s \). Let an \( s \times 1 \) vector \( f_r = (f_1(t(1)), f_1(t(2)), ..., f_1(t(s)))' \), such that \( f(t_{ij}) = n_{ijr}'f_r \), and \( \beta_r = (\beta_{ir0}, \beta_{ir1}, ..., \beta_{irp(i)})' \). Hence the semiparametric model (7) has form

\[
\eta_{ijr} = n_{ijr}'\beta_r + n_{ijr}'f_r. \quad (8)
\]

For generalization purposes, we define new vectors and matrices. Let \( y_{ij} = (y_{ij1}, y_{ij2}, ..., y_{ijq})' \), with \( E(y_{ij}) = \mu_{ij} = (\mu_{ij1}, \mu_{ij2}, ..., \mu_{ijq})' \), \( \eta_{ij} = (\eta_{ij1}, \eta_{ij2}, ..., \eta_{ijq})' \), \( X_i = [I_q \otimes x_{i1}'], N_i = [I_q \otimes N_{i1}'], \beta = (\beta_1, \beta_2, ..., \beta_p)' \), \( f = (f_1', f_2', ..., f_q')' \). Also let \( \mu_i = (\mu_{i1}', \mu_{i2}', ..., \mu_{iq}')' \), with \( E(y_i) = \mu_i = (\mu_{i1}', \mu_{i2}', ..., \mu_{iq}')' \).

References [2], [3], [7], [8], [10] and others did not mention how the GEE was obtained. They just gave the estimating equation for the parameter of interest. But [16] considered that GEE is an extension of quasi-likelihood, by introducing working correlation in the estimating equation. From this, the justification can be made by pretending that there exists a (like) quasi-likelihood function, \( \Gamma \), such that the GEE in (3) is the result of maximizing \( \Gamma \) with respect to \( \beta \), i.e. \( \partial \Gamma / \partial \beta \). If we apply this to our model in (8), the result will not be satisfactory since we can take \( \beta = 0 \) and take \( f \) that interpolates the data, that maximizes \( \Gamma \). These are not satisfactory since the estimate of \( \beta \) is the main objective of the study, and the estimate of \( f \) will be too rough or wiggly. To handle this, we can give the roughness penalty to the (like) quasi-likelihood function, \( \Omega \), such that the estimate of \( f \) will not rough but smooth enough. Assuming that \( \Gamma \) exists, we may define penalized (like) quasi-likelihood function, \( \Omega \), given by

\[
\Omega = \Gamma - (1/2)\lambda \|f''(t)\|^2 dt,
\]

and using (5), it becomes

\[
\Omega = \Gamma - (1/2)\lambda f'Kf, \quad (9)
\]

where \( \lambda \) is a smoothing parameter and matrix \( K = [I_q \otimes K'] \), with \( K \) is defined as in Section 2.2. We use profile algorithm to estimate the parametric (\( \beta \)) and nonparametric components (\( f \)). For more detail about profile algorithm, see [19], [20], [12], and [21]. Profile algorithm treats both components in different manner. The nonparametric component \( f \) is estimated by assuming that \( \beta \) is given. But to estimate \( \beta \), it is assumed that \( f \) is a function of \( \beta \), i.e \( f_{\beta} \). The estimates of \( f \) and \( \beta \) are obtained by maximizing penalized (like) quasi-likelihood function, \( \Omega \), w.r.t \( f \) and \( \beta \). It should be noted that \( \partial \Omega / \partial f = 0 \) and \( \partial \Omega / \partial \beta \neq 0 \). From (9) and following (3), the estimating equations are obtained as follows.

\[
S(f) = \frac{\partial \Omega}{\partial f} = \sum_{i=1}^{n} N_i^T A V_i^{-1}(y_i - \mu_i) - \frac{\partial}{\partial f} \left[ \frac{1}{2} f'Kf \right] = \sum_{i=1}^{n} N_i^T A V_i^{-1}(y_i - \mu_i) - \lambda Kf = 0
\]

(10)
\[ S(\beta) = \frac{\partial \Omega}{\partial \beta} \]

\[ = \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T (y_i - \mu_i) - \frac{\partial}{\partial \beta} \left[ \frac{1}{2} \lambda \bar{f}^T K \bar{f} \right] \]

\[ = \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T (y_i - \mu_i) - \bar{f}^T K \bar{f} = 0 \]

where \( A_i = \frac{\partial \mu_i}{\partial \eta_i} = \text{Diag} \{ \Sigma_1, \Sigma_2, \ldots, \Sigma_n \} \) with \( \Sigma_1 = \text{Cov}(y_i) = \text{Diag}(\mu_i) - \mu_i \mu_i^T; V_i = A_i^{1/2} R(\alpha) A_i^{1/2} \) and \( \bar{f} = \partial \bar{f} / \partial \beta \). Matrix \( A_i^{1/2} \) is obtained from decomposition of \( A_i \), such that \( A = A_i^{1/2} A_i^{1/2} \). Matrix \( R(\alpha) \) is the working correlation matrix of \( y_i \), having the form

\[
R(\alpha) = \begin{pmatrix}
I_q & \rho_{11}(\alpha)I_q & \cdots & \rho_{n1}(\alpha)I_q \\
\rho_{11}(\alpha)I_q & I_q & \cdots & \rho_{n1}(\alpha)I_q \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{1n}(\alpha)I_q & \rho_{2n}(\alpha)I_q & \cdots & I_q
\end{pmatrix},
\]

where \( \rho_{ij}(\alpha) = \text{Corr}(y_i^*, y_j^*) \) and \( \alpha \) is vector of association parameter. If we assume independence working correlation, then \( \rho_{ij}(\alpha) = 0 \).

From equation (10) and (11), the iterative procedure using Fisher scoring algorithm for \( \hat{\beta} \) and \( \hat{f} \) are

\[
\hat{\beta}^{(k+1)} = \hat{\beta}^{(k)} + \left[ \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T A_i \left(X_i + N_i \bar{f}\right) + \bar{f}^T K \bar{f} \right]^{-1} \times \left[ \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T (y_i - \mu_i) - \bar{f}^T K \bar{f} \right]
\]

\[
\hat{f}^{(k+1)} = \hat{f}^{(k)} + \left[ \sum_{i=1}^{n}N_i^T A V_{i-1}^T A N_i + \lambda K \right]^{-1} \times \left[ \sum_{i=1}^{n}N_i^T A V_{i-1}^T (y_i - \mu_i) - \lambda K \hat{f} \right]
\]

Following [2], the robust variance estimator for \( \text{Var}(\hat{\beta}) \) and \( \text{Var}(\hat{f}) \) are given by

\[
\text{Var}(\hat{\beta}) = H_{\beta}^{-1} H_{\beta} \quad \text{and} \quad \text{Var}(\hat{f}) = H_{f}^{-1} H_{f} \quad \text{and} \quad \text{Var}(\hat{f}) = H_{f}^{-1} H_{f} \]

where

\[
H_{\beta} = \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T A_i \left(X_i + N_i \bar{f}\right) + \bar{f}^T K \bar{f}
\]

\[
H_{\beta} = \sum_{i=1}^{n}(X_i + N_i \bar{f})^T A V_{i-1}^T (y_i - \mu_i) \times (y_i - \mu_i)^T V_{i-1}^T A_i \left(X_i + N_i \bar{f}\right)
\]

\[
H_{f} = \sum_{i=1}^{n}N_i^T A V_{i-1}^T A N_i + \lambda K
\]

The association parameter is estimated through another estimating equation for \( \alpha \). Defined for all \( i, j \neq j' = 1, 2, \ldots, n_i \),

\[
Z_{n_i \eta_j \eta_j'} = \frac{(y_{ij} - \mu_{ij})(y_{ij'} - \mu_{ij'})}{\left[ \mu_{ij}(1 - \mu_{ij}) \right]^{1/2} \left[ \mu_{ij'}(1 - \mu_{ij'}) \right]^{1/2}}
\]

then \( E(Z_{n_i \eta_j \eta_j'}) = \rho_{ij}(\alpha) \). In order to avoid restriction of the space of correlation, we use transformation

\[
\rho_{ij} = \frac{\exp(\tau^T \alpha) - 1}{\exp(\tau^T \alpha) + 1}
\]

where \( \tau \) is vector of covariates that affect the correlation, that can be the index of time or other covariate. Let

\[
Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{i1}, \ldots, Z_{i1}, \ldots, Z_{i1}, \ldots, Z_{i1})^T
\]

with

\[
E(Z_i) = (\rho_{12}, \rho_{13}, \ldots, \rho_{12}, \ldots, \rho_{12}, \ldots, \rho_{12}, \ldots, \rho_{12}, \ldots, \rho_{12})^T.
\]

The estimating equation for \( \alpha \) is given by

\[
S(\alpha) = \sum_{i=1}^{n} \left( \frac{\partial \rho_{ij}}{\partial \alpha} \right) W_i^T (Z_i - \rho) = 0
\]

where \( W_i = \text{Var}(Z_i) \). Since misspecification of \( W_i \) does not affect the consistency of \( \alpha \), we may use \( W_i = I \) (see [3]-[5]). The iterative procedure to estimate \( \alpha \) is

\[
\hat{\alpha}^{(k+1)} = \hat{\alpha}^{(k)} + \left[ \sum_{i=1}^{n} \left( \frac{\partial \rho_{ij}}{\partial \alpha} \right) W_i^T \left( \frac{\partial \rho_{ij}}{\partial \alpha} \right)^T \right]^{-1} \times \left[ \sum_{i=1}^{n} \left( \frac{\partial \rho_{ij}}{\partial \alpha} \right) W_i^T (Z_i - \rho) \right].
\]

The steps of GEE-Smoothing spline to estimate \( \beta \) and \( f \) using profile algorithm are as follows.

(i) Given \( \beta \), compute equation (13).

(ii) Estimate the association parameter by iterating (14) until \( \hat{\alpha} \) converges and construct \( A_i \) and \( V_i \).

(iii) Compute equation (12).

(iv) Estimate the association parameter by iterating (14) until \( \hat{\alpha} \) converges and construct \( A_i \) and \( V_i \).

(v) Repeat steps (i) - (iv).
**Table 1. Bias of the Parametric Estimates**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>30</td>
<td>AR1</td>
<td>$\beta_{01}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$\beta_{11}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$\beta_{21}$</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>AR1</td>
<td>$\beta_{02}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$\beta_{12}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$\beta_{22}$</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>AR1</td>
<td>$\beta_{03}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$\beta_{13}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$\beta_{23}$</td>
</tr>
</tbody>
</table>

**Table 2. Bias of the Nonparametric Estimates**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>30</td>
<td>AR1</td>
<td>$f_1(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_2(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_3(k)$</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>AR1</td>
<td>$f_4(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_5(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_6(k)$</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>AR1</td>
<td>$f_7(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_8(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_9(k)$</td>
</tr>
</tbody>
</table>

**Note:** * = The 95% confidence interval of the parameter does not cover the zero

**Table 3. Bias of the Nonparametric Estimates**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>EXC</td>
<td>30</td>
<td>AR1</td>
<td>$f_1(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_2(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_3(k)$</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>AR1</td>
<td>$f_4(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_5(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_6(k)$</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>AR1</td>
<td>$f_7(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXC</td>
<td>$f_8(k)$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IND</td>
<td>$f_9(k)$</td>
</tr>
</tbody>
</table>

**Note:** * = The 95% confidence interval of the parameter does not cover the zero
### Table 3. Variance of the Parametric Estimates

<table>
<thead>
<tr>
<th>True Corr.</th>
<th>Sample Size</th>
<th>Working Corr.</th>
<th>Parameter</th>
<th>$\beta_{01}$</th>
<th>$\beta_{11}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{02}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{22}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>30</td>
<td>$AR1$</td>
<td></td>
<td>0.819</td>
<td>0.826</td>
<td>0.645</td>
<td>0.274</td>
<td>0.734</td>
<td>0.528</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.812</td>
<td>0.893</td>
<td>0.651</td>
<td>0.286</td>
<td>0.807</td>
<td>0.551</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.868</td>
<td>0.972</td>
<td>0.632</td>
<td>0.296</td>
<td>0.923</td>
<td>0.536</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td></td>
<td>0.353</td>
<td>0.410</td>
<td>0.363</td>
<td>0.159</td>
<td>0.378</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.357</td>
<td>0.430</td>
<td>0.364</td>
<td>0.164</td>
<td>0.403</td>
<td>0.357</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.380</td>
<td>0.489</td>
<td>0.360</td>
<td>0.176</td>
<td>0.533</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td></td>
<td>0.158</td>
<td>0.201</td>
<td>0.155</td>
<td>0.070</td>
<td>0.167</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.160</td>
<td>0.213</td>
<td>0.156</td>
<td>0.071</td>
<td>0.176</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.191</td>
<td>0.256</td>
<td>0.161</td>
<td>0.082</td>
<td>0.255</td>
<td>0.135</td>
</tr>
<tr>
<td>EXC</td>
<td>30</td>
<td>$AR1$</td>
<td></td>
<td>0.681</td>
<td>1.211</td>
<td>0.854</td>
<td>0.293</td>
<td>0.832</td>
<td>0.789</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.664</td>
<td>1.211</td>
<td>0.863</td>
<td>0.292</td>
<td>0.855</td>
<td>0.796</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.688</td>
<td>1.246</td>
<td>0.850</td>
<td>0.295</td>
<td>0.882</td>
<td>0.786</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td></td>
<td>0.326</td>
<td>0.655</td>
<td>0.388</td>
<td>0.151</td>
<td>0.531</td>
<td>0.385</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.316</td>
<td>0.650</td>
<td>0.395</td>
<td>0.148</td>
<td>0.522</td>
<td>0.387</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.335</td>
<td>0.681</td>
<td>0.383</td>
<td>0.152</td>
<td>0.549</td>
<td>0.380</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td></td>
<td>0.141</td>
<td>0.318</td>
<td>0.181</td>
<td>0.064</td>
<td>0.220</td>
<td>0.177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.139</td>
<td>0.316</td>
<td>0.182</td>
<td>0.064</td>
<td>0.222</td>
<td>0.178</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.143</td>
<td>0.324</td>
<td>0.177</td>
<td>0.065</td>
<td>0.232</td>
<td>0.177</td>
</tr>
<tr>
<td>IND</td>
<td>30</td>
<td>$AR1$</td>
<td></td>
<td>0.544</td>
<td>0.903</td>
<td>0.319</td>
<td>0.336</td>
<td>0.892</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.539</td>
<td>0.892</td>
<td>0.319</td>
<td>0.332</td>
<td>0.881</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.540</td>
<td>0.902</td>
<td>0.318</td>
<td>0.334</td>
<td>0.886</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td></td>
<td>0.317</td>
<td>0.551</td>
<td>0.174</td>
<td>0.180</td>
<td>0.485</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.315</td>
<td>0.549</td>
<td>0.175</td>
<td>0.182</td>
<td>0.485</td>
<td>0.147</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.318</td>
<td>0.552</td>
<td>0.174</td>
<td>0.182</td>
<td>0.486</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td></td>
<td>0.132</td>
<td>0.245</td>
<td>0.084</td>
<td>0.076</td>
<td>0.220</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td></td>
<td>0.132</td>
<td>0.245</td>
<td>0.085</td>
<td>0.076</td>
<td>0.218</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td></td>
<td>0.131</td>
<td>0.245</td>
<td>0.084</td>
<td>0.076</td>
<td>0.219</td>
<td>0.069</td>
</tr>
</tbody>
</table>

### Table 4. Variance of the Nonparametric Estimates

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AR1</td>
<td>30</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td>EXC</td>
<td>30</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td>IND</td>
<td>30</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$AR1$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$EXC$</td>
<td>f(t)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$IND$</td>
<td>f(t)</td>
</tr>
</tbody>
</table>
4 Simulation Study
In order to evaluate the properties of GEE-Smoothing spline estimators, we did simulation studies by considering the bias, consistency, and the efficiency. The consistency can be evaluated using the fact that an estimator is consistent if the variance tends to zero when sample size tends to infinity. In other words, if the variance decreases when sample size increases, then the estimator is consistent. We also investigate whether using the correct correlation structure gives better efficiency compared to independence working correlation.

4.1 Scenarios
We generated correlated nominal data based on algorithm given by [22]. The algorithm employs uncertainty coefficient (U) and Goodman-Kruskall gamma as association parameter. We generated three types of nominal data sets. The first one has autoregressive lag-1 (AR1) correlation structure with the association between the j-th and j'-th time measurements is $U^{AR1}_{ij} = 0.5^{|j-j'|}$. The second one has exchangeable (EXC) correlation structure with the association between the j-th and j'-th time measurements is $U^{EXC}_{ij} = 0.3$, for all $j \neq j'$, and the third is by using independent structure (IND). We consider nominal data with levels 1, 2, and 3. Each subject was measured five times. We used logit link function in the form of (6). The systematic component has form

$$\eta_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + f_i(t_{ij})$$

where $j = 1, 2, .., 5$; $r = 1, 2$; $\beta_0 = -1$; $\beta_1 = 1$; $\beta_2 = 0.5$; $\beta_1 = 1$; $\beta_2 = -1$; $\beta_2 = 0.5$. The covariate $X_{ij1}$ is a time varying covariate and $X_{ij2}$ is a subject specific covariate. The nonparametric covariate for all subject, $t$, is constant for each time measurement, hence $t_{ij} = t_{ij}$ for all $i$ and $j$. We set $t_{ij} = t_j$ with $t_1 = -2, t_2 = -1, t_3 = 0, t_4 = 1, t_5 = 2$. The function $f_i$ is quadratic function with $f_i(t_j) = 0.5(t_j^2 - 2)$ and $f_i(t_j) = -f_i(t_j)$. The covariate $X_{ij1}$ was generated from Uniform(1) and $X_{ij2}$ was generated from Bernoulli(0.5).

We considered three levels subject numbers, $n = 30, 50$, and $100$. For each data set with combination of correlation structure and number of subject, we estimated using three working correlation, autoregressive lag-1 (AR1), exchangeable (EXC), and independence (IND).

4.1 Simulation Results
The Bias. The average of biases of parametric estimates ($\beta_{01}$, $\beta_{11}$, $\beta_{21}$, $\beta_{02}$, $\beta_{12}$, and $\beta_{22}$) are very small (see Table 1). Table 1 shows that the bias of the parametric estimates are very small. The average of biases seems unaffected by the number of subject. Increasing number of subject does not always decrease the bias. In many cases, number of subject 30 gives the smallest average bias, but in others cases, the smallest bias is given by the number of subject 100. Regarding whether the 95% confidence interval of the parameter covers zero, there is no general pattern for the estimates of intercept ($\beta_{01}$ and $\beta_{02}$), except that for independent data with large number of subject the estimates cover zero. This shows that the estimates of the intercept are generally biased, even though the biases are small.

The coefficient estimates for the time varying covariate ($\beta_{11}$ and $\beta_{12}$) show different behavior. The confidence interval for $\beta_{11}$ covers zero for independent data and small correlated data ($U_{EXC}$) with medium and large numbers of subject.

Meanwhile for data with correlation structure $U_{AR1}$, the estimates for all the working correlations and for all the numbers of subject, the confidence intervals cover zero. Whereas, the estimates of $\beta_{12}$ for correlated and independent data are unbiased. These are shown by the confidence intervals of their estimates for all working correlations and all numbers of subject covering zero. From these result it seems that the coefficient estimates of time varying covariate are unbiased.

From the result obtained, the coefficient estimates of subject specific covariate are unbiased. These are showed by the confidence intervals of $\beta_{21}$ and $\beta_{22}$ covering zero for all data structures, all working correlations, and all numbers of subject.

Table 2 gives the average of biases of the pointwise nonparametric estimates ($f_1$ and $f_2$). This table shows that the averages of bias are very small, for all data types, all working correlations, and all numbers of subject. There is no pattern of increasing or decreasing the bias with respect to increasing or decreasing the numbers of subject. There is no guarantee that increasing number of subject gives smaller bias. In many cases the average of bias for number of subject 100 is large than number of subject 30. Generally, when the number of subject is increased then the average of bias tends to a specific value, even small, but greater or smaller then zero.

The consistency. We evaluate the consistency of estimator by studying the behavior of mean square error (MSE) of the estimates with respect to increasing or increasing number of subject. The
estimator is consistent if the variance tends to zero when the number of subject tends to infinity. In other words, if the MSE decreases when number of subject increases, then the estimator is consistent. To save the space, we do not show the MSE here. The MSE of both components have the same behavior where it decreases when the number of subject increases. It does not matter, whether using correct or incorrect working correlation, this pattern holds. Thus the estimator is consistent even if there is miss-specification of the working correlation.

The efficiency. We consider whether using correct working correlation gives more efficient estimate than assuming independence. Evaluation was made by comparing the variances of the estimates of three working correlation, for their respective true correlation structure and number of subject (see Table 3 for the parametric component and Table 4 for the nonparametric component).

When data are independent, there is no general pattern of the variances of parametric components (β01, β11, β21, β02, β12 and β22). Some of the smallest variances are given by the true working correlation, but these are not general. In many cases, the smallest variances are given by incorrect working correlation. The same pattern can be observed for largest variance. The variances obtaining from three working correlation are comparable.

Even in many cases, AR1 or EXC give the same variance as IND. This result shows that when data are independent, using correct or incorrect working correlation gives similar efficiency.

When data are correlated, the efficiency behavior of the coefficients of time varying covariate (β1 and β2) and subject specific covariate are different (β31 and β32). For the respective number of subject and the true correlation structure, the largest variance for the coefficient of time varying covariate is resulted when the estimation uses independence working correlation (data are assumed independent). The smallest is resulted when the estimation uses the true working correlation. In other words, the most efficient estimate is obtained if the working correlation is the true correlation, hence the efficiency of the estimate of coefficient of time varying covariate is affected by the working correlation.

The behavior of the estimates of coefficient of subject specific covariate is different from the estimate of coefficient of time varying covariate. There is no general pattern of their variance. Some estimates with true working correlation give smaller values, but some give the larger values of variance estimate. Using independence working correlation does not always result in the largest variance estimate, in many cases the variance estimates are small. Thus it can be concluded that the efficiency of the estimate of coefficient of subject specific covariate is not affected by the working correlation.

The efficiency of nonparametric components estimates, i.e the pointwise curve estimates of fi and fj are different than parametric components. In the case of correlated data, the largest variance is given by independent working correlation, and the smallest is commonly given by the true working correlation. Whilst for independent data, the efficiency of correct or incorrect working correlation are comparable.

4 Conclusion and Discussion

GEE-Smoothing spline gives good properties when applied to semiparametric model with correlated nominal data. The parametric components are generally unbiased and the nonparametric components are biased eventhough the bias is small. The important result is that both parametric and nonparametric components are consistent, even using incorrect working correlation. This consistency property is important, since we may use this method even the true correlation is unknown. The efficiency property of the parametric estimates is varies between coefficients estimate of time varying covariate and subject specific covariate. The efficiency of the coefficient estimate of subject specific covariate is not affected by correct or incorrect the working correlation structure. Whilst for the coefficient estimate of time varying covariate, the most efficient is obtained if true working correlation is used and assuming independence gives less efficient estimate.

Our simulation study was based on condition that the nonparametric covariate for each time measurement is fixed for all subjects. This implies the nonparametric covariate is the same as time. Other simulation study might be used in the evaluation when the nonparametric covariate may vary for each time and each subject.

References:


Portable Multi-Channel Environmental Data Measurement System for Monitoring Global Warming

M. Amir Abas¹ IEEE Member, A. Khusairy Azim¹, M. Hilmi Fadzil¹, M. Dahlui², A.M.Arshad¹

Universiti Kuala Lumpur British Malaysian Institute¹, CePH Uni.Malaya²
E-mail: amirunikl@gmail.com¹, kucai_azim@gmail.com¹, helmipadel@yahoo.com¹, alizaa@bmi.unikl.edu.my¹, maznahd@ummc.edu.my²

Abstract: - This paper presents the design of data management system for environmental monitoring applications. The system uses 16-Bit Microcontroller as the main board controlling data management unit (logger), Power Supply unit, Solar Tracker block and data storage block. Three sensors were used namely temperature, humidity and ultraviolet. The system has been tested to capture data at two different conditions, industrial area and high altitude area. The measurement was carried out for 30 hours and the results were plotted using Graphic User Interface (GUI) which is also one of the system units under development.

Keywords: logger, microcontroller, solar, pyranometer

1.0 Introduction

Global warming has become the main issues in the world and urgently requires a mechanism to restore the damage that has been done. Pollution causes the increase of earth temperature which severely affects the phenomenon of glazier at north and south poles. COP15 at Copenhagen was recently held to come out with a strong stand for all the nations to agree with a new policy of environmental accord. This shows the immediate actions from all the leaders to curb the climate changes at their own country.

Many industries have been reported fail to comply with the environmental policies. Illegally dumping of toxic waste, emission of smoke to the air, release of untreated waste to the sea and rivers are among the activities that pollute the environment. Actions taken by the authority are usually late until damage has been extensive and irreversible. Environmental pollution would lead to tremendous effects on the health of the people; severe skin disease, respiratory distress, and cancers apart from causing nuisance and reduce the value of properties. The cost to treat pollution where it is possible is expensive. Hence, prevention and early detection of pollution should be practiced. People who are living in polluted and dangerous surroundings should be made aware and given the information on the situations. The information given should be evident based rather than just based on intuitive. Scientific data is certainly required of which special tools should be used for the detection, recording and analysis of the pollutions. A new portable data management system is proposed and developed with low cost for environmental activist includes student, house wife, NGO and etc to buy and use. The system has two main components, data logger and data interpreter. This paper is mainly prepared to highlight the design and the development of the system.

The other sections of this paper are explaining the build up structure of the Sun Solar Tracker, the operation UV sensor and logging technique for the data logger. At the end of the paper a measurement of Sun’ intensity is shown and analyzed to justify the effectiveness of the model developed.

2.0 Data Acquisition System

Figure 1 shows the block diagram of a standard data acquisition system. There are four main units in DAS namely Sensory Devices, Controller, Transmission medium and Data Analyser.

![Fig. 1: Data Acquisition System](image-url)
One of the main emerging components in designing a data logger is the data management of the system. There are few types of data management that have been used in DAS [6] [7] such as serial EEPROM, SRAM and MMC. Each of these is studied to identify its effectiveness in managing data for DAS particularly for remote application. The emphases to implement the effective data management are:

- To increase the speed of data captured and data storage.
- To ensure the memory storage devices capable to store huge data.
- To prolong the duration of data capture.

### 3.0 Sensory Devices

The first stage of development is to investigate the function on the three sensor devices namely temperature, humidity and Ultraviolet. The system is adaptable to accept other types of sensor such as PH, Ozone and etc based on the user’s needs. Temperature device DS1820 was used which has capability to function as one wire. The one wire feature simplifies significantly the operation and the complexity of measuring process for the temperature.

The second device used is the humidity sensor. Humidity defines the percentage of water in the air. The device HIH 4000 provides simple construction and direct connection to the controller port A/D Converter.

The third sensor is pyranometer. The pyranometer is manned to sense (Figure 3) and measure *global irradiance*, i.e. diffuse plus *direct* solar irradiance. The pyranometer is mounted on a horizontal surface or in the same plane as a solar heat collector or photovoltaic (PV) panel when the global irradiation on these surfaces is of interest [4]. Hence this UV Sensor can be used to measure UV and devices to indicate the intensity of Sun energy for solar tracker. A typical calibration factor *K of pyranometer* can be expressed as:

\[ K = 160 \text{ mV/ (kW/m}^2) \]

This value means that when the solar irradiance \( S \) is 1 kW/m\(^2\) (typical for a clear, sunny day around noon) the pyranometer will provide an output voltage around 160 mV. If the output voltage is found to be 80 mV, this indicates that the solar irradiance is about 0.5 kW/m\(^2\) = 500 W/m\(^2\). Thus:

\[ S = \frac{U}{K} \]

where *U* is the signal voltage in millivolts and *S* is measured in kW/m\(^2\). Note that forward scattering from clouds may cause values 10-20% higher under extreme conditions.

For the amplification of the voltage output of the pyranometer, an amplifier circuit was designed and attached to the ADC converter channel of the PIC18F8550. The range of the output voltage for pyranometer is in milivolt (0mV to 160mv) which is considerably small. Hence an amplifier is needed to refine the range from 0V to 5V. The gain of the amplifier was set at 26 which has matched range with the voltage of ADC input port.

### 4.0 Solar Power

The system is powered by a portable NiCAD battery 12 V DC 12 AH. It needs to be charged continuously to retain the power. Solar panel 40W is installed to perform as charger and voltage supplier to the system. Due to the position of the Sun is changing for every hour solar trackers unit is developed to ensure the solar panel is facing the Sun through out during the day time.
Pyranometer in section 3 is used as UV measuring device. The data measured is stored in the data management unit and also being used for position calculation of the solar panel. The solar panel is geared by two shafts of stepper motors. Figure 5 illustrates the strategy of tracking by the two stepper motors. The initial calibration starts with horizontal scanning or X-Stepper. The X-Stepper moves for every 15º and measure the irradiance value at each step. This process is repeated for 24 to complete the full round. The 24 readings are compared to select the highest value. The value is then computed to locate the position of the angle. Once this process is done the Y-stepper is activated. The Y-stepper begins from morning pole and moves with 15º step to the evening pole. The movement for each step is triggered for every 1 hour.

5.0 Data Measurement

First measurement was carried out to investigate the effectiveness of the measurement and data analysis. Three environment parameters in section 3 were measured at an industrial area, Bandar Baru Bangi, Selangor. The results were saved in MMC as text file using format as follow:

XX41425316244D014E01
XX41225416244D014E01
XX41025516244F014F01
XX41425516244E014D01

Figure 6 (i) (ii) and (iii) show the result of the voltage output of pyranometer against the hours. In day 1, the measurement starts at 9.00 am until 6.00 pm. Between 11.30 to12.00 pm the graph shows drastic peak of light intensity but the value is below the minimum requirement, 80 mV. Due to the cloudy conditions, there is a drop starting from 1.00 pm until the end of the measurement period. In day 2, the condition is similar as day 1. The strength of sun intensity is good in the first quarter but weak in the afternoon due to raining condition. While in day 3, the measurement took place for 12 hours from 7.00 am to 7.00 pm. The weather was good with less cloud blocking. The graph shows average sun energy from 8.00 am till 6.00 pm. The peak energy was picked up at time 1.30 pm. The drop below minimum value was logged at 6.00 pm due to the sunset.

![Figure 5: Tracking Strategy of Sun Position](image)

![Pyranometer(Vout)](image)

![Vout (Pyranometer)](image)

![PYRANOMETER OUTPUT](image)

Figure 6 (i, ii, iii): Measurement of UV for 3
The results as shown in Figure 6 indicate the typical weather in tropical country. Full sunshine is a main issue which may affect the effectiveness of the system. The average measurement of the solar irradiance is approximately 106mV or equivalent to 0.6625KW/m². Table 1 shows the relationship for pyranometer output, solar irradiance and the percentage of solar energy.

The tracking mechanism is computed using the curve of the graph of Figure 6(iii). The duration of the curve above the min line is the key factor to support the charging time, $T_C$, for the batteries. In normal circumstances the batteries require 8 hours charging time to sustain the power for 3 days.

<table>
<thead>
<tr>
<th>Pyranometer Output (mV)</th>
<th>Solar Irradiance (KW/m²)</th>
<th>Sun Energy Captured (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0625</td>
<td>6.25</td>
</tr>
<tr>
<td>50</td>
<td>0.3125</td>
<td>31.25</td>
</tr>
<tr>
<td>80</td>
<td>0.5</td>
<td>50</td>
</tr>
<tr>
<td>110</td>
<td>0.6875</td>
<td>68.75</td>
</tr>
<tr>
<td>130</td>
<td>0.8125</td>
<td>81.25</td>
</tr>
<tr>
<td>160</td>
<td>1</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 1: Pyranometer Measurements

The DAS consumes 200 mA/H and the capacity of the batteries is 12 A/H and this will be able to support 12/0.2 = 60 Hours which is approximately 3 days. The charging time, $T_C$, is depending on the output of the solar panel. If the power from the solar panel generates 40 W/H or current $I_S = 40/12 = 3.3$ A/H. Total current per day produced by the solar panel is approximately $3.3 \times 8 = 27$A and the DAS only consumes 200 mA x 8 = 1.6 A.

The ratio between DAS consumed power and the power for charging the batteries is 27:1.6 or 6 % power is utilized by the DAS during day time. During night time, for 12 hours the DAS consumes 200 mA x 12 = 2.4 A. The total power left is 27 – 1.6 – 2.4 = 23 A. The balance current is stored in the batteries. However the batteries used in the prototype design is 12 A/H (max) only.

From the calculation it shows that the solar panel has huge power to provide continuous power supply for the whole DAS system. Further more the batteries can support 3 days power supply without recharging and needs only 8 hours for full recharged.

6.0 Case Study
Monitoring Environment Effects of Timber Logging Activities at Cameron Highlands
Cameron Highlands is the largest and most famous hill resort of Malaysia with 1500 meters above sea level. Aggressive timber logging is carried out for new area of plantation. A case study was proposed with the aim to explore the effect of the logging activities to the environmental parameters. Data concerning all the environmental issues includes temperature, humidity, Ultra Violet and air quality will be measured and analyzed. Population of common disease due to sun exposure e.g. skin cancer will also be gathered. Correlation between the changes of environmental parameters issues and the population of the disease will be analyzed at the end of this case study.

The first data measurement has been carried out at Tanah Rata, Cameron Highlands on 30th Dec 2009. The measurement using the developed system was focusing on measuring three environment parameters which include temperature, humidity and ultraviolet. Figure 7 (a,b,c) show the results of the measurement for temperature, humidity and ultraviolet.
The measurement was started at 12.01 pm on 30th Dec until 18.00 pm on 31st Dec (30 Hours). The temperature has lowest value at 16ºC while the humidity (or percentage of water vapor presents in the air) has highest percentage 67% at 4.00 am and lowest percentage 40% at 1.00 pm (Figure 7a). Malaysian Meteorological Department reported the temperature for the same location was in the range of 15 ºC to 23 ºC. While figure 7c shows the irradiance graph which has short hours due to minor error during system set up. The average value from 10.00 am to 2.30 pm is 0.7kW/m² which is above the minimum value for effective charging power.

The three graphs are the outputs of Graphic User Interface (GUI) which has been developed to facilitate the analysis process for all the measured data. The system is developed using open source software, Python. The development is still in progress with additional features are upgraded to enhance the analysis capabilities.

7.0 Conclusion

Low cost data measurement system has been develop and tested. It has multi-channel for user to just plug and play. The data storage, MMC has capability to store huge size of data. A solar tracker system with autonomous feature is also constructed. This feature is certainly required for measurement at remote area without electricity power supply. The system uses batteries as power supply but it may not have sufficient power for the long operation. A mechanism to charge the batteries using solar panel is developed that could retain the power of the batteries continuously. For the data analysis off line analysis using designed GUI was used as the first option in this study. The next alternative is implementing wireless strategy for the data transfer.

Finally the developed data management system could be one of the effective tools to monitor global warming issues. Fast action is the best practice to stop the illegal activities before the environment becomes permanently damaged and harm to human and animals. Awareness campaign is also important issue to educate people for taking care and looking after the earth from devastated global warming issues.

Acknowledgement

This project is fully supported by e-science fund 01-02-13-SF0002. Full compliments are given to MOSTI for the approved fund which makes the project is carried out effectively.

References

Abstract—This paper explores the development of an optimized model to investigate the effects of peak current, pulse on time and pulse off time in electric discharge machining (EDM) performance on titanium alloy utilizing copper tungsten as electrode and positive polarity of the electrode. The experiments are carried out on Ti6Al4V. Experiments were conducted by varying the peak current, pulse on time and pulse off time and the corresponding values of surface roughness (SR) were measured. A mathematical model for correlating influences of these variables and surface finish of the workpiece is developed in this paper. Design of experiments (DOE) method and response surface methodology (RSM) techniques are implemented. The validity test of the fit and adequacy of the proposed models has been carried out through analysis of variance (ANOVA). The obtained results evidence that as the surface roughness increases with peak current and pulse on time increases. The effect of pulse off time on surface roughness changes with peak ampere. The excellent surface finish is investigated in this study at short pulse on time. The optimum machining conditions in favor of surface roughness are estimated and verified with proposed optimized results. It is observed that the developed model is within the limits of the agreeable error when experimental. This result leads to desirable surface roughness and economical industrial machining by optimizing the input parameters.

Keywords Ti-6Al-4V, surface roughness, copper tungsten, positive polarity, DOE, RSM.

1. INTRODUCTION
The selection of appropriate machining conditions for minimum surface roughness during the EDM process is based on the analysis relating the various process parameters to SR. Traditionally this is carried out by relying heavily on the operator’s experience or conservative technological data provided by the EDM equipment manufacturers, which produced inconsistent machining performance. The parameter settings given by the manufacturers are only applicable for the common steel grades. Optimization of the EDM process often proves to be difficult task owing to the many regulating machining variables. A single parameter change will influence the process in a complex way. Thus the various factors affecting the process have to be understood in order to determine the trends of the process variation. The selection of best combination of the process parameters for an optimal surface roughness involves analytical and statistical methods. In addition, the modeling of the process is also an effective way of solving the tedious problem of relating the process parameters to the surface roughness. EDM is one of the most popular non-traditional material removal processes and has become a basic machining method for the manufacturing industries of aerospace, automotive, nuclear, medical and die-mold production [1]. The EDM process normally uses thermal energy to produce heat that melts and vaporizes the workpiece by ionization within the dielectric fluid. The electrical discharges generate impulsive pressure by dielectric explosion to remove the melted material. Therefore, the amount of removed material can be effectively controlled to produce complex and precise machine components. However, the melted material is flushed away incompletely and the remaining material resolidifies to form discharge craters. As a result, machined surface has microcracks and pores caused by high temperature gradient which reduces surface finish quality. It was noticed that various machining parameters influenced surface roughness and setting possible combination of these parameters were difficult to produce optimum surface quality. The influences of the machining parameters including the pulsed current, pulse on time, pulse off time, voltage, dielectric liquid pressure and electrode material have been examined. The usage of titanium and its alloys is increasing in many industrial and commercial applications because of these materials’ excellent properties such as a high strength–weight ratio, high temperature strength and exceptional corrosion resistance [2].
In aerospace industry, titanium alloys have been widely used because of their low weight, high strength or high temperatures stability [3]. Titanium and its alloys are difficult to machine materials due to several inherent properties of the material. In spite of its more advantages and increased utility of titanium alloys, the capability to produce parts products with high productivity and good quality becomes challenging. Owing to their poor machinability, it is very difficult to machine titanium alloys economically with traditional mechanical techniques [4]. The EDM is well-established machining choice for manufacturing geometrically complex or hard material parts that are extremely difficult-to-machine by conventional machining processes [5]. Its unique feature of using thermal energy to machine electrically conductive parts regardless of hardness has been its distinctive advantage for manufacturing of mold, die, automotive, aerospace and surgical components [6]. Thus, titanium and titanium alloy, which is difficult-to-cut material, can be machine effectively by EDM [7]. Proper selection of the machining parameters can result in a higher material removal rate, better surface finish, and lower electrode wear ratio [8]. The EDM of titanium alloy (Ti–6Al–4V) with different electrode materials has been accomplished to explore the influence of EDM parameters on various aspects of the surface integrity of Ti6Al4V [2]. A study has been carried out to develop a mathematical model for optimising the EDM characteristics on matrix composite Al/SiC material [9]. They used response surface methodology to determine the optimal setting of the EDM parameters such as the metal removal rate, electrode wear ratio, gap size and the surface finish. The effect of the thermal and electrical properties of titanium alloy Ti–6Al–4V on EDM productivity has been detected [3].

Improving the surface quality is still a challenging problem that constrains the expanding application of the technology. When new and advanced materials appear in the field, it is not possible to use existing models and hence experimental investigations are always required. Undertaking frequent tests or many experimental runs is also not economically justified. The main performance characteristics in EDM process comprise material removal rate, tool wear rate and surface roughness [10]. It is difficult to acquire the accurate quantification of these performance characteristics because there are various uncertain factors and non-linear terms. Optimal selection of process parameters is very much essential as this is a costly process to increase production rate considerably by reducing the machining time. Thus, the present paper emphasizes the development of models for correlating the various machining parameters such as peak current ($I_p$), pulse on time ($t_i$) and pulse off time ($t_o$) on the most important machining criteria such as surface roughness. Machining parameters optimization for the titanium alloy material Ti-6Al-4V has been carried out using the techniques of design of experiments method and response surface methodology. The effect of input parameters on surface roughness in EDM process of Ti-6Al-4V has been analyzed.

2. EXPERIMENTAL DETAILS

2.1. Experimental Setup

Pulse on time ($t_i$) refers the duration of time (μs) in which the current is allowed to flow per cycle [11]. Pulse off time and also known as pulse interval ($t_o$) is the duration of time (μs) between the sparks. The experiments are carried out utilizing a numerical control programming electrical discharge machine known as “LN power supply AQ55L”. The EDM has the provisions of movement in three axes such as longitudinal (X-axis), lateral (Y-axis) and vertical direction of electrode (Z-axis) and has also a rotary U-axis with maximum rpm ±40. In this effort, Ti-6Al-4V was selected as the workpiece material and cylindrical Copper Tungsten (CuW) electrode were employed for machining the workpiece. The experimental setup is shown in Figure 1. The machining was usually carried out for a fixed time interval. The listing of experimental parameters is scheduled in Table 1. The weight of the workpiece and electrode before and after machining were measured by a digital balance, AND GR-200 with readability of 0.1mg and the surface roughness was assessed with Perhometer, Mahr Surf PS1. Three observations were taken for each sample and were averaged to get the value of surface roughness, $R_a$. 

108
The surface roughness of the workpiece can be expressed in different ways including arithmetic average ($R_a$), average peak to valley height ($R_z$), or peak roughness ($R_p$), etc. Generally, the SR is measured in terms of arithmetic mean ($R_a$) which according to the ISO 4987: 1999 is defined as the arithmetic average roughness of the deviations of the roughness profile from the central line along the measurement [12]. Arithmetic mean or average surface roughness, $R_a$ is considered in this study for assessment of roughness.

### Table I

**Experimental settings**

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work piece material</td>
<td>Ti-6Al-4V</td>
</tr>
<tr>
<td>Work piece size</td>
<td>25 mm x 25 mm x 20 mm</td>
</tr>
<tr>
<td>Electrode material</td>
<td>Copper Tungsten</td>
</tr>
<tr>
<td>Electrode size (diameter x length)</td>
<td>20 mm x 44 mm</td>
</tr>
<tr>
<td>Electrode polarity</td>
<td>Positive</td>
</tr>
<tr>
<td>Dielectric fluid</td>
<td>Commercial Kerosene</td>
</tr>
<tr>
<td>Applied voltage</td>
<td>120 V</td>
</tr>
<tr>
<td>Servo voltage</td>
<td>70 V</td>
</tr>
<tr>
<td>Flushing pressure</td>
<td>1.75 MPa</td>
</tr>
<tr>
<td>Machining time</td>
<td>30 minutes</td>
</tr>
</tbody>
</table>

### Table II

**Machining parameters and their levels**

<table>
<thead>
<tr>
<th>Designation</th>
<th>Process parameters</th>
<th>Lowest</th>
<th>Medium</th>
<th>Highest</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>Peak Current (A)</td>
<td>2</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>$x_2$</td>
<td>Pulse on time (µs)</td>
<td>10</td>
<td>205</td>
<td>400</td>
</tr>
<tr>
<td>$x_3$</td>
<td>Pulse off time (µs)</td>
<td>50</td>
<td>175</td>
<td>300</td>
</tr>
</tbody>
</table>

### Table III

**Design parameters**

<table>
<thead>
<tr>
<th>Experiment No.</th>
<th>Peak Current (A)</th>
<th>Pulse on time (µs)</th>
<th>Pulse off time (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>-1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>1</td>
<td>-1</td>
</tr>
<tr>
<td>10</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
</tr>
</tbody>
</table>

### 2.2. Design of Experiment

The main objective of the experimental design is studying the relations between the response as a dependent variable and the various parameter levels. It provides a prospect to study not only the individual effects of each factor but also their interactions. The design of experiments for exploring the influence of various predominant EDM process parameters as peak current, pulse on time and pulse off time on the machining characteristics such as material removal rate, surface finish ($R_a$) and electrode wear rate were modeled. In the present work experiments were designed on the basis of experimental design technique using response surface design method. The coded levels for all process parameters used are displayed in Table II. The set of designed experiments to obtain an optimal response utilizing box-behnken type of design is presented in Table III.

### 2.3. Response Surface Approach

In statistics, response surface methodology explores the relationships between several explanatory variables and one or more response variables. The main idea of RSM is to use a set of designed experiments to obtain an optimal response. In this work, RSM is utilized for establishing the relations between the different...
EDM process parameters with a variety of machining criteria and exploring their effects on SR. To perform this task second order polynomial response surface mathematical models can be developed. In the general case, the response surface is described as (1):

\[ Y = C_0 + \sum_{i=1}^{n} C_i x_i + \sum_{i=1}^{n} C_{ii} x_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij} x_i x_j \]

where, \( Y \) is the corresponding response (SR) yield by the various EDM process variables and the \( x_i \) \((1,2, \ldots, n)\) are coded levels of \( n \) quantitative process variables, the terms \( C_0 \), \( C_i \), \( C_{ii} \) and \( C_{ij} \) are the second order regression coefficients. The second term under the summation sign of this polynomial equation is attributable to linear effect, whereas the third term corresponds to the higher-order effects; the fourth term of the equation includes the interactive effects of the process parameters. Equation (1) can be rewritten according to the three variables used as:

\[ Y = C_0 + C_1 x_1 + C_2 x_2 + C_3 x_3 + C_{11} x_1^2 + C_{22} x_2^2 + C_{33} x_3^2 + C_{12} x_1 x_2 + C_{13} x_1 x_3 + C_{23} x_2 x_3 \]

where: \( x_1 \), \( x_2 \) and \( x_3 \) are peak current (\( I_p \)), pulse on time (\( t_o \)) and pulse off time (\( t_i \)) respectively.

Equation of the fitted model for SR is represented in (3):

\[ SR = 4.49862 + 1.93814 I_p + 0.60335 t_i + 0.10604 t_o - 1.15521 I_p t_i - 0.37112 t_i^2 + 0.20309 t_o^2 + 0.77866 I_p t_o + 0.06688 I_p t_o^2 - 0.22058 t_i t_o \]

2.4. Analysis of Variance (ANOVA)

The adequacy of the above three proposed models have been tested on behalf of both cases, linear and quadratic by means of analysis of variance (ANOVA) as shown in Table IV. The variance is the mean of the squared deviations about the mean or the sum of the squared deviations about the mean divided by the degrees of freedom. The fundamental technique is a partitioning of the total sum of squares and mean squares into components such as data regression and its error. The number of degrees of freedom can be partitioned in Table IV. The adequacy of a model is carried out by computing the F-ratio of the lack of fit to the pure error and comparing it with the standard value. The values of \( P (<\alpha \text{-level}) \) in the analysis ascertain that the regression model is significant. The P-value of the residual error in quadratic term 0.116 for SR is not less than \( \alpha \)-level (0.05). The results of the analysis justifying the closeness of fit of the mathematical models are enumerated. Therefore it can be concluded that the evolved models given by (3) has been adequately explained the variation in the machining parameters on SR.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Degree of freedom</th>
<th>Sum of squares</th>
<th>Mean squares</th>
<th>F-ratio</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>3</td>
<td>33.0534</td>
<td>11.017</td>
<td>12.94</td>
<td>0.001</td>
</tr>
<tr>
<td>Quadratic</td>
<td>9</td>
<td>41.2782</td>
<td>4.5865</td>
<td>20.07</td>
<td>0.002</td>
</tr>
<tr>
<td>Error</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear</td>
<td>11</td>
<td>9.3674</td>
<td>0.8516</td>
<td>22.78</td>
<td>0.043</td>
</tr>
<tr>
<td>Quadratic</td>
<td>5</td>
<td>1.1426</td>
<td>0.2285</td>
<td>7.75</td>
<td>0.116</td>
</tr>
<tr>
<td>Total</td>
<td>14</td>
<td>42.4208</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Results and Discussion

Figure 2 displays the influences of peak current and pulse on time on surface roughness. It is observed form the plots that the increase of peak current increases the surface roughness. This is due to the fact that when pulse current increase, more intensely discharges strike the surfaces and a great quantity of molten and floating metal suspended in the electrical discharge gap during EDM. As well as in a given pulse duration, the thermal energy which is induced in the workpiece through the spark is increased with pulse current. The higher the energy content of each spark, the more violent is the process, thereby generating a rougher surface. Thus increase peak ampere deteriorates the surface finish of the workpiece. The same observation has been reported by [13]. Similarly surface roughness increases as the pulse on time increase. Long pulse duration causes the more heat transfer into the sample and the dielectric fluid is unable to clear away the molten material, as the flashing pressure is the constant. In other words, while the pulse on time is increased the melting isothermals penetrate further into the interior of the material and the molten zone extends further into material and this produce a greater white layer thickness. As a result the increasing pulse on time increases the surface roughness and the fine surface finish is achieved at low pulse on time. It can also be supported by [2] and [14].
Fig. 2 The effect of peak current and pulse on time against surface roughness

Fig. 3 shows that the pulse off time increase the surface roughness started to increase little up to about 180 µs pulse interval time and hereafter decreases. If the discharge current is too high, long pulse off time increases the SR. This is due to the fact that the pulse off time must be sufficiently long to acquire a uniform erosion of the material from the surface of the workpiece and stable machining process otherwise a non uniform erosion of the workpiece surface occurs. Another reason is that the long pulse off time furnishes good cooling effect and enough time for flush away the molten material and debris from the gap between the electrode and workpiece. Thus, long pulse off time present fine surface of the workpiece and the same effect is achieved in [15]. It is apparent in this research that the optimal pulse off time on behalf of SR varies with ampere. In an average the finest surface finish is found at 280-300 µs, 230-250 µs and 180 µs pulse off time as the peak current 2-17A, 17-23A and >23A respectively.

3.1. Optimum settings and validation

The settings for titanium alloys have to be further optimized experimentally. It is also aimed to select appropriate machining conditions for the EDM process based on the analysis relating the various process parameters to SR. It is aimed to develop a methodology using an input–output pattern of data from an EDM process to solve both the modeling and optimization problems. The main objective of this research is to model EDM process for optimum operation representing a particular problem in the manufacturing environment where, it is not possible to define the optimization objective function using a smooth and continuous mathematical formula. It has been hard to establish models that accurately correlate the process variables and performance of EDM process. An attempt is fulfilled to estimate the optimum machining setting to build the best possible material removal rate, surface finish and electrode wear rate within the experimental constraints. The obtained optimum values of the parameters in EDM on Ti-6Al-4V utilizing Copper Tungsten as electrode are shown in Table V. Optimum machining parameter combinations for different EDM characteristics are tested and presented in Table VI through confirmation experiments that verify reasonably good concurrence with prediction of response surface method.

<table>
<thead>
<tr>
<th>TABLE V</th>
<th>OPTIMAL SET-UP FOR SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process parameters</td>
<td>Optimized Settings</td>
</tr>
<tr>
<td>Peak current (A)</td>
<td>2</td>
</tr>
<tr>
<td>Pulse on time (µs)</td>
<td>400</td>
</tr>
<tr>
<td>Pulse off time (µs)</td>
<td>232</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE VI</th>
<th>CONFIRMATION TEST AND COMPARISON WITH EXPERIMENTAL RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRIAL NO.</td>
<td>OPTIMUM CONDITIONS</td>
</tr>
<tr>
<td>1</td>
<td>I_p = 2 A, t_o=400 µs and t_i=232 µs</td>
</tr>
<tr>
<td>2</td>
<td>I_p = 2 A, t_o=400 µs and t_i=232 µs</td>
</tr>
</tbody>
</table>
4. Conclusion

This experiment was accomplished to investigate the influence of the peak current, pulse on time and pulse off time on EDM performance characteristics. It was also attempted to formulate mathematical model for surface roughness and finally to detect the optimal settings. The following conclusions have been drawn from the analysis:

i). As peak current increases the surface roughness of the workpiece increases. Likewise the surface roughness increases with pulse on time. The product of high ampere and high pulse on time deteriorate the more surface finish. The superior surface finish is identified in this experiment while the pulse on time <50 µs at all values of peak current. Increase pulse off time preliminary worsen the surface finish up to certain pulse interval and subsequently improves the surface finish. The converse effect of pulse off time on roughness is investigated that as too high discharge current, long pulse off time provides roughest surface finish. The optimum setting of pulse off time that generates minimum roughness is changed with peak ampere.

ii. The empirical values of the EDM parameters for optimum machining efficiency are 2 A discharges current, 400 µs pulse on time and 232 µs pulse off time in the ase of optimal surface finish.

Further study will be continue the gap between the electrode and workpiece and also employing the negative polarity of the electrode to ascertain their effect on EDM performance characteristics of Ti-6Al-4V.

ACKNOWLEDGMENT

The authors would like to thank Universiti Malaysia Pahang for providing laboratory facilities and financial support under project no. RDU090394 and Doctoral Scholarship scheme (GRS 090335).

REFERENCES

Parametric Optimization in EDM of Ti-6Al-4V using Copper Tungsten Electrode and Positive Polarity: A Statistical Approach


Manuscript received October 6, 2010. This work was supported by Universiti Malaysia Pahang under Grant RDU090394.

M.M. Rahman, K. Kadigama and Rosli A. Bakar are with the Faculty of Mechanical Engineering, Universiti Malaysia Pahang, Tun Abdul Razak Highway, 26300 Gambang, Kuantan, Pahang, MALAYSIA (phone: +6094242246; fax: +6094242202; e-mail: mustafizur@ump.edu.my; kumaranc@ump.edu.my; rosli@ump.edu.my).

Md. Ashikur Rahman Khan is PhD Scholar with Faculty of Mechanical Engineering, Universiti Malaysia Pahang, Tun Abdul Razak Highway, 26300 Gambang, Kuantan, Pahang, MALAYSIA (e-mail: ashik_nstu@yahoo.com).

M.A. Maleque with the Department of Manufacturing and Materials Engineering, Faculty of Engineering, International Islamic University of Malaysia, Jalan Gombak, 53100 Kuala Lumpur, MALAYSIA (phone: +603-61965785, Email: maleque@iiu.edu.my).
Abstract: In this paper, a fabricated voltage-controlled oscillator (VCO) with active inductors using TSMC 0.18µm CMOS process is presented. By employing an improved high-Q active inductor to improve the circuit tuning range, the measurement results of the VCO tuning range from 1.26GHz to 2.98GHz with circuit operating frequency at 2.4GHz achieves 71% tuning range and phase noise obtains -90dBc/Hz within 1 MHz frequency offset. The power consumption including buffer is 44.6mW in 1.8V power supply. Occupied chip area is around 0.585 × 0.679 mm².

Key-Words: VCO, Tuning Range, Active Inductor, Phase Noise

1 Introduction

In recent years, the rapid development of wireless communication, such as WLAN, GSM and DVB is demand for integrating systems. Indeed, the fast emerging wireless technology demands low cost and low power integrated RF transceiver. Therefore, at present the popular research for the CMOS process to implement of the radio frequency integrated circuit (RFIC) [1-2]. The benefit of using standard RF CMOS process can cause integrating the whole RF transceiver on a single chip at lower cost. However, a key building block in RF transceivers is the voltage-controlled-oscillator (VCO).

It is main role is to provide a clean source of high-frequency signals in the voltage-controlled oscillator (VCO) block for RF integrated circuits [3]. The indicators of performance requirements have to be making a VCO suitable for wireless applications which low phase noise and large frequency tuning range. In the existing CMOS VCO topologies, a passive planar spiral inductor is used in tune tank of VCO, but the spiral inductors will result in many disadvantages such as large chip area and low quality factor. These issues can be overcome by using an active inductor. An active inductor has higher quality factor and smaller chip area than a spiral inductor and electrical tunable exist in an active inductor. In this paper, we propose a CMOS wide tuning-range VCO based on an improved high-Q active inductor and a NMOS cross-coupled configuration. Using external voltage, the characteristics of VCO can be tuned. The measurement results of the VCO show that the design can obtain well performance in Bluetooth applications.

2 Architecture

2.1 Active Inductor Design

The most common one-port active inductor topology is the grounded active inductor, which is based on the "gyrator theory" containing only two transistors, which generate inductance. It is small-signal equivalent circuits based on a gyrator topology are shown in Fig.1.

Equation (1) is expression of increasing parallel conductance loss of G will reduce the Q-value of the active inductor. Therefore, in order to improve the performance such as the Q-value and the inductance (L), we using the high-Q active inductors with a feedback resistor [3]. The improved high-Q active
A Wide Tuning Range Voltage-Controlled Oscillator with Active Inductors for Bluetooth Applications

Inductor circuit is illustrated in Fig. 2. In this design, by feedback resistance method of based on gyrator theory then to increase Q-factors and reduce the loss, the active inductor is combined with two transistors, a feedback resistor, and current sources. The characteristics of the active inductor can be improved by using the feedback resistance \((R_f)\). By derived, the component value of the equivalent circuit can be expressed refer to (2).

\[
G = g_{m1} + \frac{g_{m2}}{1 + R_f g_{m2}} \\
L = \frac{C_m (1 + R_f g_{m2})}{g_{m1} g_{m2}} \\
R_f = \frac{g_{m2}}{g_{m1} g_{m2}} \\
C = \frac{g_{m2}}{g_{m2}}
\]

From equation (2), the effect of the factor, \((1+R_f g_{m2})\) designed to be a value greater than unity. This factor will result in the equivalent conductance loss \((G)\) to be minimized, as well as an increasing of the equivalent inductance \((L)\) by \((1+R_f g_{m2})\) factor. The result of scattering parameter \((S11)\) performance of the inductor is exposed in Fig. 3. It can be seen that between 1GHz and 5GHz, the curve is inclined to the outside of circle, indicating that the loss is decreased, and the smallest loss at frequency 2.4GHz.

This active inductor was fabricated by a TSMC 0.18µm RF CMOS process with using ADS EDA tool to required frequency 2.4GHz for the application of low-frequency band 2.4 GHz Bluetooth systems, as shown in Fig. 4. It has maximum 90 with scan frequency 1GHz to 5GHz at the 2.4GHz and Fig. 5 show the frequency is tuned of inductance for 3.6nH to 11nH.

**2.2 Voltage Controlled Oscillator with Active Inductors**

Fig. 6 shows the proposed VCO, it is composed with two active inductors and cross-coupled configuration. In this VCO circuit, involving transistors \((M1-M8)\) and feedback resistor \((R_{f1} \text{ and } R_{f2})\) emulate tunable active inductors. The active inductors are combined with the parasitic capacitors to form the resonator.

Transistors M1, M2 and M7, M8 are formed current source to vary the oscillated frequency. The voltage VA and VB of the current source are used to control transistors M1, M2 and M7, M8, respectively. The bias of the VA and the VB are tuned then the inductance of the active inductors will be changed. The various range of oscillated frequency will be
achieved. In the 2.4GHz oscillated frequency, the bias VA and VB are 1V and 0.667V at VDD = 1.8V, respectively. Furthermore, the oscillation frequency and tuning range will be affected by the value of the selected feedback resistance (Rf1, 2). Assume the total equivalent capacitance from the output node of the NIC is \( C_T \), and then the output oscillating frequency can be expended as refer to Eq. (4), the frequency \( \omega_0 \) is the inverse proportion of the feedback resistance \( R_f \), which oscillation frequency will be decreased if the feedback resistances are increased. The cross-coupled configuration, including transistors M9 and M10 produce negative conductance to compensate the loss of the LC-tank.

\[
\omega_0 = \frac{g_m g_{ds}}{\sqrt{(C_{p1} + C_T)(C_{p1}(1 + R_f g_{ds}))}}
\]

(4)

3 Measurement Results

The proposed circuit has been fabricated in the TSMC 0.18µm CMOS technology. Fig.7 showed the photograph of the VCO chip. The chip area including pads is 0.585 x 0.679 mm². The designed VCO including two active inductors operates at a supply voltage of 1.8 V with the total current consumption of 24.6mA.

Fig. 7 Micro photo of VCO

As the VDD = 1.8V and VA = 1V, Vb is swept from 0 to 1 V, the VCO frequency can be tuned from 1.26 to 2.98 GHz, showing in Fig.8.

Fig. 8 Measured oscillation range at \( V_{TB} = 1.8V, V_A = V_B = 0 \) to 1

Fig.9 and 10 plots the measured output power and phase noise by using the Agilent E4407B spectrum analyzer and E5052A signal source analyzer, respectively. The VCO oscillated at 2.4 GHz (VB = 0.667 V) delivers an output power of -6 dBm to the 50-Ω test instrument with a phase noise of -90 dBc/Hz at 1MHz offset frequency.

Fig.9 Measured output spectrum at 2.4GHz

Fig.10 Phase noise of the VCO at 2.4GHz

Brief characteristic of the proposed VCO of measurement results oscillators are shown in Table 1. Table 2 compares with the wideband VCOs published over the past few years. It shows that the tuning-range of the VCO with active inductor owns the wide-band of above 1 GHz.

4 Conclusion

An active inductor LC-tank voltage-controlled oscillator with a wide frequency tuning range has been proposed and implemented in the TSMC-0.18µm 1P6M RF CMOS technology. The wide-band tuning provides a frequency tuning range from 1.26GHz to 2.98GHz and output power -5.3dBm to -18.7dBm. The phase noise with the VCO tuned to 2.4GHz is -90dBc/Hz at 1 MHz frequency offset and the total current consumption of the core with output buffers are 24.6mA. The designed VCO can be used for the applications in low-frequency band of Bluetooth system.
5 Acknowledgment

The authors would like to thank the National Chip Implementation Center (CIC), Hsin-Chu, Taiwan, for chip fabrication and measurement.

### Table 1 Performance Summary of the Wide-Tuning-Range VCOs

<table>
<thead>
<tr>
<th>Process</th>
<th>Power Supply</th>
<th>Power Dissipation</th>
<th>Frequency Range</th>
<th>Tuning Range</th>
<th>Phase Noise</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSMC 0.18-μm RF CMOS technology</td>
<td>1.8V</td>
<td>44.6mW</td>
<td>1.26GHz to 2.98GHz</td>
<td>1.72GHz</td>
<td>-90 dBc/Hz at 1 MHz offset at 2.4GHz</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5.3dBm to -18.7dBm</td>
</tr>
</tbody>
</table>

### Table 2 Performance Summary of the Wide-Tuning-Range VCOs

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Process</th>
<th>Frequency (GHz)</th>
<th>Tuning range (GHz)</th>
<th><em>Phase-Noise (dBc/Hz)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>This work</td>
<td>0.18-μm CMOS</td>
<td>1.26-2.98</td>
<td>1.72</td>
<td>-90</td>
</tr>
<tr>
<td>[5] C. C. Wei</td>
<td>0.18-μm CMOS</td>
<td>2.16-6.16</td>
<td>4</td>
<td>-65 – -85</td>
</tr>
<tr>
<td>[6] L. H. Lu</td>
<td>0.18-μm CMOS</td>
<td>0.5–3.0</td>
<td>2.5</td>
<td>-101– -118</td>
</tr>
<tr>
<td>[7] R. Mukhopadhyay</td>
<td>0.18-μm CMOS</td>
<td>0.5–2.0</td>
<td>1.5</td>
<td>-78– -90</td>
</tr>
</tbody>
</table>

* dBc/Hz at 1 MHz offset frequency

References:


NON-ATOMIC SET MULTIFUNCTIONS

ALINA GAVRILUȚ

"Al.I. Cuza" University, Faculty of Mathematics
Bd. Carol I, No. 11, Iași, 700506, ROMANIA
gavrilut@uaic.ro

Abstract: In this paper, we focus on different important properties in fuzzy (set-valued) measure theory, such as, non-atomicity, diffusion, regularity, order continuity, exhaustivity, countable subadditivity, increasing/decreasing convergence, semi-convexity, Darboux property for different types of set multifunctions defined on a ring of subsets of an abstract, nonvoid space and taking values in the family of all nonvoid closed subsets of a real normed space.

Key-words: non-atomicity, atom, regularity, order continuity, set multifunction, multi(sub)measure, diffusion, semi-convexity, Darboux property, fuzziness.

1 Introduction

It is well known the theoretical and practical importance of non-additive fuzzy measure theory (and especially of some of its problems, such as, for instance, (non-)(pseudo-)atomicity, regularity, decompositions, integrability, extensions, measurability etc.).

Fuzziness has many applications in probabilities (e.g. Dempster [8], Shafer [39], Zadeh [42]), medicine (e.g. Pham, Brandl, Nguyen N.D. and Nguyen T.V. [37] in osteoporotic fractures prediction), artificial intelligence (e.g., Mastorakis [30]), mathematical economics (e.g. Shapley [1]), mathematical economics etc., the study of (pseudo-)atoms and non-atomicity for different types of set functions has intensively developed.

Because of their multiple applications, for instance in non-additive games theory (e.g., Aumann [4, 5], [7], [14-22], [24-27], [31]) and in mathematical economics, statistics, decision theory of games, biology, physics, human decision making etc.

In several recent papers (see [5, 7, 17, 20-22, 25, 31]), in the context of set multifunctions, we introduced and studied notions as atoms, pseudo-atoms, non-atomicity, finitely purely atomicity.

In this paper, we will highlight for different types of set multifunctions: atoms, non-atomicity, fuzziness, regularity, countable (sub)additivity, semi-convexity, Darboux property, extensions by preserving the properties.

2 Basic notions and results

We now introduce the notations and several definitions used throughout the paper. Then we shall present our results.

Let $T$ be an abstract, nonvoid set, $X$ a real normed space, $\mathcal{P}_0(X)$ the family of all nonvoid subsets of $X$, $\mathcal{P}_f(X)$ the family of all nonvoid, closed subsets of $X$, $\mathcal{P}_{bf}(X)$ the family of all nonvoid, closed, bounded subsets of $X$, $\mathcal{P}_{bf,c}(X)$ the family of all nonvoid, closed, bounded, convex subsets of $X$ and $h$ the Hausdorff pseudometric on $\mathcal{P}_f(X)$, which becomes a metric on $\mathcal{P}_{bf}(X)$ ([28]).

It is known (see [28]) that $h(M,N) = \max\{e(M,N), e(N,M)\}$, where $e(M,N) = \sup_{x \in M} d(x,N)$ for every $M,N \in \mathcal{P}_f(X)$ and $d(x,N) = \max\{0, d(x,N)\}$ is the distance from $x$ to $N$ with respect to the distance induced by the norm of $X$.

We define $|M| = h(M,\{0\})$, for every $M \in \mathcal{P}_f(X)$, where $0$ is the origin of $X$.

On $\mathcal{P}_0(X)$ we consider the Minkowski addition $\setminus^{\cdot}$ [28], defined by:

$$M \setminus N = \overline{M} \setminus \overline{N},$$

for every $M, N \in \mathcal{P}_0(X)$, where $\overline{M} \setminus \overline{N}$ is the closure of $M + N$ with respect to the topology induced by the norm of $X$.

By $\mathbb{N}^*$ we mean $\mathbb{N} \setminus \{0\}$ and by $\overline{1,n}$, we mean $\{1, 2, ..., n\}$.

Unless stated otherwise, in what follows, suppose $\mu : C \rightarrow \mathcal{P}_f(X)$ is a set multifunction, with $\mu(\emptyset) = \{0\}$.

We denote $|\mu|$ the extended real valued set function defined by $|\mu|(A) = |\mu(A)|$, for every $A \in C$.

The following notions extend to the set-valued case the classical corresponding notions from [9-11], [34-36]:

**Definition 2.1.** [4, 5], [7], [14-22], [24-27], [31] $\mu : C \rightarrow \mathcal{P}_f(X)$, with $\mu(\emptyset) = \{0\}$ is said to be:
1) order-continuous (briefly, $o$-continuous) with respect to h if $\lim_{n \to \infty} |\mu(A_n)| = 0$, for every sequence of sets $(A_n)_{n \in \mathbb{N}} \subseteq C$, with $A_n \neq \emptyset$.

II) exhaustive with respect to h if $\lim_{n \to \infty} |\mu(A_n)| = 0$, for every pairwise disjoint sequence of sets $(A_n)_{n \in \mathbb{N}} \subseteq C$.

III) decreasing convergent with respect to h if $\lim_{n \to \infty} h(\mu(A_n), \mu(A)) = 0$, for every decreasing sequence of sets $(A_n)_{n \in \mathbb{N}} \subseteq C$, with $\bigcap_{n=1}^{\infty} A_n = A \in C$.

III') increasing convergent with respect to h if $\lim_{n \to \infty} h(\mu(A_n), \mu(A)) = 0$, for every increasing sequence of sets $(A_n)_{n \in \mathbb{N}} \subseteq C$, with $\bigcup_{n=1}^{\infty} A_n = A \in C$.

IV) monotone (or, fuzzy) if $\mu(A) \subseteq \mu(B)$, for every $A, B \in C$, with $A \subseteq B$.

V) a multisubmeasure if it is monotone and

$\mu(A \cup B) \subseteq \mu(A) + \mu(B)$, for every $A, B \in C$, with $A \cap B = \emptyset$ (or, equivalently, for every $A, B \in C$).

VI) a multimeasure if $\mu(A \cup B) = \mu(A) + \mu(B)$, for every $A, B \in C$, with $A \cap B = \emptyset$.

VII) null-additive if $\mu(A \cup B) = \mu(A)$, for every $A, B \in C$, with $\mu(B) = \{0\}$.

VIII) null–null-additive if $\mu(A \cup B) = \{0\}$, for every $A, B \in C$, with $\mu(A) = \mu(B) = \{0\}$.

IX) $h$ – $\sigma$–subadditive if $|\sum_{n=1}^{\infty} |\mu(A_n)| | \leq \sum_{n=1}^{\infty} |\mu(A_n)|$, for every sequence of pairwise disjoint sets $(A_n)_{n \in \mathbb{N}} \subseteq C$, with $\bigcup_{n=1}^{\infty} A_n \in C$.

Definition 2.2. We consider for an arbitrary set multifunction $\mu : C \to \mathcal{P}_f(X)$, with $\mu(\emptyset) = \{0\}$, the variation $\overline{\mu}$ of $\mu$ defined by $\overline{\mu}(A) = \sup \{\sum_{i=1}^{n} |\mu(A_i)| \}$, for every $A \in C$, where the supremum is extended over all finite partitions $(A_i)_{i=1}^{n}$ of $A$.

$\mu$ is said to be of finite variation on $C$ if $\overline{\mu}(A) < \infty$, for every $A \in C$.

Remark 2.3. [14, 16-18] I) $\overline{\mu}$ is monotone and super-additive on $C$.

Also, if $\mu : C \to \mathcal{P}_f(X)$ is a multi(sub)measure, then $\overline{\mu}$ is finitely additive on $C$ and $|\mu|$ is a submeasure in Drewnowski’s sense [11] on $C$.

Moreover, if $\mu : C \to \mathcal{P}_f(X)$ is a multisubmeasure of finite variation, then the following statements are equivalent:

i) $\mu$ is $\overline{\sigma}$–subadditive;

ii) $\mu$ is order continuous;

iii) $\overline{\mu}$ is $\overline{\sigma}$–additive on $C$;

iv) $\overline{\mu}$ is order-continuous on $C$;

II) If $\mu : C \to \mathcal{P}_f(X)$ is of finite variation, then $\mu : C \to \mathcal{P}_b(X)$;

III) $\overline{\mu}(A) \geq |\mu(A)|$, for every $A \in C$.

Remark 2.4. Any fuzzy multimeasure is, particularly, a multisubmeasure.

Evidently, there are multisubmeasures that are not multimeasures.

Proposition 2.5. [14-16] Suppose $\mu : C \to \mathcal{P}_f(X)$ is a multisubmeasure. Then:

i) If $\mu$ is order continuous, then it is increasing convergent and decreasing convergent;

ii) $\mu$ is order continuous if and only if it is decreasing convergent;

iii) If $C$ is a $\sigma$-ring and $\mu : C \to \mathcal{P}_f(X)$ is order continuous, then $\mu$ is exhaustive.

Examples 2.6. i) If $\nu : C \to \mathbb{R}_+$ is a submeasure (finitely additive set function, respectively), then the set multifunction $\mu : C \to \mathcal{P}_f(\mathbb{R})$, defined by $\mu(A) = [0, \nu(A)]$, for every $A \in C$, is a multisubmeasure (multimeasure, respectively), called the multisubmeasure (multimeasure, respectively) induced by $\nu$.

ii) If $m_1, m_2 : C \to \mathbb{R}_+$, $m_1$ is a finitely additive function and $m_2$ is a submeasure (finitely additive function, respectively), then the set multifunction $\mu : C \to \mathcal{P}_f(\mathbb{R})$, defined by $\mu(A) = [-m_1(A), m_2(A)]$, for every $A \in C$, is a multisubmeasure (multimeasure, respectively).

Moreover, since $|\mu(A)| = \max\{m_1(A), m_2(A)\}$, for every $A \in C$, then $\mu$ is order-continuous (exhaustive, respectively) if and only if the same are both $m_1, m_2$.

The following notions extend to the set-valued case the classical ones from [2, 3], [29], [36], [40-41].

Definition 2.7. [5], [7], [17], [20-22], [24, 25], [31]

Let $\mu : C \to \mathcal{P}_f(X)$ be a set multifunction, with $\mu(\emptyset) = \{0\}$.

I) A set $A \in C$ is said to be an atom of $\mu$ if $\mu(A) \supsetneq \{0\}$ and for every $B \in C$, with $B \subset A$, we have $\mu(B) = \{0\}$ or $\mu(A \setminus B) = \{0\}$.

II) $\mu$ is said to be non-atomic if it has no atoms;

III) We say that $\mu$ has the Darboux property if for every $A \in C$, with $\mu(A) \supsetneq \{0\}$ and every $p \in (0, 1)$, there exists a set $B \in C$ such that $B \subset A$ and $\mu(B) = p \mu(A)$;
ii) is semi-convex if for every \( A \in \mathcal{C} \), with \( \mu(A) \supseteq \{0\} \), there is a set \( B \in \mathcal{C} \) such that \( B \subset A \) and \( \mu(B) = \frac{1}{2} \mu(A) \).

IV) \( \mu \) is said to be:

i) finitely purely atomic if there is a finite disjoint family \( (A_i)_{i=1}^n \subset \mathcal{C} \) of atoms of \( \mu \) so that \( T = \bigcup_{i=1}^n A_i \);

ii) purely atomic if there is at most a countable number of atoms \( (A_n)_n \subset \mathcal{C} \) of \( \mu \) so that \( \mu(T \setminus \bigcup_{n=1}^\infty A_n) = \{0\} \) (if \( \mathcal{C} \) is a \( \sigma \)-algebra).

**Proposition 2.8.** i) If \( \mu \) is monotone, then \( \mu \) is non-atomic if and only if for every \( A \in \mathcal{C} \), with \( \mu(A) \supseteq \{0\} \), there exists \( B \in \mathcal{C} \), with \( B \subset A \), \( \mu(B) \supseteq \{0\} \) and \( \mu(A \setminus B) \supseteq \{0\} \).

ii) [21] Any multisubmeasure having the Darboux property is non-atomic.

iii) [18] If \( \mu \) is a multisubmeasure and \( A_0 \) is an atom of \( \mu \), then \( \tau(A_0) = |\mu(A_0)| \).

iv) [6] If a set multifunction \( \mu \) has the Darboux property, then it is semi-convex.

Moreover, if \( \mathcal{C} \) is a \( \sigma \)-ring and \( \mu : \mathcal{C} \to P_{bf}(X) \) is a monotone increasing convergent multisubmeasure, then \( \mu \) has the Darboux property if and only if it is semi-convex.

In the sequel, let \( T \) be a locally compact Hausdorff space, \( \mathcal{C} \) a ring of subsets of \( T \), \( \mathcal{B}_0 \) the Baire \( \delta \)-ring generated by the \( G_\delta \)-compact subsets of \( T \) (that is, compact sets which are countable intersections of open sets) and \( \mathcal{B} \) the Borel \( \delta \)-ring generated by the compact subsets of \( T \).

In what follows, we recall different types of regularity, that we have defined and studied in [14-16], [18, 19] and [26, 27] for different types of set multifunctions with respect to the Hausdorff topology. Regularity is an important property of continuity, which connects measure theory and topology, approximating general Borel sets by more tractable sets, such as, for instance, compact and open sets.

Our study concerning regularity was inspired by a great number of papers. In this sense, we mention the contributions of Denneberg [9], Dinculeanu [10], Narukawa, Murofushi and Sugeno [32], Olejček [33], Pap [34-36], Precupanu [38].

Let \( \mu : \mathcal{C} \to P_{bf}(X) \) be an arbitrary set multifunction, with \( \mu(\emptyset) = \{0\} \).

**Definition 2.9.** 1) Let \( A \in \mathcal{C} \) be an arbitrary set. \( A \) is said to be:

i) \( \mathcal{R}^r \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exist a compact set \( K \subset A \), \( K \in \mathcal{C} \) and an open set \( D \supset A \), \( D \in \mathcal{C} \) such that \( |\mu(D)| < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset D \setminus K \);

ii) \( R^r \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exists a compact set \( K \subset A \), \( K \in \mathcal{C} \) such that \( |\mu(B)| < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset A \setminus K \);

iii) \( R^r \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exists an open set \( D \supset A \), \( D \in \mathcal{C} \) such that \( |\mu(B)| < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset D \setminus A \);

iv) \( R \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exist a compact set \( K \subset A \), \( K \in \mathcal{C} \) and an open set \( D \supset A \), \( D \in \mathcal{C} \) such that \( h(\mu(A), \mu(B)) < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset D \);

v) \( R^r \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exists a compact set \( K \subset A \), \( K \in \mathcal{C} \) and an open set \( D \supset A \), \( D \in \mathcal{C} \) such that \( h(\mu(A), \mu(B)) < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset D \);

v) \( R \)-regular with respect to \( \mu \) if for every \( \varepsilon > 0 \), there exist a compact set \( K \subset A \), \( K \in \mathcal{C} \) and an open set \( D \supset A \), \( D \in \mathcal{C} \) such that \( h(\mu(A), \mu(B)) < \varepsilon \), for every \( B \in \mathcal{C} \), \( B \subset D \);


We shall give several characterizations of the above types of regularity.

**Remark 2.10.** [18] i) Every compact \( K \in \mathcal{C} \) is \( R_1 \)-regular and \( R^r_1 \)-regular and every open \( D \in \mathcal{C} \) is \( R_r \)-regular and \( R^r_r \)-regular.

ii) a) \( \mu \) is \( R^r \)-regular if and only if for every \( A \in \mathcal{C} \) and every \( \varepsilon > 0 \), there are a compact \( K \subset C \), \( K \subset A \) and an open \( D \in \mathcal{C} \), \( D \subset A \) so that \( |\mu(D \setminus K)| < \varepsilon \).

b) \( \mu \) is \( R^r \)-regular if and only if for every \( A \in \mathcal{C} \) and every \( \varepsilon > 0 \), there is a compact \( K \subset C \), \( K \subset A \) so that \( |\mu(A \setminus K)| < \varepsilon \).

c) \( \mu \) is \( R^r_r \)-regular if and only if for every \( A \in \mathcal{C} \) and every \( \varepsilon > 0 \), there is an open \( D \in \mathcal{C} \), \( D \subset A \) so that \( |\mu(D \setminus A)| < \varepsilon \).

III) a) \( \mu \) is \( R \)-regular if and only if for every \( A \in \mathcal{C} \) and every \( \varepsilon > 0 \), there is a compact \( K \subset C \), \( K \subset A \) so that \( e(\mu(D), \mu(K)) < \varepsilon \).

b) \( \mu \) is \( R_1 \)-regular if and only if for every \( A \in \mathcal{C} \) and every \( \varepsilon > 0 \), there is a compact \( K \subset C \), \( K \subset A \) so that \( e(\mu(D), \mu(A)) < \varepsilon \).

**Theorem 2.11.** [14-16], [18], [27] Suppose \( \mu : \mathcal{C} \to P_{bf}(X) \) is a multisubmeasure. Then:
i) \( \mu \) is \( R \)-regular if and only if it is \( R_t \)-regular and \( R_r \)-regular.

ii) \( \mu \) is \( R' \)-regular if and only if it is \( R_t' \)-regular and \( R_r' \)-regular.

iii) If \( \mu \) is \( R_t \)-regular (\( R'_t \)-regular, \( R' \)-regular, respectively), then \( \mu \) is \( R_t' \)-regular (\( R'_t \)-regular, \( R' \)-regular, respectively).

iv) If \( C \) is the ring (or the \( \delta \)-ring) generated by the compact sets or by the compact, \( G_\delta \) sets, then \( \mu \) is \( R_t' \)-regular if and only if \( \mu \) is \( R'_r \)-regular if and only if \( \mu \) is \( R' \)-regular.

**Proposition 2.12.** [15, 16] i) Any \( R_t' \)-regular multisubmeasure \( \mu : C \to \mathcal{P}_f(X) \) is order continuous.

ii) A multisubmeasure \( \mu : B_0 \to \mathcal{P}_f(X) \) is \( R_t' \)-regular if and only if it is order continuous.

Diffuse set functions were studied for instance by Fernandez [12], Gardner [13] and many others. The following notion extends to the set valued case the classical definition of diffusion.

**Definition 2.13.** Let \( \mu : C \to \mathcal{P}_0(X) \) be a set multifunction, with \( \mu(\emptyset) = \{\emptyset\} \).

We say that \( \mu \) is diffused if for every \( t \in T \), with \( \{t\} \in C \), \( \mu(\{t\}) = \{\emptyset\} \).

**Remark 2.14.** If \( T \) is a locally compact Hausdorff space, then \( \{t\} \) is a compact set, hence \( \{t\} \in B \).

The same as in [17], one can easily check:

**Theorem 2.15.** Let be \( A \in B \) with \( \mu(A) \supseteq \{\emptyset\} \) and \( \mu : B \to \mathcal{P}_f(X) \) a \( R_t' \)-regular, fuzzy and null-additive set multifunction.

i) If \( A \) is an atom of \( \mu \), then there exists a compact set \( K_0 \in B \) so that \( K_0 \subseteq A \) and \( \mu(A \setminus K_0) = \{\emptyset\} \).

ii) \( A \) is an atom of \( \mu \) if and only if

\[
(\ast) \exists a \in A \text{ so that } \mu(A \setminus \{a\}) = \{\emptyset\};
\]

iii) \( \mu \) is non-atomic if and only if \( \mu \) is diffused.

Now, one can easily check that, if \( C_1 \subset C_2 \) are two rings and if \( \mu : C_2 \to \mathcal{P}_f(X) \) is a non-atomic multisubmeasure, \( \mu_{\mid C_1} \) may not be non-atomic. Although,

**Theorem 2.16.** [17] Let \( \mu : B \to \mathcal{P}_f(X) \) be a \( R' \)-regular, non-atomic Borel multisubmeasure. Then \( \mu_{\mid B_0} : B_0 \to \mathcal{P}_f(X) \) is also a \( R' \)-regular, non-atomic Baire multisubmeasure.

**Corollary 2.17.** [16, 17] If \( X \) is a Banach space and if \( \nu : B_0 \to \mathcal{P}_f(X) \) is a \( R' \)-regular Baire multisubmeasure, which has atoms, then \( \nu \) uniquely extends to a \( R' \)-regular Borel multisubmeasure \( \mu : B \to \mathcal{P}_f(X) \) which also has atoms.

In what follows, without any special assumptions, suppose \( \mathcal{A} \) is an algebra of subsets of an abstract space \( T, X \) is a real normed space, \( \mu : \mathcal{A} \to \mathcal{P}_f(X) \) is a set multifunction, with \( \mu(\emptyset) = \{0\} \) and \( f : T \to \mathbb{R} \) is a function.

**Definition 2.18.** A partition of \( T \) is a finite family \( P = \{A_i\}_{i=0}^{\infty} \subset \mathcal{A} \) such that \( A_i \cap A_j = \emptyset, i \neq j \) and \( \bigcup_{i=1}^{\infty} A_i = T \).

**Definition 2.19.** \( f \) is said to be:

I) \( \mu \)-totally-measurable on \((T, \mathcal{A}, \mu)\) if for every \( \varepsilon > 0 \) there exists a partition \( P_\varepsilon = \{A_i\}_{i=0}^{\infty} \) of \( T \) such that the following properties hold:

a) \( \mu(A_0) < \varepsilon \) and

b) \( \sup_{t,s \in A_i} |f(t) - f(s)| < \varepsilon \), for every \( i = 1, n \).

II) \( \mu \)-totally-measurable on \( B \in \mathcal{A} \) if the restriction \( f|_B \) of \( f \) to \( B \) is \( \mu \)-totally measurable on \((B, \mathcal{A}_B, \mu_B)\), where \( \mathcal{A}_B = \{A \cap B; A \in \mathcal{A}\} \) and \( \mu_B = \mu|_{\mathcal{A}_B} \).

**Example 2.20.** [23] Suppose \( T \) is a compact metric space, \( B \) is the Borel \( \delta \)-ring generated by the compact subsets of \( T \), \( f : T \to \mathbb{R} \) is continuous on \( T \) and \( \mu : B \to \mathcal{P}_f(X) \) is finitely purely atomic, null-additive, monotone and \( R_t' \)-regular.

Then \( f \) is \( \mu \)-totally-measurable on every atom \( A_i, i = 1, p \) (where \( T = \bigcup_{i=1}^{p} A_i \)).

In the following, suppose \((T, d_T)\) is a locally compact, metric space. We shall give an application of regularity:

**Theorem 2.21.** (Lusin type) [19] Let \( \mu : \mathcal{A} \to \mathcal{P}_f(X) \) be a \( R_t' \)-regular multisubmeasure and \( f : T \to \mathbb{R} \) be \( \mu \)-totally measurable on \( T \).

Then for every \( \varepsilon > 0 \), there is a compact set \( K_\varepsilon \subseteq \mathcal{A} \) so that:

i) for \( \varepsilon > 0 \), there is \( \delta(\varepsilon) > 0 \), so that for every \( t, s \in K_\varepsilon \), with \( d_T(t, s) < \delta(\varepsilon) \), we have \( |f(t) - f(s)| < \varepsilon \) and

ii) \( \mu(T \setminus K_\varepsilon) < \varepsilon \) (i.e., \( f \) is pseudo-quasi-continuous on \( T \)).

3. Concluding remarks. In this paper, we concentrate on the study of different notions, such as, non-atomicity, diffusion, regularity, order continuity, exhaustivity, countable subadditivity, increasing/decreasing convergence for different types of set multifunctions defined on a ring of subsets of an abstract, nonvoid space and taking values in the family.
of all nonvoid closed subsets of a real normed space. We synthetisize the main results in the following two schemes:

\[
\begin{align*}
\text{Darboux property} & \iff \text{semi-convexity} \\
& \iff \text{non-atomicity} \iff \text{diffusion}
\end{align*}
\]

1. \(C\) a \(\sigma\)-ring, \(\mu : C \to P_{f,c}(X)\) a monotone increasing convergent multimeasure
2. \(\mu : B \to P_f(X)\) a \(R^1\)-regular, fuzzy and null-additive set multifunction
3. \(\mu : C \to P_f(X)\) a multisubmeasure

\[
\begin{align*}
R^1\text{-regular} & \iff R^1\text{-regular} \\
& \iff \text{order continuity} \\
& \iff h-\text{regularity}
\end{align*}
\]

1. \(C = B_0\) or \(B, \mu : C \to P_f(X)\) a multisubmeasure
2. \(\mu : B_0 \to P_f(X)\) a multisubmeasure
3. \(\mu : C \to P_f(X)\) a multisubmeasure of finite variation
4. \(\mu : C \to P_f(X)\) a multisubmeasure

References:


Non-Atomic set Multifunctions


Abstract: We present some properties regarding diffusion for fuzzy set multifunctions defined on a ring of subsets of an abstract nonvoid space and taking values in the family of all nonvoid closed subsets of a real normed space.

Key-Words: diffused, fuzzy set multifunction, atom, non-atomicity, regular, multimeasure, semi-convex, null-additive.

1 Introduction

In the last years, because of its utility in a broad spectrum of areas, such as statistics, probability, economy, theory of games, artificial intelligence, computer and system sciences, biology, medicine, physics, human decision making (see, for instance, [3], [8], [27], [30], [32, 33]), a theory of fuzziness became to develop.

In recent years, we were interested in the study of different problems of the classical measure theory (regularity, fuzziness, atoms, pseudo-atoms, non-atomicity, finitely purely atomicity, integrability, continuity properties), which we treated in the set valued case (see, for instance, [1, 2], [4-7], [13-23], [28], [31]).

In this paper we investigate another important property, with many applications, the property of diffusion. Different problems concerning this property of diffusion have been studied for real valued measures for instance by [11, 12], [25, 26].

2 Basic notions

Let $T$ be an abstract nonvoid space, $X$ a real normed space, $\mathcal{P}_0(X)$ the family of all nonvoid subsets of $X$, $\mathcal{P}_f(X)$ the family of all nonvoid closed subsets of $X$, $\mathcal{P}_{bf}(X)$ the family of all nonvoid closed bounded convex subsets of $X$ and $h$ the Hausdorff-Pompeiu pseudometric on $\mathcal{P}_f(X)$. $h$ becomes a metric on $\mathcal{P}_{bf}(X)$ [24].

It is known that $h(M, N) = \max\{e(M, N), e(N, M)\}$, where $e(M, N) = \sup_{x \in M} d(x, N)$, for every $M, N \in \mathcal{P}_f(X)$ is the excess of $M$ over $N$ and $d(x, N)$ is the distance from $x$ to $N$ with respect to the distance induced by the norm of $X$.

On $\mathcal{P}_0(X)$ we consider the Minkowski addition $\overset{*}{+}$ [24], defined by:

$$M \overset{*}{+} N = \overline{M + N},$$

for every $M, N \in \mathcal{P}_0(X)$, where $\overline{M + N}$ is the closure of $M + N$ with respect to the topology induced by the norm of $X$.

We denote $|M| = h(M, \{0\})$, for every $M \in \mathcal{P}_f(X)$, where 0 is the origin of $X$.

The following notions generalize the well-known corresponding notions from the non-additive fuzzy measure theory ([9], [10], [29]).

**Definition 1** ([4, 5], [7], [13-23], [28], [31]) Consider $\mu : \mathcal{C} \to \mathcal{P}_f(X)$ a set multifunction, with $\mu(\emptyset) = \{0\}$.

$\mu$ is said to be:

I) fuzzy (or monotone) if $\mu(A) \subseteq \mu(B)$, for every $A, B \in \mathcal{C}$, with $A \subseteq B$;

II) null-additive if $\mu(A \cup B) = \mu(A)$, for every $A, B \in \mathcal{C}$, with $\mu(B) = \{0\}$;
III) null–null-additive if $\mu(A \cup B) = \{0\}$, for every $A, B \in C$, with $\mu(A) = \mu(B) = \{0\}$;

IV) decreasing convergent with respect to $h$ if $\lim_{n \to \infty} h(\mu(A_n)), \mu(A)) = 0$, for every decreasing sequence of sets $(A_n)_{n \in \mathbb{N}} \subset C$, with $\bigcap_{n=1}^{\infty} A_n = A \in C$;

V) a multimeasure if $\mu(A \cup B) = \mu(A) + \mu(B)$, for every $A, B \in C$, with $A \cap B = \emptyset$.

Remark 2 1. If $\mu : C \to \mathcal{P}_f(X)$ is null-additive, then it is null-null-additive.

The converse is not valid.

Indeed, let $T = \{a, b\}$, $\mathcal{C} = \mathcal{P}(T)$ and $\mu : C \to \mathcal{P}_f(\mathbb{R})$ defined by $\mu(T) = [0, 2], \mu(\{a\}) = [0, \frac{1}{2}]$ and $\mu(\emptyset) = \{0\}$. Then $\mu$ is null-null-additive but it is not null-additive.

II. Any multimeasure $\mu : C \to \mathcal{P}_f(X)$ is null-additive.

The converse is not true. Indeed, let $T = \{a, b\}$, $\mathcal{C} = \mathcal{P}(T)$ and $\mu : C \to \mathcal{P}_f(\mathbb{R})$ defined by $\mu(T) = [0, 2], \mu(\{a\}) = [0, \frac{1}{2}]$ and $\mu(\emptyset) = \{0\}$. Then $\mu$ is null-additive but it is not a multimeasure.

Suppose $\mu : C \to \mathcal{P}_f(X)$ is a set multifunction, with $\mu(\emptyset) = \{0\}$.

By $|\mu|$ we denote the extended real valued set function defined by $|\mu|(A) = |\mu(A)|$, for every $A \in C$.

Definition 3 For a set multifunction $\mu : C \to \mathcal{P}_f(X)$, with $\mu(\emptyset) = \{0\}$, we consider the variation $\overline{\mu}$ of $\mu$ defined by $\overline{\mu}(A) = \sup\left\{ \sum_{i=1}^{n} |\mu(A_i)| \right\}$, for every $A \in C$, where the supremum is extended over all finite partitions $\{A_i\}_{i=1}^{n}$ of $A$.

$\mu$ is said to be of finite variation on $C$ if $\overline{\mu}(A) < \infty$, for every $A \in C$.

Remark 4 1. $\overline{\mu}$ is fuzzy and super-additive on $\mathcal{P}(T)$, the family of all subgroups of $T$.

II. If $\mu$ is fuzzy, then $|\mu|$ is also fuzzy.

The converse is not true. Indeed, let $T = \{a, b\}$, $\mathcal{C} = \mathcal{P}(T)$ and $\mu : C \to \mathcal{P}_f(\mathbb{R})$ defined by $\mu(T) = \{1\}, \mu(\{a\}) = \mu(\{b\}) = [0, 1]$ and $\mu(\emptyset) = \{0\}$. We have $|\mu(A)| = 1$ if $A \neq \emptyset$ and $|\mu(\emptyset)| = 0$. Then $|\mu|$ is fuzzy, but $\mu$ is not fuzzy.

III. If $\mu$ is null-additive, then $|\mu|$ is null-additive.

The converse is not valid.

Indeed, let $T = \{a, b\}$, $\mathcal{C} = \mathcal{P}(T)$ and $\mu : C \to \mathcal{P}_f(\mathbb{R})$ defined by $\mu(T) = [0, 1], \mu(\{a\}) = 1$ and $\mu(\{b\}) = \mu(\emptyset) = \{0\}$. We have $|\mu(A)| = 1$ if $A = T$ or $A = \{a\}$ and $|\mu(A)| = 0$ if $A = \{b\}$ or $A = \emptyset$. Then $|\mu|$ is null-additive, but $\mu$ is not null-additive.

Definition 5 ([4, 5, 7, 13, 16–18, 21, 28]) If $\mu : C \to \mathcal{P}_f(X)$ is a set multifunction, with $\mu(\emptyset) = \{0\}$, then:

I) $A$ set $A \in C$ is said to be an atom of $\mu$ if $\mu(A) \supseteq \{0\}$ and for every $B \in C$, with $B \subseteq A$, we have $\mu(B) = \{0\}$ or $\mu(A\setminus B) = \{0\}$;

II) $\mu$ is said to be non-atomic if it has no atoms.

Remark 6 1. If $\mu : C \to \mathcal{P}_f(X)$ is fuzzy, then $\mu$ is non-atomic if and only if for every $A \in C$, with $\mu(A) \supseteq \{0\}$, there exists $B \in C$, with $B \subseteq A$, $\mu(B) \supseteq \{0\}$ and $\mu(A\setminus B) = \{0\}$.

II. Suppose $\mu : C \to \mathcal{P}_f(X)$ is fuzzy. If $A \in C$ is an atom of $\mu$ and $B \in C$, $B \subseteq A$ has $\mu(A) \supseteq \{0\}$, then $B$ is an atom of $\mu$ and $\mu(A\setminus B) = \{0\}$.

Definition 7 Suppose $m : C \to \mathbb{R}_+$ is an arbitrary set function, with $m(\emptyset) = 0$. We say that a set multifunction $\mu : C \to \mathcal{P}_f(\mathbb{R}_+)$ is induced by $m$ if $\mu(A) = [0, m(A)]$, for every $A \in C$.

Remark 8 Let $\mu : C \to \mathcal{P}_f(\mathbb{R}_+)$ be the set multifunction induced by a set function $m : C \to \mathbb{R}_+$.

Then:

i) $\mu$ is fuzzy if and only if $m$ is fuzzy;

ii) $\mu$ is null-additive (null-null-additive, respectively) if and only if the same is $m$;

iii) $\mu$ is a multimeasure if and only if $m$ is finitely additive;

iv) $\mu$ is non-atomic if and only if $m$ is non-atomic.

3 Diffusion of a set multifunction

We present in the sequel some properties regarding diffusion for set multifunctions.

Let $T$ be a locally compact Hausdorff space, $\mathcal{C}$ a ring of subsets of $T$, $\mathcal{B}_0$ the Baire $\delta$-ring generated by the $G_\delta$-compact subsets of $T$ (that is, compact sets which are countable intersections of open sets) and $\mathcal{B}$ the Borel $\delta$-ring generated by the compact subsets of $T$. It is known that $\mathcal{B}_0 \subset \mathcal{B}$.

Definition 9 ([28]) Let $\mu : C \to \mathcal{P}_f(X)$ be a set multifunction, with $\mu(\emptyset) = \{0\}$. We say that $\mu$ is diffused if for every $t \in T$, with $\{t\} \in C$, $\mu(\{t\}) = \{0\}$.

We recall now different types of regularity that we have defined and studied, for instance, in [14, 22, 23] for different types of set multifunctions with respect to the Hausdorff topology induced by the Hausdorff pseudo-metric.
Definition 10 Let $A \in C$ be an arbitrary set and $\mu : C \rightarrow \mathcal{P}_f(X)$ a set multifunction, with $\mu(\emptyset) = \{0\}$.

i) $A$ is said to be regular with respect to $\mu$ if for every $\varepsilon > 0$, there exists a compact set $K \subseteq A$, $K \in C$ such that $|\mu(B)| < \varepsilon$, for every $B \in C, B \subseteq A \setminus K$.

ii) $\mu$ is said to be regular if every $A \in C$ is a regular set with respect to $\mu$.

Remark 11 1. Let $\mu : C \rightarrow \mathcal{P}_f(\mathbb{R}^+)$ be the set multifunction induced by a set function $m : C \rightarrow \mathbb{R}^+$.

Also, the set functions $m : C \rightarrow \mathbb{R}^+$, induced by them.

II. If $A \subseteq C$ such that $A; K \in C$ be a regular fuzzy null-additive set $\mu(A) = \{0\}$, then $\mu(A) = \{0\}$ and this is a contradiction.

Let us prove now that every set $K \in \mathcal{K}_A$ is an atom of $\mu$. Indeed, if $K \in \mathcal{K}_A$, then $\mu(A \setminus K) = \{0\}$.

Also, for every $B \in B$, with $B \subseteq C$, since $K \subseteq A$ and $A$ is an atom of $\mu$, we get $\mu(B) = \{0\}$ or $\mu(A \setminus B) = \{0\}$.

Let us prove now that $K_1 \cap K_2 \in \mathcal{K}_A$, for every $K_1, K_2 \in \mathcal{K}_A$.

Indeed, if $K_1, K_2 \in \mathcal{K}_A$, then $K_1 \cap K_2$ is a compact set of $T$ and $\mu(A \setminus (K_1 \cap K_2)) = \{0\}$.

It recurrently results:

$$\bigcap_{i=1}^{n} K_i \in \mathcal{K}_A, \forall n \in \mathbb{N}^+, \ K_1, \ldots, K_n \in \mathcal{K}_A.$$

We prove that $K_0 = K \in \mathcal{K}_A$ is a nonvoid set.

Suppose, on the contrary, that $K_0 = \emptyset$.

There are $K_1, K_2, \ldots, K_n \in \mathcal{K}_A$ so that $\bigcap_{i=1}^{n} K_i = \emptyset$. But from (1) it results $\bigcap_{i=1}^{n} K_i \in \mathcal{K}_A$, which is a contradiction.

Now, we prove that $K_0 \in \mathcal{K}_A$. Obviously, $K_0$ is a compact set.

Let be $K \in \mathcal{K}_A$. Then $\mu(A \setminus K) = \{0\}$.

If $K_0 = K$, then $K_0 \in \mathcal{K}_A$.

If $K_0 \neq K$, then $K_0 \subseteq K$.

Because $\mu(A \setminus K_0) = \mu(K \setminus K_0)$, it remains to demonstrate that $\mu(K \setminus K_0) = \{0\}$.

Suppose that, on the contrary, $\mu(K \setminus K_0) \supseteq \{0\}$. Since $K$ is an atom of $\mu$, by Remark 6-II, it follows that $K \setminus K_0$ is an atom of $\mu$.

Because $A$ is an atom of $\mu$ and $\mu(K \setminus K_0) \supseteq \{0\}$, then $\mu(A \setminus (K \setminus K_0)) = \{0\}$.

Let us consider $C \in \mathcal{C}(K \setminus K_0)$.

Then $\mu((K \setminus K_0) \setminus C) = \{0\}$ and, since $\mu(A \setminus (K \setminus K_0)) = \{0\}$, by the null-additivity of $\mu$ we get that $\mu(A \setminus C) = \{0\}$. This implies $C \in \mathcal{C}_A$.

Therefore, $K_0 \subseteq C$, but $C \subseteq K \setminus K_0$, which is a contradiction. Therefore, $\mu(K \setminus K_0) = \{0\}$.

So, indeed, if $A \in B$ is an atom of $\mu$, we find a compact set $K_0 \in B$ so that $K_0 \subseteq A$ and $\mu(A \setminus K_0) = \{0\}$.
(ii) Let $A \in \mathcal{B}$ be an atom of $\mu$.

a) The \textit{existence part}:

We show that the set $K_0$ from the proof of i) is a singleton $\{a\}$.

Suppose that, on the contrary, there exist $a, b \in A$, with $a \neq b$ and $K_0 \supseteq \{a, b\}$.

Since $T$ is a locally compact Hausdorff space, there exists an open neighbourhood $V$ of $a$ so that $b \notin V$.

Obviously, $K_0 = (K_0 \setminus V) \cup (K_0 \cap V)$ and $K_0 \setminus V, K_0 \cap V$ are nonvoid compact subsets of $A$.

We prove that $K_0 \setminus V \in \mathcal{K}_A$ or $K_0 \cap V \in \mathcal{K}_A$.

Indeed, if $K_0 \setminus V \notin \mathcal{K}_A$ and $K_0 \cap V \notin \mathcal{K}_A$, then $\mu(A \setminus (K_0 \setminus V)) \geq \{0\}$ and $\mu(A \setminus (K_0 \cap V)) \geq \{0\}$.

Since $A$ is an atom of $\mu$, then $\mu(K_0 \setminus V) = \{0\}$ and $\mu(K_0 \cap V) = \{0\}$.

Consequently, according to the null-additivity of $\mu$, we get $\mu(K_0) = \{0\}$.

This implies $\{0\} \subseteq \mu(A) = \{0\}$, which is a contradiction.

Therefore, $K_0 \setminus V \in \mathcal{K}_A$ or $K_0 \cap V \in \mathcal{K}_A$.

Because $K_0 \subseteq K$, for every $K \in \mathcal{K}_A$, we get that $K_0 \subseteq K \setminus V$ or $K_0 \subseteq K \cap V$, which is impossible.

So, $3a \in A$ so that $\mu(A \setminus \{a\}) = \{0\}$.

b) The \textit{uniqueness part}:

Suppose that, on the contrary, there are $a, b \in A$, with $a \neq b, \mu(A \setminus \{a\}) = \{0\}$ and $\mu(A \setminus \{b\}) = \{0\}$.

Because $\{b\} \subseteq A \setminus \{a\}$ and $\mu(A \setminus \{a\}) = \{0\}$, by the fuzziness of $\mu$ we get that $\mu(\{b\}) = \{0\}$.

Then, according to the null-additivity of $\mu$, by $\mu(A \setminus \{b\}) = \{0\}$ and $\mu(\{b\}) = \{0\}$, we have $\mu(A) = \{0\}$, which is a contradiction.

Now, consider $A \in \mathcal{B}$, with $\mu(A) \supseteq \{0\}$ having property (P) and let $B \in \mathcal{B}$, with $B \subseteq A$.

If $a \notin B$, then $B \subseteq A \setminus \{a\}$. Because $\mu(A \setminus \{a\}) = \{0\}$ and $\mu$ is monotone, we have $\mu(B) = \{0\}$. If $a \in B$, then $A \setminus B \subseteq A \setminus \{a\}$, so $\mu(A \setminus B) = \{0\}$. Consequently, $A$ is an atom of $\mu$.

(iii) Let $\mu$ be diffused and suppose, on the contrary, there is an atom $A_0 \in \mathcal{C}$ of $\mu$.

By ii), there is an unique $a \in A_0$ so that $\mu(A_0 \setminus \{a\}) = \{0\}$.

On the other hand, by the diffusion of $\mu$, $\mu(A \setminus \{a\}) = \{0\}$.

The null-additivity of $\mu$ implies that $\mu(A_0) = \{0\}$, which is a contradiction since $A_0 \in \mathcal{C}$ is an atom of $\mu$. Consequently, $\mu$ is non-atomic.

Now, let $\mu$ be non-atomic and suppose, on the contrary, that $\mu$ is not diffused, so there is $t_0 \in T$ so that $\mu(\{t_0\}) \supseteq \{0\}$.

Because $\mu$ is non-atomic, there is a set $B \in \mathcal{B}$ such that $B \subseteq \{t_0\}, \mu(B) \supseteq \{0\}$ and $\mu(\{t_0\} \setminus B) \supseteq \{0\}$. Consequently, $B = \emptyset$ or $B = \{t_0\}$, which is false. The proof is thus finished.

\begin{definition}[(28)]
Let $\mu : \mathcal{C} \rightarrow \mathcal{P}_f(X)$ be a set multifunction, with $\mu(\emptyset) = \{0\}$. We say that $\mu$ is semi-convex if for every $A \in \mathcal{C}$, with $\mu(A) \supseteq \{0\}$, there exists $B \in \mathcal{C}$ so that $B \subseteq A$ and $\mu(B) = \frac{1}{2} \mu(A)$.
\end{definition}

The same as in [28] or [13] one can easily prove:

\begin{theorem}
Let $C = B_0$ (or $B$). If $\mu : \mathcal{C} \rightarrow \mathcal{P}_{bfc}(X)$ is a fuzzy decreasing convergent semi-convex multifunction, then for every $t \in T$, there exists $A_t \in \mathcal{C}$ so that $t \in A_t$ and $\mu(A_t) = \{0\}$.
\end{theorem}

\begin{theorem}[(28)]
If $\mu : B \rightarrow \mathcal{P}_f(X)$ is a diffused regular fuzzy multimeasure, then $\mu/\mathcal{B}_0$ satisfies the condition:

\[ \text{for every } t \in T, \text{there exists } A_t \in \mathcal{B}_0 \text{ so that } t \in A_t \text{ and } \mu(A_t) = \{0\}. \]

\begin{corollary}[(28)]
Let $\mu : B \rightarrow \mathcal{P}_f(X)$ be a regular fuzzy multimeasure. Then $\mu$ is diffused on $B$ if and only if for every $t \in T$, there exists $A_t \in \mathcal{B}_0$ so that $t \in A_t$ and $\mu(A_t) = \{0\}$.
\end{corollary}

4 Concluding remarks

In this paper, results concerning fuzzy diffused set multifunctions are discussed in the context of different notions such as atoms, non-atomicity, regularity, semi-convexity, decreasing convergence.

References:


The Rademacher Complexity Model over Acoustic Features for Improving Robustness in Hypernasal Speech Detection

E. DELGADO1, F. A. SEPULVEDA2, S. RÖTHLISBERGER1, C. G. CASTELLANOS2

1MIRP – Machine Intelligence and Pattern Recognition Lab, Research Center Instituto Tecnológico Metropolitano – ITM Institución Universitaria Calle 73 No. 76A-354 Vía al Volador, Medellín COLOMBIA

2SP&RG – Signal Processing and Recognition Group, DIEEyC Universidad Nacional de Colombia Sede Manizales Campus La Nubia, Km. 7 Vía al Magdalena, Manizales COLOMBIA

{edilsondelgado, sarahrothlisberger}@itm.edu.co, {fasepulvedas, cgcastellanosd}@unal.edu.co

Abstract: People with a defective velopharyngeal mechanism speak with abnormal nasal resonance (hypernasal speech). Voice analysis methods for hypernasality detection commonly use vowels and nasalized vowels. However, to obtain a more general assessment of this abnormality it is necessary to analyze stops and fricatives. This study describes a method with high generalization capability for hypernasality detection analyzing unvoiced Spanish stop consonants. The importance of phoneme-by-phoneme analysis is shown, in contrast with whole word parametrization which includes irrelevant segments from the classification point of view. Parameters that correlate the imprints of Velopharyngeal Incompetence (VPI) over voiceless stop consonants were used in the feature estimation stage. Classification was carried out using a Support Vector Machine (SVM), including the Rademacher complexity model with the aim of increasing the generalization capability. Performances of 95.2% and 92.7% were obtained in the processing and verification stages for a repeated cross-validation classifier evaluation.

Key-Words: Acoustic Analysis, Speech Analysis, Feature Selection, Support Vector Machines and Rademacher Complexity.

1 Introduction

The verbal communication process requires translation of thoughts into spoken language. A person with a physical and/or neurological impairment, may have a compromised vocal tract configuration and/or excitation, resulting in reduced speech quality. A specific example of a vocal tract dysfunction causing reduced speech quality is a defective velopharyngeal mechanism [1]. The term cleft palate refers to a malformation which affects the soft and/or hard palate, and is usually congenital [2]. Digital voice processing (DVP)-based techniques are amongst the most useful of noninvasive techniques for assessing velopharyngeal function, due to the ease of recording speech signals, which are mainly affected in two ways: 1) nasalized phonemes, and 2) weak consonants and short utterance length [3]. The most common way to detect velopharyngeal disfunction (employing DVP) is by carrying out an analysis of nasalized vowel sounds. In [2], a group of delay-based signal processing techniques was described for the analysis and detection of hypernasal speech. Experiments were carried out on the phonemes /l/, /l/, and /l/, where the results showed a high performance on hypernasality detection. The effectiveness of these delay-based acoustic measures were cross-verified with data collected in an entirely different recording environment, however, the generalization capability results of this feature set with regards to the classification accuracy were not convincing. In [1], the sensitivity of the Teager energy operator for multicomponent signals was used for detecting the hypernasal speech problem. A measurable difference was observed between the low-pass and band-pass profiles for the nasalized vowels, whereas the normal vowel, which is a single component signal, does not show a difference. Parameters such as Harmonics to Noise Ratio (HNR) [4], Normalized Noise Energy (NNE) [5], Glottal to Noise
Excitation (GNE) and so on have been proposed for the analysis of pathological voices in different studies. They were mainly designed for sustained vowels, although sometimes they have been used for voiced phonemes, as in [6]. The real problem in hypernasality detection employing DVP is the high variability within-classes, which means high complexity in the training stage and low generalization capability. In this study, unvoiced consonant analysis is proposed, which impedes the use of features previously developed in the literature for speech pathology assessment. Using parameters that correlate the imprints of Velopharyngeal Incompetence (VPI) over voiceless stop consonants such as power, duration and so on, allows for a better representation of the phenomenon currently analyzed. Additionally, finding a reduced representation space of the normal and pathological records is very important, since this procedure reduces computational complexity without loss of classification accuracy and improves the detection robustness by the Rademacher complexity model, due to the addition of an uncertainty component in the feature subset evaluation stage.

2 Materials and Methods

It is necessary to take into account the drawbacks caused by small training samples in the design of automatic classification systems. To reduce these problems, the features used must be correlated with the influence of velopharyngeal incompetence in stop consonants, and classifiers with good generalization properties should be employed [7].

2.1 Database

The sample was constituted by 88 children. Classes were balanced (44 patients with normal voice and 44 with hypernasality), and all registers were evaluated by specialists. Each recording was made up by several Spanish words, but in this study only the words coco (/koko/) and papá (/papá/) were used. Signals were acquired under low noise conditions using a dynamic, unidirectional microphone (cardioide). Signal range was between (-1, 1). A manual segmentation process was carried out to separate the stop parts of the utterances /koko/ and /papá/ resulting in various sets (two from /koko/ and two from /papá/) each formed by 88 signals.

2.2 Parametrization of plosive signals

A plosive consonant is formed by blocking the oral cavity at some point. During the articulation of most plosives the velum is raised, blocking off the nasal passages. Individuals with cleft palate have never learnt to control the movements of the velum. After reconstructive surgery or the fitting of a prosthesis, such individuals need guidance in controlling the velum to produce plosive sounds [3]. The subglottal pressure represents the energy immediately available for creating the acoustic signals of speech [9]. The pressure that builds up behind the occlusion is released suddenly as a minor explosion or popping [3]. The power of stops can help to perceive the weakness of plosive consonants in velopharyngeal patients. In this study, it is calculated using the expression: 

$$P = 10 \log \frac{P_{\text{stop}}}{P_{\text{word}}}$$

where each observation is a set of features

$$\{F_j\}_{j=1}^p$$

and $P_{\text{stop}}$ and $P_{\text{word}}$ is the power of the uttered word. Each stop segment is considered separately for the whole database. Air leakage around the blockage significantly slows down the rise in supraglottal pressure, and therefore, delays phonatory shut-down [9]. This can provoke a short utterance of consonant plosives, which in this study, is measured in seconds. Velic action allows the nasal cavities to be closed or open (or partially open, although air can leak around the velic blockage) with respect to the rest of the vocal tract, which allows sound waves to resonate within the nasal cavities, giving a distinctive nasal quality to the speech sounds produced [8]. In addition, the lower pressure of voiced stops in hypernasal speech results in reduced high-frequency energy for the burst [9]. The MFCC (Mel-Frequency Cepstral Coefficients) and DWT (Discrete Wavelet Transform) use filterbanks to obtain measures of different portions of the spectrum, so the energies at every filter could be used to model the behavior at different frequency ranges [10]. MFCC’s are currently one of the most widely used features for Automatic Speech Recognition. These features are calculated by taking the discrete cosine transform of the logarithm of energy at the output of a Mel filter. In feature estimation processes based on the Fourier transform, the extracted features have fixed time frequency resolution due to the inherent limitation of the FFT. More recently, discrete wavelet transform (DWT) and wavelet packets (WP) have been tried for feature extraction, because of their multi-resolution capabilities [11].

2.3 Feature selection

In general, given a set of observations represented for a set of features

$$\{F_j\}_{j=1}^p$$

where each observation is associated to one and only one class label from a label set $k$, the main goal of feature selection is to choose the best possible subset

$$\xi_i \subseteq \{F_j\}_{j=1}^p$$

of size $q$ from a
set of \( p \) features, where optimal and suboptimal strategies are usually considered. For the optimal case, if the cardinal of \( \xi_i \) is \( q_i \), and all the \( q \)-cardinal subsets are in \( \xi_i \), the subset \( \xi_i \) is that which optimizes an evaluation function \( f \), such as \( \frac{7}{12} \):

\[
f \left( k, \xi_i \right) = \max_{\xi \subseteq \{ F_i \}} f \left( k, \xi \right)
\]

In pattern recognition tasks, feature selection according to the evaluation function \( f \) can be carried out by \textit{wrapper type selection}, when \( f \) uses information of the classification function oriented to minimizing the classifier error, and \textit{filter type selection}, which consists in data preprocessing by optimizing \( f \) with respect to a metric (independent of the classification results), where the irrelevant, redundant and correlated variables are discarded \([12]\). Wrapper type selection procedures give better performance in cases where the number of features is lower than 50 \([13]\), while the filter type can be operated in larger spaces because its computational demand is usually lower \([7]\). Sub-optimal algorithms, although incapable of examining every feature combination, will assess a set of potentially useful feature combinations. Popular methods such as sequential forward selection (SFS) and sequential backward selection (SBS) are found. In floating search methods such as sequential forward floating selection (SFFS) and sequential backward floating selection (SBFS), the number of added and removed features can change at each step and these wrapper routines carry out the search in a considerably smaller number of subsets \([14]\).

2.4 Support vector classifiers

Support Vector Machines (SVMs) were used in this study mainly for two reasons: SVMs have a relatively good generalization capability with less amount of training data, and they have been particularly well developed for binary classification tasks. Traditional neural network approaches are more likely to suffer of poor generalization, producing models that can over-fit the data. This is a consequence of the optimization algorithms used for parameter selection and the statistical measures used to select the best model \([15]\).

For the binary classification problem, a discrimination function can be taken as \( g(x) = w^T \phi(x) + w_0 \) with decision rules \( w^T \phi(x) + w_0 \geq 0 \rightarrow x \in \omega_1 \) and \( w^T \phi(x) + w_0 \leq 0 \rightarrow x \in \omega_2 \), where \( \phi(x) : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is generally a nonlinear function which maps vector \( x \) into what is called a feature space of higher dimensionality (possibly infinite) where classes are linearly separable. The vector \( w \) defines the separating hyper-plane in such a space and \( w_0 \) represents a possible bias \([12]\).

2.5 Rademacher complexity model

Rademacher complexity is a measure proposed in \([16]\), which attempts to balance the complexity of the model with its fit to the data by minimizing the sum of the training error and a penalty term. Let \( \{ x_i, y_i \}_{i=1}^n \) be a set of training instances, where \( x_i \) is the pattern or example associated with features \( \{ F_i \}_{j=1}^q \), and \( y_i \) is the label of the example \( x_i \). Let \( h(x_i) \) be the class obtained by the classifier \( h \), trained using \( \{ x_i, y_i \}_{i=1}^n \). Then, the training error is defined as \( \hat{e}(h) = \frac{1}{n} \sum_{i=1}^n I(h(x_i) \neq y_i) \) where, \( I(h(x_i) \neq y_i) = \begin{cases} 1, & \text{when} \ h(x_i) \neq y_i \\ 0, & \text{when} \ h(x_i) = y_i \end{cases} \).

Let \( \{ \sigma_i \}_{i=1}^n \) be a sequence of Rademacher random variables (i.i.d.) independent of the data \( \{ x_i \}_{i=1}^n \) where each variable takes values +1 and -1 with probability 1/2. According to this, computation of the Rademacher complexity involves the following steps \([17]\):

- Generate \( \{ \sigma_i \}_{i=1}^n \).
- Get a new set of labels, doing \( z_i = \sigma_i y_i \).
- Train the classifier \( h_R \) using \( \{ x_i, z_i \}_{i=1}^n \).
- Compute the Rademacher penalty, given by

\[
\mathcal{R}_n = \left| \frac{1}{n} \sum_{i=1}^n \sigma_i I(h_R \neq y_i) \right|
\]

- Train the classifier \( h \), using \( \{ x_i, y_i \}_{i=1}^n \).
- Compute the training error \( \hat{e}(h) \).
- The Rademacher complexity: \( RC = \hat{e}(h) + \mathcal{R}_n \).

2.6 Proposed procedure

The representation space was composed of the features related to the plosive consonant power and its duration in relation to the word. On the other hand, feature estimation in the frequency domain was achieved by using two techniques: DWT and MFCC. Each feature was estimated for each plosive consonant at the beginning of /koko/ and /papad/. By using 3rd order spline mother wavelet, the Nyquist spectral
range was divided in 4 bands (i.e., 3 for the detail levels and 1 for the approximation level). The other features consist of estimating parameters related to 13-MFCC. Thus, the total number of extracted features for each observation was 15. With the aim of comparing these two representation forms with regards to the discriminant capability, the classification results were obtained using a SVM classifier.

Feature space reduction is carried out by using a type-wrapper algorithm for heuristic search (i.e., SFFS algorithm) with a SVM classifier (RBF-kernel) using a hypothesis test based on a distance measurement for establishing the initial conditions. Moreover, the Rademacher complexity model was included in the evaluation function $f$. With the aim of comparing the proposed model's performance, the conventional training was developed and proved under the same conditions. All data sets were divided in processing stage data (70%) and verification data (30%). That is, the processing stage data is the union of the training and validation sets, which are disjointed from the verification set.

3 Results and Discussion

The utterance /kóko/ has two plosive segments, Figures 1 and 2 show 2-dimensional scatter plots using the duration and power for each segment. Discrimination between the two classes can be observed using the first plosive segment. By contrast, in the second figure this configuration is not seen. The closure of the velopharyngeal gap is necessary to produce vowels as well as stops; but in the first segment the velopharyngeal gap begins to open, provoking a delay in the closing phase, in hypernasal speech. At the beginning of the second stop, the velum is closed as the previous phoneme is a vowel, thus the duration is more similar to the normal category, as depicted in Fig. 2. Other parameters used in this study are related to the fact that spectral components of the plosives are modified by the velopharyngeal incompetence. This set is formed by the energy for every band in a 3rd level of the DWT decomposition. In Figures 3 and 4 two of these values are shown. Once again there is greater similarity between the two classes in the second stop segment. The two sets are slightly distin-

Fig. 1: Duration vs power for the first plosive segment in the Spanish word /kóko/

Fig. 2: Duration vs power for the second plosive segment in the Spanish word /kóko/

Fig. 3: Energy in the third approximation and detail bands for the first plosive segment in the Spanish word /kóko/

guishable as can be seen in the energy-band parameter distribution, nevertheless when these parameters were evaluated (joined with Power and Duration) from the point of view of the classification rate, the performance does not reach 61%. When 13th order MFCC coefficients were calculated, instead of DWT, the performance went up to levels between 83% and 88% with an average rate of 85.7%, although this result decreased to 62.4% in the verification stage (i.e., poor generalization capability). The classifier evaluation was made by applying cross-validation for 30 folds.
Similar experiments were carried out for each word as it is shown in Table 1, where $F_{\text{p,v}}$ is the feature set related to power and duration + DWT for the processing and verification data, similarly, $F_{\text{p,v}}$ is the feature set related to power and duration + MFCC. The notation $//+//$ means that in this case the whole feature set for /koko/ and /papá/ has been considered.

Feature selection results obtained without/with the Rademacher complexity model included in the evaluation function are shown in Table 2, over the processing and verification data sets, where $\#F_{R,R}$ is the selected feature number from the $\{F_1 + F_2\}$ feature set without/with the rademacher complexity model, similarly, $A_{R,R}$ is the average classification accuracy (%) for the processing and verification data. It is remarkable that the classification results for the processing with the Rademacher complexity model are lower than for the other cases, although finally in the verification stage, the proposed training retains the classification accuracy even when the input samples are completely unknown.

### 4 Conclusions

From these experiments it can be concluded that hypernasal assessment should be determined by analyzing phoneme by phoneme, instead of complete words. The acoustic properties of the same phoneme can be completely different in different parts of the uttered word due to variability of the behavior of articulators which depend so much on the context. According to the obtained results, the Rademacher penalty adds generalization capability to the classifier, which is a necessary constraint due to the high within-class variability of speech signals. This uncertainty included in the feature selection allows for effective dimensionality reduction.

Using few features, a performance of 85.2% in the verification stage was obtained for the voiceless plosive /k/, 89.5% for the phoneme /p/ and 92.7% considering both phonemes. Thus, feature selection revealed what features contributed to the generalization capability. For example, the power has discriminant information for /p/, while the phoneme /k/ is well-represented by the duration. The other selected features were related to the high-frequency bands except one feature of low-frequency, which could probably be used as a reference for the classifier. This is in agreement with the information supplied by the clinic specialists [9].

### 5 Acknowledgements

This study was supported by the Condonable Credits program of the Departamento Administrativo de Ciencia, Tecnología e Innovación (Colciencias) in Colombia and the Research Center of the Instituto Tecnológico Metropolitano of Medellín under Grant P09225.

### References:


Intensive Quenching of Tools in Water Salt Solutions

N.I.KOBASKO¹, A.A.MOSKALENKO², E.A.MAZURENKO³, A.M.MEDVEDEV³

¹Intensive Technologies Ltd, Kyiv, Ukraine
NKobasko@aol.com

²Institute of Thermal Science and Engineering of NASU, Kyiv, Ukraine
an.moskalenko@gmail.com

³Vernadsky Institute of General and Inorganic Chemistry of NASU, Kyiv, Ukraine
eam@ukr.net, medvedev48@mail.ru

Abstract: - In the paper intensive quenching, named IQ-2 process, is widely discussed. It is shown that this process can be fulfilled by use of water salt solutions of optimal concentration (CaCl₂) with additives of Ca(OH)₂ and creation of protective anticorrosion layers on the surface of steel parts during their heating. The proposed technology can compete with the expensive technologies which use furnaces with the protective atmosphere and automated thanks filled with the intensively agitated anticorrosion salt solutions. The technology is less costly and provide in many cases the same end result in terms of increasing service life of tools. It is shown that intensive quenching of tools by water salt solutions of optimal concentration increases service life of tools by 1.5 - 2 times as compared with the oil quenching.

Key - Words: - Intensive quenching, Salt solution, Optimal concentration, Prevention from corrosion, Tools, Service life

1 Introduction

Water salt solutions of optimal concentration, especially those which prevent corrosion, are widely used for quenching of tools. As known, water salts solutions of chlorides are used as a quenchant to intensify cooling rate and achieve high mechanical properties of materials [1, 2]. In Ukraine, as a quenchant is widely used 8 -12% water solution of CaCl₂ at 20 - 40°C. To prevent corrosion, a small amount of Ca(OH)₂ is added to the solution to keep pH within 8 -12 [1]. Even with the additives of Ca(OH)₂ corrosion cannot be completely eliminated. That is why, it is proposed a method to cover tools before quenching by nano - layers of Cr₂O₃ or Al₂O₃ [3] which prevent decarburization of tools during heating and corrosion after quenching of tools in water salt solutions of CaCl₂. To interrupt intensive quenching at the end of the nucleate boiling process, a noise control system was used [4]. This system measures duration of nucleate boiling process very accurately [4]. After interruption of intensive quenching and equalizing of temperature to provide self-tempering of tools, they were intensively washed. Washing and the second step of quenching are combined to provide superstrengthening of a material [5, 6, 7]. A detail description of procedures for two step quenching of tools in water salt solutions of CaCl₂ is provided in this paper. The technology can be easily realized without use of expensive equipment to produce intensive quenching and complicated calculations to develop recipes. Instead of that, end results are the same in terms of mechanical properties and quality of the surface and increase service life of tools. It is better to use water salt solutions of optimal concentration which prevent corrosion.
completely [8]. However in many cases these solutions cannot provide appropriate critical heat flux density \( q_{\text{cr1}} \) that is why they should be intensively agitated to get high \( q_{\text{cr1}} \).

### 2 What is optimal concentration and method of quenching?

There is an optimal concentration of electrolytes [5, 9] when the first critical heat flux density has an optimal value which is maximal (see Fig. 1). We need this maximal value to prevent film boiling during quenching. When film boiling is absent, heat transfer coefficients during immersion of tools into cold water salt solutions reach 200,000 W/m\(^2\)K and higher since intensive process of nucleate boiling takes place from the very beginning of quenching. It means that surface temperature drops immediately almost to saturation temperature \( T_S \) and during nucleate boiling maintains almost at the same level [10]. The duration of this process can be calculated by equation [11, 12]:

\[
\tau_{nb} = \left[ \Omega + b \ln \frac{\vartheta_I}{\vartheta_{II}} \right] \frac{K}{a},
\]

where \( \tau_{nb} \) is duration of nucleate boiling (NB) process; \( \Omega \) and \( b \) are constants; \( \vartheta_I = T_I - T_S \); \( \vartheta_{II} = T_{II} - T_S \); \( T_I \) is wall temperature at the beginning of boiling;

\( T_{II} \) is wall temperature at the end of nucleate boiling; \( T_S \) is saturation temperature; \( K \) is Kondratjev form factor; \( a \) is thermal diffusivity.

This equation was used to calculate duration of nucleate boiling process and speed of conveyors. The speed of conveyors should provide interruption of intensive quenching write at the end of boiling process. Then steel parts are delivered after temperature equalizing to the washing and the second step of quenching. The principal scheme of the furnace with the conveyor is shown in Fig. 2. It should be noted that during calculation of speed of conveyor, the length of conveyor is taken into account which is immersion in the solution (see Eq. (2) and NB process can be measured by noise control system (see Fig. 3).

\[
W = \frac{aL}{\Omega + b \ln \frac{\vartheta_I}{\vartheta_{II}}} K
\]

**Fig. 1** First critical heat flux density \( q_{\text{cr1}} \) versus the concentration of NaCl and LiCl in water [5]

The optimal concentration of water salt solutions provides maximum critical heat flux density and maximal effective heat transfer coefficients during nucleate boiling and effectively eliminates film boiling [14]. Ref. [5] present average (effective) values of heat transfer coefficients for water and water salt solutions.

**Fig. 2** The principal scheme of the furnace with the conveyor
Fig. 3 The second version of laboratory equipment which includes personal computer and small acoustical system [4]

Fig. 4 Core Cooling rate and noise intensity in mV during quenching of cylindrical 10 × 50 mm specimen (AISI 304 steel) in 12% water CaCl₂ solution at 20°C (frequencies 2 - 3 kHz)

Fig. 5 Core cooling rate and noise intensity in mV during quenching of cylindrical 10 × 50 mm specimen (AISI 304 steel) in 8% water CaCl₂ solution at 20°C (frequencies 2 - 3 kHz)

Fig. 6 Core cooling rate of cylindrical 10 × 50 mm specimen made of AISI 304 steel and noise intensity during quenching in 6% water Na₂CO₃ solution at 20°C (1 are frequencies 100 - 1000 Hz; 2 are frequencies 1.0 - 2.0 kHz)

As we can see from Fig. 4, Fig. 5, and Fig. 6, in all three experiments film boiling was absent. Duration of transient nucleate boiling was almost 3 - 4 seconds which agrees with results presented in Table 1. It means that water salt solution (NaCl, Na₂CO₃, CaCl₂) of optimal concentration eliminate film boiling during quenching.

Fig. 7 Temperature, core cooling rate of AISI 1030 steel (cube 100×100×100 mm) and noise effect during quenching in 12% water NaCl solution at 20°C
Fig. 8  Temperature and noise effect during quenching of cylinder (50×100 mm) in 12% water NaCl solution at 20°C versus time: 1 is temperature at the core; 2 is temperature 5 mm below surface.

All data presented in Fig. 4 - Fig. 8 which were measured by noise control system (see Fig. 4) coincide very well with the data calculated by Eq. (1).

Table 1  Comparison of the measured time of the nucleate boiling process with the calculated data during quenching of pins in 30% water salt solution of CaCl₂ [1,2, 5]

<table>
<thead>
<tr>
<th>Diameter of pin in mm</th>
<th>Steel</th>
<th>Measured time of nucleate boiling in sec</th>
<th>Calculation by Eq. (1), sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>AISI E9310</td>
<td>4.3</td>
<td>4.0</td>
</tr>
<tr>
<td>12</td>
<td>SAE 4130</td>
<td>6.2</td>
<td>6.0</td>
</tr>
<tr>
<td>20</td>
<td>SAE 4130</td>
<td>15.0</td>
<td>15.0</td>
</tr>
<tr>
<td>24</td>
<td>AISI 5140</td>
<td>20.9</td>
<td>21</td>
</tr>
<tr>
<td>27</td>
<td>AISI 5136</td>
<td>28.3</td>
<td>28.0</td>
</tr>
<tr>
<td>30</td>
<td>SAE 4130</td>
<td>32.8</td>
<td>32</td>
</tr>
</tbody>
</table>

Table 2 Results of testing punches on automatic line "National - 164" with performance of 175 strikes per minute [5]

<table>
<thead>
<tr>
<th>Oil</th>
<th>IQ in water solution of CaCl₂</th>
<th>Increase, n times</th>
</tr>
</thead>
<tbody>
<tr>
<td>6460</td>
<td>15600</td>
<td>2.4</td>
</tr>
<tr>
<td>6670</td>
<td>16500</td>
<td>2.9</td>
</tr>
<tr>
<td>4000</td>
<td>12075</td>
<td>3.0</td>
</tr>
<tr>
<td>2890</td>
<td>10500</td>
<td>3.6</td>
</tr>
</tbody>
</table>

3 Developments in preventing decarburization and corrosion

Authors [3, 13] (National Academy of Sciences of Ukraine, Kyiv) have developed special metallo-organic chemical compositions containing alloying elements such as chromium (Cr), aluminum (Al) and others which during heating transform into a vapor. This vapor creates a nano-layer on the surface of steel parts consisting of chromium oxides Cr₂O₃ or aluminum oxides Al₂O₃ which prevent decarburization during heating and corrosion after quenching in water salt solutions of chlorides (see Table 3)

Table 3 Depth of oxidation of different steels depending on temperature and soak time, and existing of protective nano-layer Cr₂O₃ [3, 13]

<table>
<thead>
<tr>
<th>Steel, T₀, °C</th>
<th>Soak time, sec</th>
<th>No layer</th>
<th>Cr₂O₃ layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1045</td>
<td>850</td>
<td>30</td>
<td>0.02</td>
</tr>
<tr>
<td>50XФI</td>
<td>850</td>
<td>60</td>
<td>0.01</td>
</tr>
<tr>
<td>4X13</td>
<td>1050</td>
<td>60</td>
<td>0.02</td>
</tr>
<tr>
<td>ХВГ</td>
<td>830</td>
<td>40</td>
<td>0.02</td>
</tr>
<tr>
<td>Cast iron</td>
<td>960</td>
<td>600</td>
<td>0.2</td>
</tr>
</tbody>
</table>
The optimal interval of temperatures for nano-layer formation of $\text{Cr}_2\text{O}_3$ is $600^\circ\text{C} - 950^\circ\text{C}$ with concentration of $5 - 200 \text{ g/m}^3$ for organometallic containing chromium and $750^\circ\text{C} - 1000^\circ\text{C}$ with concentration $2.5 - 15 \text{ g/m}^3$ for organometallic containing aluminum which creates $\text{Al}_2\text{O}_3$.

Implementation of these developments into industry results in the next benefits:
- A sharp decrease in the depth of the oxidized surface layer;
- A sharp decrease in the thickness of decarburization;
- Reduction of loss of alloying components (W, Mo, V, Ti, Cr, Ni, etc);
- Saving the geometry and dimensional accuracy of parts;
- Saving of natural gas up to 40%;
- Increasing the service life of electric heaters furnaces at approximately 1.5 times.

Authors have published several technical papers and applied their method to prevent decarburization and corrosion in 10 (ten) big Co of Ukraine and other countries [3].

The organometallic compounds are available for selling.

The first work which discussed application of noise control system for quenching steel parts in water salt solutions was published in 1990 [15]. At present time steel parts are intensively quenched in water and water salts solutions which prevent corrosion [16].

4 Discussion

Water salt solutions as the quenchants are widely used in Ukraine, Russia, China and others countries. They can be used for intensive quenching to improve significantly mechanical properties of steels and eliminate oil as a quenchant. Water salt solutions of chlorides have high critical heat flux densities which completely eliminate film boiling during quenching. As a result, distortion decreases, mechanical properties increases, the environment improves. Disadvantage of chlorides is corrosion. Combining intensive quenching in chlorides with the protective $\text{Cr}_2\text{O}_3$ or $\text{Al}_2\text{O}_3$ layers, it will be possible to make next step in the heat treating industry. We can start international cooperation between countries which are using already water salt solutions as a quenchant. Especially promising are water salt solutions of $\text{CaCl}_2$ which can be used for steel superstrengthening effect obtaining [17].

5 Summary

1. Two-step quenching in cold water salt solutions of optimal concentration (8-12%) of $\text{CaCl}_2$ is discussed which prevents film boiling during quenching of steel parts.
2. The duration of the first step of quenching is measured by noise control system or calculated by generalized equation.
3. After interruption of quenching and equalizing the temperature, the steel parts are intensively washed and further quenched to room temperature simultaneously.
4. Two-step quenching in cold water salt solutions of optimal concentration (8-12%) of $\text{CaCl}_2$ is intensive quenching since heat transfer coefficient at the beginning of cooling is equal to 200 kW/m$^2$K.
5. It is shown that intensive quenching in water salt solutions increases service life of tools by 1.5 - 2 times.
6. To prevent corrosion after quenching a small amount of $\text{Ca(OH)}_2$ is added to solution to maintain pH at the level of 8 - 12.
7. To prevent oxidation and corrosion after quenching completely steel parts are heated in special atmosphere which creates nano - protective layer of $\text{Cr}_2\text{O}_3$ or $\text{Al}_2\text{O}_3$ on their surface.
8. The proposed technology can compete with the existing IQ-2 process where expensive furnaces with the protective atmosphere are used and automated tanks with anticorrosion intensively agitated salt solutions are combined with the furnace.
9. The technology can be used for quenching of tools by different companies which are making tools for himself and need technology to be less costly.

References:
**Why Database for Cooling Capacity of Various Quenchants Should be Developed?**

NIKOLAI KOBASKO
IQ Technologies Inc, Akron, USA and Intensive Technologies Ltd, Kyiv, Ukraine
NKobasko@aol.com, www.intensivequench.com

Abstract: In the paper, a necessity of the development of database for cooling capacity of quenchants is discussed. In spite of the existence of several methods for solving an inverse heat conduction problem, appropriate experimental techniques and funding, there is no database containing cooling characteristics of quenchants needed for quench process computer simulations. Two reasons explain this situation. The heat treating industry uses mostly oils for quenching alloy steels for which it is not so critical to have the above database. A standard cylindrical Inconel 600 probe (12.5 mm dia) with the thermocouple installed at the core cannot be used for evaluating real heat transfer coefficients (HTC). This probe can be used for determining only effective HTC. Taking into account environmental issues, it is desirable to substitute, when it is possible, oil quenching with intensive water quenching. Development of the proper database of heat transfer characteristics for water based quenchants will accelerate this process.

Key Words: Real and effective heat transfer coefficients, Database, Standard probe, Oil, Water as a quenchant, Environment.

1 Introduction
As known, alloy and high alloy steels are usually quenched in oils. Often a gas quenching is applied to decrease cooling rates within the martensite range. When quenching in oils, no trouble occurs because steel parts can be cooled to the room temperature without the interruption of the quench. There is no need to interrupt cooling when approaching the martensite start temperature since convective heat transfer coefficient of oil is rather low (250 - 300 W/m²K). To control and to maintain stable the cooling capacity of oils, a standard probe of 12.5 mm diameter, made of Inconel 600 with a thermocouple in the core, was designed [1]. Taking into account environmental issues, it is desirable to substitute, when it is possible, oil quenching with intensive water quenching. Development of the proper database of heat transfer characteristics for plain water and water solutions will accelerate this process.

2 Real and effective heat transfer coefficients
A real heat transfer coefficient during nucleate boiling relates to the difference \[ \Delta T = T_w - T_s \], instead of difference \[ \Delta T = T_w - T_m \] as it is considered during a convection mode of heat transfer, where: \( T_w \) is wall temperature; \( T_s \) is saturation temperature; \( T_m \) is temperature of a quenchant. As well known, the formation of nucleating centers depends on the overheat of the boundary layer which is determined by [2]:

\[
R_{cr} \approx \frac{2\sigma r^*}{r' \rho^* \Delta T},
\]

where: \( R_{cr} \) is a critical size of a bubble which is capable to grow and function; \( \sigma \) is a surface tension (N/m); \( r^* \) is a latent heat of evaporation (J/kg); \( \rho^* \) is a vapor density (kg/m³); \( \Delta T = T_w - T_s \) is a wall overheat.

Active nucleating centers are the basic carriers of the thermal energy that remove heat from the part surface and transfer it to a cooler bath. After the initiation of boiling, the bubble continues to grow (in a saturated liquid) until forces cause it to detach from the surface. After the departure, a cooler liquid from the bulk of the quench bath fills the space vacated by the bubble and the thermal layer is reformed. When
the required superheat is attained, a new bubble starts to form at the same nucleation site. Bubble dynamics include the processes of growth, bubble departure, and bubble release frequency which includes time for reformation of the thermal layer. The bubble acts like a pump to remove the hot liquid from the surface and replacing it with the cooler liquid [2]. This mechanism is the essential factor causing high intensity of heat transfer during boiling. The bath temperature has no essential effect on a value of heat transfer coefficient during nucleate boiling [2]. Therefore, when determining the heat flux density during boiling it is necessary to relate it to a difference of $T_W - T_s$, rather than to $T_S$, which can lead to large errors when calculating the part surface temperature.

To see what is happening during nucleate boiling, let's consider accurate experimental data of French (see Fig. 1 and Table 1), which were used for solving inverse heat conduction problem (IP). For this purpose, the IQLab software was applied. This software was developed by Intensive Technologies Ltd company (Kyiv, Ukraine) [4]. To solve the IP, thermal properties of materials are needed. Some of them are provided in Table 2.

**Table 1** Time required for the surface of steel spheres to cool to different temperatures when quenched from 875°C (1605°F) in 5% NaOH water solution at 20 °C and moving at 3 feet per second (0.914 m/s), according to French [3].

<table>
<thead>
<tr>
<th>Size, inch, (mm)</th>
<th>Time, s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>700°C</td>
</tr>
<tr>
<td>1'' (25.4)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
</tr>
<tr>
<td>2.5'' (63.5)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
</tr>
</tbody>
</table>

**Fig. 1** The schematic which shows how thermocouples were placed and accurately flattened to the wall of spheres and polished by French [3].

**Table 2** Thermal conductivity of silver, Inconel 600, and AISI 304 steel in W/(m K) depending on temperature (°C)

<table>
<thead>
<tr>
<th>T, °C</th>
<th>100</th>
<th>300</th>
<th>500</th>
<th>700</th>
<th>900</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silver</td>
<td>392</td>
<td>362</td>
<td>366</td>
<td>373</td>
<td>-</td>
</tr>
<tr>
<td>Inconel 600</td>
<td>14.2</td>
<td>17.8</td>
<td>21.7</td>
<td>25.9</td>
<td></td>
</tr>
<tr>
<td>Steel 304</td>
<td>17.5</td>
<td>19.6</td>
<td>23</td>
<td>26.3</td>
<td>29.3</td>
</tr>
</tbody>
</table>

**Fig. 2** Real heat transfer coefficients during quenching of spheres made of steel and quenched in agitated (0.914 m/s) 5% NaOH water solution at 20°C: a) for sphere 38.1 mm dia; b) for sphere 63.5 mm dia
As we can see from Fig. 2 and Fig. 3, the real heat transfer coefficients during nucleate boiling reach maximum values of 200,000 to 240,000 W/(m²K) in 2 s after the beginning of the quench regardless of the part size and configuration. These results were obtained for steel specimens. As known, the thermal conductivity of steel is 16 times less as compared with silver. It means that the calculated values of the heat transfer coefficient during cooling of silver probes in water and water solutions will be higher. This is because according to the Fourier law, the heat flux is directly proportional to the material thermal conductivity. Assume that heat transfer coefficients during cooling of silver spherical probes (20 mm in dia) are the same, then we can calculate Biot number, which corresponds to the maximum value shown in Fig. 2 and Fig. 3, i.e.:

\[ Bi = \frac{240,000 W/m^2 K \times 0.01 m}{366 W/m K} = 6.6. \]

During testing of silver probes, it is accepted by investigators that the probe core temperature and surface temperature are equal. It is almost true when Bi < 0.2. However, Bi > 6.6 and it means that the probe core temperature and the probe surface temperature are not equal. According to the experimental and calculated data this difference is up to 100°C and the error can exceed 100% because Biot number is 33 times higher. This problem was widely discussed in book [5].

The main difference is in heat transfer evaluation is. The real heat transfer coefficient is calculated as

\[ \alpha_{nb} = \frac{q}{T_w - T_s}. \quad (2) \]

The effective heat transfer coefficient during boiling is calculated as

\[ \alpha_{eff} = \frac{q}{T_w - T_m}. \quad (3) \]

In practice, quenchants are tested using standard probes, for example, silver sphere probes or cylindrical probes made of Inconel 600. The thermocouples, as it was mentioned, are located at the core. It is considered that temperature field throughout the probe cross section for silver is uniform, i.e., \( Bi \leq 0.2 \).

This assumption was confirmed by the calculation of the heat transfer coefficient from Eq. (3), where \( q \) is a heat flux density on the probe surface; \( T_w \) is the wall temperature; \( T_m \) is the temperature of the quenchant. This is a generally accepted approach for the film and nucleate boiling heat transfer evaluation. However, film and nucleate boiling heat transfer coefficients should be determined from Eq. (2).

It means that, in reality, heat transfer coefficients during nucleate boiling are much higher. Let us illustrate the above mentioned phenomenon by the following example. Let \( T_w = 110°C \); \( T_s = 100°C \); \( T_m = 20°C \). If \( \alpha \) is calculated by equation (3) then:

\[ \alpha_1 = \frac{q}{110°C - 20°C} = \frac{q}{90°C}; \]

If the calculation is performed using equation (2) then:

\[ \alpha_2 = \frac{q}{110°C - 100°C} = \frac{q}{10°C}. \]

Now consider by how many times \( \alpha_2 \) differs from \( \alpha_1 \) by dividing one value by the other:

\[ \frac{\alpha_2}{\alpha_1} = \frac{q}{10°C} \cdot \frac{90°C}{q} = 9, \]

i.e., \( \alpha_2 \) is 9 times greater than \( \alpha_1 \), which means that, using silver probes during testing, subsequent miscalculations will appear. Therefore, it seems like Biot number \( Bi \leq 0.2 \).
However, actually it is greater than approximately 6.6, that is why the sphere core temperature differs from the sphere surface temperature.

![Fig. 4](https://via.placeholder.com/150)

**Fig. 4** Heat flux density and Kondratjev number $\text{Kn}$ vs. time for 1 inch (25.4 mm) cylindrical steel probe quenched in 5% NaOH solution

The software IQLab was used also for calculations of the effective heat transfer coefficients, generalized Biot numbers and Kondratjev numbers $\text{Kn}$. Fig. 4 shows that after 2 seconds of cooling, the Kondratjev number is a linear function of time. In coordinates $\text{Kn}$ vs. $\text{Fo}$ for different diameters of cylinders, the established function is the same (idem). According to the results shown in Fig. 5, it is possible to say that $\text{Kn} = \text{idem}$ (4). It means that heat transfer coefficients for boiling processes can be generalized in forms (4). This simplifies significantly cooling time calculations during the transient nucleate boiling process. Fig. 6 explains why it happens. With increasing of the size of the cylinder the heat flux decreases proportionally to the cylinder size. The average values of $\text{Bi}_V$ and $\text{Kn}$ remain the same (see Fig. 5 and Fig. 6). The main conclusion from the above calculations is that the standard Inconel 600 probe can be used for effective HTC calculations. However, one should keep in mind that such data can be used for cooling rate calculations at the core of steel parts and cannot be used for temperature fields calculations.

![Fig. 5](https://via.placeholder.com/150)

**Fig. 5** Kondratjev number ($\text{Kn}$) vs. Fourier number ($\text{Fo}$) for cylindrical steel probes 20, 30, and 40 mm dia. and quenched in 5% NaOH solution [6]

With increasing of the size of the cylinder the heat flux decreases proportionally to the cylinder size. The average values of $\text{Bi}_V$ and $\text{Kn}$ remain the same (see Fig. 5 and Fig. 6). The main conclusion from the above calculations is that the standard Inconel 600 probe can be used for effective HTC calculations. However, one should keep in mind that such data can be used for cooling rate calculations at the core of steel parts and cannot be used for temperature fields calculations.

![Fig. 6](https://via.placeholder.com/150)

**Fig. 6** Temperature field distribution in cylindrical probes 20mm and 40 mm dia. during quenching in 5% NaOH solution at 20 °C [6]

### 3 Critical heat flux densities and their impact on understanding the boiling processes

It is important to keep in mind that upon the immersion of a steel part into the quenchant, the initial heat flux density $q$ can be:

$$q >> q_{cr1}; \ q \approx q_{cr1} \text{ or } q << q_{cr1}.$$
In the first case, \( q >> q_{cr1} \), a full film boiling is observed. A transition boiling is observed when \( q \approx q_{cr1} \). In the last case, \( q << q_{cr1} \), film boiling is absent and the main mode of heat transfer is nucleate boiling. Each of these three cases will produce different values of \( \alpha = f(T_{sf}) \) versus the part surface temperature. Therefore, there is no unique interrelationship of the heat transfer coefficient \( \alpha \) as a function of the surface temperature.

To predict what kind of heat transfer mode one can expect, we must know critical heat flux densities. When evaluating critical heat flux densities, only film boiling heat transfer coefficients are used, which are small enough (300 - 1000 W/m²K). In this case, the condition \( \text{Bi} < 0.2 \) will be always satisfied. It means that the surface temperature \( T_W \) during testing will be equal to the core temperature \( T_{core} \) of the probe only at the end of film boiling. The probe presented in Fig. 7 can be used for the critical heat flux evaluation [7]. Rounded ends of cylinder will provide the second type of heat transfer mode and accurate estimations of critical heat flux densities.

Fig. 7 Shape and dimensions of the silver cylindrical probe with rounded ends that can be recommended for evaluations of critical heat flux densities [7].

4 Discussion

There are many institutes and universities which investigate cooling capacity of quenchants and use noise control systems to evaluate boiling processes. Unfortunately, there is no database designed for cooling capacity of quenchants which can be used by engineers for the cooling recipes development. We have only cooling curves at the core of standard probes. These data cannot be used for solving the inverse heat conduction problem for getting real heat transfer coefficients. Using these curves, it is possible to calculate only average effective heat transfer coefficients suitable for cooling time calculations at the core and not at the surface.

At present time, it is possible to design DATABASE for cooling capacity of quenchants since everything is prepared for this purpose:
1. Liscic probe is available, which should be standardized since it is very suitable for solving the inverse heat conduction problem and for providing very accurate data [8].
2. Small silver probes and the method for calculations of critical heat flux densities are available, which should be standardized for critical heat flux densities evaluation [9].
3. The noise control system and method for processing of data obtained are available which should be standardized for transient boiling processes investigations [10].

Also an international team is organized to design the above DATABASE (see www.worldses.org/projects/Heat_and_Mass_Transfer.doc)

If such DATABASE is designed, it will be applied for solving important problem [11-15] and for accurate computer simulations of technological processes [16, 17].

Switching from oils to plain water as a quenchant will make environment cleaner. The database, if designed, will accelerate this process.

5 Summary
1. There are two approaches in HTC evaluation. The first approach provides real heat transfer coefficients, which can be used to calculate temperature fields and residual stress distribution. The second approach provides effective heat transfer
coefficients, which can be used to calculate the part core cooling rate. It cannot be used to calculate correctly the temperature field in steel parts during nucleate boiling process.

2. The Inconel 600 cylindrical probe is used to control stability of cooling capacity of quenchants. It can be also used to obtain effective heat transfer coefficients.

3. There is a need to create DATABASE which includes critical heat flux densities, real heat transfer coefficients, and initial heat flux densities during immersion of steel parts into the quench bath. If designed, the DATABASE will accelerate in many cases switching from oil to plain water as a quenchant and will make environment cleaner.

References:
New Approach in Manufacturing of High Quality Superstrengthened Steels

NIKOLAI KOBASKO

IQ Technologies inc, Akron, USA, Intensive Technologies Ltd, Kyiv, UKRAINE

Abstract: In the paper the new approaches in designing of high quality and high superstrengthened steels are discussed. The first approach consists in designing the optimal hardened layer, optimal residual stress distribution in steel part after quenching with the superstrengthened material in outer layers, which appear to be a result of intensive quenching technologies. The second approach consists in enhancing the possibilities of high-temperature (HTMT) and low temperature (LTMT) thermomechanical heat treatment combined with the intensive quenching (IQ) processes. Especially, the second approach is very effective when high carbon steels are used, containing up to 1.8% of carbon. The proposed new approaches allow the metallurgists to make environment cleaner due to water as a quenchant, use plain carbon steels instead of alloy steels due to superstrengthening phenomenon and compressive residual stresses at the surface of components, increase productivity due to high intensity of cooling.

Key - Words: New approaches, High strengthening, Plain carbon steels, Compressive residual stresses, Service life, Productivity, Environment.

1 Introduction
As a rule, plain carbon steels are quenched in water and alloy steels are quenched in oils. High carbon steels and high alloy steels are quenched in oils or polymer solutions of high concentration that often create environment problem. Due to slow cooling in oils, mechanical properties of steel after quenching don’t satisfy the industrial requirements. In this case, metallurgists must add more alloys to steel to provide hardenability that makes material more expensive. In fact, components of machines can be made of plain carbon steels when after machining they are intensively quenched. In the paper two approaches of strengthening of steels are discussed and illustrated by results of experiments and calculations.

2 Optimization of residual stresses by adjusting chemical composition of steel
In heat treatment handbooks, a chemical composition of steel is usually provided for each CCT diagram. Having a digital DATABASE that includes CCT diagrams with steel chemistries will simplify the process of determining an optimal chemical composition of steel for a specific part to maximize residual surface compressive stresses. It has been established that the optimal residual stress distribution takes place when the following ratio is satisfied [1, 2, and 3]:

\[
\frac{D_{opt}}{D} = 0.35,
\]

where \( D \) is a critical size of the steel part of any configuration; \( D_{opt} \) is the real size of the part made of steel having the optimal chemical composition.

An ideal critical diameter or critical size of steel parts of any configuration can be calculated by Eq. 2:

\[
D_{l} = \left( \frac{a b \tau_{M}}{\Omega + \ln \theta} \right)^{0.5},
\]

where \( D \) is the part characteristic size (diameter of cylinder, sphere, thickness of plate, etc.); \( a \) is an average thermal diffusivity (m\(^2\)/s); \( b \) is a constant which depends on the shape of the steel part; \( \Omega \) is an average Kondratjev number (a dimensionless value); \( \tau_{m} \) is the time where the part core cooling curve intercepts with the 50% martensite transformation curve at the level of the martensite start temperature ( Fig. 1); \( \theta = \frac{T_{0} - T_{m}}{T_{M} - T_{m}} \). The results of calculations can be used to optimize
chemical composition of steel for developing high residual surface compressive stresses after quenching [4, 5, 6].

3 Designing of high quality of steel components by interruption of intensive quenching in proper time

The method consists in interruption of intensive quenching of steel parts when compressive residual stresses achieve their maximum value at the surface and hardened layer is optimal [5, 6]. The method can be illustrated by quenching the small rollers (20 mm in diameter and 20 mm in height) in condition $\text{Bi} \rightarrow \infty$. Assume that roller is made of AISI 52100 steel, CCT diagram of which is shown in Fig. 2. The rollers are heated to 860°C and then intensively quenched in condition $\text{Bi} \rightarrow \infty$ in water at 20°C ($\text{Kn} = 1$). It has been established that compressive residual stresses are established when core temperature drops approximately to 450 – 600°C. The cooling time can be calculated using generalized equation [6, 7]:

$$
\tau = \left[ 0.72 + \ln \left( \frac{860°C - 20°C}{450°C - 20°C} \right) \right] \frac{12.12 \times 10^{-6} m^2}{4.6 \times 10^{-8} m^2/s} = 3s,
$$

where $K = 12.12 \times 10^{-6} m^2$; $a = 4.6 \times 10^{-8} m^2/s$; $K$ is Kondratjev form factor in m$^2$ and $a$ is thermal diffusivity of material in m$^2$/s. The results of FEM calculations are shown in Fig. 4 and Fig. 5. Calculations were made by software HART [8]. At that moment hoop stresses at the surface of roller are very high which reach up to - 1400 MPa (see Fig. 4). At the core of roller small tensile stresses are observed (Fig. 4).
The illustrated method of quenching can be easily applied to simple configuration of steel parts like rollers, cylinders, plates, and others machine components and constructions. However, when steel parts or components of constructions are complicated configurations like gears, optimization of chemical composition of steel is best a way for creation compressive residual stresses at the surface [1, 2, 3]. As known, alloy content in steel is always presented as a range. For the minimum value of the alloy content, \( \frac{D_{I_u}}{D_{opt}} = 0.2 \), and for the maximum value of the alloy content, the ratio should correspond to \( \frac{D_{I_u}}{D_{opt}} = 0.5 \), i.e.

\[
0.2 < \frac{D_{I_u}}{D_{opt}} < 0.5.
\]

Below is an example of the application of the described method. Let’s consider a cylinder of 80mm diameter and 80mm height (it could be a stamp). A CCT diagram presented in Fig. 5 provides the ratio \( \frac{D_{I_u}}{D_{opt}} \) of approximately 0.2. This CCT diagram corresponds to a modified 52100 steel containing less amount of manganese. Fig. 6, 7 present the results of FEM calculations of the temperature stress conditions for the above cylindrical part after 50 seconds of intensive quenching using from 860°C to 477°C. The heat transfer coefficient on the part surface during the quench was assumed of 40,000W/m²K. Note that 50 seconds is the time when the quench should be interrupted to obtain maximum residual surface compressive stresses and superstrengthened material [9, 10 11].

**Fig. 5** CCT diagram for steel that provides \( \frac{D_{I_u}}{D_{opt}} = 0.2 \) for a cylinder of 80 mm diameter and 80 mm height.

**Fig. 6** Temperature distribution (in °K) in the finite cylinder of 80mm diameter and 80mm height after 50 seconds of intensive quenching.

Note that the finite cylinder of 80mm diameter and 80mm height is cooled from 860°C to 477°C under the same as above heat transfer coefficient for the following time:

\[
r = \left( 0.72 + \ln \left( \frac{860 - 20}{477 - 20} \right) \right) \frac{1939 \times 10^6 m^2/s}{5.36 \times 10^6 m^2/s} \approx 48s
\]

that agrees very well with the FEM computer calculations [12, 13].

**Fig. 7** Hoop stress in the finite cylinder of 80mm diameter and 80mm height after 50 seconds of intensive quenching.
4 Duration of transient nucleate boiling process and its regularities

The transient nucleate boiling and self-regulated thermal processes were investigated since 1968 and their results were published in [14, 15]. Duration of self-regulated thermal process differs insignificantly from the time of transient nucleate boiling (within 0.5–1 second). The notion of self-regulated thermal process was proposed in 1968 [16] and it means that wall temperature of steel part is kept at the level of saturation point varying insignificantly during all process of transient nucleate boiling. The equation for determining the duration of transient nucleate boiling (self-regulated thermal process) was firstly received by generalization of experimental data [15, 16, 17] and then derived from the analytical equation and has the form [15]:

\[ \tau_{nb} = \frac{\Omega k_F k_W D^2}{a}, \] (4)

where value \(\Omega\) depends on initial temperature of a steel part and condition of cooling. For initial temperatures 850°C it can be within 3.6–4.17 (see Table 1). Koefficient \(k_F\) depends on configuration of steel part (see Table 2). For plate-shaped forms \(k_F = 0.1013\); for cylinder-shaped form \(k_F = 0.0432\); for spherical-shaped forms \(k_F = 0.0253\); \(k_W\) is dimensionless coefficient which depends on liquid flow velocity. For motionless liquid \(k_W = 1\). For high flow velocity of liquid which prevents nucleate boiling \(k_W = 0\). That is why for different condition we have \(0 \leq k_W \leq 1\). \(D\) is thickness of the component: diameter of cylinder, sphere or thickness of the plate; \(a\) is thermal diffusivity of a material.

The discovered law (4) of transient nucleate boiling process can be used to delay martensite transformation in order to improve significantly mechanical properties of steel by applying law-temperature thermomechanical heat treatment (LTMT). Fig. 8 explains what are high-temperature (HTMT) and low-temperature thermomechanical heat treatment (LTMT).

5 Combining thermomechanical heat treatment with the intensive quenching process could be very beneficial

There are not enough data providing the impact of intensive quenching combined with the thermomechanical treatment (TMT) process on mechanical properties of a material. This is because the TMT process is used mostly for alloy steels to make low-temperature thermomechanical treatment reliable. As a rule, alloy steels are quenched in oil. Let’s compare the mechanical properties of AISI 5140 steel with those of AISI 1040 steel, both subjected to TMT. The AISI 5140, after TMT, was quenched in oil; the AISI 1040 was quenched in water. The martensite start temperature for both steels was about 350°C. The heat transfer coefficient of oil within a range of temperature of 100–350°C is 300 W/m2K, and the heat transfer coefficient within
the same interval for water is about 4,000 W/m2K due to the boiling process, which occurs above 100°C. This means that the cooling rate within the martensite range differs significantly between the two. The equation for cooling rate evaluation, depending on heat transfer coefficients, is well known [4, 6]:

\[ v = \frac{abKn}{D^2} (T - T_m) \]  

(5)

where \( v \) is the cooling rate in °C/s; \( a \) is the thermal diffusivity of a material in m²/s; \( b \) is the coefficient depending on the configuration of steel parts; \( Kn \) is the Kondratjev number (a dimensionless value); \( D \) is size (diameter or thickness) in m; \( T \) is temperature in °C; and \( T_m \) is bath temperature in °C. For a cylindrical 10-mm specimen, when quenching in oil (300 W/m²K) and water (4,000 W/m²K), the Kondratjev numbers are 0.05 and 0.4. According to Eq 5, the cooling rate within the martensite range in cold agitated water is eight times faster than with the still oil. Mechanical properties of alloy steels, subjected to TMT and quenched in oil, are less as compared with the mechanical properties of plain carbon steel, subjected to TMT and quenched in cold water. It is well known that alloying increases the mechanical properties of steel considerably, depending on the content of the alloying elements. It was established that HTMT increases yield strength of low alloy steels by 15% and and impact strength by 250% [4]. The chemical composition of steels AISI 1040, AISI 5140, and 40KhN are presented in [20]. Benefits of intensive quenching are properly discussed in the publications [21 – 25].

7 Discussion

The interruption of the quench process is used to create a hardened shell at the surface of steel parts. The hardened shell provides high surface residual compressive stresses. This approach is suitable when quenching steel parts of simple configuration. When quenching steel parts of complicated shapes, consisting of thin and thick sections, it is impossible to create a uniform hardened shell throughout the part surface area since the thin section of the part can be quenched through. Surface tensile stresses will be developed in the thin section of the part. In this case, it is better to optimize chemical composition of steel which will create uniform hardened shell in the both the thick and the thin sections of the part. Calculations showed that very high surface compressive residual stresses (up to 1300 - 1900 MPa) can be developed even in the thin sections of the part, when the steel part made of steel of optimal chemistry and is quenched very intensively. For example, for bearing products, these results are true for modified AISI 52100 steel with small amount of Mn and Cr and containing about 1% weight of carbon.

It should be noted that it is extremely important to create a global database for CCT diagrams which should contain chemical composition of steels and their mechanical properties. These data can be used as a basis for optimization of the residual stress distribution and for the selection of optimal steel for every specific part to increase its service life. Both the surface compressive residual stresses and the superstrengthening of steel allow using less costly materials for designing components of machines. The advanced methods of calculations and new methods of quenching are discussed in publications [22 –25].

6 Conclusions

1. The optimal quenched layer and intensive quenching under the condition of \( 0.8 \leq Kn \leq 1 \) provide the optimal residual stress distribution (high compressive residual stresses at the surface and small tensile stresses at the part core) and make possible to use a superstrengthening phenomenon in practice [5].
2. High compressive residual surface stresses and superstrengthened material developed in the part surface layer increase significantly the service life of steel parts while decreasing their cost.
3. On the basis of discovered law of transient nucleate boiling process, it is possible to apply LTMT to plain high carbon steels to increase significantly their mechanical properties.
4. The presented approaches can be used to make environment cleaner using plain water as a quenchant instead of hazardous oils, and to use plain carbon steels instead of alloy steels. Producing of alloy steels is more harmful for environment as compared with the plain carbon steels.

References:

2. Kobasko, N.I., Current and residual stresses during quenching of steel parts, in Finite
153

New Approach in Manufacturing of High Quality Superstrengthened Steels

5. N.I.Kobasko, US Patent No. 6,364,974 B1


Abstract: - A low cost, black nickel solar absorber coating used in thermal solar collectors on nanostructured high thermal conductor copper was analyzed using XPS (X-ray photoelectron spectroscopy) and MFM (magnetic force microscopy) techniques. Copper substrate nanostructuration was realized by electrochemical polishing to improve MFM analysis. Black nickel solar absorber coating was electrodeposited on nanostructured C81100 copper due to their good optical properties and low cost. XPS results after 15 minutes of sputtering identified black nickel as a Cermet mainly composed of two metallic nickel atoms for each nickel and zinc sulfide, i.e. 2 Ni: NiS: ZnS and traces of thiocyanate. The MFM black nickel lock-in-presents a estimate mean diameter particle size in of (31 ± 16) nm and mode diameter particle size of (6 ± 16) nm. This nanometric particle was attributed to the ferromagnetic metallic nickel found in XPS analysis. Thus, MFM validate that black nickel solar absorber coating is a nanometric graded cermet of NiS-ZnS with 13 % of metallic nickel nanoparticles by chemical composition of XPS results and MFM.

Key-Words: - solar absorber coating; X-ray photoelectron spectroscopy; magnetic force microscopy; nanometric structure

1 Introduction

The demand for energy is rising all over the world, with a doubling every five to seven years in underdeveloped countries, seven to ten years in developing countries and ten to fifteen years in advanced countries. Renewable energy from natural resources such as sunlight, wind, rain, tides, and geothermal heat present solutions to decrease critical climate changes due to emissions of harmful
pollutants from conventional energy consumption (petrol, coal, and others). However, compared to other renewable energy resources, solar energy has probably the greatest potential, because the energy coming from the sun is the most powerful (a solar energy per m² per day has 7 kWh energy content), cleanest, and stable source [1, 2]. Sunlight is an extremely energetic electromagnetic radiation with wavelengths from the near infrared to the ultraviolet (wavelength between 300-2500 nm). The amount of sunlight striking the Earth’s atmosphere continuously is 1.75 x 10^5 TW, with 60% passing through the atmospheric cloud cover 1.05 x 10^5 TW reaches the Earth’s surface continuously. If the irradiance on only 1% of the Earth’s surface could be converted into electric energy with a 10% efficiency, it would provide a 105 TW of energy, whereas the total global energy needs for 2050 are projected to be only about 25–30 TW. The present state of solar energy technologies is such that solar-cell efficiencies have reached over 20% and solar thermal systems 40%—60% are efficient. Solar PV panels have come down in cost from about $30/W to about $3/W in the last three decades. At $3/W panel cost, the overall system cost is around $6/W, which is still too high to compete with other resources on the electricity grid. However, there are many off-grid applications where solar PV is already cost-effective, especially with governmental incentives, such as feed-in laws [1, 2].

A solar collector is an apparatus that collects the sunlight energy then alters this energy into a more usable or storable energy form. There are wide ranges of different solar collectors; most consist of a flat copper plate with a coated black paint on the absorber surface with water tubes attached to the absorber plate in which the energy is transferred. Technologies that use solar energy in commercial applications include concentrate solar power, photovoltaic, solar heating [3,4]. Solar thermal devices use selective absorber surfaces to produce heat from solar radiation for different applications, from domestic to industrial, as well as generate electricity. The photothermal effect is a phenomenon that induces the absorption of radiation by electromagnetic radiation producing photoexcitation and heating on an absorber surface. The absorber surface used in thermal solar collectors requires high absorptance in the solar spectrum and low emittance, i.e., high reflectance, in the infrared wavelength range or a lower thermal radiation from the heated absorber surfaces.

For solar collectors, the cost and effective utilization of solar energy requires an efficient solar absorber tandem coating (multi-layer coating) with a high sunlight absorptance coefficient such as low energy losses [5-8]. Heat treatment studies help to identify the efficiency of the solar absorber tandem coatings (efficiency of converting the sunlight to other desirable energy form or work) and energy losses connecting different parameters such as the microstructure of a material, the chemical composition, diffusivity, conductivity, and emittance [2, 9-11].

Cermet (ceramic–metal composite) films, also known as composite films, have optical properties appropriate for high solar absorption. They basically consist of a metal oxide coating containing metal nanoparticles embedded in a dielectric host on a metallic substrate. It has been suggested that these films strongly absorb in the solar region due to inter-band transitions in the metal in combination with small particle resonance, while they are almost transparent in the thermal IR region. Several techniques, such as vacuum techniques (sputtering, electron beam, chemical vapor deposition, etc.), sol-gel, electroless deposition (catalytic reduction process) are currently used to produce solar absorber surfaces [12]. However, the desired characteristics of the solar absorber coating may be better controlled by direct electrodeposition such as their nanostructuration [13]. The electrodeposition advantages against vacuum techniques are basically as follows [14]:

a. Rapidity;
b. Low cost;
c. Free from porosity;
d. High purity;
e. Industrial applicability;
f. Ability to produce structural features with sizes ranging from nm to μm;
g. Easy to control alloy composition.

Advantages of copper substrate for selective coatings in solar applications are as follows: longer life or durability in piping systems suitable for heating fluids up to 85°C, higher thermal and electrical conductivity (lower energy losses), low price (5 €/kg, New York Mercantile Exchange-2010) and low maintenance costs. Beside these advantages, copper tubes are the standard plumbing material for potable water and heating systems in most European countries [10]. Other choices are chemical affinity and well know processes and reactions [15, 16]. A thin layer of nickel is plated to the copper to improve wear resistance and to prevent the oxidation of copper, besides promoting better black nickel adhesion. Nickel can be plated onto copper surfaces by either electroless plating (a catalytic reduction process) or electrolytic plating techniques [17]. Black nickel is a Cermet type of solar absorber coating that can be applied onto various metal
substrates using an electroplating technique [12, 17]. The dark blackening of the black nickel absorber is attributed to the deposition of sulfide particles together with the Ni-Zn alloy [12]. The ceramic matrix was composed of NiS and ZnS with a highly absorbent coating in the solar region while fine nickel-zinc particles in ceramic matrix offer an IR-reflective property. The black appearance of black nickel deposits attributed to decreased reflectivity by reducing the effective electron density from the formation of a cermet composed in a dispersion of metal in a dielectric matrix. Thus, when the metal contents of the deposit are small, the optical constants gave low reflectivity at sufficiently large thicknesses [12].

2 Experiments

The nanostructuration in copper samples of 1 x 1 x 0.1 cm$^3$ was done by mechanical polishing (P1200) followed by three cycles of electropolishing assays were all performed using a Autolab™ type III potentiostat. The electrolytic cell for electropolishing was composited of 85 %v/v H3PO4 as electrolytic solution, at room temperature (25 °C), at an electrical potential of 1.75 V [16, 17].

The electrodeposition of nickel Watts underlayer on copper bulk was performed to give black nickel better adhesion. The nickel Watts underlayer was electrodeposited using a sulfate Watts bath under the chosen plating conditions of 0.02 A/cm² for 5 minutes (room temperature) [17].

To plate black nickel on nickel Watts coating, a bath composed of nickel, nickel sulfate, nickel ammonium sulfate, zinc sulfate, and sodium thiocyanate was used following Fichlock procedure [17]. The black nickel coating was electrodeposited out using a sulfate Fishlock’s bath under the chosen plating conditions of 0.002 A/cm² for 15 minutes (room temperature). AFM were performed by intermittent contact mode operation on a JPK™ atomic force microscope that was employed to obtain topography and lock-in phase images using the tips such as NSC16/AlBS. Magnetic images were employed to obtain magnetic information from bulk by using the tips such as one coated with Co-Cr. Topography and lock-in phase images were analyzed by JPK™ image processing, v.2 for the surface roughness. SPIP™, v. 5.0.1.0 (Image Metrology A/S, Denmark) was used for the diameter particle size.

Chemical composition was obtained by X-ray photoelectron spectroscopy (XPS) analysis. XPS was performed with the OMICRON™ system (Omicron NanoTechnology GmbH) equipped with EAC2000 SPHERA™ and DAR400 DAR 400 twin-anode X-ray source. Data analysis was performed with Casa™ software.

The thickness of nickel Watts and black nickel layer used were measured as a result of the cross-section photomicrography image of the back scattered electron imaging. Scanning Electron Microscopy (SEM) and Energy Dispersive X-ray (EDS) spectroscopy were performed using JEOL JSM-6460 LV and EDX Noran in 200 kV (EDS), respectively. A digital micrometer with a precision measurements of 0.1 µm was used to measured the thickness of the black nickel (black nickel/nickel Watts/copper) and the nickel Watts (nickel Watts/copper) systems.

3 Results and Discussion

3.1 Crystallographic and Chemical analysis: X-ray diffraction analysis and X-Ray Photoelectron Spectroscopy (XPS)

XPS survey spectrum of the black nickel films after fifteen minutes of sputtering is presented in Figure 1 as well as the comparative evaluation of the atomic composition. The C1s (284.6 eV) peak was used to adjust the position of the XPS spectra. The black nickel survey scan exhibits the main core-level peaks for O 1s, C 1s, N 1s, S 2p, Zn 2p, and Ni 2p, centered at binding energies 531.6, 284.5, 397.9–400, 162.7, 1021.6, and 852.9-855.7 eV, respectively.

![Figure 4 - XPS survey Image of Black nickel after 5 minutes of sputtering.](image-url)
In order to get information on the chemical state of the different species found on the film, high resolution spectra were obtained for nickel, zinc, sulfur, oxygen, and carbon with sputtering of 5 minutes, table 1. The survey XPS spectrum in figure 4 detected the presence of C (40.65 %atomic), O (13.08 %atomic), Ni (21.57 %atomic), Zn (7.33 %atomic), S (14.78 %atomic), and N (2.59 %atomic). The high resolution sulfur spectrum in table 1 after 15 minutes of sputtering detected the presence of sulfide compounds like NiS (Ni2+ found in XPS Ni high resolution) and ZnS (ZnS compound founded in XPS Zn high resolution). The high resolution nickel spectrum with 15 minutes of sputtering found the presence of Ni0 (60.8 % atomic) and as Ni2+ (39.2% atomic). Satellites values (Sat) were added to the main near peaks, table 1. By observing the range between nitrogen 1s and sulfur 2p in figure 4 to survey XPS, in terms of binding energy position, a small amount of sulfur can be assumed as the presence of thiocyanate compounds, (Ni1s-S2p) = 235.2 eV, according to Walton [18]. The presence of nitrogen compounds as thiocyanate is below 5 % atomic by XPS survey. Thus, analysis of the XPS spectrum indicates that black nickel bulk (after 15 minutes of XPS sputtering) is mainly composed of a comparative atomic amount of the twice metallic nickel to each nickel sulfide and zinc sulfide, that is 2 Ni: NiS: ZnS.

Table 1. XPS high resolution of Ni, S, and Zn after 5 minutes of sputtering.

<table>
<thead>
<tr>
<th>XPS High resolution</th>
<th>Name</th>
<th>Position (eV)</th>
<th>At.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nickel</td>
<td>Ni 2p3/2 (0)</td>
<td>852.86</td>
<td>27.60%</td>
</tr>
<tr>
<td></td>
<td>Ni 2p1/2 (0)</td>
<td>870.26</td>
<td>15.46%</td>
</tr>
<tr>
<td></td>
<td>SAT (Ni)</td>
<td>858.12</td>
<td>5.08%</td>
</tr>
<tr>
<td></td>
<td>SAT (Ni)</td>
<td>875.46</td>
<td>12.64%</td>
</tr>
<tr>
<td></td>
<td>Ni 2p3/2 (2+)</td>
<td>854.72</td>
<td>12.00%</td>
</tr>
<tr>
<td></td>
<td>Ni 2p1/2 (2+)</td>
<td>870.12</td>
<td>6.00%</td>
</tr>
<tr>
<td></td>
<td>Sat (Ni 2+)</td>
<td>860.06</td>
<td>17.71%</td>
</tr>
<tr>
<td></td>
<td>Sat (Ni 2+)</td>
<td>879.22</td>
<td>3.52%</td>
</tr>
<tr>
<td>Zinc</td>
<td>Zn 2p1/2 (ZnS)</td>
<td>1022</td>
<td>63.31%</td>
</tr>
<tr>
<td></td>
<td>Zn 2p3/2 (ZnS)</td>
<td>1045.1</td>
<td>34.69%</td>
</tr>
<tr>
<td>Sulfur</td>
<td>S2- (sulfide)</td>
<td>161.78</td>
<td>38.13%</td>
</tr>
<tr>
<td></td>
<td>Poli S2-</td>
<td>161.94</td>
<td>61.87%</td>
</tr>
</tbody>
</table>

3.2 Morphological Analysis: Magnetic Force Microscopy

The mean size of the metallic nickel grains was indirectly interpreted by associating the MFM image and the grain size analysis to the black nickel absorber. In figure 2, magnetic domain structure is presented by the MFM image with lifting-in of 35nm. This image presents an estimated a JPK™ root mean square roughness (RMS) of 10.920 nm, a mean diameter SPIP™ particle size of (31 ± 16) nm and a SPIP™ mode diameter particle size of (6 ± 16) nm. This nanometric particles were attributed to the ferromagnetic metallic. Figure 2 lock-in-phase trace shows a one-domain structure, “round-up” points nanometrically distributed lock-in phase image as observed by Xiansong Liu [19]. Lock-in phase at lifting in 35 nm of the surface correspond well to a presence of ferromagnetic metallic nickel.

Figure 2 - black nickel MFM µm images 5 x 5 with lifting-in of 35 nm: the height (topography profile) and the lock-in phase (trace and retrace) with oscilloscope in a MFM scanning.

4 Conclusions

Surface Analysis by XPS after 15 minutes of sputtering identified black nickel solar absorber coating as a Cermet mainly composed of two metallic nickel atoms for each nickel and zinc sulfide, i.e. 2 Ni: NiS: ZnS.

MFM 5 x 5 µm lifting-in 35 nm images of topography and lock-in-phase from the cermet surface of the black nickel nanostructured present ferromagnetic nanoparticles with a estimate mean
diameter particle size of $(31 \pm 16)$ nm and mode diameter particle size of $(6 \pm 16)$ nm. This nanometric particle was attributed to the ferromagnetic metallic nickel found in XPS analysis. Thus, MFM validate that black nickel is a graded cermet of NiS-ZnS with 13 %atomic of metallic nickel nanoparticles by chemical composition of XPS.

5 Acknowledgements

Thanks to PEMM-COPPE-UFRJ and DIMAT-INMETRO-XERÉM help in the experiments, to CNPq-Brazil for financial support.

References:
Constructive Principles of the Italian System Used
To build the bastioned fortification of Oradea Fortress

VALENTIN CAPOTESCU.
Department of Civil Engineering
Politehnica University of Timişoara
2nd T. Lalescu, 300223 – Timişoara
ROMANIA
capotescuvalentin@yahoo.com,valentin.capotescu@ct.upt.ro, http://www.ct.upt.ro

Abstract: This paper presents a constructive concept of the Italian system of fortification and demonstrates that the fortification of the Fortress Oradea is belonging to this principle. The Italian system appears in the same time with the Renaissance. This is the very first modern bastioned fortification which is using the sharp salient-angle (the feather model). Renaissance’s Principles arrived in Transylvania also with the Italian architects and engineers brought here by Caterina de Medici. This paper presents a detailed analyse of the Oradea fortress, and contain a demonstration and explanation of the constructive concept which was used.

Key-Words: - Military Architecture Theory, Fortifications, bastioned fortress, rampart, bastion, artillery, fortification, Renaissance, artillery.

1 Introduction
The construction of the star shape fortress Oradea starts around 1540 and is was made according with the Italian System trace, drafted by Pietro Cataneo which were later used by Italian military architects who worked in Oradea (fig.1).

The architects chose a pentagonal shape pattern with 5 bastions, adapted for artillery, on each corner. The project was assigned to the Italian architect Domenico di Bologna who has visited the court og Gheorghe Martinuzzi. Other Italian architects such as Ottavio Baldigara, Domenico dei Ridolfini da Camerino, Simone Genga, Achille Tarducci da Carinaldo, Giovan Marco Isolano da Bologna, Maurizio Veneiro or Giacomo Resti got involved in this huge project for building this fortress.

The fortress was erected in 2 large stages: the 1st stage was between 1518-1550 under the leadership of Gabriel Bethlen and Rackoczi family; the 2nd stage was between 1569-1598 under the leadership of Ioan Sigismund of Bathory and of Bathory family. The fortress is surrounded by a very large ditch between 60 and 80 meters wide and 6 meters deep, supplied with warm water from Peta river that did not freeze in winter.

Fig.1. Cataneo’s System
2 Material and method.

Constructive principles of the Italian system

The bastion had the shape of the feather with a salient angle, the flanks formed a salient angle generally of 60°, being ended and protected by orillons. The orillons were linked with the curtain wall through the lateral flanks of the bastion that, connecting the bastion to the curtain side, were further strengthened by a wall endowed with a platform and a parapet, for a better lateral flanking of the curtain wall. A parallel fire was thus established.

![Fig.2 Cross-section](image1)

![Fig.3 Cross-section](image2)

Fig.2. Bastion with a cavalier, Fig.3. Bastion without a cavalier.

By placing a cavalier or a lunette on the bastion, the fortified element became even harder to conquer, and very resistant to enemy fire. Cataneo’s concept that was used at Oradea is explained in figure 4. Figure’s 4 legend: g-glacis, W-re-entering places of arms, a-orillon, b-ravelin, c-cavalier.

In order to understand this concept from fig.4, we will analyse thoroughly all the elements of the fortress of Oradea and we will compare them with the theoretical pattern.

The defensive structure of the Oradea fortress is analysed in figure 5. One can notice that the fortification consist of 2 fortified pentagonal precincts (fig.5 upper right)- the red pentagon that represents the castle and bastion precinct represented in yellow. The defence ditch surrounds the fortress on all sides of the starred perimeter and it is represented in blue. In fig.5-left corner, the fortress ditch is represented in red and the old glacis which does not exist anymore, is represented in yellow. Thus, one can notice the way in which the urban structure got too close to the fortress entering even the protection area. The fortification had 2 interior concentric yards, the 1st being the castle yard-represented in yellow in fig.5 (3rd figure on the column), while the 2nd was the bastioned fortification yard which included the first one as well, represented in green (last figure in the column). In fig.6 is presented the study of the fortress today-stage, and an epoch stamp (16th century). Notice in fig.7,8,9 the way the system works. In those figures is analysed the different types of fire.
Constructive Principles of the Italian System used to Build the Bastioned Fortification of Oradea Fortress

Fig. 5. Analytic study of the component elements of Oradea Fortress

Fig. 6. Present study of the Oradea Fortress and its representation in an epoch stamp.

Fig. 7. Carrying out the cross fire and flanking fire from the lateral flanks and bastions.

Fig. 8. Carrying out the cross fire from the lateral flanks of the bastion in combination with the cross-fire of the first precinct towers.
Fig. 9. The demonstration of the Italian geometrical pattern over the fortress ground-plane (the composition axis are represented in red and O1, O2, O3 are the centres of the geometrical plan of the fortress).

The demonstration of the Italian geometrical trace of the bastion 5 with explanation of the drawing method and the component elements: AB, AC-bastion flanks, CD, BF-bastion orillons, DE, FG-side flanks.

Figure 9 points out the concept of ideal pentagonal fortification, the 3 geometrical centres being collinear and equally apart from one another. This manner in which a bastion was built (it has an identical pattern for all 5 bastions), is represented in fig. 10. The cannons could be placed in two ways:- 1 one the bastion platform, firing from the the parapet, or 2- inside the bastion, but only in lateral flanks, in order to create a flanking-fire for the curtain, and the rest of the bastion was filled with earth. The main strong point of this type of fortification is to give up passive defence and adopt the active defence. The advantages of the Italian system are: usage of cross-fire, the adoption of the starred plan and the salient angle-sharp shape of the bastion which removes the dead angle of the medieval bastion in circular plan (fig. 11).

Fig. 10. The demonstration of the Italian geometrical trace of the bastion 5 with explanation of the drawing method and the component elements: AB, AC-bastion flanks, CD, BF-bastion orillons, DE, FG-side flanks.

Fig. 11. The demonstration of the dead-angle elimination in the Italian trace.

3 Problem Solution
Reshaping the protection area is utterly necessary for the rescue of the monument. A simple restoration is not enough, if a method required for the revival and urban inclusion of the fortress can not be found. It is also important to acknowledge the value of the monument since it is the only one of this type in Romania, and the only example of Renaissance military architecture of Italian type in Romania. (there was another square bastioned fortification built on the Italian system at Alba Iulia, but in the 18th century was part-demolish and included inside the modern Vauban-type fortification of Alba-Iulia- Carolina Fortress).
4 Conclusion:
In order to restore the citadel’s old look, and to bring it to an acceptable shape, it is necessary to bring life inside it. You can’t bring life inside if is not restored properly. Knowing the past, helps you knowing to restore the present and preserve it for the future. This case in particular is the single one in Romania, so it is necesarely to save the monument asap. A quick, fast and professional intervention is required.

This paper explains the building method of Oradea Fortress and why it does belong to the Italian School, and it demonstrates the concept and analyses the reconversion and resquing possibilities of this monument; this act can only be made after a good knowledge of the monument past and the way in which it was built.

5 Acknowledgements
The research works was based on the book wrote by the author, and elements which are part from some other private collections and Museum archives, from Romania. The author would also like to thank to Prof.dr.ing. Stoian Valeriu, and Prof.dr.arh.T. Octavian Gheorghiu for providing historic material from their personal collection.

This work was partially supported by the strategic grant POSDRU 6/1.5/S/13, (2008) of the Ministry of Labour, Family and Social Protection, Romania, co-financed by the European Social Fund – Investing in People.

References:
[17] CESARE BRANDI- Teoria Restaurarii
[20] LENDY (Captain) Treatise on Fortification London 1862
[22] VITRUVIUS The Ten Books of Architecture
[23] MACHIABELLI- Dialogues on the Art of War 1573
[24] The Codex of Francesco Laparelli MS 1566
[26] FRANCESCO DI GIORGIO Trattati di Architettura, Ingegneria e Arte Militare
[27] LANTERI GIACOMO Due dialoghi Venice 1557
[28] CORNEWLEYE ROBERT The Maner of Fortification Of Cities, Townes, Castelles and other Places MS 1559
[29] MAGGI GIROLAMO & CASTRIOTTO JACOMO Della fortificatione della citta Venice 1564
[30] LUPICINI ANTONIO Architettura Militare...Florence 1582
[31] RUSCELLI GIROLAMO Perecetti della militia moderna tanto per mare quanto per terra Venice 1583
[32] BUSCA GABRIELLE Della espugnazione et difesa delle fortezze Turin 1585
[33] IVE PAUL. The practise of Fortification London 1589
[34] DE MAECHI FRANCESCO Architettura Militare die Capitano F de M...Brescia 1599
[35] STEVIN SIMON- Of Besieging Towns and Fortresses 1608-1615
[36] SARDI PIETRO Corona imperiale dell`architettura militare Venice 1618
[37] FLORIANI PEITRO PAOLO Difesa et offesa delle piazze macerata 1638
[38] VAUBAN SEBATIEN- Manuel of Siege and Fortification
[39] SAXE MAURICE (Field MARSHAL)- Reveries or,memoirs concerning the art of war by Maurice Count de Saxe Marshal-General of the armies of France 1776, French-Edinburgh
[40] PASLEY C.W. A course of elementary Fortification, including rules, deduced from experiment for determining the strength ov revetments; treated on the principle of peculiar perspicuity 1851 London
[41] YULE HENRY (LIEUT) Fortification for Officers of the Army and students of Military History, London 1851.
[42] DUEKER ALBRECHT- Etliche underricht zu befestigung der stet schloss und Flecken Nurenberg 1527
[43] COEHORN MENNO VAN- Nieuwe Vestingbouw op een natte of lage Horisont Welke op drydeleijen manieren getoont word in `t Fortificeren der binnengrote van der Franske Royale seshoek. London 1705
A Robust Control by Extended Static Output Feedback for Discrete-Time Uncertain Linear Systems

ADDISON RIOS-BOLIVAR
Departamento de Sistemas de Control
Facultad de Ingeniería, ULA
VENEZUELA
ilich@ula.ve

FLOR NARCISO
Departamento de Computación
Facultad de Ingeniería, ULA
VENEZUELA
fnarciso@ula.ve

Abstract: - This paper presents a novel approach for the controller synthesis in discrete-time linear systems (DTLS) with uncertainty polytopical type, by means of the extended static feedback of the measured output. The method consists of designing feedback gains for the injection of the output and its derivative, which corresponds to the control signal. Conditions for the existence of such controllers are established. The uncertainty parameters are supposed to belong to a polytope. The stabilization problem is formulated in the context of linear matrix inequalities (LMI). Multiobjective performance indices are also considered, which are described as LMIs. The extended $H_2/H_\infty$ conditions are obtained by means of the well known projection lemma. Performance of the proposed synthesis technique is illustrated by a numerical example.


1 Introduction

In any process of synthesis of control systems, the final goal is the practical implementation of the designed controller. When considering the synthesis of control systems for practical purposes, it is necessary to look for greater flexibility and simplicity of implementation [15]. Hence, the output feedback control has been a topic of much research interest.

As it is known, one of the controllers that offer these characteristics are built by using states feedback, which have the disadvantage that not always is possible to have all the states available. However, the static output feedback allows an implementation without many requirements. The major drawbacks for the control synthesis of static output feedback (SOF) are the conditions of existence of such controllers [17].

The SOF problem may be described as following: Given a linear system, which is generally time-invariant, find a gain of static output feedback so that, in closed loop, the controlled system exhibits the desired behavior or determine the nonexistence of such gain [5]. Although there exist several theoretical conditions for the existence of controllers by SOF, do not exist practical algorithms for the solution with few requirements [16].

1.1 Problem Formulation

Consider the DTLS system defined by

$$x_{k+1} = Ax_k + Bu_k; \quad y_k = Cx_k$$

where $A \in \mathbb{R}^{n \times n}$, $B \in \mathbb{R}^{n \times m}$ and $C \in \mathbb{R}^{p \times n}$. The problem is described by: given the system (1) with $(A, B)$ stabilizable, find a control of the form

$$u_k = Ky_k$$

where $K \in \mathbb{R}^{m \times p}$ is the static feedback gain to be built, so that the closed loop system is stable. This means that if there exists $K$, then the closed-loop dynamic matrix $A + BK$ must be stable.

**Problem 1:** Given the system (1), with $(A, B)$ stabilizable, find $K$ for the control (2) such that the matrix $A + BK$ has all its eigenvalues in the stable region.

The selection of the gain $K$ must be allowed to satisfy the performance requirements according to the design objectives of the control system. The first detail to define is the existence of such gain of static feedback. For this, several results have been presented for both the continuous-time and discrete-time systems: [1, 3,
The established conditions do not show directly the solution algorithms. Thus, several methods for the solution of the SOF problem have been proposed [2, 7].

On the other hand, in the particular case of DTLS, the SOF problem in systems with uncertainties has also been considered [10, 16], in which it is emphasized that this is a non-convex NP-hard optimization problem. This problem can be represented, for many design objectives, as an optimization problem with bilinear matrix inequality (BMI). In this case, the reported solutions maintain the problem of exigent requirements from the computational point of view, limiting the practical application of SOF techniques in industrial processes.

**Notation.** $\mathbb{R}$ is the set of real numbers. For a matrix $A$, $A^T$ denotes its transpose. $tr(A)$ defines the trace of the matrix $A$. $I_n$ is the Hilbert space of vectorial signals defined on $(-\infty, \infty)$, such that $\|x\|_2 = (\sum_{n} |x_n|^2)^{1/2} \leq \gamma \forall x \in l_2$. $diag(A, B)$ is a diagonal matrix with inputs $A$ and $R$ on its diagonal. In symmetric matrices partitions $*$ denotes each of its symmetric blocks. $I$ defines the identity matrix of appropriate dimension.

### 2. Preliminaries

In this section some preliminary results are presented to give additional performance conditions in the analysis of discrete-time linear systems (DTLS). Consider the DTLS system

$$G(z) := \begin{cases} 
  x_{k+1} = Ax_k + Bu_k \\
  y_k = Cx_k + Du_k
\end{cases}$$

(3)

where, $x_k \in \mathbb{R}^n$ are the states, $u_k \in \mathbb{R}^m$ are the controls and $y_k \in \mathbb{R}^p$ are the measured variables. The matrices $A, B, C, D$ are well known and they have the appropriate dimensions.

In order to study the performance of this system, in the optimal robust control theory, it is common applying criteria based on norms. Thus, there exist improved versions of the **Bounded Real Lemma** to reduce conservatism [8, 12, 18], or in the case of performance in $H_\infty$ [12, 13], this in order to reduce conservatism. Next, some of these methods will be described, which are the basis for the development of the results to be shown in this paper.

#### 2.1 Extended LMI conditions for DTLS

**Lemma 2.1** ($H_\infty$ norm relaxed in discrete time):

Consider a system as given by (4). The following statements, with $P = P^T > 0$, are equivalents

I. $A$ is stable and $\|C(zI - A)^{-1}B\|_\infty < \gamma$.

II. Exist $P$ and $Z$, such that

$$\begin{bmatrix}
  P & PA & PB \\
  A^TP & P & 0 \\
  B^TP & 0 & I
\end{bmatrix} > 0,$$

$$Z = C, D^T$$

(4)

$$\begin{bmatrix}
  Z & C & D^T \\
  C^T & P & 0 \\
  D^T & 0 & I
\end{bmatrix} > 0, \quad tr(Z) < \mu.$$

III. Exist $P, Z$ and $G$, such that

$$\begin{bmatrix}
  G + G^T - P & GA & GB \\
  A^TG & P & 0 \\
  B^TG & 0 & I
\end{bmatrix} > 0,$$

$$\begin{bmatrix}
  Z & C & D^T \\
  C^T & P & 0 \\
  D^T & 0 & I
\end{bmatrix} > 0, \quad tr(Z) < \mu.$$

**Proof:** The proof is based on the implementation of the projection lemma. This proof can be evaluated in [4].
These results will be useful for the design of robust controllers by static feedback of extended output, such as it will be shown below.

3. Control by Static Feedback of the Extended Output

Consider the system (1). The control is defined by:
\[ u_k = K_0 y_k + K_1 y_{k+1}, \]
where \( K_0 \) and \( K_1 \) are the feedback gains to be determined for output and its discrete derivatives. In this case, the discrete derivative of the output is used in the context of derivative action for PID controllers. Thus, the control is given by
\[ u_k = (I - K_1 CB)^{-1} (K_0 C + K_1 CA) x_k \] (9)

As may be seen, the existence of the control depends on the invertibility of the matrix \( I - K_1 CB \), which is a condition less strong than those established in the Problem 1 of SOF. Thus, for the control by static feedback of extended output the following problem is established:

**Problem 2:** Given the system (1) with \((A, B)\) stabilizable, find \( K_0 \) and \( K_1 \) for the control (14) such that the matrix \( A + B (I - K_1 CB)^{-1} (K_0 C + K_1 CA) \) has all its eigenvalues in the stable semi-plane.

**Lemma 3.1** Given \( M = I - K_1 CB \), there exists a control by static feedback of extended output with the form
\[ u_k = M^T (K_0 C + K_1 CA) x_k \] (10)
if and only if \( M \) have generalized inverse ((Moore-Penrose pseudoinverse), given by \( M^+ \)).

In fact, if \( M^T \) is the transpose matrix of \( M \), which have complete range for columns, then the pseudoinverse Moore-Penrose Matrix is \( M^+ = (M^T M)^{-1} M^T \), which, if exist, will allow to compute the control law. This condition debilitates and generalizes those established in [9], giving a solution to the problem of SOF [14].

3.1. LMI Formulation

Consider the problem of stabilization by the feedback of extended output in the context of LMIs. That is, given the system (1), find \( K_0 \) and \( K_1 \) so that the closed-loop system is stable in the sense of Lyapunov (bilinear matrix inequalities) are satisfied:
\[ P > 0, \quad P - A^T P A_c > 0; \] (11)
where \( A_c \) is the closed-loop dynamic matrix.

**Theorem 4.1** Let the system given by (1) with the stabilizable pair \((A, B)\), there exists a control by extended SOF of the form (9) that stabilizes the closed-loop system, if there exists \( M^3 \) and the symmetric matrix \( P > 0 \), and the matrices \( Y, Z \) such that the following LMI is satisfied
\[
\begin{bmatrix}
    P & AP + BY + BZ^T \\
    PA^T + Y^T B^T + Z^T B^T & P
\end{bmatrix} > 0.
\] (12)
where the feedback gains are obtained
\[ K_0 = Y^T P^{-1} C^T (CC^T)^{-1} \]
\[ K_1 = \bar{V}^T P^{-1} A^T (C A A^T C^T)^{-1} \] (13)
where \( \bar{V} = \bar{I} + ZP^{-1} A^T (C A A^T C^T)^{-1} CB \).

**Proof:** See [15].

It is important to note that conditions for calculating \( K_0 \) and \( K_1 \) are equivalent to those in [9]. However, in this case are less restrictive for the types of systems since the stabilization can be achieved through the proper calculation of \( K_0 \) and \( K_1 \), whose design condition extends the class of systems that can be controlled by this technique.

3.2 Control in \( H_\infty - H_\gamma \)

Consider the LTI system defined by
\[
\begin{align*}
\dot{x}_k &= A x_k + B_a u_k + B_d \omega_k \\
z_k &= C_x x_k + D_d \omega_k \\
y_k &= C_y x_k
\end{align*}
\] (15)
where \( \omega_k \in l_2 \) is an unknown disturbance. \( z_k \) is the controlled output and \( y_k \) is the measured output.

**Problem 3:** Given the system (15), with \((A, B_a)\) stabilizable, find \( K_0 \) and \( K_1 \) for the control (8) such that the dynamic matrix \( A_c \), in closed loop be stable, and the norm-2 or norm-\( \infty \) of the transfer function of the disturbance to controlled output be minimum in some sense, that is \( || H_{z\omega}(s) ||_2 < \mu \) or \( || H_{z\omega}(s) ||_\infty < \gamma \)

In [15], an approach for control of discrete-time linear systems with \( H_2/H_\infty \) performance indexes has been presented. There, the robustness problem has not been considered.
4. Robust Control

So far the design of controllers that minimize the effect of disturbances on the controlled outputs has been considered, such as it has been shown in [14]. From now on, the interest is focused on the design of robust controllers for processes where there is uncertainty in the dynamic matrix of the plant. Consider the uncertain system

\[
\begin{align*}
 x_{k+1} &= A(\alpha)x_k + B(\alpha)u_k + B_u u_k \\
 z_k &= C_z(\alpha)x_k + D_\omega(\alpha)\omega_k
\end{align*}
\]  

(16)

where \( x_k \in \mathbb{R}^n \) are the states, \( \omega_k \in \mathbb{R}^q \) are unknown disturbances, \( u_k \in \mathbb{R}^m \) are the control inputs, \( z_k \in \mathbb{R}^p \) are the controlled outputs, \( y_k \in \mathbb{R}^p \) are the measured outputs. Additionally, the unknown matrices of the system form uncertain polytopic, that is, they belong to a convex hull, thus

\[
\Omega_N = \{ A, B, C_z, D_\omega \} \triangleq \bigoplus_{i=1}^{N} \alpha_i (A_i, B_i, C_i, D_i), \quad \sum_{i=1}^{N} \alpha_i = 1
\]

(17)

Given that \( \alpha \) is restricted to a unit set since \( \alpha_i \geq 0, \sum_{i=1}^{N} \alpha_i = 1 \), then the matrices \( (A, B, C_z, D_\omega)\) are affine functions of the vector of uncertain parameters \( \alpha \in \mathbb{R}^N \), which is described by the convex combination of vertex matrices \( A_i, B_i, C_i, D_i, i = 1, 2, \ldots, N \).

As can be seen, the matrices \( B_u \) and \( C_y \) are considered known. This has correspondence with the fact that they represent, from the practical point of view, the fundamental and precise elements for the implementation of control systems, which are the actuators and sensors, respectively.

Now applying the control by static feedback of the extended output and assuming that \( \mathcal{M} = I - K_1 C_y B_u \), \( u_k \) is given by

\[
\begin{align*}
 u_k &= \mathcal{M}^T (K_0 C_y + K_1 C_y A(\alpha)) x_k + \\
 &+ \mathcal{M}^T K_1 C_y B_u(\alpha) \omega_k
\end{align*}
\]

(18)

then, the closed-loop matrices are

\[
\begin{align*}
 A_c &= A(\alpha) + B_u \mathcal{M}^T K_0 C_y + B_u \mathcal{M}^T K_1 C_y A(\alpha) \\
 B_c &= B_c(\alpha) + B_u \mathcal{M}^T K_0 C_y + B_u \mathcal{M}^T K_1 C_y B_u(\alpha), \\
 C_c &= C_c(\alpha), \quad D_c = D_\omega(\alpha).
\end{align*}
\]

4.1 \( \mathcal{H}_\infty \) Robust Control

From the Lemma 2.1 the following result is obtained:

**Theorem 4.1** Consider the system (16) on the polytope (17). There exists a control law by static feedback of the extended output of the form (18), which guarantees a performance in \( \mathcal{H}_2 \) for the system in closed loop, fulfilling the formal conditions, if the following are satisfied

\[
\begin{bmatrix}
 G + G^T & P_i \\
 A_i^T G + C_i^T Y^T + A_i^T C_i^T Z^T & P_i \\
 B_i^T G + B_i^T C_i^T Z^T & 0 & I
\end{bmatrix}^T > 0
\]

(19)

usual for \( i = 1, \ldots, N \), where \( G \in \mathbb{R}^{n \times n}, Y \in \mathbb{R}^{n \times p_i}, Z \in \mathbb{R}^{n \times m}, P_i = P_i^T > 0 \in \mathbb{R}^{n \times n}. \) Thus, the static feedback gains of the extended output are obtained from

\[
K_0 = \mathbb{V}^T B_i^T G^{-1} Y, \quad K_1 = \mathbb{V}^T B_i^T G^{-1} Z \quad \text{(20)}
\]

Proof: Let’s assume that a feasible solution exists for the problem of static feedback of extended output, according to control (18). Then, from the Lemma 2.1 and the closed-loop dynamic matrix, if \( Y = G^T B_u \mathcal{M}^T K_1 \) and \( Z = G^T B_u \mathcal{M}^T K_0 \), then by applying the Lemma 2.1 the shown result is reached.

4.2 \( \mathcal{H}_\infty \) Robust Control

From the Lemma 2.2 the following result is obtained:

**Theorem 4.2** Consider the system (16) on the polytope (17). There exists a control law by static feedback of the extended output of the form (18), which guarantees a sub-optimal performance in \( \mathcal{H}_\infty \) for the closed loop system, which fulfill the formal conditions, if the following is satisfied

\[
\begin{bmatrix}
 G + G^T & P_i \\
 A_i^T G + C_i^T Y^T + A_i^T C_i^T Z^T & P_i \\
 B_i^T G + B_i^T C_i^T Z^T & 0 & I
\end{bmatrix}^T > 0
\]

(21)

with \( \mathbb{V} = \mathcal{M}^T = I + B_i^T G^{-1} Z C_y B_u \).

Proof: Let’s assume that a feasible solution exists for the problem of static feedback of extended output, according to control (18). Then, from the Lemma 2.2
and the closed-loop dynamic matrix, and if \( Y = G^T B_s M^* K_0 \) and \( Z = G^T B_s M^T K_1 \), then by applying the Lemma 2.2 the shown conclusion is reached.

The results obtained in \( H_{\infty} \) and \( H_\sigma \) can be combined to achieve controller with multi-objective. In the presented case, due to the extended versions of the characterizations of the \( H_{\sigma} \)–\( H_{\infty} \) norms as LMIs, the products that involve Lyapunov matrices are not generated, and the design matrices \( K_0, K_1 \) will not depend on them, avoiding the need to use the same Lyapunov matrix for all the specifications, thereby the conservatism is reduced.

5. Numerical Evaluation

Consider the system defined by [4]:

\[
\begin{align*}
x_{k+1} &= \begin{bmatrix} 0 & -0.5 \\ 2.5 & 1 + \alpha \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u_k + \begin{bmatrix} -6 & 0 \\ 1 & 0 \end{bmatrix} \omega_k \\
z_k &= \begin{bmatrix} -100 & 10 \end{bmatrix} x_k + \begin{bmatrix} 0 & 1 \end{bmatrix} \omega_k, \quad y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} x_k;
\end{align*}
\]

which is unstable, with \( |\alpha| \leq 0.45 \). Therefore, there exist values of \( \alpha \) where the system is unstable. This structured uncertainty form a two-vertices polytope. Applying the Theorem 4.2 the following results are obtained: \( K_0 = -0.8334 \), \( K_1 = 0.3333 \) for a value of \( \mu = 8.0007 \). For these gains, the closed-loop dynamic matrix is

\[
\mathcal{A}_c = \begin{bmatrix} -1.2501 & -0.7500 \\ 1.2499 & 0.7500 + \alpha \end{bmatrix};
\]

which, for the range of values \( \alpha \), is stable, as shown in Figure 1.

To emphasize, at the vertices of the polytope and for the central value, \( \alpha = 0 \), these matrices are:

\[
\mathcal{A}_{c_1} = \begin{bmatrix} -1.2501 & -0.7500 \\ 1.2499 & 0.3000 \end{bmatrix}, \quad \mathcal{A}_{c_0} = \begin{bmatrix} -1.2501 & -0.7500 \\ 1.2499 & 0.7500 \end{bmatrix};
\]

\[
\mathcal{A}_{c_2} = \begin{bmatrix} -1.2501 & -0.7500 \\ 1.2499 & 1.2000 \end{bmatrix};
\]

which, as can be verified, are all stable.

For the simulation, first, it has been considered a value of \( \alpha = 0 \), next for \( k = 150 \) the new value is \( \alpha = -0.25 \). As can be seen in Figure 2, the system is controlled, even in the presence of the variability of the parameter, according to the established for the feedback robust of the extended output.

6. Conclusions

From the output and its discrete derivative, the problem of control by static feedback of the output in discrete-time linear systems with polytopic uncertainties has been addressed and solved. Using both signals, two feedback gains have been designed under the premises of robust stability and disturbance management. The solution allows extending the space...
of the systems that can be stabilized by the traditional techniques of static feedback of the output. The conditions for the problem solution have been presented, and it has been solved for the control in $H_\infty - H_\infty$, using linear matrix inequalities (LMIs). The conditions are analogous to the well known problems of static feedback of the output. However, the spectrum of the uncertain discrete-time linear systems that can be controlled by this type of feedback has been extended. The design of the controller by static feedback of extended output is robust since linear systems with polytopic uncertainties have been considered. The feedback gains of the extended output are obtained by solving the relaxed LMIs, which are less conservative. This is verified through a numerical example.

Acknowledgments
The authors want to thank the CDCHT-ULA for its financial support through the project No. I-1206-09-02-B.

References
Optimization of Limit Characteristics in Opened Queueing Networks

GURAMI TSITSIASHVILI
Institute for Applied Mathematics, Far Eastern Branch of RAS
Laboratory of Probability Methods and Systems Analysis
Radio Street 7, 690041 Vladivostok
RUSSIA
guram@iam.dvo.ru

Abstract: A model of an aggregated queuing network and an optimization of its limit characteristics are considered. In the aggregated queuing network numbers of servers in different nodes and an intensity of Poisson input flow are proportional to a large parameter $n$. Each node of the aggregated opened queuing network is described by stationary occupancy of its servers. For $n \to \infty$ these stationary occupancies tend by a probability to some deterministic values. A maximization of a minimum of these limit occupancies consists of a finding of a route for some one-variable function and includes a definition of permissible solutions of some auxiliary transportation problem. An ability to handle customers in a superposition of opened networks is calculated. Network superpositions may be interpreted as recursively defined structures widely used in a nanotechnology.

Key-Words: Opened queueing network, Route matrix, Limit occupancy of servers.

1 Introduction

A model of an aggregated opened queueing network is considered. This model is constructed from $n$ copies of an opened queuing network. Each copy $G$ consists of the nodes $0, 1, \ldots, m$, $m > 1$ where the nodes $1, \ldots, m$ are one server queueing systems with service intensities $\mu_1, \ldots, \mu_m$. The node 0 is a source for customers arriving the network $G$ with the intensity $\lambda$ and an outflow for customers leaving $G$. In the network $G$ customers circulate in an accordance with the indecomposable probability transition (route) matrix $\Theta = [\theta_{ij}]_{i,j=0}^m$.

The aggregated opened queuing network $G_n$ has the Poisson input flow with the intensity $n \lambda$ and $n$-server queueing systems in the nodes 1, $\ldots, m$. Each server in the node $i$ has the service intensity $\mu_i$.

A main problem of this paper is to minimize (to maximize) a maximum (a minimum) of stationary occupancy of servers in the nodes 1, $\ldots, m$ for $n \to \infty$. Such limit characteristics are very convenient for next optimization because it is possible to prove that the stationary occupancies tend by a probability to some deterministic quantities. And the minimum (the maximum) of these quantities may be optimized by sufficiently simple procedure from [1]. The optimization procedure consists of a finding of a route for some one-variable function and includes a definition of permissible solutions of some auxiliary transportation problem.

As customers motions in the network $G_n$ are described by the indecomposable route matrix $\Theta = [\theta_{ij}]_{i,j=0}^m$ so the solution $(1, \Lambda)$, $\Lambda = (\lambda_1, \ldots, \lambda_m)$ of the system

\[(1, \Lambda) = (1, \Lambda)\Theta \tag{1}\]

exists, is single and consists of positive components [2].

Suppose that

\[\rho_i = \frac{\lambda \lambda_i}{\mu_i} < 1, \ i = 1, \ldots, m \tag{2}\]

then a vector of numbers of customers in the nodes 1, $\ldots, m$ of the aggregated network $G_n$ is described by the ergodic [3, chapter 2] discrete Markov process $y(t) = (y_1(t), \ldots, y_m(t))$ with the states set $Y = \{n = (n_1, \ldots, n_m) : n_i \geq 0, \ i = 1, \ldots, m\}$. The process $y(t)$ limit distribution [2, § 2]

\[\Pi(n) = \prod_{i=1}^m \pi_i(n_i), \tag{3}\]

\[\pi_i(n_i) = C_i^{-1} a_i(n_i), \ C_i = \sum_{n_i=0}^\infty a_i(n_i), \tag{4}\]

\[a_i(0) = 1, \ a_i(k) = \prod_{1 \leq j \leq k} \frac{n \rho_i}{\min(j, n)}, \ 0 < k < \infty. \]

Here $\pi_i(n_i)$ is a limit distribution of a stationary number of customers in the node $i$. 

2 Law of large numbers for distribution $\Pi(n)$

**Theorem 1** If the condition (2) is true then for any $i$ and for any $\varepsilon$, $0<\varepsilon<1/2$ we have that

$$\lim_{n\to\infty} \sum_{n\rho_i(1-2\varepsilon)<k<n\rho_i(1+2\varepsilon), 1\leq i \leq m} \Pi(n) = 1.$$  \hspace{1cm} (5)

and a convergence in (5) is geometric.

**Proof:** Fix $\varepsilon$, $0<\varepsilon<1/2$ then from the definition (4)

$$\pi_i(k) \leq \frac{\pi_i(k)}{\pi_i([n\rho_i])}$$

and consequently

$$\sum_{0\leq k \leq n\rho_i(1-2\varepsilon)} \pi_i(k) \leq \sum_{0\leq k \leq n\rho_i(1-2\varepsilon)} \frac{\pi_i(k)}{\pi_i([n\rho_i])} \leq \sum_{0\leq k \leq n\rho_i(1-2\varepsilon)} \prod_{k<j\leq n\rho_i} \frac{j}{n\rho_i} \leq \sum_{0\leq k \leq n\rho_i(1-2\varepsilon)} \prod_{n\rho_i(1-2\varepsilon)<j\leq n\rho_i(1-\varepsilon)} \frac{j}{n\rho_i}.$$

As a result obtain that for fixed $\varepsilon$, $0<\varepsilon<1/2$, and for $n\to\infty$

$$\sum_{0\leq k \leq n\rho_i(1-2\varepsilon)} \pi_i(k) \leq n\rho_i(1-\varepsilon)^{n\rho_i\varepsilon} \to 0.$$  \hspace{1cm} (6)

Analogously it is easy to prove that

$$\sum_{n\rho_i(1+2\varepsilon)\leq k \leq n, 1\leq i \leq m} \Pi(n) \leq (1-\rho_i)(1+\varepsilon)^{-n\rho_i\varepsilon} \to 0$$  \hspace{1cm} (7)

and

$$\pi_i(k) \leq (1+\varepsilon)^{-n\rho_i\varepsilon} \rho_i^{k-n}, \; k>n$$  \hspace{1cm} (8)

so from the condition (2)

$$\sum_{n<k<\infty} \pi_i(k) \leq \frac{(1+\varepsilon)^{-n\rho_i\varepsilon}}{1-\rho_i} \to 0.$$  \hspace{1cm} (9)

The limits (6) - (9) have a geometric (by $n$) rate convergence and lead to the formula

$$\lim_{n\to\infty} \sum_{n\rho_i(1-2\varepsilon)<k<n\rho_i(1+2\varepsilon)} \pi_i(k) = 1.$$  \hspace{1cm} (10)

Then the formulas (3), (10) give the limit (5) with a geometric rate convergence.

The theorem 1 establishes that the vector $(\rho_1, \ldots, \rho_m)$ characterizes the limit occupancies of the nodes $1, \ldots, m$ in the aggregated network $G_n$, $n \to \infty$.

3 Optimization of limit network characteristics

Fix the intensities $\lambda, \mu_1, \ldots, \mu_m$ and introduce the functions

$$\Phi(\Lambda) = \max_{1\leq i \leq m} \rho_i, \; \Psi(\Lambda) = \min_{1\leq i \leq m} \rho_i.$$  

Define the equalities

$$\rho_1 = \ldots = \rho_m.$$  \hspace{1cm} (11)

Suppose that the set $M$ consists of vectors $\Lambda = (\lambda_1, \ldots, \lambda_m)$ so that

$$\lambda_1 > 0, \ldots, \lambda_m > 0, \; \sum_{k=1}^{m} \lambda_k \geq 1,$$  \hspace{1cm} (12)

and satisfies the conditions:

a) the set $M$ is convex and closed and has a smooth boundary,

b) $\exists \Lambda^* \in M$ so that the minimum $\Phi(\Lambda)$ by all $\Lambda \in M$ which satisfy the equalities (11) is in the point $\Lambda^*$,

c) $\exists \Lambda_\varepsilon \in M$ so that the maximum $\Psi(\Lambda)$ by all $\Lambda \in M$ which satisfy the equalities (11) is in the point $\Lambda_\varepsilon$,

d) the inclusion $M \subseteq \{\Lambda : \Phi(\Lambda) < 1\}$ is true.

**Minimization of maximum of vector components.**

Consider a problem of a minimization of the function

$$\Phi(\Lambda) \implies \min, \; \Lambda \in M.$$  \hspace{1cm} (13)

**Theorem 2** If a tangent plane $L$ to $M$ in the point $\Lambda^*$ may be represented by the equation $\Lambda \cdot A = d$ in which the vector $A = (a_1, \ldots, a_m)$ satisfies the conditions $a_1 > 0, \ldots, a_m > 0$ then a single solution of the problem (13) is the point $\Lambda^* = (\lambda_1^*, \ldots, \lambda_m^*)$.

**Corollary 1** The route matrix $\Theta$ is a solution of the problem (13) if and only if the matrix $\Theta = ||\theta_{i,j}||_m$ is a permissible solution of the transportation problem

$$\sum_{j=0}^{m} \theta_{i,j} = \sum_{j=0}^{m} \theta_{j,i} = \lambda_i^*, \; 0 \leq i \leq m.$$  \hspace{1cm} (14)

The condition (12) guarantees a solvability of the transportation problem (14).

**Maximization of minimum of vector components.**

Consider a problem of a maximization of the function

$$\Psi(\Lambda) \implies \max, \; \Lambda \in M.$$  \hspace{1cm} (15)

**Theorem 3** If a tangent plane $L$ to $M$ in the point $\Lambda_\varepsilon$ may be represented by the equation $\Lambda \cdot C = f$ in which the vector $C = (c_1, \ldots, c_m)$ satisfies the conditions $c_1 > 0, \ldots, c_m > 0$ then a single solution of the problem (15) is the point $\Lambda_\varepsilon = (\lambda_1^*, \ldots, \lambda_m^*)$. 

172
Corollary 2 The route matrix $\Theta$ is a solution of the problem (15) if and only if the matrix $\Theta = [\{\theta_{i,j}\}]_{i,j=0}^m$ is a permissible solution of the transportation problem

$$\sum_{j=0}^m \theta_{i,j} = \sum_{j=0}^m \theta_{j,i} = \lambda_i, \ 0 \leq i, j \leq m. \ \ (16)$$

The condition (12) guarantees a solvability of the transportation problem (16).

Remark 1 The statements of the theorem 2 and of the corollary 1 are proved in [11] where a similar optimization procedure appears in a maximization of an ability to handle customers in an opened queueing network. This approach is suggested in [4] for deterministic problems of vector optimization and gives a simple calculation procedure. The statements of the theorem 3 and of the corollary 2 may be proved analogously.

4 Cooperative effects in opened queueing networks

Results of previous section showed that the condition (11) of equal nodes load is very important in considered problems of occupancy optimization. But there are some more properties of aggregated opened queueing networks connected with mean stationary queue length

$$N_i^{(n)} = \sum_{k>n} \pi_i(k)(k-n), \ 1 \leq i \leq m,$$

or mean stationary waiting time

$$W_i^{(n)} = \frac{1}{n\lambda} \sum_{k\geq n} \pi_i(k)(k-n+1), \ 1 \leq i \leq m,$$

in the network nodes. As the condition (11) is true then $\rho_1 = \ldots = \rho_m = \rho$.

Suppose that $m = 1, n = 1$, if $\rho_1 \rightarrow 1$ then $N(1), W(1) \rightarrow \infty$. But if $m = 1, n \rightarrow \infty$ and $\rho_1 = \rho_1(n), 1-\rho_1(n) \sim n^{-\alpha}$ then for $0 < \alpha < 1$ we have $W_1(n) \rightarrow \infty$ and for $1 < \alpha$ we have $W_1(n) \rightarrow 0$ as $n \rightarrow \infty$. Analogously if $m = 1, n \rightarrow \infty$ and $\rho_1 = \rho_1(n), 1-\rho_1(n) \sim n^{-\alpha}$ then for $0 < \alpha < 1/2$ we have $W_1(n) \rightarrow \infty$ and for $1/2 < \alpha$ we have $W_1(n) \rightarrow 0$ as $n \rightarrow \infty$. These cooperative effects have been established for one node queueing network in [5, Corollaries 1, 2]. But for opened queueing network with arbitrary number of nodes these results may be generalized in the case of equal nodes load or in the case of approximately equal nodes load as follows.

Theorem 4 Suppose that the condition (11) is true and $\rho = \rho(n) = \text{const} < 1, n > 0$, then

$$N_i^{(n)} \rightarrow 0, \ W_i^{(n)} \rightarrow 0, \ n \rightarrow \infty, \ 1 \leq i \leq m, \ \ (17)$$

with a geometric rate of convergence.

Proof: The formula (17) is a corollary of the inequality (7).

The conditions of the theorem 4 may be realized in the following way:

$$\frac{\lambda_1}{\mu_1} = \ldots = \frac{\lambda_m}{\mu_m} = a, \ \rho = \lambda a < 1.$$

Theorem 5 Suppose that the condition (11) is true. If $1-\rho(n) \sim n^{-\alpha}$ then for $0 < \alpha < 1/2$ we have $N_i^{(n)} \rightarrow 0$, as $n \rightarrow \infty$, and for $\alpha > 1/2$ $N_i^{(n)} \rightarrow \infty$, as $n \rightarrow \infty$; $1 \leq i \leq m$.

Theorem 6 Suppose that the condition (11) is true. If $1-\rho(n) \sim n^{-\alpha}$ then for $0 < \alpha < 1$ we have $W_i^{(n)} \rightarrow 0$, as $n \rightarrow \infty$, and for $\alpha > 1$ $W_i^{(n)} \rightarrow \infty$, as $n \rightarrow \infty$; $1 \leq i \leq m$.

The theorems 5, 6 may be proved using the formulas (3), (4) and [5, Corollaries 1, 2]. The conditions of the theorems 5, 6 may be realized in the following way:

$$\frac{\lambda_1}{\mu_1} = \ldots = \frac{\lambda_m}{\mu_m} = a, \ \lambda = \lambda(n) = \frac{1-n^{-\alpha}}{a}.$$

Theorem 7 If $1-\rho_i(n) \sim n^{-\alpha_i}$ then for $0 < \alpha_i < 1/2$ we have $N_i^{(n)} \rightarrow 0$, as $n \rightarrow \infty$, and for $\alpha_i > 1/2$ $N_i^{(n)} \rightarrow \infty$, as $n \rightarrow \infty$; $1 \leq i \leq m$.

Theorem 8 If $1-\rho_i(n) \sim n^{-\alpha_i}$ then for $0 < \alpha_i < 1$ $W_i^{(n)} \rightarrow 0$, as $n \rightarrow \infty$, and for $\alpha_i > 1$ $W_i^{(n)} \rightarrow \infty$, as $n \rightarrow \infty$; $1 \leq i \leq m$.

The theorems 7, 8 may be proved using the formulas (3), (4) and [5, Corollaries 1, 2]. The conditions of the theorems 7, 8 may be realized in the following way:

$$\lambda(n) \equiv \lambda, \ \mu_i(n) = \frac{\lambda \lambda_i}{1-n^{-\alpha_i}}.$$

Remark that if we consider multi phase queueing system with loaded nodes then from the theorem 8 it is possible to obtain in last nodes a very small queue and in first nodes - a very large queue. This property may be interpreted as a cooperative network property. An existence of such network properties were noted by some specialists in computer science.
5 Superposition of Queueing Networks

In this section a problem of a calculation of limit distributions in a superposition of queueing networks is solved. A superposition of queueing network is an open or a closed network in which some nodes are replaced by opened networks. Limit distributions in these superpositions are defined by limit distributions in initial networks using product formulas.

Such constructions allow to calculate their distributions in a few steps of networks superposition. In this section an ability to handle customers in a superposition of networks is calculated. Network superpositions may be interpreted as recursively defined structures. Such structures now are widely used in a nanotechnology.

5.1 Product Theorem

Consider opened Jackson networks $G, G'$ described in Subsection 1.1 with the sets of onserver nodes $\{g_0,g_1,\ldots,g_m\}, \{g'_0,g'_1,\ldots,g'_r\}$, with the input Poisson flows (with single intensities) and with the service intensities $\mu_1,\ldots,\mu_m$ and $\mu'_1,\ldots,\mu'_r$. Denote by $\Theta = [\theta_{ij}]_{i,j=0}^m$, $\Theta' = [\theta'_{ij}]_{i,j=0}^r$ the route matrices of the networks $G, G'$. Define the superposition $\mathcal{G} = G \otimes G'$ of the networks $G, G'$ by a replacement of the node $g_m$ in $G$ by the network $G'$. Here an input flow (output flow) of the network $G'$ is created from customers arriving (departing) to the node (from the node) $g_m$. In the network $\mathcal{G}$ the input flow is Poisson with the single intensity, the node set is

$$\{\bar{\varphi}_0, \bar{\varphi}_1, \ldots, \bar{\varphi}_{m+r-1}\} = \{g_0,g_1,\ldots,g_m-1, g'_1,\ldots, g'_r\}$$

and the route matrix $\bar{\bar{\Theta}} = [\bar{\theta}_{ij}]_{i,j=0}^{m+r-1}$ is defined from the formulas $\bar{\theta}_{ij} = \theta_{ij}$, $i,j = 0, 1, \ldots, m - 1$,

$$\bar{\varphi}_{m-1+i, m-1+j} = \theta_{im}\theta'_{oj}, \quad i = 0, 1, \ldots, m - 1, \quad j = 1, \ldots, r,$$

$$\bar{\varphi}_{m-1+i, m-1+j} = \theta'_{ij}\theta_{mj}, \quad i = 1, \ldots, r, \quad j = 0, 1, \ldots, m - 1.$$

Lemma 1 If $\theta_{mm} = 0$ then the matrix $\bar{\bar{\Theta}} = [\bar{\theta}_{ij}]_{i,j=0}^{m+r-1}$ satisfies properties of a route matrix and is indivisible.

Proof: The matrix $\bar{\bar{\Theta}}$ is route as all its elements are nonnegative and for $i = 0, 1, \ldots, m - 1$ and for $i = 1, \ldots, r$

$$\sum_{j=0}^{m+r-1} \bar{\varphi}_{m-1+i,j} = \sum_{j=0}^{m-1} \theta_{ij} + \sum_{j=1}^{r} \theta'_{ij} = 1 - \theta_{im} + \theta_{im} = 1,$$

$$\sum_{j=0}^{m+r-1} \bar{\varphi}_{m-1+i,j} = \sum_{j=0}^{m-1} \theta_{ij} + \sum_{j=1}^{r} \theta'_{ij} = 1 - \theta_{im} + \theta_{im} = 1,$$

and for $i = 1, \ldots, r$

$$\sum_{j=0}^{m+r-1} \bar{\varphi}_{m-1+i,j} = \sum_{j=0}^{m-1} \theta_{ij} + \sum_{j=1}^{r} \theta'_{ij} = 1 - \theta_{im} + \theta_{im} = 1.$$

Show that for $\forall i, j \in \{0, 1, \ldots, m + r - 1\}$

$$\exists \kappa_1, \ldots, \kappa_s \in \{1, \ldots, m + r - 1\} : \bar{\theta}_{\kappa_1}, \bar{\theta}_{\kappa_2}, \ldots, \bar{\theta}_{\kappa_s} > 0. \quad (18)$$

If $i,j \in \{0, 1, \ldots, m - 1\}$ then the formulas $(18)$ may be obtained from the matrix $\Theta$ (from the matrix $\Theta'$) indivisibility. If $i \in \{0, 1, \ldots, m - 1\}$, $j = \{m, \ldots, m + r - 1\}$ then from the matrix $\Theta$ indivisibility $\exists i_1, \ldots, i_s \in \{1, \ldots, m - 1\}$

$$\bar{\theta}_{i_1} > 0, \ldots, \bar{\theta}_{i_{s-1}} > 0, \bar{\theta}_{i_s} > 0,$$

and from the matrix $\Theta'$ indivisibility $\exists j_1, \ldots, j_n \in \{1, \ldots, r\}$

$$\theta'_{0, m-1+j_1} > 0, \theta'_{m-1+j_1, m-1+j_2} > 0, \ldots, \theta'_{m-1+j_n, j_m+1} > 0.$$

As $\bar{\theta}_{i_1} > 0, \ldots, \bar{\theta}_{i_{s-1}} > 0, \bar{\theta}_{i_s} > 0,$

$$\bar{\theta}_{i_1} > 0, \ldots, \bar{\theta}_{i_{s-1}} > 0, \bar{\theta}_{i_s} > 0,$$

the case $i \in \{m, \ldots, m + r - 1\}$ is considered similar.

From Lemma 1 of [2, § 2] there is the single vector $\bar{\Lambda}$ so that

$$\bar{\Lambda} = \bar{\Theta} \bar{\Lambda} = (1, \lambda_1, \ldots, \lambda_{m+r-1}). \quad (19)$$

Lemma 2 If $\theta_{mm} = 0$ then

$$\bar{\lambda}_i = \lambda_i, \quad i = 1, \ldots, m - 1, \quad \bar{\lambda}_{m-1+i} = \lambda_m \lambda'_i, \quad i = 1, \ldots, r.$$

Here $\Lambda = (1, \lambda_1, \lambda_2, \ldots, \lambda_m), \Lambda' = (1, \lambda'_1, \lambda'_2, \ldots, \lambda'_r)$ are solutions of the systems $\Lambda = \Lambda \Theta, \Lambda' = \Lambda' \Theta'$.

Proof: Indeed if $i \in \{1, \ldots, m - 1\}$ then from the formula $(19)$ obtain

$$\lambda_i = \lambda_i = \bar{\theta}_{0i} + \sum_{j=1}^{m+r-1} \bar{\lambda}_j \bar{\theta}_{ji} = \theta_{0i} + \sum_{j=1}^{m-1} \lambda_j \theta_{ji} + \sum_{j=1}^{r} \lambda_m \lambda'_j \theta'_{0j} \theta_{mi} = \lambda_i.
If \( i \in \{1, \ldots, r\} \) then
\[
\lambda_m \lambda'_i = \lambda_{m-1+i} = \theta_0 m_{-1+i} + \sum_{j=1}^{m+r-1} \lambda_j \theta_j m_{-1+i} = \theta'_0 \theta_{0m} + \sum_{j=1}^{m-1} \lambda_j \theta'_j m_\theta + \sum_{j=1}^{r} \lambda_m \lambda'_j \theta'_j m_\theta = \theta'_0 \theta_{0m} + \theta'_0 \lambda_m - \lambda_{m \theta_{0m}} + \lambda_m \lambda'_i - \lambda_m \theta'_0 = \lambda_m \lambda'_i.
\]

Describe a dynamics of the network \( G \) by the discrete Markov process \( \Psi(t) \) with the state set \( \Psi = \{ \Psi = (n_1, \ldots, n_{m-1}, n_1', \ldots, n_r') : n_{11}, \ldots, n_{m-1}, n_1', \ldots, n_r' \geq 0 \} \).

Theorem 9 If \( \theta_{mm} = 0 \) and \( \rho_i = \lambda_i / \mu_i < 1, \ i = 1, \ldots, m-1, \ \rho'_i = \lambda_m \lambda'_i / \mu'_i < 1, \ i = 1, \ldots r, \) then the Markov process \( \Psi(t) \) is ergodic and its limit distribution \( P(\Psi) \), \( \Psi \in \Psi \), has the form
\[
P(\Psi) = \prod_{i=1}^{m-1} a_i(n_i) \prod_{i=1}^r a'_i(n'_i),
\]

\[
a_i(n_i) = (1 - \rho_i) \rho_i^{n_i}, \ i \in \{1, \ldots, m\},
\]

\[
a'_i(n'_i) = (1 - \rho'_i) \rho'_i^{n'_i}, \ i \in \{1, \ldots r\}.
\]

Proof: The network \( G \) is Jackson network and so the theorem statement is a corollary of the lemmas 1, 2 and [2, theorem 2.1]. \( \square \)

5.2 Abilities to Handle Customers of Opened Networks

Suppose that the network \( G \) satisfies the theorem 9 conditions. Construct the following sequence of the opened networks. Replace each node of the network \( G \) by the network \( G \) and denote this network by \( G^{(1)} \). In the network \( G^{(1)} \) replace each node of the network \( G \) and obtain the network \( G^{(2)} \) and so on. After \( n \) steps obtain the network \( G^{(n)} \) with \( m^{n+1} \) nodes among which there are \( m^n \) nodes with the service intensity \( \mu_k \). Denote the route matrix of the network \( G^{(i)} \) by \( \Theta^{(i)} \) and suppose that the single solution of the system \( \Delta^{(i)} = \Delta^{(i)}(\Theta^{(i)}) \) is \( \Delta^{(i)} = (1, \lambda_1^{(i)}, \ldots, \lambda_{m^{(i)}}^{(i)}, 1) \).

If \( \theta_{ii} = 0, \ i = 1, \ldots, m, \) then from the lemma 2 the network \( G^{(1)} \) satisfies the formulas

\[
\lambda_k^{(1)} = \lambda_k \lambda_1^{(1)}, \ \lambda_{m+k}^{(1)} = \lambda_k \lambda_2^{(1)}, \ldots, \lambda_{m^{(1)}-m+k}^{(1)} = \lambda_k \lambda_m^{(1)}.
\]

\( k = 1, \ldots, m, \) and the network \( G^{(2)} \) satisfies the formulas

\[
\lambda_k^{(2)} = \lambda_k \lambda_1^{(2)}, \ \lambda_{m+k}^{(2)} = \lambda_k \lambda_2^{(2)}, \ldots, \lambda_{m^{(2)}-m+k}^{(2)} = \lambda_k \lambda_m^{(2)}.
\]

Then the network \( G^{(n)} \) satisfies the formulas

\[
\lambda_k^{(n)} = \lambda_k \lambda_1^{(n-1)}, \ \lambda_{m+k}^{(n)} = \lambda_k \lambda_2^{(n-1)}, \ldots, \lambda_{m^{(n)}-m+k}^{(n)} = \lambda_k \lambda_m^{(n)},
\]

\( k = 1, \ldots, m, \) so in the network \( G^{(n)} \) all nodes with the service intensities \( \mu_k \) correspond to the vector \( \Lambda^{(n)} \) components of the form

\[
\lambda_k^{(n-1)}, \text{where} \ \lambda_j^{(n-1)} = \lambda_1^{h_1} \lambda_2^{h_2} \ldots \lambda_n^{h_n},
\]

\( h_1 + \ldots + h_n = n. \)

Calculate now the ability to handle customers \( a_n \) of the network \( G^{(n)} \). From [1] obtain

\[
a_n = \min_{1 \leq k \leq m} \frac{\mu_k}{\lambda_k \lambda_j^{(n-1)}} = \min_{1 \leq k \leq m} \frac{\mu_k}{\lambda_k (n \lambda_j)^n}.
\]

If \( \min_{1 \leq j \leq \lambda_j} \frac{1}{n \lambda_j} < 1 \) then \( a_n \to 0, \ n \to \infty, \) if \( \min_{1 \leq j \leq \lambda_j} \frac{1}{n \lambda_j} > 1 \) then \( a_n \to \infty, \ n \to \infty, \) if \( \min_{1 \leq j \leq \lambda_j} \frac{1}{n \lambda_j} = 1 \) then \( a_n \equiv \min_{1 \leq k \leq m} \frac{\mu_k}{\lambda_k}. \)

Acknowledgements: The research was supported by the Far Eastern Branch of RAS (grant No. 09-I-12-07).

References:


Design of Control Systems Based on Vector Error

Oleg N. Agamalov

Tashlyk Pump-Storage Power Plant, Ukraine (e-mail: aon@ukrsat.mk.ua).

Abstract: a new approach to the design of control systems is offered. This approach based on definition the vector error allowing combining in real time a stage of identification of unknown plant and calculation of control. The measurement method of a vector error with use of Hilbert transform also it is shown.

Keywords: Vector error, identification, nonlinear system, Hilbert transform, stability.

1. INTRODUCTION

Development of control systems is directed on design adaptive and robust controllers allowing providing required quality of control for unknown, nonlinear plants at various conditions and restrictions. The review of the literature on control systems shows, that alongside with modern directions adaptive and robust controllers [1-9], such as fuzzy control, artificial neural networks, genetic algorithms, $H^2$, $H^\infty$, $\mu$-synthesis, etc. for plant control in the industry, power systems, transport and other still remains to the most widespread classical PID controller structures [10-16]. It is caused by simplicity of its design and tuning. However for PID control cannot be taken into account the basic features of nonlinear systems:

- The superposition principle is not carried out. The research of nonlinear systems at several influences cannot be reduced to research at the sum of single influence;
- The plant stability to depend at initial deviation from position of balance;
- At the fixed external influences some positions of balance are possible.

Development of control systems both on the basis of modern control algorithms and with use of the PID control demands an exact model of plant. For this purpose it is necessary to execute identification of plant: defined its dynamic behaviors - change of an outputs $y$ at all possible changes of an inputs $u$, noises $n$ and disturbances $d$. Performance of the given problem theoretical is possible, however in practice in all cases is impracticable on restrictions of technological character and impossibility to predict $n$ and $d$.

Therefore control systems development, providing quality, stability and robust control achievable on the basis of last modern algorithms and having simplicity of a design inherent in the PID control, is perspective. Performance of the given requirements probably if identification of plant to carry out continuously and changing parameters or even structure of controller depending at received results. Taking into account complexity of sensors installation for measurement state variables of plant, identification is preferable to carry out used information about output change of plant $\Delta y$ depending on an input change of plant $\Delta u$ at real parameters $n$ and $d$. For this purpose it is necessary to measure not only $y$ and $u$, but also time (dynamic) relations between changes of an input and an output under influence $n$ and $d$.

The decision of the given scientific problem can be achieved if an error of control system to present as a vector [17, 18]. The Vector Error of control system is a difference vector between an input vector of controller (vector of reference $r$) and an output vector of plant that allows identifying control system (controller + plant) at influences of noises $n$ and disturbances $d$ and to take into account delays and the unknown order of plant. On each step of control change of the real component of vector error defines intensity of a control while the argument or a phase of a vector error defines character of a feedback in control system - positive, negative or its absence on some steps of control - i.e. feedforward control.

In paper is considered the opportunity application a Hilbert transform for measurement vector error of control system. All input and an output of control system are represented as analytical signals. It is allows to apply the Hilbert transform. Using the given transform, differences of instant phases for all required combinations of inputs - outputs of plant can be determined. On the basis of the received differences of instant phases of vector errors and their components are calculated the optimum control for each of inputs.

The paper includes the following sections. In the second section definition of a vector error of control system and the technical method of its measurement is shown. This method based on representation of inputs - outputs of plant as analytical signals and use of Hilbert transform for calculation of differences of their instant phases. In the third section the algorithm and block diagram of a vector controller and definition stability of control system with a vector controller is considered. Examples of application an offered vector controller for control of the first order plant with a transport delay and also more complex example as an automatic voltage regulator (AVR) of the synchronous generator with use of SIMULINK are resulted in the fourth section. Conclusions are presented in Sec.5.
2. THE VECTOR ERROR OF CONTROL SYSTEM

Modern control systems for maintenance their adaptive and robust properties should be created on the basis of the algorithms which are not demanding the detailed aprioristic information about plant, capable to make identification, structural and parametrical optimization based on continuous measurement inputs and outputs of plant. The control system with disturbance \( d \) and noise \( n \) is given in Fig. 1.

![Control system with disturbance and noise](image)

where \( r \) - reference, \( e = r - y_2 \) - a control error, \( u \) - a control input, \( y \) - a plant output, dependent at state variables \( x \), \( d \) - disturbance, \( n \) - noise, \( y_2 \) - a measured output of plant.

For identification based on the information about inputs and outputs it is determined the vector error \( e \) which is function not only difference of the modules reference \( r \) and a measured output of plant \( y_2 \), but also the difference phases between them \( e = r - y_2 = e_R + j e_I \). Thus the module \( e \) and argument \( \Delta \phi \) a vector error, its real \( e_R \) and imaginary \( e_I \) components will be defined according to geometrical ratio:

\[
e^2 = r^2 + y_2^2 - 2 \cdot r \cdot y_2 \cdot \cos(\Delta \phi),
\]

\[
\Delta \phi = \arctg \left( \frac{e_I}{r + e_R} \right) = \arcsin \left( \frac{r + e_R}{y_2} \right),
\]

\[
e_R = r - y_2 \cdot \cos(\Delta \phi),
\]

\[
e_I = y_2 \cdot \sin(\Delta \phi),
\]

where \( r \) - the module of a reference vector concerning which displacement of an output of plant is determined, \( y_2 \) - the module of an output of plant, \( \Delta \phi \) - difference phase between a reference vector and an output vector. Thus, the real part of a vector error \( e_R \) takes into account dynamics of plant \( y_2 \) and noise, \( e_I \) - a control error, \( e_R \) - a control inputs, \( y_2 \) - an output, \( \Delta \phi \) - a differences of instant phases. The algorithm of vector error control system shown in Fig. 2:

![Graphic interpretation of a vector error](image)

The difference of instant phases between an output and an input can be determined with use of Hilbert transform, representing an input and an output as analytical signals. As is known [19, 20], the analytical signal represents the sum of two orthogonal signals with components are shifted on 90º to each other. For analytical signal can be determined the instant phase and instant frequency. The imaginary part of an analytical signal \( Z(t) \) is analytically connected with its real part \( \text{Re}[Z(t)] = s(t) \) through Hilbert transform HT:

\[
\text{Im}[Z(t)] = s'(t) = HT[s(t)]
\]

and accordingly the analytical signal is represented as: \( Z(t) = s(t) + j s'(t) \). The difference of instant phases of two signals \( s_1(t) \) and \( s_2(t) \) can be determined with use of Hilbert transform as [21]:

\[
\Delta \phi_{12} = \phi_1(t) - \phi_2(t) = \arctan \frac{\bar{s}_1(t) \cdot s_2(t) - s_1(t) \cdot \bar{s}_2(t)}{s_1(t) \cdot s_2(t) + \bar{s}_1(t) \cdot \bar{s}_2(t)}.
\]

For control system with single input and single output (SISO) a difference of their instant phases:

\[
\Delta \phi_m = \phi_m(t) - \phi_m(t) = \arctan \frac{\bar{y}_m(t) \cdot r(t) - y_m(t) \cdot \bar{r}(t)}{y_m(t) \cdot r(t) + \bar{y}_m(t) \cdot \bar{r}(t)}.
\]

The calculated differences of instant phases (3), (4) can be used for identification of plant, calculation of a vector error and its components according to (1) and control of plant.

3. CONTROL SYSTEM WITH VECTOR CONTROLLER

Control system with a vector controller based on the Hilbert transform shown in Fig. 3:

![Control system with vector controller](image)

where \( C \) - controller, \( P \) - plant, \( \text{HT} \) - block the Hilbert transform and calculate of a difference instant phases between input and output, \( r \) - a reference vector, \( e_R \) - a real component of the vector error, \( u \) - a control inputs, \( y_2 \) - an output, \( \Delta \phi_{mk} \) - a differences of instant phases. The algorithm of vector controller given in Fig. 4:
The Calculation of an instantaneous difference of phases $\Delta \varphi_{\text{inst}}$

The Calculation of a control $u$

The Calculation of vector $r$

The Calculation of Hilbert transform $HT$

The Calculation of real component of vector error $E_r$

The Calculation of an output vector $Y$

Fig. 4. Algorithm of a vector controller

The control with the account (1), can be submitted as:

$$u(t) = K_v e_y(t) = K_v (r - y_e \cdot \cos(\Delta \varphi)),$$  \hspace{1cm} (5)

where $K_v$ – the gain of a vector controller.

Let’s consider a nonlinear control system, Fig.5 ($y_e = y$):

Fig. 5. The nonlinear control system.

$$\dot{y}(t) = y(t) \cdot f(*) + u(t).$$  \hspace{1cm} (6)

With the account (5) dynamics of nonlinear control system can be submitted as:

$$\dot{y}(t) = y(t) \cdot f(*) + K_v (r - y_e \cdot \cos(\Delta \varphi)).$$  \hspace{1cm} (7)

The gain $K_v$ is determined from the stability analysis of control system. For this purpose we shall define the Lyapunov’s function in the square-law form concerning the real component of vector error and a control input, Fig. 6:

Fig. 6. The Lyapunov’s function determined by values of the real component of vector error and a control input.

$$L(e_y, u) = e_y^2 + u^2 = \left(1 + K_v^2 K_e \right) e_y^2 \left(1 + K_v^2 e_y^2 \right) \cdot \left(\frac{y^2 + y^2 \cos^2(\Delta \phi) - 2 y y \cos(\Delta \phi)}{r + y \cdot \cos(\Delta \phi)}\right).$$  \hspace{1cm} (8)

The Lyapunov’s function is positively determined. We shall calculate a derivative of Lyapunov’s function (8) for control system (6) along trajectories of change the real component of vector error $e_y$ and a control input $u$:

$$L(e_y, u) = 2 \left(1 + K_v^2 \right) \cdot \left(\frac{r \cdot y \cdot \sin(\Delta \phi) - r \cdot y \cdot \cos(\Delta \phi) + y \cdot \cos^2(\Delta \phi) - y^2 \cdot \cos(\Delta \phi) \cdot \sin(\Delta \phi)}{r + y \cdot \cos(\Delta \phi)}\right).$$  \hspace{1cm} (9)

According to Lyapunov’s theorem [1,2] stability of control system is provided, if the derivative of function (9) will be negatively determined:

$$\left(\frac{r \cdot y \cdot \sin(\Delta \phi) - r \cdot y \cdot \cos(\Delta \phi) + y \cdot \cos^2(\Delta \phi) - y^2 \cdot \cos(\Delta \phi) \cdot \sin(\Delta \phi)}{r + y \cdot \cos(\Delta \phi)}\right) < 0$$

Substituted value (7), we shall receive a stability condition of nonlinear control system with a vector controller:

$$K_v \left(\frac{r \cdot f(*) + y \cdot \sin(\Delta \phi) - (\cos(\Delta \phi) + r \cdot \tan(\Delta \phi))}{r + y \cdot \cos(\Delta \phi)}\right)$$

We research an inequality (10) at change of a difference of phases $\Delta \varphi$:
In case $rf(*) \gg 1$ inequalities (11) determining stability of nonlinear control system can be simplified:

$$\Delta \phi \rightarrow 0 \Rightarrow K_v < \frac{y - r \cdot f(*) - 1}{r},$$

$$\Delta \phi \rightarrow \pm \frac{\pi}{2} \Rightarrow K_v < \infty,$$

$$\Delta \phi \rightarrow \pm \pi \Rightarrow K_v < \frac{y - r \cdot f(*) + 1}{r - y}.$$  \tag{11}

Inequalities (10) - (12) allow defining allowable area of values the gain of vector controller $K_v$ for nonlinear control system proceeding only from the measured values of inputs - outputs and a difference of phases between them.

Further are considered an example of application a vector controller for control of the first order plant with a transport delay, and also more complex example of application as an automatic voltage regulator (AVR) of the synchronous generator.

4. THE APPLICATIONS OF VECTOR CONTROLLER

4.1 The control system for first order plant with a transport delay.

Let’s consider the first order plant with a transport delay:

$$W(s) = \frac{K_p}{1 + sT} e^{-sL},$$ \tag{13}

where $s$ - complex frequency, $K_p = 1$ – gain of plant, $T = 1$ - time constant of plant, $L = 0.3$ - transport delay of plant. The optimum values of the PID controller for this plant have been determined used method CHR (Chien, Hrones, Reswick) [22] and the manual adjustment: $K = 3$, $T_i = 1$, $T_d = 0.0452$. The control system with the plant (13) it has been simulated with use of SIMULINK and it is shown on Fig. 7:

The measurement noise of control system is submitted as block “Band-Limited White Noise” at: Noise power = 0.001, Sample time = 0.01, Seed = 23341. Disturbance of control system is submitted as block “Random Number” with parameters: Mean = 0, Variance = 0.001, Initial seed = 0. The PID controller is shown on Fig. 8:

Properties of the control system were investigated at submission of measurement noise and disturbance at reference $r = 1$ and the submission a step signal at the time $t = 15s$. Received transient for plant with the nominal parameters is shown on Fig. 9:
The properties of control system were investigated at change parameters of plant (16) in range $K_p = 0.5 - 2$, $T = 0.5 - 2$, $L = 0.3 - 1$. The received results, Figs. 10-14, shown low adaptive and robust properties of control system with a PID controller adjusted for plant with nominal parameters:

**Fig. 10.** The transient for plant (13) at $L = 0.7$.

**Fig. 11.** The transient for plant (13) at $K_p = 0.5$.

**Fig. 12.** The transient for plant (13) at $K_p = 2$.

**Fig. 13.** The transient for plant (13) at $T = 0.5$.

**Fig. 14.** The transient for plant (13) at $T = 2$.

Further it is considered the control system for plant (13) and the vector controller (5), shown on Fig. 15, 16:
Fig. 15. The control system with a vector controller.

For measurement a difference of phases between a reference and an output according to (1) and (3) the circuit shown on Fig. 17 has been developed:

Fig. 16. The vector controller for plant (13)

Fig. 17. The circuit for measurement a difference of phases based on the Hilbert transforms.

Properties of control system with a vector controller were investigated under the same conditions as for control system with the adjusted PID controller. Value of gain $K_V = 2.5$ was accepted proceeding from (10) - (12). The transient for plant with nominal parameters is shown on fig. 18:

Fig. 18. The transient for control system with vector controller and plant with the nominal parameters.

Properties the control system with a vector controller was investigated at similar changes of plant (13). The received results, Figs. 19-23, have shown high adaptive and robust properties control system with the vector controller (5):

Fig. 19. The transient for plant (13) at $L = 0.7$ with vector controller.
Fig. 20. The transient for plant (13) at $K_P = 0.5$ with vector controller.

Fig. 21. The transient for plant (13) at $K_P = 2$ with vector controller.

Fig. 22. The transient for plant (13) at $T = 0.5$ with vector controller.

Fig. 23. The transient for plant (13) at $T = 2$ with vector controller.

The control system with vector controller for first order plant with a transport delay provides comprehensible quality of control and stability at change of plant parameters in significant ranges. The divergence of plant output from required at change of gain plant $K_P$ can be eliminated by introduction the adaptive linear gain $K_V$ from an plant output as the given divergence has linear character.

4.2 The vector automatic voltage regulator of the synchronous generator.

The offered vector controller is applied to model « synchronous machine – infinite bus» (SMIB) as an automatic voltage regulator (AVR) of the synchronous generator. The model "power_turbine.mdl" SIMULINK represents work of the three-phase synchronous hydrogenerator with nominal parameters: full capacity 200 MVA, rated voltage 13.8 kV, nominal frequency 112.5 min$^{-1}$ connected to a power system by power 10000 MVA and voltage 220 kV through the block transformer and a long transmission line with a ratio of inductive and active resistance $X/R = 10$. A three-phase short circuit in a network 220 kV by duration 0.2s and also submission a step signals on inputs references of terminal voltage $V_t$ and active power $P_{ref}$ was simulated. The size of step signals varied in a range of 0.1-0.3 p.u. for reference of terminal voltage and 0.05-0.15 p.u. for active power. AVR represents a vector controller with input $\Delta V = V_r - V_t$, where $V_t$ - a vector of a terminal voltage, $V_r$ - a vector of reference. For measurement of phase difference $\delta$ between of this vectors simulated the equation (3), shown on Fig. 17. The given block allows to develop vector AVR integrated it in structure of excitation system standard IEEE type 1 [23]. The excitation system shown on Fig. 24.
The experiments with given model have shown, that stability of the synchronous hydrogenerator is provided at three-phase short circuit with duration 0.25s for a wide range of changes a model parameters: reference a terminal voltage $V_r = 0.7-1.1$ p.u., active power $P = 0.8-1.3$ p.u., frequency $f = 0.9-1.1$ p.u., power of system $S = 2000-10000$ MVA, ratio $X/R = 1-100$. Changes frequency and power of the hydrogenerator in a transient shown in Figs 25-26:

In comparison with AVR standard IEEE type 1, a vector controller possesses the greater robust to change of parameters. The problem of a vector controller it is an adaptation of gain $K_V$ to change of parameters plant. The given question defines a direction of the further researches.

5. CONCLUSIONS
1. The new approach for development of control system with use of a vector controller is considered. Vector controller based on Hilbert transform and allows executing identification and control of plant not only depending on value, but also difference phases between its input (reference) and an output.
2. Conditions of stability nonlinear control system with a vector controller are determined.
3. Applications of a vector controller for control the first order plant with a transport delay and as AVR of the synchronous generator shown robust properties of the offered algorithm.

REFERENCES

[8] Chih-Min Lin, Wei-Liang Chin and Chung-Li, Adaptive Hierarchical Fuzzy Sliding-Mode Control for a Class of Cou-


THE POSSIBLE CAUSES OF TAX EVASION IN ROMANIA

COMANICIU CARMEN
Faculty of Economic Sciences
"Lucian Blaga" University of Sibiu
Sibiu, Calea Dumbravii, no. 17
ROMANIA

carmen_comaniciu@yahoo.com  http://www.economice.ulbsibiu.ro

Abstract: Only a thorough and systematic analysis of mechanisms escapist phenomenon may lead to the choice of methods, techniques and measures capable of predictable and combat tax evasion. What causes that generate tax evasion phenomenon? and What are the possibilities to combat tax evasion? Here are two questions dependency. If the answer to first question is complete and relevant, response is obtained for the second question, because it can be discovered ways to combat and eradicate the phenomenon escapist. Through this article, we will try to enumerate the possible causes of tax evasion in Romania, while specifying the possible solutions for their elimination.

Key-Words: causes of tax evasion, tax legislation, misunderstanding, high tax pressure, ineffective and inefficient, fiscal control, tax education

1 Introduction
When burdens are too high and push too much on the material taxable, taxpayers tend is to "escape". Human nature always put first individual interests to the detriment of general interest. In this regard, taxpayers tried and tried to evade tax payment, using various means, as yet be considered mandatory levies bring damage to property and earnings. Although escapist phenomenon can not be stopped completely, there are techniques and methods of control and mitigation, so, the injury of public money to be minimized. The phenomenon of tax evasion is treated mostly from the problematization of the concept, continuing with forms of manifestation and implementation mechanisms. Taxpayers desire to be sheltered from tax obligations there have been since established for the first time taxes and there, as long as the purpose and role of taxation is not understood, and action is not reflected in measures to establish a optimal tax system.

Decrease tax evasion phenomenon can be achieved only under conditions which take into account the existence of interdependencies among elements characterizes tax evasion (definition of tax evasion, manifestations, causes of tax evasion, effects of tax evasion, measures to prevent and combat tax evasion) and can anticipate the action of the active and ingenious in the scope of coverage.

2 Problem Formulation
Only a thorough and systematic analysis of mechanisms escapist phenomenon in Romania may lead to the choice of methods, techniques and measures capable of predictable and combat tax evasion.

Taxpayers tend to escape the tax burden, there is and sometimes acquire large scale, by the circumvention of tax laws, reason to have always sought solutions to remove this phenomenon, primarily due to effects generated.

Evasion of tax payment existed, exists and will exist. Only forms and procedures for commission change over time. Fiscal fraud, manifested in deliberate circumvention of tax liabilities, is as aggravating circumventio, being studied, analyzed, legislated and punished whhervver it is detected.

What causes that generate tax evasion phenomenon in Romania? and What are the possibilities to combat tax evasion in Romania? Here are two questions dependency. If the answer to first question is complete and relevant, response is obtained for the second question, because it can be discovered ways to combat and eradicate the phenomenon escapist. It is true that tax evasion will not disappear entirely, but techniques, tools and methods used by the tax unit will lead to a restriction of this phenomenon, not allowing his development, both horizontally and vertically.

The tax evasion causes are multiple, resulting in reduced taxable material and entirely escape taxation. We will try to enumerate the possible causes of tax evasion in Romania, while specifying the possible solutions for their elimination:

- The existence of tax legislation with gaps, inconsistencies, ambiguities, flaws and repeated changes. One of the basic principles of taxation
refers to provisions in tax legislation. In this context, tax law must be clear, concise, do not give rise to any arbitrary and known in advance by taxpayers. Failure to observe this principle entails the ability of the taxpayer, so, it seeks in the tax law the shall agree on matters of taxation. A tax law on which there hung qualities such as clarity, relevance, appropriateness, etc., will be difficult to enforce both by taxpayers and by organizations concerned with the supervision of tax payment obligations.

Line separating areas of tax evasion is difficult to draw, starting from simple mistakes and knowingly breach reaching tax provisions. At the risk of incurring sanctions, such penalties, increases and fines, taxpayers using all means which may attract a reduced payment of tax liabilities. [1]

Taking advantage of loopholes in the tax system, taxpayers seeking items masking taxable material. From this perspective, an incomplete declaration can be classified as income tax evasion. Also, income from financial activities held abroad and unreported to national tax authorities fall into the category of acts of tax evasion. The examples could continue, this is why it takes "skill" of the tax authority to detect what acts lead to real tax fraud. [2]

In Romania, in a given period of time, number of fiscal normative acts is very high (for example for the period 23.06.2010 to 20.08.2010 were issued 34 tax laws [3]), resulting impossibility of appropriate applying from taxpayers. Basic Law of the Romanian tax is Tax Code, published in the Official Gazette. 927/23.12.2003. From publication and until 14.07.2010, it has undergone many changes, through 71 acts. [4] Also, through the Government Emergency Ordinance amending Law 571/2003 regarding the Fiscal Code, valid from 1 July 2010, the 32 points of order make amendments to 17 articles of the Tax Code. [5]

These are just some of the reasons that led us to consider the Romanian tax legislation as a cause of tax evasion. Responsible for elimination of this tax evasion case is the legislature.

- Misunderstanding of the tax authorities of the reasons that determine the taxpayer to evade the tax system. Taxpayer behavior before taxes must be analyzed in terms of social and psychological variables.

Knowledge and understanding of this behavior for authority with responsibilities in tax will generate optimal solutions, for change, modification and improvement behavior of all "actors on tax stage", so, place and role of taxation to be fully understood.

Variables considered by psychologists in terms of taxpayer behavior are stigma, reputation and social norms. For this reason, tax authorities must find a correlation between the threat of punishment for acts of tax evasion and the taxpayer's intention to tax evasion, so this leads to discourage taxpayers to behave properly and establish a sense of guilt. Recourse to the conscience and civic education, if properly targeted, can positively change the five positions of the taxpayer's motivation, namely: commitment, capitulation, resistance, non commitment and play. [6]

Number, size and major changes of taxes, will always cause taxpayers to choose or to migrate to that legal form, assumed in their point of view, the best option tax. In this respect, requires that any micro or macro decision must to consider the structure of tax payers, because is possible that a measure of fiscal policy that aimed to increase tax revenues, leading to a reduction in their.

In this respect, an example is the introduction in Romania the minimum tax, since 1 may 2009, according to the portion of total income corresponding previous fiscal year. The reaction of politicians, analysts and business to that tax change came immediately after approving the Emergency Ordinance, particularly on the legal form of organization for taxpayers. Thus, on 16.04.2009, stated that "at least 20% of the total number of 478.173 registered microenterprises in the year 2007 will turn in authorized individuals, to escape the flat tax" [7]. In July 2009, details were as follows: "Following the introduction of the minimum tax, the number of authorized individuals registered at the first five months was 26.254, up 32.6% over the same period last year ... This increase was due mainly to the introduction of the minimum tax, because many businesses were closed, preferring to become authorized individuals ... Number of firms that entered into insolvency in the first five months of this year increased by 60,5% from 6.035 to 9.687. Meanwhile, the number of registrations during January to May was 54.225, down 12% from 66.124 companies in the same period last year and canceled number rose to 21.345". [8]

Increasing the number of authorized individuals and legal entities decrease from may 2009, is mainly due to introducing flat tax, because, while companies with legal personality required to pay the standard rate, whether done or not income, individuals pay tax depending on income, in real system or by the rule of income. Underline again, the need for prior approval of an amendment in the tax system, to undertake analysis of the impact of that changes to the structure of taxpayer.

- The existence of high tax pressure. Ratio of all taxes and social contributions actually received
by government and gross domestic product is compulsory levies or tax burden rate broadly. [9]

In light of the tax burden, it can be addressed both in terms of flow and in terms of indices. In terms of flow, fiscal pressure is the amount of tax liability on that income support, both at the individual level and at sectoral and global level. [10]

In terms of indices, fiscal pressure reflecting flow ratio of tax revenue and creating economic flows, in the monetary expression. [11]

In the contributor language, term fiscal pressure is associated with term fiscal burden, which includes many tax obligations, tax rates and high taxation, tax laws doubt, pressures on the taxpayer, differences in tax, etc.

The large number of acts and deeds of tax evasion can be explained starting from the classical definition of fiscal pressure and taking into account what is perceived by the taxpayer through this term.

Often a simple enumeration and summation of its tax rates leading some taxpayers to believe that there is a high level of fiscal pressure and act within the scope of tax evasion.

According to the fourth study conducted by World Bank experts and representatives of PricewaterhouseCoopers, to the taxation of 183 states, in the year 2009, the situation for Romania is as follows: Depending on the number of payments made by a company, at 2009, Romania ranks 182, with 113 payments. Biggest deviation from EU average in terms of total payments of tax liabilities recorded in Romania. If the EU average is 18 payments of duty, their number in Romania is 6.28 higher. Within the European Union, Romania ranks 14, with an average number of hours per year required for tax obligations of 202 hours. [12]

Presentation of these studies (achieved either nationally or internationally) highlights the need to refine the work done at institutions with tasks tax, because, state placement of one of the last positions, in terms of taxation, may entail, decrease trust in unit tax and the existence of escapist behavior from taxpayers.

Applying a tax system based on equity and efficiency will lead to understanding the role of taxes and contributions to the budget revenues. Although at first sight, implies a tax rate increase to attract additional resources from state budget, whether this increase is not fully justified will generate increased tax evasion.

Most significant factors influencing the tax burden are: performance of the economy at a time; reduced efficiency in the use of public financial resources obtained on account of taxes and contributions; the low level of understanding of taxpayers on budget needs; level of democracy achieved by the population, etc.

Based on these factors, we consider the relationship between taxation and fiscal policy should be based on economic and social normality and reasoning.

- **Ineffective and inefficient in use of public financial resources.** To combat tax evasion, the authorities must consider how public goods and services are offered, quantitatively but also qualitatively.

Taxpayer behavior is often inadequate because does not receive the expected benefit and even promised, in return for taxes paid. In this context, public authorities must show the interdependence between fiscal policy and budget policy, showing where the taxpayer's money are. [13]

To support that inefficient use of public money becomes an issue of tax evasion in Romania, we will proceed to identify negative examples: a) Legislation on public procurement offers many wickets for targeted transfer of money from the public to the private, and authorities have not yet required to reduce the effectiveness of the phenomenon. [14]; b) From 1 January to 31 October 2009, were completed 1.924 actions by Romanian Court of Accounts, which: 1.267 financial audits, 246 performance audits and 411 control action. The results of these activities have resulted in providing a number of 225 certificates of compliance; issuing a number of 1.177 decision; arrangement a number of 6.021 measures to remove irregularities and deficiencies noted. [15] The data presented are a significant number of large deviations, which means that there is inefficiency in use resource in the public institutions, institutions have to provide quality services to all citizens.

- **Existence precarious financial status among taxpayers.** In the case of a financial jam, taxpayers among legal persons found how hard it is to work without cash, the amount theoretically available on paper, under any contract, but which can not be used to pay suppliers, employees and taxes.

For this reason, companies choose their priorities and challenges that make such payments, taxes and duties remain the last.

Extending financial blockade in the economy is another factor that increases the escapist phenomenon, because it represents the total of all debts recorded between public institutions and firms, between firms and suppliers or between companies and banks.

Economic specialists claim that the State would be the first actor should act to settle the economic and financial deadlock, regardless of political color, but using the system of taxes as the main economic and financial leverage.
• **Inefficiency guidance and assistance to taxpayers.** Guidance and assistance to taxpayers is to provide general information about the way in which taxpayers must do to qualify for a tax law or to comply with a duty. Scope of guidance and assistance activities include taxes, social contributions and other revenues administered by the tax administration.[16]

Sometimes, the lack of professionalism of the tax authorities representatives and differences in tax laws lead in providing information to taxpayers likely to induce the latter to choose the path that requires the lowest payment of taxes and fees.

• **Inefficiency of fiscal control.** Effectiveness of fiscal control subjects must be considered control techniques and procedures, normative-legal means or how to recover tax revenues are found to be evading the payment.

From the theoretical point of view, effectiveness of fiscal supervision must continually reported in its mission to encourage voluntary tax compliance of taxpayers and tax compliance.[17] If at implementing this goal does not exist, we face a series of "small or large arrangements" between taxpayers and inspectors, leading to increased tax evasion.

Analyzing the data presented by the National Agency for Fiscal Administration in late 2009, it is enough negative aspects regarding the degree of budgetary revenue collection and degree of voluntary compliance in paying taxes and contributions. Regarding the degree of compliance of taxpayers to declare and pay tax obligations, in the year 2009 have given less encouraging, especially for payment of taxes and contributions, namely: degree of voluntary compliance for declaration - 83.9%; degree of voluntary compliance for payment (number) - 64.3%; degree of voluntary compliance for payment (value) - 77.4%. [18]

High fines and imprisonment no longer scare taxpayers, because the number of crimes of tax evasion is high, injury was significantly, and the number of sectors with high rates of tax evasion is increasing. In Romania, in the year 2009, Romanian Police found 9,293 acts of tax evasion, in that 7,514 people were investigated which produced a loss estimated at 1,754,663,040 lei. [19] At EU level, for 2009, Romania has ranked second in the tax evasion top (with 42.7% of unreported taxable income), being surpassed by Italy (51%), followed by Bulgaria (39,2%) Estonia (37,6%), Slovakia (34,1%). [20]

Control is the key element in managing any socio-economic system and exercise preventive fiscal control function is a clear way to prevent tax evasion.

• **The existence of a low fiscal education.** Taxation is the essential component of social and economic life of a nation. In order to understand this appearance, requires an fiscal education, in that, tax is not considered injury to a income or wealth of the taxpayer, but a tool through which to develop the personality of each citizen.

Fiscal education must be made with regard to age groups, so, the information submitted must be understood and over time deepen the role of taxation.

Fiscal education must be considered an essential part of education for review. Thus, for children of primary school, penetrate into the mysteries of taxes should be done with fables, such "King's Artax and taxes" or "Tax evasion, a temptation as old as the world". Pupils are able to assimilate knowledge about contribution in terms of capacity, report income - tax, income characteristics. For students, information submitted relating to the tax area may include characteristics of the tax system and tax policy, components of the tax system, tax returns, role and functions of the fiscal system, etc. [21]

## 3 Problem Solution

Through direct or indirect effects of tax evasion, fight against it must continue to exist, because the damage of this phenomenon are immense.

At first glance, tax evasion prejudice only to the state, taxpayers are are those who earn. In fact, those who feel or will feel the effect of tax evasion (on short or long term) are, equally, fraudsters and non-fraudsters. Effects of tax evasion directly reflected on the levels of tax revenues, lead to distortions in the market mechanism and can contribute to social inequity.

Willful theft of tax obligations for a certain period, bring benefits for fraudsters, in the same period, for non-escapist, this removal leads to the increased fiscal pressure. Thus, by redistributing the tax burden, the inevitable will happen, taxpayers morale will destroy.

Measures taken by public authorities for combating escapist will have indirect effects on all taxpayers, starting from the increasing of tax law complexity and reaching potential erosion of civil rights or legal professional privilege.

The direct effect of tax evasion for the economy, having regard to the reduction of budget revenues, will be reflected in unemployment, rising inflation, increasing interest rates, non-financing sectors of national interest. In this context, nation as a whole will be "sick" and suffering will be for all taxpayers. [22]
4 Conclusion
In the current period, Romanian tax system needs improvement.

Identifying opportunities for improvement in the tax area in Romania, requires consideration of the following issues: establish an optimal correlation between the structure of taxpayers and changes in tax system; tax return analysis; strengthening the role of fiscal management to achieve fiscal activity performance; increase voluntary compliance; improve the image of romanian taxation; appropriate use of numerical fiscal rules; efficient use of an Integrated Financial Management System.

The benefit of these training opportunities will be felt, only insofar as, will diminish the acts and deeds of tax evasion.

"The vicious circle of taxation" creates "pain". State "complain" frequently for low level of public resources, and taxpayers "complain" for the extent of tax liability and the way is spent public money. It can break this vicious circle? The answer is yes, provided that, the rationality level of taxation offers better responsiveness to the taxpayer burden.

Fără a avea pretenŃia că au fost epuizate toate cauzele ce generează apariŃia evaziunii fiscale, considerăm demn de luat în seamă, eliminarea cauzelor specificate, pentru a reuşi distrugerea cercului vicios al fiscalităŃii din românia.

Without claiming that have been exhausted all causes that generate the appearance of tax evasion, we consider that is required to eliminate the causes specified, to succeed destroy the "vicious circle" of taxation in Romania.

References:
[16] Order no. 1.338/30.09.2008 of MEF on procedure guidance and assistance of taxpayers by tax authorities
Swarm Intelligence for Mixed Pixel Resolution

V.K. PANCHAL*, NITISH GUPTA†

*Defence Terrain Research Laboratory, Defence Research and Development Organisation
†Guru Gobind Singh Indraprastha University
New Delhi
INDIA

*vkpans@gmail.com, †ntshgpt@gmail.com

Abstract: - Mixed pixels are usually the biggest reason for lowered success in classification accuracy. Aiming at the characteristics of remote sensing image classification, the mixed pixel problem is one of the main factors that affect the improvement of classification precision in remote sensing classification. How to decompose the mixed pixels precisely and effectively for multispectral/hyper spectral remote sensing images is a critical issue for the quantitative remote sensing research. As Remote sensing data is widely used for the classification of types of land cover such as vegetation, water body etc. thus Conflicts are one of the most characteristic attributes in satellite remote sensing multilayer imagery. Conflict occurs in tagging class label to mixed pixels that encompass spectral response of different land cover on the ground element. In this paper we attempted to present a new approach for resolving the mixed pixels using Biogeography based optimization. The paper deals with the idea of tagging the mixed pixel to a particular class by finding the best suitable class for it using the concept of immigration and emigration.

Key-Words: - Geographical Information systems, Decision Support, Process Modeling, Artificial Intelligence, Mixed Pixel, Biogeography based optimization, Remote Sensing.

1 Introduction
Remote sensing [1] data can be put to use in classifying the features in an image into distinct categories. The categorized images can then be used in different ways – a farmer may use thematic maps to monitor the health of his crops without going out to the field. A geologist may use the images to study the types of minerals or rock structure found in a certain area. A biologist may want to study the variety of plants in a certain location. Remote sensing with multi-spectral satellite imagery works on the concept that different features of the land cover reflect electromagnetic radiation over a range of wavelengths in their own characteristic way according to their chemical composition and physical state. The problem arises when the terrain features do not correspond to the pure object or feature and may not have a prior spectral signature library. Pixels which are in the interface region of two classes e.g. vegetation & Water body, cannot be clubbed together to a single category, say vegetation, because these pixels do not follow any one particular class’s spectral signature. These types of pixels are generally known as mixed pixels. Assigning a class tag to the mixed pixels is cause of conflict [2] in the expert’s mind and it is the focus point of the study. Mixed pixel resolution of remote sensing images is one of valid assistance means for improving quality of feature extraction from the images. A new algorithm implement mixed pixel resolution of remote sensing images and takes full advantage of neighbourhood information, makes the resolution result more human, and presents better robustness to environment.

The paper is organized into four sections. The section following the introduction illustrates the problem formulation which deals with the concepts of Biogeography based optimization and mixed pixels origin and then methodology. The third section describes problem solution in which the new algorithm for mixed pixel resolution is being proposed. The fourth section show experiment and summarizes the important findings for this paper, is to observe this mixed pixel scenario and attempt to resolve the mixed pixel and tag them with a specific class.

2 Problem Formulation

2.1 Biogeography Based Optimization
Biogeography is the study of the distribution of biodiversity spatially and temporally. Over areal ecological changes, it is also tied to the concepts of
species. Immigration is the introduction of new people into a habitat or population. It is a biological concept and is important in population ecology. A habitat (which is Latin for "it inhabits") is an ecological or environmental area that is inhabited by a particular species of animal, plant or other type of organism. It is the natural environment in which an organism lives, or the physical environment that surrounds (influences and is utilized by) a species population.

Biogeography unfolds the geographical distribution of biological organisms. The mindset of the engineer is that we can learn from nature. This motivates the application of biogeography to optimization problems. The science of biogeography can be traced to the Work of nineteenth century naturalists such as Alfred Wallace [3] and Charles Darwin [4]. In the early 1960s, Robert MacArthur and Edward Wilson began working together on mathematical models of biogeography, their work culminating with the classic 1967 publication The Theory of Island Biogeography [5]. Their interest was primarily focused on the distribution of species among neighbouring islands. The application of biogeography to engineering is similar to what has occurred in the past few decades with genetic algorithms (GAs), neural networks, fuzzy logic, particle swarm optimization (PSO), and other areas of computer intelligence.

The term “island” here is used descriptively rather than literally. That is, an island is any habitat that is geographically isolated from other habitats. We therefore use the more generic term “habitat” in this paper (rather than “island”) [6]. Geographical areas that are well suited as residences for biological species are said to have a high habitat suitability index (HSI) [7]. Features that correlate with HSI include such factors as rainfall, diversity of vegetation, diversity of topographic features, land area, and temperature. The variables that characterize habitability are called suitability index Variables (SIVs). SIVs can be considered the independent variables of the habitat, and HSI can be considered the dependent variable.

Habitats with a high HSI have many species that emigrate to nearby habitats, simply by virtue of the large number of species that they host. Habitats with a high HSI have a low species immigration rate because they are already nearly saturated with species. Habitats with a low HSI have a high species immigration rate because of their sparse populations. Biogeography is nature’s way of distributing species, and is analogous to general problem solutions. Suppose that we are presented with a problem and some candidate solutions.

A good solution is analogous to an island with a high HSI, and a poor solution represents an island with a low HSI. High HSI solutions resist change more than low HSI solutions. We call this new approach to problem solving biogeography-based optimization (BBO)[8].

2.2 Mixed Pixel
Each pixel [9] of remote sensing image contains the information from multifarious ground objects due to the difference from the resolution of remote sensing image, called Mixed pixel.

The basic building block of the study is the concept of mixed pixel, and what way the concept representing vital knowledge is elaborated using rough sets. A fundamental assumption that is commonly made in remote sensing is that each pixel in the image represents the area on the Earth’s surface that contains a single class. This is often not the case, with the mixed pixels containing presence of more than one class.

Origin of the mixed pixels may be attributed to the following:
1. Mixed caused by the presence of small, sub-pixel targets within the area it represents (Fig 1.i).
2. Mixing as a result of the pixel straddling the boundary of discrete thematic classes (Fig 1.ii).
3. Mixing due to gradual transition observed between continuous thematic classes (Fig 1.iii).
4. Mixing problem due to the contribution of a target (black spot) outside the area represented by a pure but influenced by its point spread function. (Fig 1.iv)

Also the pixel appears superficially be pure. The portion of mixed pixels in an image is often large in remote sensing studies. The exact proportion of mixed pixels in an image is an interactive function of properties of sensor (e.g. the class composition, spatial arrangement) and so the mixed pixel problem is a contextual issue as well. To tag the mixed pixel is really a matter of concern and requires strong analysis of terrain features before tagging it to a particular class.
2.3 Methodology
The image is classified into pure pixels corresponding to different land features like water, vegetation, rocky, barren and urban. Hence beforehand we had the data set of mixed pixels and pure pixels. Our objective is to resolve the mixed pixel and tag it to the specific class to which it revolves on the basis of the seven band digital data (DN values). We have taken a multi-spectral multi-resolution & multi-sensor image of Alwar area in Rajasthan. The area is selected since it contains a good variety of land use features like urban, water body, rocky, vegetation & barren areas (Fig.2).

The multi-spectral geo-referenced image-set consists of satellite images of dimension 472 X 546 pixels. The 4-spectral bands are in the visible bands namely: red, green, near- infrared (NIR) and middle infra-red (MIR) from LISS-III sensor of Indian Remote Sensing sat satellite Resourcesat-1. Also, two SAR images namely: low incidence S1 beam -20° - 27° (RS1) and other is High incidence S7 beam 45° - 49°, (RS2) of the same area taken from Canadian Radarsat-1 satellite. The seventh band is digital elevation model (DEM) of the area. The ground is resolution of the image from LISS-III and Radarsat-1 image is 23.5m and 10m respectively. The DEM dataset is also generated from SAR interferometry using RS1 and RS2 and have 25-meter resolution.

We are having spectral signatures from seven bands namely Red, Green, NIR, MIR, RS1, RS2 and DEM of the mixed classes (water-vegetation) (Table.1) and training data set of pure pixels (water) (Table.2) and training data set of pure pixels(vegetation) (Table 3) provided by the expert. The table has, therefore 7 attributes i.e. Red, Green, Near-Infrared, Middle Infrared, Radarsat-1, Radarsat-2, DEM.

3 Problem Solution
3.1 Swarm Based Mixed Pixel Resolution Algorithm
3.1.1 Definition
Definition 1: A habitat is a vector of integers that represents a feasible solution to some problem (Here habitat is the data set of pure pixels being classified).

Definition 2: A suitability index variable (SIV) is an integer that is allowed in a habitat, is the set of all integers that are allowed in a habitat. (SIV’s are the seven band DN values i.e. Red, Green, Near-
Infrared, Middle Infrared, Radarsat-1, Radarsat-2, DEM).

Definition 3: A habitat suitability index HSI: is a measure of the goodness of the solution that is represented by the habitat. (HIS is the mean of standard deviation of each band represented as column in table).

Definition 4: An ecosystem is a group of habitats. The size of an ecosystem is constant (A set mixed pixels with each mixed pixel being considered as habitat).

Definition 5: Immigration rate is a monotonically Non increasing function of HSI. is proportional to the likelihood that SIVs from neighbouring habitats will migrate into habitat.

Definition 6: Emigration rate is a monotonically Non-decreasing function of HSI. is proportional to the likelihood that SIVs from habitat will migrate into neighbouring habitats.

3.1.2. Algorithm
Frame work of swarm intelligence based mixed pixel resolution algorithm.

Input- Dataset of Pure and mixed pixels of land features.

Output- All mixed pixels are classified.

[Begin]

Initialization
Condition= no of different sets of mixed pixel /* Data set of Water-vegetation, urban-rocky, urban- Vegetation are taken*/

Reading training data set of all pure pixels.
Reading training data set of all mixed pixels.

While (condition! = 0)
{
    /* taking one unique set of mixed pixel for each Iteration for condition*/
    Loop= no of pixels in mixed training data set taken.
    Original_HSI_1= mean (standard deviation of each Band DN values of pure pixel data set of class_1 of which a mixed pixel corresponds).
    Original_HSI_2= mean (standard deviation of each band DN values of pure pixel data set of class_2 of which a mixed pixel corresponds).
    for(i=0; i< Loop; i++)
    {
        Add pixel [Loop] from mixed pixel to tables of both the pure pixel of which the Mixed pixel corresponds. /* Emigration*/
        Calculate New_HSI_1, New_HSI_2 /* after Addition*/
        Deviation_1=Original_HSI_1- New_HSI_1;
        Deviation_2=Original_HSI_2- New_HSI_2;
        If (Deviation_1<Deviation_2)
        {
            Classify Pixel [Loop] as Class 1;
            /*Immigration*/
        }
        Else
        Classify Pixel [loop] as class_2;
        /*Immigration*/
    }
}
[End]

4 Conclusion

4.1 Experiment

Alwar district is situated in the northeast of Rajasthan between 27o4’ and 28o4’ north Latitudes and 76o7’ and 77o13’ east Longitude. Its greatest length from south to north is about 137 K.M. and greatest breadth from east to west about 110 K.M. The image of the district with dimensions 472 X 546 is taken. On the basis of spectral behaviour observed by the sensors of satellite, The image is being classified into set of pure pixels of various land features and different pairs of mixed pixels depending upon the land featured encountered in image.

Now the swarm based mixed pixel resolution algorithm is applied on training data set of pure pixels and mixed pixels. On the basis of DN values of bands of mixed pixels, these are tagged under unique class which shows the less deviation after adding that mixed pixel in it.

The Figure 3 and Figure 4 shows the deviation of water-vegetation mixed pixels (9 pixels of distinguished values from data set of mixed pixel are taken) from pure water and pure vegetation respectively. Hence which ever shows the less deviation from the pure pixel class, that pixel is classified and iteration is carried out till all the
mixed pixels are being tagged (Table 4). Similarly the case of urban-rocky mixed pixels is taken and graph is plotted and all the mixed pixel cases are solved using this algorithm.

Judging from the experimental results, the mixed pixel is resolved and hence the whole image is classified totally.

Due to constraint of space, only the water-vegetation mixed pixels resolution is shown (9 pixels).

![Figure 3. Mixed Pixel in Pure water and their respective deviation](image)

![Figure 4. Mixed Pixel in pure vegetation and their respective deviation](image)

<table>
<thead>
<tr>
<th>Pixel</th>
<th>Deviation from Pure water dataset</th>
<th>Deviation from Pure vegetation dataset</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.106</td>
<td>0.5928</td>
<td>Water</td>
</tr>
<tr>
<td>2</td>
<td>2.183</td>
<td>0.578</td>
<td>Vegetation</td>
</tr>
<tr>
<td>3</td>
<td>0.92</td>
<td>0.5785</td>
<td>Vegetation</td>
</tr>
<tr>
<td>4</td>
<td>0.92</td>
<td>0.2513</td>
<td>Vegetation</td>
</tr>
<tr>
<td>5</td>
<td>0.92</td>
<td>0.6721</td>
<td>Water</td>
</tr>
<tr>
<td>6</td>
<td>2.158</td>
<td>0.0002</td>
<td>Vegetation</td>
</tr>
<tr>
<td>7</td>
<td>6.0004</td>
<td>0.7543</td>
<td>Water</td>
</tr>
<tr>
<td>8</td>
<td>2.181</td>
<td>0.5805</td>
<td>Vegetation</td>
</tr>
<tr>
<td>9</td>
<td>0.9216</td>
<td>0.6199</td>
<td>Water</td>
</tr>
</tbody>
</table>

Table 4. Mixed pixels are being classified (assumed reflectance values being considered on the basis of common platform of each land feature like vegetation is being observed as on the platform of photo-synthesis).

1. The mixed pixels are resolved using the BBO technique considering all the permissible values of seven bands, which are used for observing the land feature.
2. The new algorithm is based on the local search of individuals in swarm so the method utilizes adequately neighbouring information of mixed pixels in range.
3. Spectral bands have definite contribution towards the mixed pixel resolution.

4.2 Conclusion

Swarm based mixed pixel resolution algorithm is put forward based on the mechanism of Biogeography based optimization in view that every pixel of particular land feature comes under the a range of DN (reflected) values or they are related on the basis of closeness and similarity between them.

References:
Modulation Transfer Function Compensation through a modified Wiener Filter for spatial image quality improvement

JESSICA WONG SOO MEE
Malaysia Space Centre
National Space Agency of Malaysia (ANGKASA)
Lot 2233, Jalan Turi, Sg. Lang, 42700, Banting, Selangor
MALAYSIA
jessica@angkasa.gov.my   http://www.angkasa.gov.my

Abstract: - The knowledge of Modulation Transfer Function (MTF) for a degraded image provides a mean to compensate for the image degradation, which improves the image quality in terms of sharpness. MTF compensating is in the image restoration; hence it is fundamentally an ill-posed problem. This paper proposes a stable and flexible filtering technique that executes an optimal tradeoff between sharpness and noise to warrant an acceptable result of image restoration. The MTF compensating is performed through a modified Wiener filter, and can be reduced to a well-posed problem by incorporating the regularization method. The modified Wiener filter employed the L-curve method for selection of optimal regularization parameter, and introduces an inverse control and smooth control parameter that allows freedom of tuning when the importance of image sharpness versus noise is being compromised. The data sets used in the analysis were synthetically blurred remotely-sensed images simulated from Level-2A product of IKONOS. The results by modified Wiener filter were analyzed and presented, they were found to be able to reduce the Mean Square Error by more than 50% with respect to the original image, and it results in a significant improvement of spatial image quality.

Key-Words: - Modulation Transfer Function, Spatial image quality, Wiener filter, Image restoration, Ill-posed problem, L-curve

1 Introduction
Spatial image quality is one of the key parameters for characterizing and validating image data. The technical characteristics of image data must be well understood before image analysis can be done, since the quality of the analysis depends on the quality of the data. Image quality can be measured by a vast number of factors, such as contrast, brightness, noise variance, sharpness, radiometric resolution, granularity, modulation, contrast transfer function and many more [1]. Among other measures, image sharpness is vital for characterizing images, for much of the information of an image resides in its edges. Image blurriness, which limits the visibility of details, can be objectively measured by the point spread function (PSF), or its amplitude spectrum, which is the modulation transfer function (MTF). Spatial image degradation can happen in many ways. For satellite imaging, image acquisition occurs while orbiting the earth, and due to the satellite’s altitude determination control for maneuvering, the instantaneous field of view (IFOV) of the imaging system can be greater or lesser than the nominal resolution at any point in time during image capturing. This results in MTF degradation proportional to the ratio of IFOV and Ground Sample Distance (GSD) [2]. Moreover, regardless of how well an image system is fabricated, it will inevitably suffer from some degree of blur. These sorts of degradations need to be compensated, and they can be compensated through the knowledge of the degradation function which is the MTF for that image restoration. MTF compensating is in the image restoration; hence it is fundamentally an ill-posed problem. This paper proposes a stable and flexible filtering technique that executes an optimal tradeoff between sharpness and noise to warrant an acceptable result of image restoration.

2 MTF Compensation
MTF Compensation (MTFC) is a restoration technique; Restoration techniques find filters that apply inverse process in order to recover its original image. In image restoration, the main challenge is to prevent noise of input data from being amplified to unacceptable artifacts in the restored image. Image restoration is ill-posed, which means the MTFC is also an ill-posed problem, meaning that it does not have a unique solution, even with the absence of noise [3]. However, scientists and engineers are usually less concerned with existence and uniqueness of the solution to the ill-posed problem and worry more about the stability of the solution. With the concept of regularization, the ill-posed problem of the
existing MTFC can be reduced to a well-posed problem by introducing a regularization parameter as a priori constraints to produce a stable and flexible deconvolution filter for image restoration. In this case, it is stable in the sense that it will always produce an acceptable image restoration result, while also being flexible by allowing freedom of tuning when the importance of image sharpness versus noise is compromised.

2.1 Deconvolution Filter
As a deconvolution filter, Wiener filtering is an inverse problem, therefore it is likely to be ill-posed. Regularization is necessary while solving inverse problem because of the “naïve” least squares solution by Wiener filter. By using regularization, the error contribution can be damped and the residual norm can be kept in reasonable size. In other words, by incorporating regularization parameter into the Wiener filter, it is possible to control the power of inverse filtering, and subsequently minimizes the amplification of noise in the image. The Wiener filter has been used by other researchers [4] [5] as an MTFC technique to compensate for the degradation of remotely-sensed images. And it also has been used by Shacham et al.[6] as blind restoration technique to compensate for the degradation of astronomical data. However, none of them applied the regularization method in their filters, thus a modified Wiener filter based on the idea in the regularization method is employed in this work.

2.1.1 The Modified Wiener Filter
In this paper, the proposed deconvolution filter considers images and noise to be a random process, and is rooted in one of the most widely used deconvolution filters, which is the Wiener filter. In order to regularize the inverse problem of the Wiener filter, the concept of the Tikhonov regularization was adopted. According to Wiener’s theory [7], given a degraded image \(g(x, y)\), the degraded image can be restored by \(W(u, v)\) in a Fourier sense, as shown by the following equations:

\[
W(u, v) = \frac{H^*(u, v)S_f(u, v)}{|H(u, v)|^2S_f(u, v) + S_n(u, v)}
\]

\[
= \frac{H^*(u, v)}{|H(u, v)|^2} \frac{1}{S_f(u, v)/S_n(u, v)} \quad \text{(2)}
\]

\[
= \frac{1}{H(u, v)} \frac{|H(u, v)|^2}{|H(u, v)|^2 + \frac{S_n(u, v)}{S_f(u, v)} \quad \text{(3)}
\]

where \(H(u, v)\) represents the degradation function; \(S_n(u, v)\) represents power spectrum of the noise; \(S_f(u, v)\) represents power spectrum of the undistorted image and the pair \((u, v)\) represents the location in the spatial frequency domain.

From equation (3), the Wiener filter can be perceived to have two separate parts: an inverse filtering part and a noise smoothing part. It not only performs the deconvolution by inverse filtering but also removes the noise with a compression operation. These properties were advanced to solve the research problem.

In the absence of noise (i.e. \(S_n(u, v) = 0\)), the Wiener filter reduces to the inverse filter. On the other hand, in the presence of noise (i.e. \(S_n(u, v) > 0\)), the denominator of (2) is never zero. In the limiting case \(S_n(u, v) \rightarrow 0\), then

\[
W(u, v) = \begin{cases} \frac{1}{|H(u, v)|} & H(u, v) \neq 0 \\ 0 & H(u, v) = 0 \end{cases} \quad \text{(4)}
\]

thus no ill-posed problems arise in Wiener filtering. The term \(\frac{S_n(u, v)}{S_f(u, v)}\) in (3) can be regarded as a term that smooths the inverse filter \(\frac{1}{|H(u, v)|}\). However, for frequencies at which \(S_f(u, v) \leq S_n(u, v)\), estimation of \(W(u, v)\) becomes nearly equal to \(\frac{1}{|H(u, v)|}\), which means the Wiener filter behaves as an inverse at the frequencies \((u, v)\). At high spatial frequencies, \(S_f(u, v)\) is often much larger than \(S_n(u, v)\), thus condition \(S_f(u, v) \geq S_n(u, v)\) holds true. For degradation such as blur, \(|H(u, v)|\) at high spatial frequencies is relatively small. When \((u, v)\) approaches a region with a large value of \(\sqrt{u^2 + v^2}\), \(|H(u, v)|\) is nearly zero. Therefore, when \(S_f(u, v) \geq S_n(u, v)\) and \(|H(u, v)| \approx 0\), the Wiener estimate \(w(u, v) = \frac{1}{|H(u, v)|}\) becomes ill-posed. Consequently, this estimate will lead to a poor restoration result.

For the ill-posed inverse problem to be solved, it is necessary to regularize the solution by introducing a priori constraints [8]. This can be done by using a numerical regularization parameter to control the power of the inverse filter, thereby penalizing those \((u, v)\) with very high spectral frequencies. The \(\frac{S_n(u, v)}{S_f(u, v)}\) in equation (3) is assumed to be constantly proportional to the inverse of the signal-to-noise ratio (SNR) of the image. This leads to the following modification of \(W(u, v)\) in equation (3):

\[
W(u, v) = \frac{1}{H(u, v)} \frac{|H(u, v)|^2}{|H(u, v)|^2 + \frac{\lambda}{\text{SNR}}} \quad \text{(5)}
\]
Thus, the frequency domain solution to optimize the degradation problem is expressed as

\[
\hat{F}(u,v) = \left[ \frac{1}{H(u,v)} \right] \cdot \left[ \frac{H(u,v)^2 + \frac{1}{\text{SNR}}}{H(u,v)^2} \right] G(u,v) \quad (6)
\]

where \( \hat{F}(u,v) \) represents the restored image, \( G(u,v) \) represents the degraded image, and \( \lambda \) represents the regularization parameter. The regularization parameter was determined by using the L-curve method. The L-curve is a parametric plot of the size of the regularized solution and its corresponding residual \[9\]. The underlying idea is that a good method for choosing the regularization parameter for discrete ill-posed problems must incorporate information about the solution size, in addition to using information about the residual size. The L-curve method requires both the regularized solution and its corresponding residual to find the corner of the L-shape curve for the optimal selection of regularization. With the underlying degradation is the linear, position-invariant degradation that is described in a spatial domain by

\[
g = Hf + \eta \quad (7)
\]

where \( g \) represents the degraded image, \( f \) represents the original image, matrix \( H \) represents the degradation function, and \( \eta \) represents the additive white noise. In this case, the knowledge of noise variance is not required. Thus the regularization problem can be expressed using the Tikhonov regularization as

\[
g - \hat{Hf} = \lambda \left\| \hat{\tilde{r}} \right\| \quad (8)
\]

Let \( \left\| \tilde{r} \right\| \) be the “residual” vector described by

\[
\left\| \tilde{r} \right\|^2 = \left\| g - \hat{Hf} \right\|^2 \quad (9)
\]

The \( \left\| \tilde{r} \right\|^2 \) can be calculated using the following equation:

\[
\left\| \tilde{r} \right\|^2 = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} r^2(x,y) \quad (10)
\]

where \( r^2(x,y) = 3^{-1} \left[ G(u,v) - H(u,v) \hat{F}(u,v) \right] \) (11) substituting the right side of equation (11) for \( \hat{F}(u,v) \) in equation (6) yields

\[
r^2(x,y) = 3^{-1} \left[ \frac{G(u,v)}{H(u,v)^2 + \frac{1}{\text{SNR}}} \right] \quad (12)
\]

where \( 3^{-1} \) represents inverse Fourier transform. Meanwhile, \( \left\| \tilde{r} \right\|^2 \) can be calculated using the following equation:

\[
\left\| \tilde{r} \right\|^2 = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} r^2(x,y) \quad (13)
\]

where \( \tilde{r}^2(x,y) = 3^{-1} \left[ \hat{F}(u,v) \right] \)

The procedures for determining the regularization parameter are as follows:

- **Step 1:** Specify an initial value of \( \lambda \)
- **Step 2:** Calculate \( \left\| \tilde{r} \right\|^2 \) and \( \left\| \hat{\tilde{r}} \right\|^2 \)
- **Step 3:** Gradually increase \( \lambda \) until \( \lambda = 1 \);
- **Step 4:** Convert both \( \left\| \tilde{r} \right\|^2 \) and \( \left\| \hat{\tilde{r}} \right\|^2 \) to logarithmic functions.
- **Step 5:** Plot the L-curve in log-log scale (see Fig.1).
- **Step 6:** Trace the transition of the shape of the corner of the L-curve plot, and mark the optimal \( \lambda \),
- **Step 7:** Use the optimal \( \lambda \) to compute the optimal estimate \( \hat{F}(u,v) \)

As mentioned earlier, the Wiener filter can be perceived to have two separate parts: an inverse filtering part which contributes to the sharpness for deblurring, and a noise smoothing part for denoising. Therefore it is possible to control the regularization and inverse power independently. Subsequently, it allows more degrees of freedom for image restoration that incorporate both deblurring and denoising techniques in one single filter. Hence the proposed filter is as expressed as:

\[
W(u,v) = \left( \frac{1}{H(u,v)} \right)^{\beta} \left( \frac{H(u,v)^2 + \frac{1}{\text{SNR}}} {H(u,v)} \right)^{\gamma} \quad (14)
\]

where \( \beta \) and \( \gamma \) represent the inverse control parameter and the smooth control parameter respectively.
3 Experimental Results
The original image used in this work is presented in Fig. 2; it is one part of Level 2A product of IKONOS, containing 7800 x 7800 pixels.

![Fig. 2. The original image](image)

Fig. 3 presents one of the blurred version of the images obtained by using the inverse Fourier transform, by multiplying the original (in Fourier sense) with a Gaussian filter. The Gaussian filter function, \( H_G(u,v) \) which approximates the optical point spread function in the frequency domain is expressed as:

\[
H_G(u,v) = e^{- \frac{(u^2+v^2)}{2\sigma^2}}
\]  

(15)

where \( \sigma \) is the standard deviation of Gaussian curve. The standard deviation, \( \sigma \) of Gaussian function for this blurred image was set to 80. The data sets for this work consist of sub-scenes extracted from this synthetically blurred image.

![Fig. 3. A synthetically blurred image](image)

The proposed MTFC algorithm was applied to data sets and evaluated quantitatively using two metric, namely Mean-Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), which are defined as follows:

\[
MSE = \frac{1}{MN} \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} [f(x,y) - \hat{f}(x,y)]^2
\]  

(16)

\[
PSNR = -10 \log_{10} \frac{MSE}{S^2}
\]  

(17)

where \( f, \hat{f} \) and \( S \) represent the original image, the restored images and the maximum pixel value of the image, respectively.

Two versions of blurred images were used to present the experimental result of MTFC. The standard deviation, \( \sigma \) of Gaussian function for these images is 50 and 80, respectively (see Fig. 4(a) and (b)). The degradation function, \( H(u,v) \), which is the MTF of the degraded image, was modeled using the measured MTF by edge input method. The measured (estimated) MTF were compensated using the modified Wiener filter as described in Equation (14) to restore these images. The regularization parameter, \( \lambda \) was tuned between \( 10^{-6} \) and \( 10^{2} \). Experimental result shows that regularization parameter, \( \lambda = 10^{-1} \) is relatively good to optimize the regularization of restoration problem. The L-curve plot for regularization process of the blurred image in Fig. 4(a) and (b) is presented in Fig. 4(c) and (d), respectively.

By visual observation, it is clear that the restored images in Fig. 4(e) and (f) are considerably better than their respective blurred image shown in Fig. 4(a) and (b). It can be observed that the blurred images are sharpened without the side effect of noise amplification. Typically, blurred images are assumed to have a lower quality as compared to the restored images. Therefore the MTF for the blurred images are expected to plummet to lower MTF values as compared to the restored images. Under this circumstance, MTF area under the MTF curve after restoration will be larger than those before restoration. This supposition was confirmed by the two MTF plots depicted in Fig. 4.
One criterion for the restoration quality is the ratio between the areas under the MTF curve after and before restoration. The MTF area (MTFA) ratio for the restored image in Figure 4(e) and (f) are 2.08 and 1.84, respectively. Table 1 shows other forms of quantitative results that quantified the image quality improvement.

<table>
<thead>
<tr>
<th>Performance Measure</th>
<th>Fig. 4(a)</th>
<th>Fig. 4(e)</th>
<th>Fig. 4(b)</th>
<th>Fig. 4(f)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR (dB)</td>
<td>23.17</td>
<td>26.70</td>
<td>25.29</td>
<td>29.07</td>
</tr>
<tr>
<td>MSE</td>
<td>313.24</td>
<td>139.09</td>
<td>192.37</td>
<td>80.48</td>
</tr>
</tbody>
</table>

For comparison purpose, the classical Wiener filter was implemented using Equation (3). Both classical and modified Wiener filter were applied to the synthetically blurred image shown in Figure 4(a). The restored images obtained from both classical and modified Wiener filter are presented in Figure 5(a) and 5(b), respectively. Qualitative measures based on visual observation found that 5(b) displays a more enhanced details as compared to 5(a). Fig. 6 shows the comparison of MTF plot between these two filters; MTF area under the MTF curve for the modified Wiener filter are obviously larger than the classical Wiener filter, which means that it produces a better image quality.
Fig. 6. MTF Comparison between classical and the modified Wiener filter

Table 2 summarizes the performance assessment of these filters; the results demonstrated that the modified Wiener filter was indeed able to produce a better quality of restored image.

Table 2: Performance assessment of classical and the modified Wiener filter

<table>
<thead>
<tr>
<th>Filter</th>
<th>PSNR (dB)</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classical Wiener Filter</td>
<td>23.78</td>
<td>283.98</td>
</tr>
<tr>
<td>Modified Wiener Filter</td>
<td>26.70</td>
<td>139.09</td>
</tr>
</tbody>
</table>

4 Conclusion

This paper describes a stable and flexible MTFC technique through Wiener filter that aims to execute an optimal tradeoff between sharpness and noise to warrant an acceptable result of image restoration for spatial image quality improvement. The technique is based on derivation of the spatial frequency response to edge target input, which is the MTF measurement. The derived MTF measurement was used to model the degradation function to compensate for image degradation. The main novelty this filter is the incorporation of L-curve regularization method into the classical Wiener filter, which then becomes the modified Wiener filter to overcome the ill-posed problem of image restoration. The algorithm was applied to synthetically blurred images. Restoration results, using the estimated MTF and L-curve regularization show significantly improved of spatial quality image, where the restored image by the modified Wiener filter has the Mean Square Error (MSE) reduced by more than 50% with respect to the original image. The ability of the modified Wiener filter to invert blur and suppress noise amplification at the same time demonstrated the robustness of the proposed MTF compensation technique.

References:


Abstract: - Speed and efficiency of sorting algorithms are essential for high speed data processing. FPGA based hardware accelerators show better performance than the general purpose processors. Similarly traditional algorithms may not be always efficient on FPGAs. Sorting networks have come as suitable alternatives which can be implemented on FPGAs efficiently. Each application has its own constraint on latency and throughput. A careful selection of a sorting network with suitable number of pipeline stages performs at higher throughput, without contributing much latency.

Key-Words: - FPGA, Sorting networks.

1 Introduction
With the demand for high speed network and computing, speed and parallel algorithms have become essential tools for development. Many of these operations were performed by a general purpose processor [1]. But now days due to the availability of FPGAs, many researchers try to implement various algorithms on FPGAs more efficiently [2] [3]. FPGAs are often used as hardware accelerators.

One of the commonly used operations in high speed data processing is data sorting. The most commonly used sorting algorithm is Bubble sorting. For efficient and reduced operations implementation of sorting, Batcher proposed a technique of sorting using sorting networks [4] [5] [6]. Many of these are implemented of FPGAs and general purpose processors [7] [8] [9].

In this paper we evaluate various sorting networks based on the complexity and speed, focusing on FPGA implementation. For the analysis purposes all the networks are configured to accept eight unsorted numbers, of eight bit wide each. The result will be the sorted numbers.

2 Sorting Networks

2.1 Bubble Sort
In bubble sort the adjacent pair of data elements are compared and swapped if they are found in wrong order, and this process is repeated until the last two elements of the array are compared. With each pass in the bubble sort, by compare and swap process the smaller elements bubble or move up to their designated locations in the array.

2.2 Shell Sort
Shell Sort is one of the oldest sorting algorithms, which arranges the data sequence in a two dimensional array and then sorts the columns of the array [10], which results in the data sequence being partially sorted. As the process repeats the array becomes narrow. Each time the number of columns will keep decreasing. In the end, the array will have a single column. Shell sort method actually groups the data at each step, rather than sorting the data by itself. At each step, either insertion sort or bubble sort is used to arrange the data. The number of times the data elements need to be rearranged is reduced in this type of sorting method.

Fig.1. Shell Sort, n=8.

2.3 Odd-even transposition sort
In general, odd-even transposition sort compares the adjacent pair of data in an array to be sorted and, if a pair is found to be in the wrong order then those elements are swapped. In the first step, odd index and the adjacent even index elements are compared and are swapped, if found in wrong order. In the next step, even index and the adjacent odd index elements are compared and are swapped, if found in wrong order. This process continues with alternating (odd, even) and (even, odd) phases, until no swapping of data elements are required. Thus the resultant array is a sorted one. This network comprises of the same number of comparator stages as the number of inputs. In each stage either odd or even
index positions are compared with their respective neighbors. Each stage alternates between even and odd.

Fig.2. Odd-even transposition Sort, n=8.

2.4 Bitonic Sort

Bitonic Sort is a sorting algorithm which is designed for parallel machines. On any arbitrary sequence to be sorted, bitonic sort produces a bitonic sequence by sorting the two halves of the input sequence in opposite directions.

Fig.3. Bitonic Sort, n=8.

A bitonic sequence is one, which consists of two sub-sequences, one that monotonically increasing and the other monotonically decreasing. Hence for any arbitrary sequence of length n, in the bitonic sort, first two n/2-element sorts are performed, one increasing and the other is decreasing. This results in an n-element bitonic sequence. This entire sequence is now bitonically sorted to produce a sorted (monotonic) sequence.

2.5 Odd-even merge sort

The earlier odd-even transposition sorting algorithm has a complexity of \(O(n^2)\). With such a complexity, for any large sequence with sequence length n the number of steps to perform a complete sort will be very high in real time situations. Odd- even merge sort solves this problem. In odd-even merge sort, all the odd index elements and even index elements are sorted separately and then merged; this step is repeated until we get a completely sorted sequence. Odd-even merge sort is also called as optimal sorting algorithm.

Fig.4. Odd-even merge sort, n=8.

A summary of complexities of above mentioned given in Table 1.

<table>
<thead>
<tr>
<th>Sort Algorithm</th>
<th>Comparator Stages</th>
<th>Comparator Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd-even transposition sort</td>
<td>(O(n))</td>
<td>(O(n^2))</td>
</tr>
<tr>
<td>Bubble Sort</td>
<td>(O(n))</td>
<td>(O(n^2))</td>
</tr>
<tr>
<td>Bitonic Sort</td>
<td>(O(\log(n)^2))</td>
<td>(O(n \cdot \log(n)^2))</td>
</tr>
<tr>
<td>Odd-Even mergesort</td>
<td>(O(\log(n)^2))</td>
<td>(O(n \cdot \log(n)^2))</td>
</tr>
<tr>
<td>Shellsort</td>
<td>(O(\log(n)^2))</td>
<td>(O(n \cdot \log(n)^2))</td>
</tr>
</tbody>
</table>

Where \(n\) is the number of inputs.

3 FPGA Implementation

All the five algorithms are described by Verilog HDL language in two different approaches. One approach is without any pipeline registers and other is with pipeline registers. Each module accepts eight parallel input data of width eight bits each and the clock. The output is the sorted data. The functionality of each of the designed modules is verified by the simulation in ModelSim, as shown in Fig.5.
Each module is then synthesized and implemented on a using Xilinx ISE [11]. Then each module is tested by downloading the bit file to Xilinx Virtex-5 LX50 FPGA.

<table>
<thead>
<tr>
<th>Sort Network</th>
<th>LUT</th>
<th>FLOP</th>
<th>Frequency (MHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd-even transposition sort</td>
<td>1237</td>
<td>128</td>
<td>72</td>
</tr>
<tr>
<td>Bubble Sort</td>
<td>1331</td>
<td>128</td>
<td>46</td>
</tr>
<tr>
<td>Bitonic Sort</td>
<td>800</td>
<td>128</td>
<td>85</td>
</tr>
<tr>
<td>Odd-Even mergesort</td>
<td>654</td>
<td>128</td>
<td>101</td>
</tr>
<tr>
<td>Shellsort</td>
<td>885</td>
<td>128</td>
<td>61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sort Network</th>
<th>LUT</th>
<th>FLOP</th>
<th>Frequency (MHz)</th>
<th>Pipeline Stages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odd-even transposition sort</td>
<td>632</td>
<td>512</td>
<td>353</td>
<td>7</td>
</tr>
<tr>
<td>Bubble Sort</td>
<td>650</td>
<td>512</td>
<td>363</td>
<td>12</td>
</tr>
<tr>
<td>Bitonic Sort</td>
<td>540</td>
<td>448</td>
<td>353</td>
<td>5</td>
</tr>
<tr>
<td>Odd-Even mergesort</td>
<td>425</td>
<td>368</td>
<td>368</td>
<td>5</td>
</tr>
<tr>
<td>Shellsort</td>
<td>537</td>
<td>448</td>
<td>360</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2 & Table 3 summarize performance of the non pipeline and pipeline stages of designs.

It is observed that pipeline stages based algorithms have better operating frequency than the non pipelined stage algorithms. On the contrary, pipeline stages consume more FPGA resources than the non pipeline stages. The Odd-even merge sort consumes fewer resources and gives optimal operating frequency performance compared to the bitonic sort which not only consumes more resources but also has a poor operating frequency performance. Both modules have a throughput of sorting eight data per clock with a latency of five.

4 Conclusion

It is often critical do decide the best sorting algorithm for a given application. This is based on the tradeoff between pipeline stages, area and speed. It is observed that by adding five pipeline registers for odd even merge sort, throughput can be increased significantly without much increase in area (FPGA resource). To achieve similar performance, other sorting algorithms require more number of pipeline stages.

References:
Prediction of Wear Mechanisms in High Speed Steel Hobs Using Artificial Neural network

M.JALALI AZIZPOUR, H.MOHAMMADI MAJD

PTRI OF ACECR

IRAN

m_jpour@yahoo.com

Abstract— In this paper back-propagation artificial neural network (BPANN) is employed to predict the wear of the gear hobbing tools. The wear of high speed steel hobs during hobbing has been studied. The wear mechanisms are strongly influenced by the choice of cutting speed. At moderate and high cutting speeds three major wear mechanisms were identified: abrasion, mild adhesive and severe adhesive. The microstructure and wear behavior of two high speed steel grades (M2 and ASP30) has been compared. In contrast, a variation in chemical composition or microstructure of HSS tool material generally did not change the dominant wear mechanism. However, the tool material properties determine the resistance against the operating wear mechanism and consequently the tool life. The metallographic analysis and wear measurement at the tip of hob teeth included scanning electron microscopy and stereoscope microscopy. Comparing experimental and BPANN results, an acceptable correlation was found.

Keywords— Back-propagation artificial neural network abrasion, adhesion, cutting speed, hobbing, wear mechanism

1. Introduction

Gear hobbing is a widely used method in mass gear production. A hob has a large number of cutting edges arranged spirally around the tool body. Gear hobbing remains a cutting technology where high speed steel continues to find wide application in modern manufacturing practices [1], [2]. Gear hobbing has complicated process kinematics; chip formation and Tool wear mechanisms. Many researchers investigated the wear behavior and tool life of turning and milling cutting tools [3]-[6]. To understand the wear mechanisms in Gear hobbing, it is necessary to have a brief understanding of the Hobbing tribosystem includes hob, Gear, Cutting operation and sever contact in the tool-chip interface. Among the various hobbing parameters cutting speed has the most effective role on wear behavior [7].

In this study the effect of cutting parameters on the wear mechanisms of HSS hobs has been investigated at industrial conditions. The type of high speed steel influences the speed that will be used and the wear of hob. Two grades of HSS (AISI M2 and ASP30) are selected for this purpose. Moreover back-propagation artificial neural network (BPANN) is employed to predict the wear of the gear hobbing tools.

2. Experimental procedure

The main shaft of a tractor (Fig.1) is selected as gear blank. Table 1 shows the detail of hob, gear and hobbing condition. The vertical hobbing machining Rh6/1623 is used. Cutting speed that has direct relationship with temperature can change the predominant wear mechanisms and wear behavior of hob as well as gear. Scanning electron microscopy (SEM) was used to study the worn surface and microstructure. Stereoscope (Nikon-type 104) was used to measure the flank and crater wear on the flank and rake face. Surface profilometery is used to measure the roughness of gear that produced by high speed hobbing process.
3. Result of experiments

Among the variable hobbing parameters, only the cutting speed can affect the wear behavior strangely [7]. Cutting speed has direct relationship with Tool-chip interface temperature [7], [8]. At lower hobbing speed where the temperature is not high enough, a stable built-up edge (BUE) protects the cutting edge against the abrasive and adhesive wear. However, the formation and rebound mechanism of BUE Causes a sudden failure in cutting edges in chipping form. The relative motion between Tool and blank at such condition is stick–slip. At higher cutting speeds, this relative motion changes to slip so that the BUE will be unstable to play a wear particle (debris) role, causes three body abrasion wear. Fig.2 shows the SEM topography of BUE in the cutting edge. The segregated carbides and hard partied from tool and blank have the same effects.

As shown in Fig.3 increasing the hobbing speed leads to high abrasion wear. The peak of flank wear is at 65 m/min. At V= 65m/min the flank wear is 0.3 mm. in this condition each of the adhesive and abrasive wear is present but the abrasion is the predominant wear mechanism. Fig. 4 shows the SEM topography of the adhesive wear that exists at each condition. This adhesive component often is referred to as mild adhesive wear.

Two body abrasion wear is counteracted with high yield strength (high hardness) as well as large carbide volume of the HSS for three body abrasion [7],[10]. At such condition also, the segregated BUE have a scratch action. The presence of BUE as a thermal barrier layer protects the rake face against the high temperature. At cutting speeds higher than 65 m/min wear particles (debris) begin to soften, and therefore lose their abrasive role at flank wear. Declining in the flank wear curve is because of phenomenon described above. Softening and rebounding of the thermal barrier layer leads to heat transfer from cutting zoon to rake face that softens the hob tooth and forms a crater in rake face. This adhesive component often is referred to as mild adhesive wear. If the hob is used upper than its limit of heat resistance, sever adhesive wear may result a large scale plastic flow of surface material in direction of the chip flow. Severe adhesive wear are primarily resisted by HSS material through its high yield strength at elevated temperatures (High hot hardness). At V=90 m/min the severe adhesion wear is predominant.

Adhesion wear mechanism is identified by deep craters. The depth of crater increases always proportionally with cutting speed. It will be useful to find an optimum condition in which, the flank wear is decreased but the thermal barrier role of BUE remains exist. We can select V=50 m/min or higher cutting speed such as V=95 m/min for economic aspects.

Macroscopic fracture of the tool can occur is a rather scare event. More common it localized chipping of tool edge. Smooth tool surface, high fracture Toughness promoted by fine grained structure of both matrix and hard phased is counteractive tool properties against this wear mechanism. This type of wear can occurs at low or high cutting speeds.

---

**Table 1: Hobbing tribosystem**

<table>
<thead>
<tr>
<th>Gear</th>
<th>Hob</th>
<th>conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material :21</td>
<td>ASP30&amp; M2</td>
<td></td>
</tr>
<tr>
<td>NiCrMo2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hardness :170 BHN</td>
<td>1 st</td>
<td></td>
</tr>
<tr>
<td>Spure Gear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>d_k=50.8 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L_cut=256 mm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of teeth: 18</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cutting fluid</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telos</td>
<td>Continuous</td>
</tr>
<tr>
<td></td>
<td>1 st</td>
<td>shifting</td>
</tr>
<tr>
<td></td>
<td>GL=∞</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D_f= 100 mm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L= 100 mm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hardness: 65</td>
<td>RC</td>
</tr>
<tr>
<td></td>
<td>speed : variable</td>
<td>Climb</td>
</tr>
<tr>
<td></td>
<td>Axial feed rate :</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cut depth :</td>
<td>5.5mm</td>
</tr>
</tbody>
</table>

---

**Fig 1:** Main shaft of a tractor as gear blank

**Fig 2:** SEM topography of built-up edge
Fig. 3: Flank and crater wear vs. cutting speed

Fig. 4: SEM topography of adhesion wear mechanisms

Fig. 5: SEM micrograph of ASP30 hob microstructure

Fig. 6: SEM micrograph of M2 hob microstructure

Fig. 7: SEM topography of severe abrasion wear mechanisms of M2 hob

4. Neural networks

An artificial neural network is a parallel distributed information processing system. It stores the samples with distributed coding, thus forming a trainable nonlinear system. The main idea behind a neural network approach resembles the human brain functioning. Given the input and the expected outputs, the program is self adaptive to the environment so as to respond to different inputs rationally. The objective of this paper is to investigate the prediction of wear in gear hobbing, by training a BPANN.
The neuron can be classified into three types: input, output, hidden neurons. Input neurons are the ones that receive input from the environment, such as cutting speed in this study (Table.2). Output neurons are those that send the signals out of the system, like flank wear, crater wear and wear mechanisms. As the activation function, Sig activation function has been used, which is continuous, nonlinear, monotonic non-decreasing and S shaped: [12]

\[ f(x) = \frac{1}{1 + e^{-\beta x}} \]

In this study, the back propagation, which is a widely used algorithm, is used in the training step. Back propagation is a systematic method for training multilayer artificial neural networks. It has a strong mathematical foundation based on gradient descent learning. Elman BP network train with the back propagation algorithm is used. Elman networks are back propagation networks, with the addition of a feedback connection from the output of the hidden layer to its input. This feedback path allows Elman networks to learn to recognize and generate temporal patterns, as well as spatial patterns [13]. For an Elman to have the best chance at learning a problem it needs more hidden neurons in its hidden layer than are actually required for a solution by another method.

This model has four layers including, an input layer, two hidden layer and an output layer. In this work, different number of hidden units has been employed to obtain the optimum number of hidden units. The experiments show that number of 20 units in the hidden layer is enough to reach the desired accuracy.

Training of the neural network was done in MATLAB, using Sig and TRAINLM function. TRAINLM is a network training function that updates weights and bias values in a back propagation algorithm according to Levenberg–Marquardt optimization. Levenberg–Marquardt algorithm is a highly efficient method for solving non-linear optimization problems [14], [15].

5. Conclusions

In this work, a two-layer back propagation network is developed to best fit this nonlinear engineering problem. Through comparison between the targeted value and training results with different neuron numbers in the hidden layers, an appropriate number of 20 is suitable to set up this network. For this nonlinear engineering problem, the appropriate algorithm is Levenberg -Marquardt because it can reach high accuracy. The error between the predicted value and targeted one is little. Using this network can save much time. To obtain the most proper network a net with different structures has been trained. For this purpose first a 1-20-20-3 network with 100 epochs was selected to modify the weights. The average of mean square error (MSE) was recorded. Other networks have been investigated. According to comparison among the structures it was be seen that the best network is 1-20-20-3. After 160 epochs the mean of errors converged to 2.3e-6. (Table.3). 70 percent of experimental data was used for training the network, 15 percent for test and 15 percent to verify the trained network.

Wear mechanisms of HSS hobs has been investigated. Tribological and metallurgical analyses are employed for this purpose. A summery of conclusions is as follow:

- In HSS hobs at moderate cutting speed, the predominant mechanism is abrasion wear at the flank of cutting edge. The maximum of flank wear is at \( V=65 \text{m/min} \).
- Adhesive wear can exist at each hobbing condition but at high speed hobbing the predominant wear mechanism is severe adhesive wear.
- The presence or absence of built-up edge is important to change the predominant wear mechanism.
- The hob made by ASP30 has high hardness, excellent hot hardness and wear resistance compared with M2.
- The grade of tool material can not change the wear mechanism, but wear resistance of hob can be improved by correct selecting of tool material considering cutting condition to be used and machining economy.
• The error between the predicted value and targeted one is near to zero and the network can predict the wear with high accuracy.

Table 2: Experimental results

<table>
<thead>
<tr>
<th>m/min</th>
<th>V B</th>
<th>K T</th>
<th>Predominant Wear mechanisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.05</td>
<td>0.005</td>
<td>Chipping</td>
</tr>
<tr>
<td>15</td>
<td>0.05</td>
<td>0.005</td>
<td>Chipping</td>
</tr>
<tr>
<td>20</td>
<td>0.08</td>
<td>0.005</td>
<td>Chipping</td>
</tr>
<tr>
<td>25</td>
<td>0.09</td>
<td>0.005</td>
<td>Adhesion (BUE)</td>
</tr>
<tr>
<td>30</td>
<td>0.1</td>
<td>0.005</td>
<td>Abrasion</td>
</tr>
<tr>
<td>35</td>
<td>0.12</td>
<td>0.007</td>
<td>Abrasion</td>
</tr>
<tr>
<td>40</td>
<td>0.15</td>
<td>0.01</td>
<td>Abrasion</td>
</tr>
<tr>
<td>45</td>
<td>0.17</td>
<td>0.01</td>
<td>Abrasion</td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>0.015</td>
<td>Abrasion</td>
</tr>
<tr>
<td>55</td>
<td>0.22</td>
<td>0.02</td>
<td>Abrasion</td>
</tr>
<tr>
<td>60</td>
<td>0.25</td>
<td>0.025</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>62</td>
<td>0.25</td>
<td>0.03</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>64</td>
<td>0.28</td>
<td>0.03</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>65</td>
<td>0.3</td>
<td>0.03</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>67</td>
<td>0.27</td>
<td>0.03</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>70</td>
<td>0.25</td>
<td>0.035</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>75</td>
<td>0.23</td>
<td>0.04</td>
<td>Severe abrasion</td>
</tr>
<tr>
<td>80</td>
<td>0.21</td>
<td>0.04</td>
<td>Adhesion</td>
</tr>
<tr>
<td>85</td>
<td>0.18</td>
<td>0.05</td>
<td>severe adhesion</td>
</tr>
<tr>
<td>88</td>
<td>0.13</td>
<td>0.05</td>
<td>severe adhesion</td>
</tr>
<tr>
<td>90</td>
<td>0.13</td>
<td>0.05</td>
<td>severe adhesion</td>
</tr>
<tr>
<td>95</td>
<td>0.1</td>
<td>0.06</td>
<td>severe adhesion</td>
</tr>
<tr>
<td>98</td>
<td>0.1</td>
<td>0.06</td>
<td>severe adhesion</td>
</tr>
<tr>
<td>100</td>
<td>0.07</td>
<td>0.07</td>
<td>severe adhesion</td>
</tr>
</tbody>
</table>

Table 3: Test conditions

<table>
<thead>
<tr>
<th>cut speed (input)</th>
<th>Flank wear V B</th>
<th>Crater wear K T</th>
<th>Network output output K T</th>
<th>Network output output V B</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.05</td>
<td>0.005</td>
<td>-0.00378</td>
<td>0.0365</td>
</tr>
<tr>
<td>20</td>
<td>0.08</td>
<td>0.005</td>
<td>0.004054</td>
<td>0.0798</td>
</tr>
<tr>
<td>30</td>
<td>0.10</td>
<td>0.005</td>
<td>0.006915</td>
<td>0.99</td>
</tr>
<tr>
<td>40</td>
<td>0.15</td>
<td>0.010</td>
<td>0.008219</td>
<td>0.1305</td>
</tr>
<tr>
<td>50</td>
<td>0.20</td>
<td>0.015</td>
<td>0.018519</td>
<td>0.18</td>
</tr>
<tr>
<td>60</td>
<td>0.25</td>
<td>0.025</td>
<td>0.023029</td>
<td>0.24</td>
</tr>
<tr>
<td>70</td>
<td>0.25</td>
<td>0.035</td>
<td>0.034289</td>
<td>0.29</td>
</tr>
<tr>
<td>80</td>
<td>0.21</td>
<td>0.040</td>
<td>0.043573</td>
<td>0.3</td>
</tr>
<tr>
<td>90</td>
<td>0.130</td>
<td>0.050</td>
<td>0.049961</td>
<td>0.15</td>
</tr>
<tr>
<td>100</td>
<td>0.070</td>
<td>0.070</td>
<td>0.0699</td>
<td>0.09</td>
</tr>
</tbody>
</table>

References:
Charging for Reactive Power Employing Generation Distribution

FRANCISCO DAVID MOYA CHAVES
Engineer - Electrical School
La Salle University
Cra 2 N 10 – 70
Bogotá, Colombia
frmoya@unisalle.edu.co

Abstract: This paper presents an analysis to calculate relative sensitivity by using five different methodologies based on calculations of relative sensitivity that achieved a proper relationship between loads and generation nodes. The calculation of sensitivity using five methodologies on the IEEE-30 is estimated.

Key-Words: Reactive Power, Relative Sensitivity, Generation Distribution.

1 Introduction

Several reasons have driven the interest in Distributed Generation (DG). A new energy source will emerge, hydrothermal, photovoltaic, biomass, etc. Associated with the consumer, it will certainly have an important role, since it is the only way to ensure the implementation of additional capacity at short notice and at competitive costs. Moreover, this future generation of electricity should also suit the needs of energy markets, while respecting the unique features of your electrical system, introducing efficiency, reliability and flexibility, and looking at the same time to respond to the challenges of ever increasing efficiency of utilization of energy resources and minimizing environmental impact during the process.

New developments in technology have placed these DG plants as a concrete alternative for the supply of electricity and heat, installing generators at the point of final consumption or near. Since DG is located, by definition, close to the electric charges, their generating units, besides supplying the energy locally, have proper conditions to play an important role in all of the Interconnected System, even when it stops (emergency generators, for example), they increase power reserves among these loads, and therefore reduce the risk of instability and increase the reliability of the supply process. To the extent that only some segments of the economy have the capacity to produce electricity in a competitive way, a significant portion of consumers will continue depending on the Central Generation (CG) of power, so it can be said that DG is not competing, but does complement and enhance the system based on CG, both existing in the future.

As the costs of various technologies have DG decreasing trends (McDonald, 2004), the system moves towards a situation in which, facing a high cost of transmission, the DG may increase its stake in the Colombian generation park. In this context, it is urgent to understand this process and facilitate and encourage wider use of DG in the Colombian context. The CG has dominated the production of power, due to economies of scale, the savings that offset the construction of complex systems of interconnection, where DG was used to increase both the efficiency of certain processes regarding their safety and / or to meet the individual systems [1].

Currently, electric systems are changing and with the DG growing up, the distribution systems are turning from passive to active entities [2]. The new and future DG schemes are allowing the establishment of a wide range of new energy suppliers and complementary services such as auxiliary services in which the reactive power supply takes part. DG is a new participant in the current and future electric systems in terms of programming and operating reserves dispatch [3].

Dispatch of centralized units requires a higher generation level, which should supply energy to all the network points, including all the load power and the network losses in the whole system, the transmission system and the final points related to the distribution system [4].

On the other hand, delivering active power by the generators in the distribution system, close to the load, reduces losses and, consequently, improves voltage levels and increases the security of the system.

This work shows some technical advantages when DG is used in the dispatch of reactive power reserves. In the following section, the formulation of the
problem is described. Then, the methodology implemented is explained. Also, a methodology is proposed to establish the maximum DG level that can be installed.

It is also showed an application aiming at analyzing what happens when DG is used to supply reactive power. Additionally, DG technologies that attend more efficiently the management of reserves are described. Finally, the conclusions of this work are presented.

2 DG Technologies to Provide Reactive Power

There are different kinds of technologies that can be used for DG applications (wind power generation, photovoltaic power, cogeneration, etc). Every technology presents a different production profile, therefore, the load flows and losses impact depending on each particular technology. Generally, the DG profile varies constantly. This variation can be produced by the primary resource (wind, sun, water, etc) or due to process features (e.g. cogeneration process).

As the DG production profile is adapted to the demand profile of the buses where the DG is installed, the electric parameters will be better or worse. For instance, if the DG production in each bus is exactly equal to the demand in each bus, the losses will be cancelled since all the demand is supplied locally and there are not flows within the network. Therefore, the more appropriate GD technologies for supplying reactive power have constant production profiles.

In this category the generators which supply almost constant power at all time can be classified (higher utilization factor, generally classified from 0 up to 1, where 1 is the maximum value). In some cases the DG production can be practically constant for instance combustible cells, biomass plants and gas microturbines. These technologies are well conditioned to the demand profile. In [5] it was shown that the wind and photovoltaic power generation are ill conditioned to suit the demand profile. Table 1 shows the behavior of those technologies related to the demand and the reactive power supply.

3 Sensitivity Calculations in Electric Systems

The calculation of relative sensitivity is a tool to identify devices that have an impact on some sets of load nodes in electric systems. These devices allow the achievement of a better contribution in the delivery or absorption of reactive power above certain node. The ability of a generator to participate in the voltage regulation above a load node depends on two concepts: the influence of reactive power generated on the level of voltage and reactive power available from the generator [6].

<table>
<thead>
<tr>
<th>Supply Technologies</th>
<th>Demand following</th>
<th>Reactive Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Gas Turbine</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Micro-turbine</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Steam Turbine</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Combined Cycle</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Electrochemical Devices</td>
<td>Very good</td>
<td>Very good</td>
</tr>
<tr>
<td>Small hydro</td>
<td>Very Bad</td>
<td>Normal</td>
</tr>
<tr>
<td>Wind turbines</td>
<td>Very Bad</td>
<td>Very Bad</td>
</tr>
<tr>
<td>Mini-Wind</td>
<td>Very Bad</td>
<td>Very Bad</td>
</tr>
<tr>
<td>PhotoVoltaic</td>
<td>Very Bad</td>
<td>Very Bad</td>
</tr>
<tr>
<td>PhotoThermal</td>
<td>Good</td>
<td>Normal</td>
</tr>
<tr>
<td>Fuel Cell</td>
<td>Good</td>
<td>Bad</td>
</tr>
</tbody>
</table>

3.1 Methodology 1

The incremental ability of a generator to influence a voltage in a specific operating capacity is defined by the Jacobian matrix. The marginal capacity of a generator is the difference between maximum generating capacity of reactive power and reactive power current output, which is:

\[ Q^\text{marginal} = Q^\text{maxima} - Q^\text{actual} \] (1)

It takes into account the available margin of each generator and not the total range (Qmaxima – Qminima). If the calculation of sensitivity is defined based on the delivery of reactive power available from each generator, then changes in voltage due to the injection provided by a given generator m is:

\[ \begin{bmatrix} \Delta \theta \\ \Delta V \end{bmatrix} = \left( J^{-1} e_i^Q \right) Q^\text{marginal} \] (2)

Where \( Q^\text{marginal}_i \) is the marginal reactive power generation of m, and it is a null vector, except at the position corresponding to local injection at the node i.
After evaluating $\Delta V_i$ for all participants $i$ with relation to the generator $m$, the sensitivity $S_i$ is calculated using the formula:

$$S_i = 100\frac{\Delta V_i}{\sum_{i=1}^{n} \Delta V_i}$$ (3)

### 3.2 Methodology 2

The work in [5] describes another study to calculate the sensitivity coefficient which relates the change in voltage in a bar $i$ respect to change of generation (or absorption) of reactive power at node $j$. The simplifications of a decoupled power flow follow:

$$\begin{bmatrix} \Delta Q \\
\Delta V \end{bmatrix} = \begin{bmatrix} -B_{i1} - B_{i2} - \cdots & \Delta V_j \\
-B_{2i} - B_{22} - \cdots & - \end{bmatrix} \begin{bmatrix} \Delta V_i \\
\Delta V_j \end{bmatrix}$$ (4)

Where $B$ is the susceptance matrix. By inverting matrix $B$ it is possible to find the vector of voltage variations $[\Delta V]$ as follows:

$$[\Delta V] = [X][\Delta Q]$$ (5)

Therefore a coefficient which relates voltage changes with variations in the reactive power generation is given by:

$$\lambda = \frac{\Delta V}{\Delta Q_j}$$ (6)

Where $V_i$ is voltage on bar $i$ and $Q_i$ is the reactive power at node $j$. This ratio means the way it affects the reactive power generated in node $j$, the voltage at the load node $i$.

### 3.3 Methodology 3

To determine which of the generators in the system creates a better supply in the supply of reagents needed for the load, a calculation of sensitivities was carried out, causing a variation in each of the individual generators, and by seeing the effect of this variation in each generator, we charge.

By injecting of DG in power distribution network, the impact depends on the behavior of the energy it injects or consumes. There may be three situations:

a) Distributed Generator provides active and reactive energy,

b) Distributed Generator provides active power and reactive power consumed,

c) Distributed Generator provides active energy and does not supply or consume reactive power.

Thus, the calculation of sensitivity depends on the variations of active and reactive energy and the coefficient sensitivity calculated as follows:

$$\lambda_j = \frac{\Delta V_i}{\Delta P_j} + \frac{\Delta V_i}{\Delta Q_j}$$ (7)

To determine the influence of a particular generator on each of the network nodes in terms of percentage, use the following equation:

$$\lambda_j = \frac{\Delta V_i}{\Delta P_j} + \frac{\Delta V_i}{\Delta Q_j}$$ (8)

$$VSI = \frac{\partial V_i}{\partial P_j} + (1 - w) \frac{\partial V_i}{\partial Q_j}$$ (9)

### 3.4 Methodology 4: Voltage Sensitivity

To assess network capability to absorb available distributed resources safely, a steady-state system representation in the form of power flow equations was used in [7]. The Jacobian inverse power flow relates changes in power injections to changes in angles and voltages by using the following equation:

$$\begin{bmatrix} \Delta \delta \\
\Delta V \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 V}{\partial S \partial V} & \frac{\partial V}{\partial V} \\
\frac{\partial V}{\partial S} & \frac{\partial^2 V}{\partial V \partial V} \end{bmatrix}^{-1} \begin{bmatrix} \Delta P \\
\Delta Q \end{bmatrix}$$ (10)

In order to determine the most suitable sites for DGs, two sensitivity-based approaches related to voltage control and power loss are proposed. Both a voltage sensitivity index (VSI) and a loss sensitivity index (LSI) are defined and used to identify and rank the
nodes within the network with respect to receiving new generation. It is assumed that generators can connect to any point in the network subjected to security constraints and are not restricted in their location by generator controllers or existing protection devices.

Assuming that angle-related problems are not a concern, the voltage sensitivity can be defined as:

\[
[\Delta V_j] = \left[ \frac{\partial V_j}{\partial P_i} \right] [\Delta Q_i] + \left[ \frac{\partial V_j}{\partial Q_i} \right] + [\Delta P]
\]

From (I), for each stem node, there is an associated real power sensitivity \( \frac{\Delta V_j}{\Delta P_i} \) and reactive power sensitivity \( \frac{\Delta V_j}{\Delta Q_i} \).

These values can be used to rank the overall voltage sensitivity of each node to real or reactive power injection. A Voltage Sensitivity Index (VSI) used in ranking is defined in [7].

The diagonal elements of the Jacobian matrix represent the sensitivity of one bus voltage magnitude to the injection of power at the same bus, whereas the off diagonal elements represent the sensitivity to power injected at other buses. Since the purpose of adding dispersed generation is to bring about an improvement in network performance, the effect of power injection at a single bus on the voltage sensitivities of the whole network must be considered. This is achieved by expressing the VSI for each node as a Euclidean norm normalized across all load buses. The value of the weighting factor \( w \) will depend on the X/R ratio of the network under consideration. The nodes are ranked according to the VSI value and the ranked set is used to define the optimum sites to accept injection of P and/or Q.

3.5 Methodology 5: Loss Sensitivity Index

The majority of power losses are ohmic in nature caused by power flow through lines and transformers, i.e.,

\[
P \text{ loss} = P(\delta, V) \quad (12)
\]

\[
Q \text{ loss} = Q(\delta, V) \quad (13)
\]

Combining equations (1) and (4) results in:

\[
\begin{align*}
\frac{\partial P_{\text{loss}}}{\partial P} & = J^T_{\delta} \frac{\partial P_{\text{loss}}}{\partial \delta} \\
\frac{\partial P_{\text{loss}}}{\partial Q} & = J^T_{\delta} \frac{\partial P_{\text{loss}}}{\partial V}
\end{align*}
\]

The Loss Sensitivity Index (LSI) is defined as:

\[
LSI = w \left[ \frac{\partial P_{\text{loss}}}{\partial P} \right] + (1 - w) \left[ \frac{\partial P_{\text{loss}}}{\partial Q} \right] \quad (15)
\]

4 Charges per user in compensation costs

If the power factor measured at the premises of the consumer is less than 0.92 the cost of reactive consumption will be charged, due to the difference between the minimum allowed value and the value calculated in the cycle. The premium is calculated by the following formula:

\[
FER = CA \times \left[ \frac{F \text{ Pr}}{F \text{ Pr m}} - 1 \right] \times TCA \quad (16)
\]

FER = value of total revenues corresponding to the consumption of reactive power surplus to the amount allowed by the power factor of reference in the billing period,

AC = active power consumption measured during the billing period,

FPR = reference power factor equal to 0.92,

FPm = power factor of electrical inductive medium of consumer unit, calculated for the billing period,

ACT = active energy tariff applicable to the supply.

Equation 16 shows that \( \left[ \frac{F \text{ Pr}}{F \text{ Pr m}} - 1 \right] \) is a relationship between the cost of active energy and the cost of reactive power. For the purpose of simple calculation of the cost of reactive power in the simulations, we used the relationship mentioned, with an FPm = 0.8, thus, it has become a factor related to the active energy of 0.15. Thus, according to the cost for hydropower in [8], it has a production cost of active energy of 64.78 ($ MW / hr), and using the relationship mentioned above, there is a cost of reactive power of 9.72 ($ MVAr / hr).
By simulating these two scenarios, the sensitivity coefficients can be calculated. Table 2 shows a declining scale, the results for the most influenced nodes. The calculation of sensitivity allows the calculation to charge consumers, so each of us can be benefited from the generation of energy:

\[
\text{Payment of user } i = \lambda_i \times \text{cost energy } Q
\]  

(17)

### TABLE 2
Cost of active and reactive power with GD

<table>
<thead>
<tr>
<th>Nº nó</th>
<th>V1</th>
<th>V2</th>
<th>ΔV</th>
<th>λ</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.994</td>
<td>1.199</td>
<td>0.205</td>
<td>0.026</td>
</tr>
<tr>
<td>99</td>
<td>0.994</td>
<td>1.190</td>
<td>0.196</td>
<td>0.024</td>
</tr>
<tr>
<td>98</td>
<td>0.994</td>
<td>1.189</td>
<td>0.195</td>
<td>0.024</td>
</tr>
<tr>
<td>97</td>
<td>0.994</td>
<td>1.186</td>
<td>0.191</td>
<td>0.024</td>
</tr>
<tr>
<td>96</td>
<td>0.994</td>
<td>1.160</td>
<td>0.166</td>
<td>0.021</td>
</tr>
<tr>
<td>95</td>
<td>0.995</td>
<td>1.102</td>
<td>0.107</td>
<td>0.013</td>
</tr>
<tr>
<td>94</td>
<td>0.997</td>
<td>1.069</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>93</td>
<td>0.980</td>
<td>1.069</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>92</td>
<td>0.982</td>
<td>1.071</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>91</td>
<td>0.983</td>
<td>1.072</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>90</td>
<td>0.984</td>
<td>1.073</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>89</td>
<td>0.984</td>
<td>1.073</td>
<td>0.089</td>
<td>0.011</td>
</tr>
<tr>
<td>88</td>
<td>0.960</td>
<td>1.051</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>87</td>
<td>0.960</td>
<td>1.051</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>86</td>
<td>0.960</td>
<td>1.051</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>85</td>
<td>0.961</td>
<td>1.052</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>84</td>
<td>0.966</td>
<td>1.056</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>83</td>
<td>0.966</td>
<td>1.056</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>82</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>81</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>80</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>79</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>78</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>77</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
<tr>
<td>76</td>
<td>0.966</td>
<td>1.057</td>
<td>0.091</td>
<td>0.011</td>
</tr>
</tbody>
</table>

### 5 Case Study

In this work, a simulation for different scenarios is implemented aiming at comparing technical centralized generation aspects and non-centralized generation when a DG is involved.

These scenarios were computed by using a load flow (static model) for an electric system which takes into account voltage levels for transmission, sub transmission and distribution in order to assess the DG impact on every voltage level. The IEEE30 bus system was used to carry out the study. Two buses on the medium voltage level were separated for generating the distribution network (low voltage level). A distribution system of 70 buses, presented in Figure 1, was connected to buses 30 and 24.

For computing, the optimal flow, the Newton-Raphson and Fast-Decoupled Newton methods were used. In order to develop the simulations, the program for power system analysis MATPOWER was used. Table 3 and Table 4 shows cost of active and reactive power without and with GD respectively.

### TABLE 3
Cost of active and reactive power without GD

<table>
<thead>
<tr>
<th>Nº nó</th>
<th>V1</th>
<th>V2</th>
<th>ΔV</th>
<th>λ</th>
<th>P (MW)</th>
<th>Cost of GenerationP ($/hr)</th>
<th>Q (MVAr)</th>
<th>Cost of GenerationQ ($/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.994</td>
<td>1.199</td>
<td>0.205</td>
<td>0.026</td>
<td></td>
<td>8.1</td>
<td>64,78</td>
<td>0.92</td>
</tr>
<tr>
<td>99</td>
<td>0.994</td>
<td>1.190</td>
<td>0.196</td>
<td>0.024</td>
<td></td>
<td>43</td>
<td>64,78</td>
<td>15.33</td>
</tr>
<tr>
<td>98</td>
<td>0.994</td>
<td>1.189</td>
<td>0.195</td>
<td>0.024</td>
<td></td>
<td>100</td>
<td>64,78</td>
<td>23.99</td>
</tr>
<tr>
<td>97</td>
<td>0.994</td>
<td>1.186</td>
<td>0.191</td>
<td>0.024</td>
<td></td>
<td>73.29</td>
<td>64,78</td>
<td>40</td>
</tr>
<tr>
<td>96</td>
<td>0.994</td>
<td>1.160</td>
<td>0.166</td>
<td>0.024</td>
<td></td>
<td>29.67</td>
<td>64,78</td>
<td>1.07</td>
</tr>
<tr>
<td>95</td>
<td>0.995</td>
<td>1.102</td>
<td>0.107</td>
<td>0.013</td>
<td></td>
<td>30.92</td>
<td>64,78</td>
<td>-6</td>
</tr>
</tbody>
</table>

### TABLE 4
DG technologies behaviors for reactive power supplying

<table>
<thead>
<tr>
<th>Nº nó</th>
<th>V1</th>
<th>V2</th>
<th>ΔV</th>
<th>λ</th>
<th>P (MW)</th>
<th>Cost of GenerationP ($/hr)</th>
<th>Q (MVAr)</th>
<th>Cost of GenerationQ ($/hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>7.98</td>
<td>64,78</td>
<td>0.63</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G2</td>
<td>42.52</td>
<td>64,78</td>
<td>14.47</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G3</td>
<td>100</td>
<td>64,78</td>
<td>23.52</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G4</td>
<td>71.59</td>
<td>64,78</td>
<td>40</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G5</td>
<td>29.52</td>
<td>64,78</td>
<td>1.34</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6</td>
<td>31.27</td>
<td>64,78</td>
<td>-5.25</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G7</td>
<td>2</td>
<td>64,78</td>
<td>0.12</td>
<td>9,72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### 4 Conclusions

This paper has presented an analysis to calculate relative sensitivity by using five different methodologies based on calculations of relative sensitivity to achieve a proper relationship between loads and generation nodes, as well as the delivery of
reactive power between areas, allowing a procedure for awarding bids to supply reactive power to be implemented.

This paper illustrates how the information on the sensitivity provides valuable information for the installation of strategic small generators. Reactive power shows economic and operational advantages.

The DG impact on the system depends generally on the production profile and how it is adaptable to the demand profile. The DG technologies that can accomplish a better role in the operating reserves delivery are those with a constant production profile and also those that can begin to increase output delivery immediately in response to changes in interconnection frequency to correct generation/load imbalances caused by generation or transmission outages.

It is possible to analyze that the losses active and reactive increase in times of lower demand with the use of GD, otherwise in times of higher demand (peak hour) where the losses decrease. The use of GD for reduction of losses is of the utmost importance in reducing reactive power losses, and it is motive enough to remunerate the DG for the compensation and control of reactive in the electric systems.

References:


Biography

Francisco D. Moya Ch received the P.E. in Electrical Engineering from the National University of Colombia and the M.S. in electrical Engineering from University Los Andes, Colombia, in 2000 and 2005, respectively. He received the Ph.D. degree in electrical engineering at UNICAMP (University of Campinas), Brazil. He is an Associate Professor at the Department of Electrical Engineering, Engineering School from Universidad de La Salle. His main research interests are in power quality, distributed energy systems, energy efficiency and conservation, renewable energy and environmental policy.
Modeling and Prediction of Ship Squat Using ANFIS and GMDH-type Neural Network

K.SALMALIAN\textsuperscript{1}, M.SOLEIMANI\textsuperscript{2}

DEPARTMENT OF MECHANICAL ENGINEERING
\textsuperscript{1} Islamic Azad University, Langaroud Branch
\textsuperscript{2} Islamic Azad University, Lahijan Branch
P.O Box 44715-1333, Langaroud, Guilan
IRAN
\textsuperscript{1}kavehsalmalian@iaul.ac.ir, \textsuperscript{2}majidsoleimani@iau-lahijan.ac.ir

Abstract: - The reduction of the distance between ship floor and seabed, while the ship is moving forward, is called squat. In this research the squat is determined for vessels with Series-60 hull forms in various depths by experimental methods and then different numerical methods are employed for squat modeling. For this reason, a set of facilities for testing the ship movement in shallow waters is prepared. A series of models of the vessel is manufactured and many tests are carried out. The aim of the present study is to demonstrate the usefulness of Adaptive-network-based fuzzy inference system (ANFIS) for modeling and predicting squat parameter for ships in shallow waters. It is also shown how dimensionless squat ($S^*$) varies with the variation of important parameters namely, block coefficient ($CB$), dimensionless distance between the seabed and ship floor ($\delta$) and hydraulic Froude Number ($Fn_h$). The results obtained through ANFIS are also compared with those of a multiple linear regression and GMDH-type neural network with multi-layered feed forward back propagation algorithm. The results show that the ANFIS-based squat has higher predictability function than other numerical methods.

Key-Words: - ANFIS, GMDH, Squat, Shallow Water, Physical Model.

1 Introduction

To keep a safe passage of a ship in shallow water, it is important to study the relation between ship behaviors and the water depth in shallow waters. Grounding of ships in shallow waters may result in severe damage to the ship and, in extreme cases, may lead to the complete loss of port or channel capabilities. Squat is defined as the increase of draught of vessel due to its forward movement in shallow water [1-3]. One can calculate the squat in shallow water by a number of methods such as analytical method [4], numerical and experimental methods [5] and [6]. Due to the existence of complicated three-dimensional fluid flow around the ship in shallow water, experimental methods are the most viable option and the most accurate method.

Kreitner [7] was the first one who calculated the squat of a given vessel by fundamental equations of fluid mechanics. Havelock [8] obtained the squat of a boat with elliptic hull form through analytical equations. Constantine [9] obtained some equations for the squat by one-dimensional hydraulics theory. Rubin and Naghdi [10] calculated the value of the squat for a certain ship by two-dimensional hydraulics method and verified their results by experimental means. Barras [11] introduced experimental relations for real vessels that now have applications for determining the squat in shallow and narrow canals. Tuck, Taylor and Millward [12] obtained other experimental relations for calculating the squat that has limited application and are not suitable for all ships and velocities. Due to the effects of the squat in shallow waters, the resistance of ships increases. This can be used for testing the physical model and estimating the extra force needed for ships movement in shallow waters.

Commonly, there are many parameters influencing the squat such as block coefficient, the dimensionless distance between the seabed and the floor of ships, and hydraulic Froude Number. Due to the complexity of squat assessment and interrelationships among the influencing parameters, it is difficult to develop a parametric model solution especially for three-dimensional problem. So, a series of experiments must be carried out and then modeling should be done. Soft computing, such as ANFIS or ANN, could be a good approach for modeling and prediction of the squat as an important phenomenon in shallow water for ships. Artificial neural networks (ANNs) have become popular
because of their high computational rates, robustness and ability to learn, and they have been used in diverse applications in power systems, manufacturing, optimization, medicine, signal processing, control, robotics, and social/psychological sciences [13, 14]. Group Method of Data Handling (GMDH) algorithm is a self-organizing approach through which gradually complicated models are generated, based on the evaluation of their performances on a set of multi-input-single-output data pairs \((X_i, y_i)\) \((i=1, 2, ..., M)\). The GMDH was firstly developed by Ivakhnenko [15] as a multivariate analysis method for complex systems' modeling and identification. This way, GMDH was used to circumvent the difficulty of considering the a priori knowledge of the mathematical model of the process. In other words, GMDH can be used to model complex systems without having specific knowledge of them. The main idea of GMDH is to build an analytical function in a feed forward network based on a quadratic node transfer function [16] whose coefficients are obtained using regression techniques. In fact, real GMDH algorithm in which model coefficients are estimated by means of the least squares method has been classified into complete induction and incomplete induction, which represent the combinatorial (COMBI) and multilayered iterative algorithms (MIA), respectively [17]. In recent years, however, the use of such self-organizing network leads to successful application of the GMDH-type algorithm in a broad range area in engineering, science, and economics [15-21].

Fuzzy logic is a problem-solving technique that derives its power from its ability to draw conclusions and generate responses based on vague, ambiguous, incomplete and imprecise information. To simulate this process of human reasoning, it applies the mathematical theory of fuzzy sets, first defined by Lotfi zadeh, in 1965 [22]. ANFIS, developed in the early 90s by Jang [23], incorporates the concept of fuzzy logic into the neural networks to facilitate learning and adaptation.

In this paper, first, the results of experimental method for the squat of commercial vessels in shallow waters are applied. Then, the use of ANFIS with different type of membership functions for modeling and prediction of squat parameters in shallow waters for ships has been demonstrated. After that, a comparison is made between two GMDH-type neural networks and multiple-linear regression models with the best ANFIS model, with respect to Mean Square Error (MSE) of modeling and prediction, on two predefined datasets namely, Training set and Testing set.

2 Experimental Procedure

2.1 The principles of the model testing of the squat

For assessing the ship behavior by a model test, one should establish geometrical and kinematical similarities between the ship and its model. Consequently, a dynamic similarity will take place which is the result of the model testing. In most of the cases, the geometrical similarity is defined as scaling down the ship dimensions and water depth by a certain value.

\[
\lambda = \frac{L_s}{L_m} = \frac{h_s}{h_m}
\]

where \(\lambda\) is called scale, \(L_s\) denotes ship length, \(L_m\) is model length, \(h_s\) is water depth for ship and \(h_m\) is water depth for model.

In a case of squat model testing, the kinematical similarity is defined as the Froude Number of a model and the corresponding ship are to be the same and the model is to be large enough, the Reynolds Number of which falls in the turbulent region.

\[
Fn_m = Fn_s \Rightarrow \frac{V_m}{\sqrt{gh_m}} = \frac{V_s}{\sqrt{gh_s}}
\]

\[
\Rightarrow V_m = V_s \frac{h_m}{h_s} = \frac{1}{\sqrt{\lambda}} V_s
\]

\[
Re_m = \frac{V_m L_m}{\nu} > Re_{Critical}
\]

where subscript \(m\) refers to model and \(s\) refers to ship, \(Fn\) is Froude Number based on water depth, \(V\) is speed, \(g\) is gravitational acceleration and \(\nu\) denotes kinematic viscosity. The dynamic similarity is achieved if one builds a model and shallow water, which are geometrically similar to the ship and its water depth, equation (1), and model moves forward with the speed of \(V_m = \frac{1}{\sqrt{\lambda}} V_s\), then the dynamic similarity will be achieved as follows:

\[
S_s = \lambda S_m
\]

Practically, there may exist some errors which are called the scale effect. Having enough experience in
model testing with particular apparatuses, the scale effect can be minimized or deducted from the test results.

2.2 Model properties and laboratory preparation

Several tests are planned for the experimental analysis of the ship squat in shallow water. These tests are carried out at the marine laboratory of Sharif University of Technology. To perform the tests, two models are built, a shallow water tank is prepared, and measurement and data recording tools are provided.

Two models with series 60 hull form and block coefficient of CB=0.7 and CB=0.75 and with the scale of 1:40 and 1:70 are precisely manufactured. The main particulars of the models are presented in Table 1.

<table>
<thead>
<tr>
<th>No.</th>
<th>Length (m)</th>
<th>Beam (m)</th>
<th>CB</th>
<th>Model Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.50</td>
<td>0.31</td>
<td>0.70</td>
<td>commercial</td>
</tr>
<tr>
<td>2</td>
<td>2.38</td>
<td>0.323</td>
<td>0.75</td>
<td>commercial</td>
</tr>
</tbody>
</table>

The models are shown in Figures 1 and 2.

Tests are carried out for two models at four different depths at several different speeds. Tests are continued until the model floor hits the seabed. That is why the test for small δ is cut at low Froude Number where for large δ is continued up to Froude Number of about 0.6.

Moreover, due to interactions between the tank bottom and the ship floor, the increase in the speed of the model leads to an increase in the squat. It is also found that as, δ, the dimensionless gap between the seabed and model floor increases, consequently the squat decreases.

2.4 THE TEST RESULTS

In order to introduce a comprehensive equation for the squat, the authors introduce some new dimensionless parameters. These parameters are $S^*$, dimensionless squat, and $\delta$, dimensionless gap between the seabed and the model floor. The above-mentioned dimensionless parameters in conjunction with hydraulics Froude Number, $Fn_h$, are used for further analysis. They are as follow:

$$S^* = \frac{S}{h}, \delta = \frac{h-T}{h} \text{ and } Fn_h = \frac{V}{\sqrt{gh}} \quad (4)$$

Tests are carried out for two models at four different depths at several different speeds. Tests are continued until the model floor hits the seabed. That is why the test for small $\delta$ is cut at low Froude Number where for large $\delta$ is continued up to Froude Number of about 0.6.

Moreover, due to interactions between the tank bottom and the ship floor, the increase in the speed of the model leads to an increase in the squat. It is also found that as, $\delta$, the dimensionless gap between the seabed and model floor increases, consequently the squat decreases.

3 Modeling Using GMDH-type Neural

By means of GMDH algorithm, a model can be represented as a set of neurons in which different pairs of each layer are connected through a quadratic polynomial, and thus produce new neurons in the next layer. Such representation can be used in modeling to map inputs to outputs. The formal definition of the identification problem is to
find a function $\hat{f}$ so that it can be approximately used instead of the actual one, $f$ in order to predict output, $\hat{y}$, for a given input vector, $X = (x_1, x_2, x_3, ..., x_n)$, as close as possible to its actual output, $y$. Therefore, given $M$, to be the observation of multi-input-single-output data pairs so that

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \quad (i=1,2, ..., M), \quad (5)$$

it is now possible to train a GMDH-type neural network to predict the output values, $\hat{y}_i$, for any given input vector, $X = (x_{i1}, x_{i2}, x_{i3}, ..., x_{in})$, that is

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) \quad (i = 1, 2, ..., M). \quad (6)$$

The problem is now to determine a GMDH-type neural network so that the square of the difference between the actual output and the predicted one is minimised, that is,

$$\sum_{i=1}^{M} [\hat{f}(x_{i1}, x_{i2}, x_{i3}, ..., x_{in}) - y_i]^2 \rightarrow \min \quad (7)$$

The full form of the mathematical description can be presented by a system of partial quadratic polynomials consisting of only two variables (neurons) in the form of:

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2. \quad (8)$$

Consequently, the coefficients of each quadratic function $G_i$ are obtained to optimally fit the output in the whole set of input-output data pair, that is,

$$\sum_{i=1}^{M} (\hat{y}_i - G_i(x_p, x_q))^2 \rightarrow \min. \quad (9)$$

In the basic form of the GMDH algorithm, all the possibilities of two independent variables out of total $n$ input variables are taken in order to construct the regression polynomial in the form of equation (8) that best fits the dependent observations ($y_i$, $i=1, 2, ..., M$) in a least-squares sense.

Consequently, \( \binom{n}{2} = \frac{n(n-1)}{2} \) neurons will be built up in the first hidden layer of the feed forward network from the observations \( \{ \langle y_i, x_{ip}, x_{iq} \rangle : (i=1,2, ..., M) \} \) for different $p,q \in \{1,2, ..., n\}$ [19]. In other words, it is now possible to construct $M$ data triples \( \{ \langle y_i, x_{ip}, x_{iq} \rangle : (i=1,2, ..., M) \} \) from the observation, using such $p,q \in \{1,2, ..., n\}$ in the form of:

$$\begin{bmatrix}
  x_{1p} & x_{1q} & y_1 \\
  x_{2p} & x_{2q} & y_2 \\
  \vdots & \vdots & \vdots \\
  x_{Mp} & x_{Mq} & y_M
\end{bmatrix}$$

Using the quadratic sub-expression in the form of equation (8) for each row of $M$ data triples, the following matrix equation can be readily obtained as:

$$Aa = Y \quad (10)$$

where $a$ is the vector of unknown coefficients of the quadratic polynomial as in equation (8)

$$a = \{a_0, a_1, ..., a_5\}^T \quad (11)$$

and

$$Y = \{y_1, y_2, y_3, ..., y_M\}^T \quad (12)$$

is the vector of output's value from the observation. It can be readily seen that

$$A = \begin{bmatrix}
  1 & x_{1p} & x_{1q} & x_{1p} x_{1q} & x_{1p}^2 & x_{1q}^2 \\
  1 & x_{2p} & x_{2q} & x_{2p} x_{2q} & x_{2p}^2 & x_{2q}^2 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  1 & x_{Mp} & x_{Mq} & x_{Mp} x_{Mq} & x_{Mp}^2 & x_{Mq}^2
\end{bmatrix}. \quad (13)$$

The least-squares technique from multiple-regression analysis leads to the solution of the normal equations in the form of:

$$a = (A^T A)^{-1} A^T Y \quad (14)$$

which determines the vector of the best coefficients of the quadratic equation (8) for the whole set of $M$ data triples. It should be noted that this procedure is repeated for each neuron of the next hidden layer, according to the connectivity topology of the network. However, such solution directly taken from solving normal equations (SNE) is rather susceptible to round off errors and, more importantly, to the singularity of these equations.
The evolutionary methods such as genetic algorithms have been widely used in different aspects of design in neural networks because of their unique capabilities of finding a global optimum in highly multi-modal and/or non-differentiable search space [24, 25]. In this work, the design of architecture is performed using Genetic Algorithm (GA). The incorporation of genetic algorithm into the design of such GMDH-type neural networks starts by representing each network as a string of concatenated sub-strings of alphabetical digits. The fitness, \( \Phi \), of each string of the symbolic digits which represents a GMDH-type neural network to the model squat is evaluated in the form of:

\[
\Phi = \frac{1}{E}
\]  

(15)

where \( E \) is the mean square of the error given by equation (9) is minimized through the evolutionary process by maximizing the fitness, \( \Phi \). The evolutionary process starts by randomly generating an initial population of symbolic strings, each as a candidate solution. Using the aforementioned genetic operations of roulette wheel selection, crossover and mutation, the entire populations of symbolic strings is to improve gradually. In this way, GMDH-type neural network models of ship squat with progressively increasing fitness, \( \Phi \), are produced until no further significant improvement is achievable. It should be noted that such an evolutionary process was used in conjunction with the normal equation approach for the coefficients of the quadratic polynomials involved in the design of the GMDH-type networks. The details of several types of GMDH neural networks are available in [26].

4 Adaptive-network-based Fuzzy Inference Systems

In this study, the use of ANFIS is adopted in modeling ship squat in shallow water. ANFIS was first introduced by Jang [27]. The model is based on Takagi–Sugeno inference model [28, 29]. ANFIS uses a hybrid learning algorithm to identify the consequent parameters of Sugeno-type fuzzy inference systems. Furthermore, the Sugeno fuzzy model is assumed to have two inputs, \( m \) and \( n \), and one output, \( f \). For a first-order Sugeno fuzzy model, a typical rule set with two fuzzy if–then rules can be expressed as:

Rule 1: If \( (m \text{ is } A_i) \) and \( (n \text{ is } B_j) \) then

\[
f_i = p_{1i}m + q_{1j}n + r_i
\]

(16)

Rule 2: If \( (m \text{ is } A_2) \) and \( (n \text{ is } B_j) \) then

\[
f_j = p_{2i}m + q_{2j}n + r_i
\]

(17)

where \( p_{1i}, p_{2j}, q_{1j}, q_{2j}, r_i \) and \( r_j \) are linear parameters and \( A_i, A_j, B_i \) and \( B_j \) are nonlinear parameters.

The entire system consists of five layers, fuzzy layer, product layer, normalized layer, de-fuzzy layer and total output layer. The relationship between input and output of each layer is discussed in the following sections.

Layer 1 is the fuzzy layer, in which \( m \) and \( n \) are the input of nodes \( A_i, B_i \) and \( A_j, B_j \) respectively. \( A_i, A_j, B_i \) and \( B_j \) are the linguistic labels used in the fuzzy theory for dividing the membership functions. The membership relationship between the output and input functions of this layer can be expressed as below:

\[
\begin{align*}
O_{ij} &= \mu_{A_i}(m), \quad i = 1, 2 \\
O_{ij} &= \mu_{B_j}(n), \quad j = 1, 2
\end{align*}
\]

(18)

where \( O_{ij} \) and \( O_{ij} \) denote the output functions, and \( \mu_{A_i} \) and \( \mu_{B_j} \) denote the membership functions.

Layer 2 is the product layer that consists of two nodes labeled \( \Pi \). The output \( W_i \) and \( W_j \) are the weight functions of the next layer. The output of this layer is the product of the input signal, which is defined as follows:

\[
O_{2,i} = w_i = \mu_{A_i}(m)\mu_{B_i}(n), \quad i = 1, 2
\]

(19)
where $O_{2,j}^*$ is the output of Layer 2.

The third layer is the normalized layer, whose nodes are labeled $N$. The function of this layer is to normalize the weight function in the following process:

$$\hat{O}_{3,j} = w = \frac{w_j}{w_1 + w_2}, \quad i = 1, 2 \quad (20)$$

where $O_{3,j}$ is the output of Layer 3.

The fourth layer is the defuzzification layer. The nodes in this layer are adaptive nodes. The relationship between the inputs and outputs of this layer can be defined as the following:

$$\hat{O}_{4,j} = w_j (p_i m + q_i n + r_i) \quad i = 1, 2 \quad (21)$$

where $O_{4,j}$ is the output of Layer 4, and $p_i$, $q_i$, and $r_i$ are the linear parameters of the node.

The fifth layer is the output layer, whose node is labelled $\sum$. The output of this layer is composed of all the ingredients of the inputs, which represents the results of the cleaning rates. The output can be expressed as below:

$$\hat{O}_{5,j} = \sum_i w_i f_i = \sum_i w_i f_i, \quad i = 1, 2 \quad (22)$$

where $O_{5,j}$ is the output of Layer 5.

## 5 ANN and ANFIS Squat Modeling

### 5.1 Data Preparation

The parameters of the interest in these multi-input single-output systems, both GMDH-type neural network and ANFIS, that affect the ship squat are block coefficient (CB), the dimensionless distance between the seabed and the ship’s floor ($\delta$) and Hydraulics Froude Number ($Fn_h$), that is described in Section 2 in detail. There has been a total number of 82 input-output experimental data considering 3 input parameters, namely CB, $\delta$, $Fn_h$ and one output namely, dimensionless squat ($S^*$). In order to demonstrate the prediction ability of both GMDH-type neural networks and ANFIS, the data have been divided into two different sets, namely training and testing sets. For dimensionless squat, training set which consists of 60 out of 82 inputs-output data is used for training both the neural network and ANFIS models. The testing set, which consists of 22 unforeseen inputs-output data samples, is merely used for testing to show the prediction ability of such GMDH-type neural network and ANFIS models during the training process.

### 5.2 GMDH-type Neural Network modeling of ship squat in shallow waters

The GMDH-type neural network is now used for such input-output data to find the polynomial model of dimensionless squat in respect to their effective input parameters. In order to genetically design such GMDH-type neural network described in previous section, a population of 50 individuals with a crossover probability of 0.7 and mutation probability of 0.07 has been used in 160 generation which no further improvement has been achieved for such population size. The structure of the evolved GMDH models with 2 and 3 hidden layers is shown in Figure 3.

Corresponding MSE are calculated as 0.000193 and 0.000114 for training and testing set, respectively. The MSE for GMDH model with 3 hidden layers is 0.000168 and 0.000096 for training and testing sets, respectively. One can define the maximum hidden layers of such GMDH neural networks higher than three until both training and testing errors decrease with increasing the hidden layers. When the number of hidden layers increase, the corresponding testing error increases (in spite of decreasing the training error), too. This is related to a model that has been made with less a hidden layer. When the over-fitting phenomenon happens, consequently, a model with less hidden layer must be chosen as the best model for modeling each process. For the squat modeling with GMDH neural networks, when four hidden layers are predefined

![Fig. 3 Structures of GMDH neural network models for dimensionless squat : Two and Three hidden layers with three and six neurons respectively.](image)
for training, the testing error of this trained model is greater than the model with three hidden layers. For these reason, three hidden layers’ model has been chosen as the best GMDH model for ship squat in shallow water. The good behavior of GMDH-type neural network model with 3 hidden layers is also depicted in Figure 4.

![Figure 4](image)

**Fig. 4 The variation of Dimensionless Squat with input data samples (GMDH with 3 Hidden Layers)**

### 5.3 ANFIS modeling of ship squat in shallow water

The computation of data for ANFIS is conducted using MATLAB. The ANFIS training includes hybrid method. The parameters of the membership functions are optimized in the identification dataset through back propagation while the consequent parameters are calculated using a linear least squares method.

In order to model such three-input single-output set of data, an ANFIS with two linguistic terms in each antecedent which is equivalent to two membership functions for each input variable, is considered. Different models are built, using triangular, bell-shape, Gaussian and trapezoidal membership functions. It should be noted that the number of parameters in each vector of coefficients in the concluding part of each TSK-type fuzzy rule is four, according to the assumed linear relationship of input variables in the consequents. Consequently, $2^3 = 8$ TSK-type fuzzy rules are identified using ANFIS, given in the MATLAB fuzzy logic toolbox. Compared results of such models with different types of membership functions aspect to MSE of training and testing set have been shown in Table 2.

<table>
<thead>
<tr>
<th>Mem. Function</th>
<th>Mean Square Training Error</th>
<th>Mean Square Prediction Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>gbellmf</td>
<td>2.20381E-05</td>
<td>5.15621E-05</td>
</tr>
</tbody>
</table>

It can be seen that the triangular membership functions result in the best values for both training and prediction errors. The triangular curve is a function of a vector $x$, and depends on three scalar parameters $a, b$ and $c$ as given below:

$$f(x,a,b,c) = \max \left( \min \left( \frac{x-a}{b-a}, \frac{x-c}{c-b} \right), 0 \right)$$

Parameters $a$ and $b$ locate the feet of triangle and the parameter $c$ locates the peak. Figure 5 demonstrates the training and prediction behaviors of the ANFIS model obtained using triangular membership functions (trimf).

![Figure 5](image)

**Fig. 5 The variation of Dimensionless Squat with input data samples**

The ANFIS parameter values for inputs premise parameters of this model have been shown in Table 3.

<table>
<thead>
<tr>
<th>Mem. Function(trmf)</th>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CB</td>
<td>0.65</td>
<td>0.7</td>
<td>0.7499</td>
</tr>
<tr>
<td>A2</td>
<td>0.7001</td>
<td>0.75</td>
<td>0.8</td>
</tr>
<tr>
<td>A3</td>
<td>0.04951</td>
<td>0.09037</td>
<td>0.1268</td>
</tr>
<tr>
<td>A4</td>
<td>-0.0504</td>
<td>0.2294</td>
<td>0.2858</td>
</tr>
<tr>
<td>$F_{nh}$</td>
<td>-0.4134</td>
<td>0.2265</td>
<td>0.5151</td>
</tr>
<tr>
<td>A5</td>
<td>0.08751</td>
<td>0.6278</td>
<td>1.071</td>
</tr>
</tbody>
</table>

Table 3. ANFIS Parameters Values for Inputs Premise Parameters
The obtained set of TSK-type fuzzy rules for modeling of a dimensionless ship squat in shallow water are as follows:

Rule 1: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = 0.02153 \ CB + 0.002798 \ \delta - 0.5927 \ Fn_{b} + 0.03075. \]

Rule 2: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = 0.06426 \ CB + 0.008354 \ \delta + 2.103 \ Fn_{b} + 0.09181. \]

Rule 3: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = -0.0265 \ CB - 0.03465 \ \delta + 0.772 \ Fn_{b} - 0.03786. \]

Rule 4: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = -0.2024 \ CB - 0.6669 \ \delta + 1.459 \ Fn_{b} - 0.2892. \]

Rule 5: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = 0.1065 \ CB + 0.01292 \ \delta - 3.077 \ Fn_{b} + 0.142. \]

Rule 6: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = 0.8043 \ CB + 0.09759 \ \delta + 2.329 \ Fn_{b} + 1.072. \]

Rule 7: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = -0.03658 \ CB + 0.09796 \ \delta + 0.8211 \ Fn_{b} - 0.04877. \]

Rule 8: If \( CB = \alpha_{i} \) and \( \delta = \alpha_{i} \) and \( Fn_{b} = \alpha_{i} \), then
\[ S' = -0.1287 \ CB - 1.7400 \ \delta + 1.603 \ Fn_{b} - 0.1717. \]

6 Comparison of the ANFIS results with other techniques

In two former sections the GMDH4-type neural network and ANFIS, methods are used for modeling and prediction of ship squat in shallow water. In this section, initially, a simple method called multiple linear regressions is used and, eventually, the results of these three methods have been compared. In this part of study, the comparison of these three approaches is made on the basis of the accuracy, R-square fitting parameter (that is, the square of the correlation between the experimental dimensionless squat values and the predicted dimensionless squat values) and train and test performance.

Multiple linear regression analysis is usually used to summarize data as well as to study the relations between variables [30].

The simple equation obtained for dimensionless squat with multiple regression analysis is:

\[ S' = -0.433 + 0.538 \ CB - 0.260 \ \delta + 0.620 \ Fn_{b} \]  \hspace{1cm} (23)

The comparison of the GMDH-type neural networks (with 2 and 3 hidden layers), ANFIS (with 8 rules and triangular membership function) and multiple linear regression models is presented in Table 7. The comparison shows that:

(i) the minimum R-square is obtained with the multiple regression model,
(ii) the R-square of ANFIS is higher than that of GMDH-type neural network models,
(iii) the training and test errors of ANFIS are smaller than those of GMDH and regression models.

In brief, it may be stated that ANFIS yields most accurate results.

7 Conclusions

This paper is conducted to demonstrate the usefulness of the artificial intelligence techniques for the prediction of ship squat in shallow water. GMDH-type neural networks and ANFIS are applied for modeling the ship squat that varies with three effective parameters, namely block coefficient (CB), dimensionless distance between the seabed and ship floor (\( \delta \)) and Froude Number (\( Fn_{b} \)) that were investigated experimentally. The GMDH-type neural network model with 3 hidden layer and 6 neurons within those layers is selected as the optimum and best network for modeling and prediction of ship squat among other GMDH-type models because of minimum MSE on two predefined sets, the training set and testing set. Eventually, the accuracy of predictions and the adaptability of the ANFIS have been examined. The ANFIS indicated that it is capable to learn the training dataset and accurately predict the output of test data. Triangular membership functions are chosen as the best membership function for ANFIS training of the experimental data. The results obtained with ANFIS and GMDH-type neural networks are compared with each other. These results were also compared with the multiple linear regression method. The comparison showed that the ANFIS performed better than GMDH-type neural networks and multiple linear regressions.

8 Acknowledgment

The authors wish to express their appreciations to the helpful suggestions and comments of Professor Nader Nariman-Zadeh from Mechanical Department of University of Guilan.

References:

Modeling and Prediction of Ship Squat using ANFIS and GMDH-type Neural Network


Modelling on Pedestrian Accidents in Malaysia

NOORIZAM DAUD and HASLINDA AB. MALEK
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam
MALAYSIA
noorizam@tmsk.uitm.edu.my, linda_gurl04@yahoo.com
http://www.uitm.edu.my

Abstract: - Malaysia has been experiencing rapid growth in population, economy and motorization. This increase in population and motorization has led to the increase in the number of pedestrian accidents on road networks. Recently, road safety has become a major concern for many countries including Malaysia. Many researches pertaining to the pedestrian accident data has been conducted in aiming to reduce the rate of pedestrian accidents in Malaysia. The main aim of this study is to identify which distribution is the best fitted for pedestrian accident data obtained from 30 locations in Malaysia from year 2006 to 2008. It is found that Negative Binomial distribution provides a better fit for daytime and night-time of pedestrian accident data as compared to the Poisson distribution. This study also emphasized on the identification of pedestrian accident blackspots using hierarchical Bayesian approach based on the correlation between the mean number of pedestrian accidents during night-time and daytime. This study reveals that Tapah emerged as the top ranked pedestrian accident location in this period.

Key-Words: - Pedestrian Accidents, Negative Binomial, Poisson, Blackspots, Hierarchical Bayesian, correlation.

1 Introduction

The pedestrians are known as the most vulnerable of all road users. Malaysia has been experiencing rapid growth in population, economy and motorization. This increase in population and motorization has led to the increase in the number of pedestrian accidents on road networks. The children and elderly people are considered to be at a higher risk regarding pedestrian crashes [9]. The pedestrian accidents contribute to the health and social problem in this country. Moreover, road accidents are one of the major contributors of human deaths in Malaysia. Pedestrian exposure research has been conducted for decades; however, the crashes between pedestrians and motor vehicles become more serious problem. Therefore, this paper concentrate on finding the best distribution model of pedestrian accidents and use it in identifying the potential pedestrian accident blackspots. The identification of pedestrian accident prone locations could help the road authority to make the decision regarding the road safety especially in order to improve the pedestrians’ safety and hence, it could reduce the number of such pedestrian accidents in the near future.

Many research related to the accident data has been conducted in order to help in reducing the rate of accidents. Among the studies conducted on the distribution related to the accident data are Poisson distribution, Binomial distribution, Negative Binomial distribution, Truncated Negative Binomial distribution and Log Series distribution. Besides, many researchers claimed that among the most popular distributions applied in road safety analysis are the standard Poisson and Negative Binomial (NB) distributions [1, 6, 8]. Furthermore, some of the researchers have proved that the accident data can be modelled by a Poisson distribution, for example Zahavi [13] has shown that the Poisson distribution is the best fitted model for the monthly accident data at Israel from 1959 until 1961. The same result found by Chapman [3] for daily accident occurred at New South Wales, Australia from 1958 until 1963. On the other hand, Weber [12] has proved that the Negative Binomial distribution could fit the accident data for 148,000 drivers in California from 1961 until 1963. He concluded that the number of accidents for a certain period of time can be modelled by a Negative Binomial distribution. The same result found by Wan and Kamarulzaman [11]. They have fitted a Poisson and Negative Binomial distributions for accident data obtained from 25 locations in United Kingdom from year 1974 to 1992. Based on the comparison of the fitted model, they concluded that Negative Binomial distribution provides a better fit for the accident data as compared to the Poisson distribution. However, Senn and Collie [10] have fitted the Bivariate Negative Binomial
distribution to the accident data for two periods where the first period covers from year 1979 until 1980 and the second period covers from year 1981 until 1982. They concluded that the Bivariate Negative Binomial distribution is appropriate to model accident data with such features. In addition to that, Maher [5] also suggested that this model could be extended to the data on two types of accident (slight and severe) and data on two period of time (night-time and daytime).

More recently, Bayesian techniques are used to tackle problems in traffic safety. A number of authors have used the empirical Bayesian approach, combined with multivariate regression models to estimate the safety at various types of facilities. This approach was first proposed by Bonneson et al.[2] to estimate the safety at two way stop controlled intersections on rural highways. On the other hand, Noorizam and Kamarulzaman [7] have used a hierarchical Bayesian approach in order to rank the accident blackspots based on the correlation exists between the mean number of fatal and serious injury motor vehicle accidents. Furthermore, Higle and Hecht [4] have conducted a controlled experiment to compare the efficiency of different Bayesian and conventional hazardous location identification methods. It was concluded that Bayesian identification methods generally perform better than the conventional methods in correctly identifying hazardous locations.

2 Methodology
This section will discuss about data used in the study and also the method used in searching the suitable distribution to fit this pedestrian accident data. In addition, the discussion also covers on the hierarchical Bayesian approach by considering the correlation exists between the mean number of night-time and daytime pedestrian accidents.

2.1 Data Collection
The data used in this study was obtained from the Malaysian Institute of Road Safety Research (MIROS). The number of pedestrian accidents occurred at 30 different locations in Malaysia from the year 2006 until 2008. There are three variables measured:
1) The number of pedestrian accidents in each location.
2) The location of pedestrian accidents.
3) The time of pedestrian accidents (night-time or daytime).

2.2 FITTING DISTRIBUTION TO THE PEDESTRIAN ACCIDENT DATA
Here, Poisson and Negative Binomial distributions would be considered in order to identify the appropriate distribution to model the night-time and daytime pedestrian accident data. The parameters for both distributions are estimated using the method of maximum likelihood. The Chi-square goodness of fit test will be used to determine the appropriate distribution to the data.

2.3 HIERARCHICAL BAYESIAN APPROACH BASED ON THE CORRELATION
Consider two discrete random variables $X_i$ and $Y_i$ each representing the number of pedestrian accidents in night-time and the number of pedestrian accidents in daytime occurring at a certain location. Since both $X_i$ and $Y_i$ satisfy the characteristics of a Poisson process, it is reasonable to assume that both variables are following Poisson distribution with mean $\lambda_1$ and $\lambda_2$, respectively. Assuming $\lambda_1$ follows Gamma distribution with parameter $\alpha$ and $\beta$ whereas $\lambda_2$ follows Gamma distribution with parameter $\alpha$ and $\gamma$. The $\lambda_1$ and $\lambda_2$ each were having means $\mu_1 = \frac{\alpha + \beta}{\gamma}$ and $\mu_2 = \frac{\alpha + \beta_2}{\gamma}$.

According to Maher [5], $\lambda_1$ and $\lambda_2$ can be assumed to be

$\lambda_1 = U + V$
$\lambda_2 = U + W$

where;
$U \sim \text{Gamma}(\alpha, \gamma), V \sim \text{Gamma}(\beta_1, \gamma)$ and $W \sim \text{Gamma}(\beta_2, \gamma)$.

Assuming $U$ and $V$, and also $U$ and $W$ are independent and because of $U$ is common to both, $\lambda_1$ and $\lambda_2$ are positively correlated with correlation coefficient denoted as $\rho$. This value of correlation coefficient, $\rho$ can be obtained by using the following formula:

$$\rho = \frac{\alpha}{\sqrt{(\alpha + \beta)(\alpha + \beta_2)}}$$

(1)

Maher [5] also has suggested the application of Bivariate Negative Binomial distribution to be used to analyze traffic accidents during night-time and daytime. The application of this distribution is based on the assumption that the positive correlation exists.
between the mean number of pedestrian accidents during night-time and daytime.

The Bivariate Negative Binomial distribution as suggested by Maher [5] is given below:

\[
P(r_1, r_2) = \frac{\Gamma(r_1 + r_2)}{\Gamma(r_1) \Gamma(r_2)} \left( \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2 + \theta} \right)^{r_1 + r_2} \left( \frac{\lambda_1}{\lambda_1 + \lambda_2 + \theta} \right)^{r_1} \left( \frac{\lambda_2}{\lambda_1 + \lambda_2 + \theta} \right)^{r_2}
\]  

in which \(f_0(u), f_0(v), f_0(w)\) are the probability density functions of three gamma variables \(U, V,\) and \(W.\)

### 3 Analysis and Findings

Based on the parameter estimations, the corresponding estimated values for Poisson and Negative Binomial distributions using method of maximum likelihood are given in Table 1.

Table 1: Summary of the parameter estimates

<table>
<thead>
<tr>
<th>Time of Accidents</th>
<th>Distribution</th>
<th>Estimated Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Night-time</td>
<td>(X\sim\text{Poi}(\lambda_2))</td>
<td>(\hat{\lambda}_2 = 1.99)</td>
</tr>
<tr>
<td>Daytime</td>
<td>(Y\sim\text{Poi}(\lambda_2))</td>
<td>(\hat{\lambda}_2 = 3.53)</td>
</tr>
<tr>
<td>Night-time</td>
<td>(X\sim\text{NB}(k, \theta))</td>
<td>(k = 1.187) (\theta = 0.374)</td>
</tr>
<tr>
<td>Daytime</td>
<td>(Y\sim\text{NB}(k, \theta))</td>
<td>(k = 2.654) (\theta = 0.429)</td>
</tr>
</tbody>
</table>

Table 2 shows the summary of the Chi-square goodness of fit test. Based on the comparison of the fitted model, it is found that Negative Binomial distribution provides a better fit for night-time and daytime of pedestrian accident data as compared to the Poisson distribution.

In this study, the correlation between the mean number of pedestrian accidents during night-time and daytime was found to be 0.41. Based on the p-value (0.024), at the 5% level of significance, we could conclude that there is a significant correlation between the mean number of pedestrian accidents during night-time and daytime. Hence, the Bivariate Negative Binomial is appropriate to be used in analyzing this data where it parameters need to be estimated. By using the value of correlation coefficient, \(r = 0.41\), mean of night-time accidents, \(\mu_1 = 2\), mean of daytime accidents, \(\mu_2 = 3.5\) and \(\gamma = 1\), the parameters of Bivariate Negative Binomial which are \(\alpha, \beta_1, \beta_2\) can be found using the simultaneous equations technique. Each posterior mean of \(\lambda_1\) and \(\lambda_2\) are obtained using a written MATLAB program. The top ten pedestrian accident locations identified to be potential hazardous pedestrian accident locations is given in Table 3.

Table 2: Summary of the chi-square goodness of fit test

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Chi-square Value</th>
<th>Chi-square Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X\sim\text{Poi}(1.99))</td>
<td>29.186</td>
<td>((x^2_{0.053} = 7.013))</td>
</tr>
<tr>
<td>(Y\sim\text{Poi}(3.53))</td>
<td>47.761</td>
<td>((x^2_{0.053} = 9.236))</td>
</tr>
<tr>
<td>(X\sim\text{NB}(1.187, 0.374))</td>
<td>3.420</td>
<td>((x^2_{0.053} = 7.813))</td>
</tr>
<tr>
<td>(Y\sim\text{NB}(2.654, 0.429))</td>
<td>11.659</td>
<td>((x^2_{0.053} = 12.592))</td>
</tr>
</tbody>
</table>

Table 3: Posterior mean with the correlation coefficient 0.41

<table>
<thead>
<tr>
<th>Location</th>
<th>(X_i)</th>
<th>(Y_i)</th>
<th>Posterior Mean</th>
<th>Ranking</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tapah</td>
<td>17</td>
<td>17</td>
<td>10.358</td>
<td>R(1)</td>
</tr>
<tr>
<td>Mersing</td>
<td>14</td>
<td>19</td>
<td>10.061</td>
<td>R(2)</td>
</tr>
<tr>
<td>Barat daya</td>
<td>9</td>
<td>17</td>
<td>8.081</td>
<td>R(3)</td>
</tr>
<tr>
<td>Kuala Lipis</td>
<td>7</td>
<td>18</td>
<td>7.753</td>
<td>R(4)</td>
</tr>
<tr>
<td>Langkawi</td>
<td>7</td>
<td>16</td>
<td>7.237</td>
<td>R(5.5)</td>
</tr>
<tr>
<td>Pekan</td>
<td>6</td>
<td>17</td>
<td>7.202</td>
<td>R(7)</td>
</tr>
<tr>
<td>P. Besar</td>
<td>11</td>
<td>10</td>
<td>6.763</td>
<td>R(8)</td>
</tr>
<tr>
<td>Kulai Jaya</td>
<td>13</td>
<td>8</td>
<td>6.730</td>
<td>R(9)</td>
</tr>
<tr>
<td>Machang</td>
<td>8</td>
<td>13</td>
<td>6.742</td>
<td>R(10)</td>
</tr>
</tbody>
</table>

### 4 Conclusion

Negative Binomial distribution provides a better fit for night-time and daytime of pedestrian accident data as compared to the Poisson distribution. Since the mean number of pedestrian accidents during night-time and daytime are found to have a significant positive correlation, a Bivariate Negative Binomial distribution is used to model the night-time and daytime pedestrian accidents. Based on the overall posterior mean value, Tapah emerged as the top ranked pedestrian accident location in this study.
References:


Deep and surface learning of elementary calculus concepts in a blended learning course

R. Haripersad
Department of Mathematics, Statistics and Physics
Durban University of Technology
Kwa-Zulu Natal
South Africa
rajeshh@dut.ac.za

Abstract

This paper focuses on students understanding of elementary calculus in a blended learning course at a University of Technology. Conventional lectures were integrated with the computer laboratory teaching environment to promote interactive and discovery learning. Projects were designed to support the development of calculus frames in conjunction with a theoretical framework that was used in analyzing students understanding of integral calculus concepts. The students in the blended learning mathematics course (experimental group) was also compared to students that were traditionally lectured (control group). Both groups were assessed by the modified Orton’s battery of tests on integral calculus. The experimental group exhibited deep learning of concepts, while the control group possessed more surface structures.

Keywords: Blended Learning, Calculus, Deep and Surface Structures

1 Introduction

Calculus is fundamental to further study of mathematics at a University of Technology. However, students that study engineering at the UOT, enter with low symbols in mathematics. They possess many misconceptions and have poor pre-knowledge frames in basic mathematics and calculus. Naidoo [1] deduced that, first year mathematics students that are taught traditionally, study by rules. Lecturers tend to teach mechanistically and do standard type problems and solutions. Tall [2] states that students develop coping strategies, like computational and manipulative skills, when faced with conceptual difficulties.

Research in teaching and learning using the computer laboratory method gave a measure of success, Naidoo [3], especially in graph construction and numerical solutions. Although students were performing better, they still made errors that include, amongst others inability to conclude that sequences converge; problems with rate of change of a curve; students lacked the ability to interpret symbols.

Silverberg [4], studies performed by a Calculus Reform Group showed a measure of success. His analysis of grades of traditional and reform cohorts produced the most compelling results. Significant improvements in results were noted between cohorts in contrast to the reform group that performed better after some time.

The need for alternate methods of instruction to enhance teaching and learning of calculus is essential. A Blended Learning (traditional and computer laboratory teaching) mathematics course was developed and implemented in an attempt to improve student’s understanding of elementary calculus.
2 Blended Learning

According to Singh [5], blended learning mixes various event-based activities, including face-to-face classrooms, live e-learning, and self-paced Web Based Learning (WBL). Blended Learning (BL) often is a mix of traditional instructor-led training, synchronous online conferencing or training, asynchronous self-paced study.

He propagates this type of learning since learning styles of each learner tend to be different, and hence, “a single mode of instructional delivery may not provide sufficient choices, engagement, social contact, relevance, and context needed to facilitate successful learning and performance”.

Blended learning is viewed as midway along a continuum that at one extreme has conventional face-to-face instruction, and on the other end totally WBL. It is self-paced, collaborative or an inquiry-based study. Blended learning should not be an “add on” to instruction, but as an integrated component of the course [6]. The current trend of research is to explore environments with a better balance between two extremes.

By using blended learning, we expect to enable students to easily move between

- listening to a lecture;
- engaging in class discussion;
- working collaboratively;
- using available software to investigate concepts or solve problems
- accessing their archived work

In this environment, linguistic, cultural, social and economic groups can interact within a group and among each other. The WBL was blended into the traditional lectures using the Wrap-Around Model of Mason [7].

The Blended Learning engineering mathematics one course was designed using ‘Rule of Three’ guiding principles which consist of graphical, numerical and analytical methods that are used to teach calculus concepts. The aim is to produce a course where the three points of view are balanced, and where students see each major idea from several angles which is necessary for an engineer, [8].

3 Deep and Surface Structures

Deep and surface level procedures of learning has been identified in many studies. Matz [9], states that surface level procedures are ordinary rules of algebra while deep learning serves the purpose of creating and modifying superficial-level rules or changing the control structure.

The blended mathematics course is guided by Campbell [10] who outlines the following methods to promote deep learning:
- encourage faculty/student interaction
- encourage student/student interaction
- use active and interactive teaching methods
- make links with existing student knowledge
- discussing/teaching learning skills explicitly
- link topics to student’s lives and career aspirations
- encourage collaborative projects

Blended Learning discourages surface learning which focuses on comprehension and reproduction of knowledge (rote learning) as follows:
- excessive amount of material – Blended Learning releases the bare minimum until after interaction with students whose pre-conceptual frames determines the amount of learning material exposed to the student
- high lecture contact hours
lack of opportunity to pursue subject in depth – Blended mathematics gives the opportunity to students to investigate the concepts by using the quiz or project or enrichment materials

Ramsden, [11], summarized the deep and surface approach to learning as follows:

<table>
<thead>
<tr>
<th>Deep</th>
<th>Surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus is on “what is signified”</td>
<td>Focus is on the “signs”</td>
</tr>
<tr>
<td>Relates previous knowledge to new knowledge</td>
<td>Focus is on unrelated parts of the task</td>
</tr>
<tr>
<td>Relates knowledge from different courses</td>
<td>Information for assessment is simply memorized</td>
</tr>
<tr>
<td>Relates theoretical ideas to everyday experience</td>
<td>Facts and concepts are associated unreflectively</td>
</tr>
<tr>
<td>Relates and distinguishes evidence and argument</td>
<td>Principles are not distinguished from examples</td>
</tr>
<tr>
<td>Organizes and structures content to coherent whole</td>
<td>Task is treated as an external imposition</td>
</tr>
<tr>
<td>Emphasis is internal, from within the student</td>
<td>Emphasis is external, from demands of assessment</td>
</tr>
</tbody>
</table>

The point is how to get students to use the deep approaches rather than the surface approaches, [12]. What students do when learning and why they do it is described as a ‘congruent motive-strategy package’, [13]. Case & Gunstone, [14], describe metacognitive development as the move to greater knowledge, awareness and control of one’s own learning.

Flavell, [15], describes metacognition as ‘one’s knowledge concerning one’s own cognitive processes and products or anything related to them’. It also includes ‘the active monitoring and consequent regulation and orchestration’ of information processing activities. The computer is a means of getting students to use deep approaches in their search for solutions. In the case of calculus solutions it also provides a visual aid to enhance comprehension.
4 Methodology and Qualitative Analysis

The qualitative analysis considered various sections in elementary calculus: sequences, limits and infinity, symbolism, area and integration to classify errors made by students. Students had to perform the project tasks that contributed to their part assessment for the mathematics course. Further investigations was performed to find out the strategy used by students with respect to the use of deep and surface structures to relate to the tasks presented to them.

Task A, required students to retrieve the pre-knowledge frames: area of trapezium = \( \frac{1}{2} (a + b) h \), area of rectangle = \( l \times b \) and area of triangle = \( \frac{1}{2}(b \times h) \). Then using the appropriate values in the respective formulae they should find that area = \( \frac{1}{2} (a + b) h = [l \times b + \frac{1}{2}(b \times h)] \). They should predict that this method gives the area under the line.

In Task B, the Zoom graph toolbar in Mathematica allows students to experiment with concepts by magnifying graphs, changing variables, etc. From this exercise they will see that as the number of rectangles under the graph is increased, the more accurate the area under the graph - the error is reduced.

Task C, tested the students understanding of area. Even though the result of finding the definite integral \( \int_{a}^{b} f(x)dx \) is 0, there is clearly a region enclosed by the curve and the \( x \) axis. Learners had to reason that part of the curve lies below the \( x \) axis and if the Riemann integral is used the total area will not be equivalent to zero. Therefore, if area is considered positive both under and above the graph then the definite integral and Riemann sum should be equivalent to each other.

We further used a modified Orton’s test (1983) to elicit responses from a control (traditional learning group) and the experimental group. The tasks were modified to include heights of rectangles, area under graph, summation of area of rectangles and limit of sequence equals area under graph. This was related to 10 descriptive items and included in the project tasks. The control and experimental group were interviewed on the tasks output. Table 1 & 2 categorizes the learning into deep, surface and intermediate structures from the data.

<table>
<thead>
<tr>
<th>Items</th>
<th>Deep</th>
<th>Intermediate</th>
<th>Surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>82%</td>
<td>6%</td>
<td>12%</td>
</tr>
<tr>
<td>2</td>
<td>76%</td>
<td>12%</td>
<td>12%</td>
</tr>
<tr>
<td>3</td>
<td>46%</td>
<td>42%</td>
<td>12%</td>
</tr>
<tr>
<td>4</td>
<td>48%</td>
<td>45%</td>
<td>7%</td>
</tr>
<tr>
<td>5</td>
<td>82%</td>
<td>12%</td>
<td>9%</td>
</tr>
<tr>
<td>6</td>
<td>58%</td>
<td>3%</td>
<td>39%</td>
</tr>
<tr>
<td>7</td>
<td>39%</td>
<td>10%</td>
<td>51%</td>
</tr>
<tr>
<td>8</td>
<td>46%</td>
<td>13%</td>
<td>41%</td>
</tr>
<tr>
<td>9</td>
<td>31%</td>
<td>13%</td>
<td>56%</td>
</tr>
<tr>
<td>10</td>
<td>33%</td>
<td>33%</td>
<td>33%</td>
</tr>
</tbody>
</table>
Table 2: Cognitive level: Control Group

<table>
<thead>
<tr>
<th>Items</th>
<th>Deep</th>
<th>intermediate</th>
<th>Surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79%</td>
<td>8%</td>
<td>13%</td>
</tr>
<tr>
<td>2</td>
<td>69%</td>
<td>13%</td>
<td>18%</td>
</tr>
<tr>
<td>3</td>
<td>13%</td>
<td>41%</td>
<td>46%</td>
</tr>
<tr>
<td>4</td>
<td>44%</td>
<td>32%</td>
<td>24%</td>
</tr>
<tr>
<td>5</td>
<td>25%</td>
<td>12%</td>
<td>63%</td>
</tr>
<tr>
<td>6</td>
<td>28%</td>
<td>2%</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>15%</td>
<td>11%</td>
<td>74%</td>
</tr>
<tr>
<td>8</td>
<td>36%</td>
<td>6%</td>
<td>58%</td>
</tr>
<tr>
<td>9</td>
<td>27%</td>
<td>9%</td>
<td>64%</td>
</tr>
<tr>
<td>10</td>
<td>8%</td>
<td>53%</td>
<td>39%</td>
</tr>
</tbody>
</table>

5 Project Tasks

Discuss the following problems by using the chat-room tool on the course. When solving, show all techniques, numerical tables, graphs and explain your answers fully.

**TASK A**

Find the area under $y = 6x + 3$ between $x = 0$ and $x = 2$ using the formulae for area of trapezium.

Can you verify this answer by using the formulae for area of a rectangle and area of a triangle?

Is it possible to find the area under a curve using this method, explain

**TASK B**

Using the Mathematica Graph toolbar (appendix 3) and the method of sums of rectangles under the curve $y = x^2$ above $x$ axis, between $x = 2$ and $x = 5$, calculate the area if

i) there is one rectangle

ii) there are three rectangles

iii) there are six rectangles
iv) Explain what happens as the number of rectangles under the curve is increased. Give a possible reason.

**TASK C**

For the graph of \( y = x^3 \), find the area enclosed by the graph and the \( x \) axis between \( x = -2 \) and \( x = 2 \)

**Appendix 1: Exemplars from Project Work**

**TASK A**

Area of trapezium = \( \frac{1}{2} (a + b)h = \frac{1}{2} (3 + 15)2 = 18 \)

Area under graph = area of rectangle + area of triangle

\[
= 1 \times b + \frac{1}{2} (b \times h) = (3 \times 2) + \frac{1}{2} (2 \times 15) = 6 + 15 = 21
\]

…. but I think I made an error since Area of trapezium should equal area under graph

[Intermediate Structure - student is able to see the relationship between the area under the graph and the area of trapezium however made an executive error of substituting the incorrect height for the triangle, thereby resulting in the incorrect answer].

**TASK B**

i) Area of one rectangle under curve = 1

\[
x \times h = 3 \times (2)^2 = 12
\]

ii) Area of 3 rectangles under curve = \( \sum (l \times h) = 1 \times ((2)^2 + (3)^2 + (4)^2) = 29 \)

iii) Area of three rectangle under curve

\[
\sum (l \times h) = \frac{1}{2} x ((2)^2) + (2,5)^2 + (3)^2 + (3,5)^2 + (4)^2 + (4,5)^2 = 34
\]

iv) Using the Zoom Graph toolbar, we see that as the number of rectangles are increased, the area becomes closer and closer to the actual area under the graph which is given by

\[
area = \int_2^5 x^2 \, dx = \frac{x^3}{3}\bigg|_2^5 = (125/3) - (8/3) = 39
\]

[Deep Structure – student exhibits all the appropriate pre-knowledge and lecture frames]
R. Haripersad

TASK C
i) 
\[
area = \int_{-2}^{2} x^3 \, dx = \left[ x^4 / 4 \right]_{-2}^{2} = (16 / 4) - (16 / 4) = 0
\]

ii) First, partition the intervals [-2,0] and [0,2] into n subintervals, each of length
\[
\Delta x = \frac{b-a}{n} = \frac{2-0}{n} = \frac{2}{n}
\]
Let’s work with interval [0,2]. The implication for interval [-2,0] are the same.
Because \( f \) is increasing in the interval [0,2], the minimum value on each subinterval occurs at the left endpoint, and the maximum value occurs at the right endpoint.

Using the left endpoint, \( m_i = \frac{2(i-1)}{n} \), the lower sum is
\[
s(n) = \sum_{i=1}^{n} f(m_i)(x_i - x_{i-1})
= \sum_{i=1}^{n} f\left(\frac{2(i-1)}{n}\right) \frac{2}{n}
= \sum_{i=1}^{n} \left[\frac{16}{n^3}\right] (i^3 - 3i^2 + 3i - 1)
= \frac{16}{n^2} \sum_{i=1}^{n} (i^3 - 3i^2 + 3i - 1)
= \frac{16}{n^2} \left(\frac{n^4}{4} + 4n^3 + 7n^2 + 8n\right)
= \frac{16}{n^2} \left(\frac{n^4}{4} + \frac{28}{n^2} + \frac{32}{n^3}\right)
= 4 - \frac{16}{n^2} + \frac{28}{n^3} + \frac{32}{n^4}
\]

Using the right endpoints, \( M_i = \frac{2i}{n} \), the upper sum is
\[
S(n) = \sum_{i=1}^{n} f(M_i)(x_i - x_{i-1})
= \sum_{i=1}^{n} f\left(\frac{2i}{n}\right) \frac{2}{n}
= \sum_{i=1}^{n} \left[\frac{16}{n^3}\right] (i^3 - 3i^2 + 3i - 1)
= \frac{16}{n^2} \sum_{i=1}^{n} (i^3 - 3i^2 + 3i - 1)
= \frac{16}{n^2} \left(\frac{n^4}{4} + 4n^3 + 7n^2 + 8n\right)
= \frac{16}{n^2} \left(\frac{n^4}{4} + \frac{28}{n^2} + \frac{32}{n^3}\right)
= 4 + \frac{8}{n} + \frac{4}{n^3}
\]
The example illustrates that for any value of n, the lower sum is less than (or equal to) the upper sum.

\[
s(n) = 4 - \frac{16}{n^2} + \frac{28}{n^3} + \frac{32}{n^4}
< 4 + \frac{8}{n} + \frac{4}{n^3} = S(n)
\]
Also, the difference between these two sums lessens as n increases, so if we take the limit as \( n \to \infty \), both the lower sum and upper sum approach 4

\[
limit_{n \to \infty} s(n) = \lim_{n \to \infty} \left(4 - \frac{16}{n^2} + \frac{28}{n^3} + \frac{32}{n^4}\right) = 4
\]
\[
limit_{n \to \infty} S(n) = \lim_{n \to \infty} \left(4 + \frac{8}{n} + \frac{4}{n^3}\right) = 4
\]
So the area for the region [-2, 2] which are intervals [-2,0] and [0,2] is 4 + 4 = 8

iii) The definite integral and the Riemann Sum should have the same answer, but it is not the case since we are determining areas below and above the graph.

[Arbitrary Error/Deep structure – student was able to calculate the areas using the Riemann Sum and definite integral but did not consider that because areas below the axis are taken as negative, so the integral from -2 to 2 gives the area difference and not the sum].
Appendix 2: Mathematica Demonstration

The area under a curve can be approximated by a Riemann sum. The definite integral is the limit of that area as the width of the largest rectangle tends to zero. Observe that as the number of rectangles is increased, the estimated area approaches the actual area.
References


http://www.maa.org/saum/maanotes 49/1245.html


Towards a General Legal Framework of an International Objective Environmental Liability?

CRISTIAN MAREŞ
Law Department, Faculty ofJuridical, Social and Political Sciences
“Valahia” University of Târgoviște
8-10 Gaesti Șos., Târgoviște, Dâmbovița County
ROMANIA
cristianmares@yahoo.fr

Abstract: Given that at international level the mechanism of international liability was considered as a possible way of avoiding the environmental damages and considering also the important number of accidents having a major impact on the environment, the codification of the rules governing the international liability is a subject which has interested the majority of the jurists and of the international professionals. States are still reluctant to accept and apply international rules that would enforce liability for activities whose harmful environmental consequences are likely to be limited to their own territory. Therefore, the civil liability imposed on operators is preferred to the objective liability of States. Although there are different activities which present a significant risk for the environment, the economic interests play a major political role.

Keywords: international liability, international responsibility, objective liability, absolute liability, transboundary harm, damage, precautionary principle, prevention principle, polluter pays principle

1 Introduction
The liability institution in the international environment law is under a crystallization period with regard to the legal regulation and jurisprudence, at present being manifested as a hot area, because of the fact that nature and its resources are severely affected, both at international level, and globally.

Since the very beginning we should mention that there is not any international legal instrument or document that would regulate a general, unitary and coherent legal status of international liability for the damages caused to the environment. In the contemporary international law, the rules on the liability institution for internationally wrongful acts and for harmful consequences resulting from activities that are not forbidden by international law, are regulated by the three drafts with a general character elaborated by the International Law Commission (hereinafter the “ILC”), respectively the Draft Articles on Responsibility of States for Internationally Wrongful Acts, Draft Articles on Prevention of Transboundary Harm from Hazardous Activities, both adopted on the occasion of ITS fifty-third session, in 2001, Draft Principles on the Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities, adopted at ITS fifty-eighth session in 2006, and the Lugano Convention on Civil Liability for Damage Resulting from Activities Dangerous to the Environment as of June 20, 1993 (hereinafter the “Lugano Convention”) [1] that represents the most elaborated international treaty until present, that takes into consideration liability and repair of damages caused to the environment.

The U.N. General Assembly, by its Resolution No.56 as of December 12, 2001, expressed its appreciation for the ILC activity that supplied this Draft Articles on Prevention of Transboundary Harm from Hazardous Activities and requested the ILC to go on working for the finishing of the second part of the activity related to the environmental transboundary harm, namely on liability, taking into account the relationship between prevention and liability [2].

Considering the Resolution No.56 as of December 12, 2001, at its fifty-eighth session in 2006, ILC adopted, on the second lecture, the text of the Draft Principles on the Allocation of Loss in the Case of Transboundary Harm Arising out of Hazardous Activities. Subsequently ILC recommended to the General Assembly of the UN to adopt a resolution by which to appropriate the Draft Principles and requested to the states to act at national and international levels for their implementation.
2 International Responsibility – International Liability

Following the analysis of these legal documents we made a distinction between responsibility for internationally wrongful acts and liability for legal activities, concluding that the basis of the two types of liability is different. In case of states’ responsibility for internationally wrongful acts, the fundamental basis is represented by the breach of an obligation that rests with a state towards another state, group of states or towards the international community in its entirety. In the other case, of the liability for legal activities, the fundamental basis is the risk involved by an activity which may cause harm to another state or to a zone being under its jurisdiction or control. Therefore, in case of liability for activities not forbidden by international law, we have in view the accomplishment of legal activities, that are necessary for the economical and social development of a state, but due to their nature, to the used technology, they present a risk to cause harms to other states.

With regard to the two types of liability, for internationally wrongful acts and for harmful consequences resulting from activities that are not forbidden by international law, these may be simultaneously engaged, meaning that a state, on whose territory, under whose jurisdiction or control there is developed an activity that presents the risk to cause a transboundary harm, that breaches its due diligence obligation, may be sued, both for the compensation of the harms caused on the grounds of the liability for legal activities, and on the grounds of the responsibility for internationally wrongful acts, as well.

3 Objective Liability

Although these international documents, regulating an objective liability, do not provide in an express way the condition of the existence of any fault for the involvement of international liability for the harms caused to the environment, according to the traditional conception this liability is subjective, fault representing one of the conditions of its involvement.

There are, yet, fields where liability is objective, having as a specific ground the idea of risk and of guarantee, or even absolute. Therefore, in the present international law, the objective liability of states in the environmental law applies and is the rule, being regulated in a conventional way, mainly, in three fields:

3.1 Objective Liability in the Field of Nuclear Energy

International conventions provide the existence of an objective liability for the exploitation of nuclear installations. For example, the Paris Convention on Civil Liability in the Field of Nuclear Energy as of July 29, 1960, adopted under the auspices of the Nuclear Energy Agency of the Organization for Economic Cooperation and Development, the Vienna Convention on Civil Liability for Nuclear Damages as of May 21, 1963, adopted under the auspices of the International Atomic Energy Agency, as subsequently amended and completed and the Bruxelles Convention on Civil Liability in the Field of Maritime Carriage of Nuclear Materials as of December 17, 1971, adopted under the auspices of the International Atomic Energy Agency, of the Organization for Economic Cooperation and Development and of the International Maritime Organization.

The Vienna Convention on Civil Liability for Nuclear Damages provides, expressly, the operator’s objective liability.

Liability is channeled exclusively to the operators of the nuclear installations. This is the person designated or acknowledged, from the beginning, by the national authorities, as being the liable person for any accident occurred to an installation or during the transport to or from that installation. That person is also liable for the accidents that take place during the nuclear material transport. Therefore, the initial purpose of the Bruxelles Convention of 1971, is to exonerate any person that transports nuclear material that could be held liable on the grounds of an international convention in the field of maritime transport, of the liability for nuclear damages in cases the operator of a nuclear installation is held to be liable according to the Paris and Vienna Conventions.

Initially, art.I para.1 let.k of the Vienna Convention defined the notion of damage without also including the damages caused to environment. Subsequently, the nuclear accidents from Three Mile Island and Chernobyl raised the problem of damages caused to environment, of economic losses and of the cost of preventive measures and of the possibility of their compensation. Therefore, by the Protocol amending the Vienna Convention, adopted in Vienna on September 12, 1997, there were included in the notion of „nuclear damage“ the cost to recover the environment, economic losses, loss of profit, cost of preventive measures and any other economic loss, but only to the extent these are determined by the law of the competent court of law.
The operator can be exonerated from liability, wholly or partially, if he (she) proves that:

i) the nuclear damage resulted, in totality or in part, from a serious negligence of the person that suffered it or that this person acted or omitted to act, with the intention to cause a damage;

ii) the nuclear damage caused by a nuclear accident resulting directly from acts of armed conflict, hostilities, civil war or insurrection;

iii) the damage is caused by a nuclear accident resulting directly from a nuclear cataclysm with exceptional character.

### 3.2 Objective Liability in the Field of Maritime Activities

In case of marine oil pollution, the international conventions provide an objective liability of the owners of ships that carry oil in bulk as cargo or of those who exploit marine drilling installations. For example, the *International Convention on Civil Liability for Oil Pollution Damage* as of 1969 and the *International Convention on the Establishment of an International Fund for Compensation for Oil Pollution Damage (FUND)*, as of 1971. In 1992, upon their amendment by 2 Protocols, they are named the *Convention of 1992 on Civil Liability* and the *Convention of 1992 on Fund creation*.

The Convention of 1992 on Civil Liability governs the liability of ship owners for the damages produced by oil pollution. This establishes the principle of their objective liability and requires an insurance system – mandatory liability. The owner of a ship has the right to limit its liability to an amount depending on the capacity of its ship. This Convention applies to all seagoing vessels actually carrying oil in bulk as cargo, but only ships carrying more than 2,000 tons of oil are required to maintain insurance in respect of oil pollution damage.

The Convention of 1992 on Fund creation [3], which completes the Convention of 1992 on Civil Liability, establishes a regime of indemnification of the affected victims when the indemnification provided by the applicable Convention on civil liability is insufficient. The International Fund for Compensation of 1992 for the damages caused by Oil Pollution (FIPOL 1992) was created following the Convention of 1992 on Fund creation.

According to art.III para.1 of the Convention of 1992 on Civil Liability, a ship owner has an objective liability for the damages produced by oil pollution, at the time of an incident, or, where the incident consists of a series of occurrences, at the time of the first such occurrence, shall be liable for any pollution damage caused by the ship as a result of the incident. He is not exonerated from the liability unless he proves that the damage by pollution:

i) resulted from an act of war, hostilities, civil war, insurrection or a natural phenomenon of an exceptional, inevitable and irresistible character; or

ii) was wholly caused by an act or omission done with intent to cause damage by a third party; or

iii) was wholly caused by the negligence or other wrongful act of any Government or other authority responsible for the maintenance of lights or other navigational aids in the exercise of that function.

If the owner proves that the pollution damage resulted wholly or partially either from an act or omission done with intent to cause damage by the person who suffered the damage or from the negligence of that person, the owner may be exonerated wholly or partially from his liability to such person.

### 3.3 Objective Liability in the Field of Outer Space Activities

In the field of outer space activities of states or of international organizations, for example, the *Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies*, as of January 27, 1967, hereinafter the “Treaty on Outer Space”; The *Convention on International Liability for Damages Caused by Space Objects*, as of March 29, 1972. Therefore, in the outer space law, liability for the damages caused by space objects is grounded on the risk theory, considering that the activity of states to launch space objects has a legal character *per se*.

The objective liability has a narrower application field for the moment, but there are important opinion trends on world level that incline towards the acceptance of a more extended applicability for this form to engage states’ responsibility. Given the increasing frequency of ecological accidents and of the acts of environmental degradation, we consider that it would be recommended to simplify the involvement of the liability of that who caused an environmental damage, on the purpose to limit as much as possible the events that have as a consequence a hardly reversible degradation of it. In this respect, for the involvement of the material liability of a state there would be sufficient only the existence of a major material damage produced by this, or by an individual or legal entity legally related to it.
4 Absolute Liability
Absolute liability, expressly regulated at international level, although, in essence, this does not suppose the existence of any fault, being also an objective liability, is distinct from the latter by the impossibility to invoke any exoneration reason. The treaty that includes it as a standard is the Convention on International Liability for Damages Caused by Space Objects, as of March 29, 1972, but in relationship with the damages caused on the Earth surface or to aircraft in flight. Therefore, according to art.2 of this Convention a State which launches or procures the launching of a space object, or a State from whose territory or facility a space object is launched, shall be absolutely liable to pay compensation for damage caused by its space object on the surface of the Earth or to aircraft in flight. These provisions were applied effectively in 1979, when the Soviet satellite “Cosmos 954” disintegrated in the Canadian atmospheric space, radioactive remainders of it being spread on the territory of Canada. The Canadian government requested, back then, to the U.S.S.R., the reimbursement of the expenses performed for the localization and recovery of the elements that could be hazardous, without invoking, in this approach, the effective fault of the former Soviet Union and succeeded in obtaining a part of the requested amount. On April 02, 1981, following the negotiations between Canada and the Soviet Union, there was concluded an agreement between the two parties, by which Moscow accepted to pay the amount of 3 million Canadian dollars.

5 Conclusions
Although there are several international documents and legal instruments that regulate the legal status of international liability for damages caused to the environment, we cannot say that there is a general, unitary and coherent international legal status applicable to the environment harms. In this respect, de lege ferenda, there is imposed the adoption of an international legal instrument that would regulate a clear general status of environment liability applicable at international level. This international legal instrument may have as its starting point the three drafts for the encoding of international liability, elaborated by ILC, the Lugano Convention, other international conventions and even the Strasbourg Convention on environmental protection by criminal law as of November 04, 1998. Towards the objective or subjective character of the international liability for the damages caused to the environment, we propose de lege ferenda the generalization of the collective objective liability for environment at international level, each time there is a present harm or an imminent threat with a harm, being established a limited number of causes that would exonerate from liability.

In this respect, considering irreversibility, and the devastating effects of some damages caused to the environment, we formulate another proposal of de lege ferenda, respectively that exoneration from liability would be expressly limited only to the following three cases: (i) an armed conflict, hostilities, a civil war or an insurrection; (ii) a natural phenomenon of an exceptional, inevitable and irresistible character; (iii) the act of a third person for which the active subject is not held liable, unless the latter took all the corresponding preventive measures that imposed in that case, but nevertheless the damage or imminent threat of its occurrence comes.

Into an integrated approach of the existing environmental problems at international level, it is necessary to analyze the intrinsic relationship between the prevention mechanisms and the liability mechanism. If in the case of the prevention mechanisms we talk about the techniques and instruments that regulate the economic activities to prevent the occurrence of a harm, in the case of the liability mechanisms we refer to the measures taken in case of an imminent threat with a harm or after the occurrence of an accident or incident that may produce a harm, in order to avoid its production or to diminish its effects. From the performed analysis results that there cannot be supported the precedence of a mechanism against the other, each of them fulfilling a well determined part in the environment protection activity.

According to the French professors Geneviève Viney and Philippe Kourilski “the precaution principle is the attitude that any person must adopt that makes a decision on an activity about which we may reasonably suppose to present a serious hazard for the health of the present and future generations or for the environment. These persons, especially the public authorities, must give priority to health and security imperatives on economic freedoms... and to reduce the risk to an acceptable risk for a bearable economic cost” [4].

The fundamental orientation of the precaution principle is at first sight enough simple and direct. If an activity threatens the environment or human health, precautions are imposed, even if this threat is not evidenced from the scientific point of view.
To a more close analysis, the principle is difficult to understand, not offering but a basis which must be improved in the national and international legal systems.

The Earth Summit that took place at Rio de Janeiro, in 1992, must have given the occasion to UN to mark a new stage in the development of the environment law and a global concern for the future of the planet. It could have been the occasion to unify the definitions of the precaution principle that, in whole, were not provided but within non binding instruments from the legal point of view and to precise the senses of the principle to make it operational, not just for the fight against pollutions, to which it was restricted, but more generally, for the entire environmental protection. But the reluctance, especially of the USA, that saw in the precaution principle a brake for the technological innovation, did not allow this thing. However, there cannot be retained a prevention of any scientific innovation, by the application of the precaution principle. Professors Kourilsky and Viney remind that „there is no a priori opposition between precaution and technological progress. The precaution principle invites to a reflection upon the conditions in which this progress is performed rather than to an inhibition of any innovation.”

The best solution to respond to the present problems, with regard to the reduction and, finally, elimination of the negative impact of human activities upon the environment consists in our appreciation in the combination of techniques and instruments with preventive and precautionary character. It is necessary to grant all the support to the process of research and innovation, to find the best methods, means, mechanisms, techniques and instruments to approach the environment protection problem.

Referring to the difference between the prevention principle and that of precaution, this is based first of all on the risk knowledge degree. The risk concept within the prevention principle is reported to the risks whose cause-effect relationship is known, unlike the common risks to the precaution principle, that are unknown.

Prevention involves both the risk evaluation to avoid the hazards, and the actions based on the knowledge of the present situation, to prevent environmental degradation.

This principle supposes actions against the causes that produce pollution or degradation and activities to limit the destructive or noxious effects for the environmental factors [5].

In international law, some authors [6] have established that the moment of the prevention principle occurrence is identified with the sentence of the Arbitral Tribunal as of March 11, 1941, in Trail Smelter Case [7]. The Tribunal held that the decisions of the Supreme Court of the United States (taken as a guide in the field of international law in so far as they had dealt with controversies between the various federal states of the US) provided an adequate basis for holding that under the principles of international law, as well as the law of the United States, no State has the right to use or permit the use of its territory in such a manner as to cause injury by fumes in or to the territory of another or the properties or persons therein, when the case is of serious consequence and the injury is established by clear and convincing evidence. In this case, the Tribunal found therefore that Canada was responsible in international law for the conduct of the Trail Smelter. It had a duty to see to it that conduct was in accordance with Canada’s obligations under international law. Accordingly the Trail Smelter would be required to refrain from causing any damage through fumes in the US. The Tribunal decided, finally, that, to prevent future damage, a regime of control, which it stipulated, would be applied to the operations of the Smelter.

In this respect we propose de lege ferenda – the inclusion of the observance and application of the precaution and prevention principles as obligations in the charge of all the international law subjects, in order to completely eliminate and minimize both the unknown, and the known risks, considering the effects of the environmental liability, put into effect in the preventive and remedying measures, regulated by the international legal instruments with general character. Therefore, we have in view the replacement of the polluter pays principle, with the precaution and prevention principles that correspond to an objective environment liability.

As regards the active liability subject, the state has a residual liability, the obligation of harm repair belonging in first instance to the operator. The state has the obligation to take the necessary measures to ensure a prompt and adequate compensation for the victims of transboundary damages caused by the hazardous activities performed on its territory or under its jurisdiction or control. This principle responds and reflects at the same time the request and the consensus of the international community that the state of origin takes all the measures so that adequate compensation measures in case of damage would be available as a part of the mechanisms that regulate hazardous activities performed under its jurisdiction or control.

In this respect, a joint and several liability reflects the best the particularities of the liability for
damages caused to the environment, leading eventually to a better protection of it considering the following reasons:

i) it is very difficult to establish the causal link between the actions of different operators or the incidents that took place and the environmental harm, moreover when this evidence is in the charge of the passive subject;

ii) often the effects of the operators’ actions or incidents are cumulated being possible to determine either an increase of the caused harm, or the occurrence of connected harms. In this case, it is difficult to establish to what extent each operator bears the liability for the cumulated effects or for the connected damages.

iii) It is possible that the damage would produce at far distances or after a certain time, that makes it almost impossible to precisely determine the part of liability that comes to the different operators.

With respect to the consequences of the environmental liability for transboundary harms, we can conclude that the preventive and the repairing measures were recently acknowledged as elements of harm, their cost being included within the notion of harm, following to be borne by the active subject of liability or paid as a part of the request for compensations further to the occurrence of a transboundary harm.

Failure of the State to comply with its legal obligations to take appropriate measures to prevent harm by non-State actors within its jurisdiction is often the basis of claims against States under the law of State responsibility.

Finally, we express our conviction that besides the already announced de lege ferenda proposals, there also imposes the regulation by imperative norms of a general regime of liability for the harms caused to the common heritage of mankind, by a liability without the existence of a harm and of fault, following the breach of the precaution principle by any international law subject. Also, we consider as necessary the creation of an authority that would act in the mankind’s name, in the case of precaution obligation breach.

In November 1967, ambassador Arvid Prado [8], a permanent representative of Malta to the United Nations, urged the delegates to consider the oceans resources outside the national jurisdictions as being the common heritage of mankind. As already provided in this declaration, the idea of heritage of mankind exceeds the notion of res communis. If a space or anything else is considered as such, it cannot be appropriated. It is comprised in the idea of res communis, but it belongs to mankind and, as such, it must be managed and controlled by humanity. Officially, in international law, the concept of common heritage of mankind was formulated during the United Nations Conference on Sea Law, at the beginning of the 1970s. Subsequently, there were officially proclaimed, as being a part of the common heritage of mankind, the Moon and its natural resources – Agreement Governing the Activities of States on the Moon and Other Celestial Bodies, as of December 05, 1979 – and the mineral resources of marine beds – United Nations Convention of Montego Bay on the Law of the Sea as of December 10, 1982.

References:

[1] It has not entered into force yet, considering that until present no state has ratified it yet, although it was signed by 9 states: Cyprus, Finland, Greece, Iceland, Italy, Liechtenstein, Luxemburg, the Netherlands and Portugal.


Measuring and Improvement the Productivity by PMCI Model in Fuzzy Environment (Case Study)

GHASSEM HAJ EBRAHIMI, M.ALI SOBHANALLAH, A.MIRZAZADEH
Department of Industrial Engineering
Tarbiat Moallem, Tehran, Iran
ghassemhajebrahimi@googlemail.com

Abstract: -The measurement is the main part and as a point of some person is the beginning point of the scientific process in the productivity management. If we want to show exploitation in organization culture, the main condition will be tooled for control and observation for progress, providing feedback, determining measurable goals and assessment of the operation of organization PMCI is productivity approach, which includes the process of the feedback system, improvement of productivity finally determining the goals of organization by the unique measurable system. This paper uses fuzzy systems, by considering the structure of measurement system and also uncertainty in efficiency effect and effectiveness effect that previously they were used brain storm methods and AHP.

Key-Words: productivity; pmci; fuzzy system; expert system

1 Introduction
The productivity is a meaning which is considered by experts of economics and politics and increasing it is been necessary for improving the level of life and making more welfare society that is national goal for all the worlds. By improvement knowledge of human in economic and social problems, the mining of productivity is changed from the point of shape and content and continuously is made new sides. The productivity is used by depending on the kind of persons and in connection to their experience and knowledge. The main duty of management, the earning of productivity is more and this mater is considered by the manager for management and reengineering of organization.

There are several definitions for productivity; however, each of them can be correct and skilful. But, if it can't express the main and basic meaning as a basis of culture, consciousness and even customs and belief of the different person, it will be imperfect. It's not doubt that whatever the groups have more culture and knowledge, we need deeper definition. The editing of successful method for improvement productivity is needed to arrangement systemic way for measuring of productivity a macro level and in micro level, too. For evaluation of productivity is used various measures, and their measures are different from one organization to another and from one industry to another.

2 Problem Formulation
The PMCI model is one of the productivity measurement techniques in organization which the process of this system is shown below.

1. Organization objects
2. Measurement system
3. Feedback system
4. Increasing of Productivity
5. Organization objects

Its clear the process begins by determining the goals of organization, in the basic of the goals is determined the measurement system the compatible with them, then the data of it report to members of organization officially and regularly, this feed back is the main discussion about the quality of the improvement of productivity.

In according to goals if productivity increases the goals of organization will be fulfilment. It’s the
basis of process that the process of event is from down to up. So after approval, the project is planned by senior manager and creating the productivity measurement system begin from down and ends to up hierarchy. It's obviously that the drawing team include the manager of meetings, the maker of facilities, the agents of fart and they plan the measurement system. When they agree on system, they offer the results to senior manager for discussion and final approval. If the plan is approval, it will be got ready to execution and providing feed back.

The process of the establishment of system, the processes of the establishment of this system are:
1) The determining of important exits
2) The determining the indictor of exits,
3) The contingency relationship (the determining efficiency)
4) The performance and providing feed back report.

2.1 The first level, the determining the important exits with necessary particular
Each unit has served goals and duties which we expect these duties do to fulfilment goals the name. The measurement of activities is productions or feed back in this system. Because the productivity of organization is depends on quality of producing these feedbacks, the first phase is the determining of outputs whit necessary particular. The designing team must provide a list important outputs whit the necessary particular.

2.2 The second level the determining of related particulars
When, the outputs whit the necessary particulars is determined, then there are determined the related particulars of outputs. One particular is a criterion which shows how it produces the output unit. The members of team discuss together for determining the related particulars and determine several particular.

2.3 The third level: the contingency relationship
(the determining efficiency) the other stages is the establishment of contingency relationship after approval outputs and particulars. The contingency is relationship between the value of particular and the efficiency of it. This relationship says that how much the different quantity are owned share in total function of unit. This relationship has showed in plan coordinate that horizontal axis shows different quantities of particulars, and vertical axis is the share of efficiency in the function of unit. In horizontal axis is recorded the amount of particular from worst member to the best manner. The efficiency of them is determined in vertical axis. The changing of efficiency locates from 100 to -100.

These changing, have zero point that shows the natural and normal efficiency level, it's not good nor bad, this level is balanced.

When the worst and the best axis level are determined, then they determine the real function which the amount of particulars is related to efficiency. First, the zero point is determined, then the efficiency level of max and min of particulars is specified.

It's necessary to say that: firstly the whole trend of relation indicates the importance of related particular; the sharp trend made many changes in efficiency and a slow trend because little changes. Secondly, the relation is nonlinear and it's necessary for accuracy of relation of organization function. These characteristics are very important. Often in many measurement systems, productivity combines all of the particulars and earns one number. These means all of the function has some value. It's clear, this number don't reflect the reality of organization. Different functions don't have the same importance.

2.4 The forth level: providing feedback reports.
Inserting the amount of particular in a formed feedback system is the last stage. In this part, firstly, is collected the data in a specific period of time. Then, the efficiency numbers is determined for each particular in the base of the contingency relationship. Then the total productivity is earned by adding the whole quantities of each output. Each of them has private meaning. Zero number means that unit estimates expectation. This productivity is neither good nor bad. If efficiency is positive, the unit function will go up more than our expectation. When it is negative, it will have lower function.

On of the advantage of it is the possibility of collecting numbers. In fact, different function is done by units and its name is the whole particular of
productivity. The contingency relationship which has reflected the importance of related particular is called as unit efficiency. This characteristic provides the possibility of collecting function to find the whole function.

3 The comparison of different units
Another characteristic is this system can comprise the units productivity together even those do different duties. This tasks place the possible number of whole efficiency each unit by base of maximum. If its percent is more, will obtain the best productivity. The productivity of different unit which do different function, comprise together by earning the percent of real function these max value. Each unit has the highest productivity, if it has max percent.

3.1 The combination of total units
Also, in this system, the function of whole port is earned by combination function of each unit. For example, if a unit has 5 team for earning productivity, the productivity of each unit is obtained and we add them together and earn the port productivity. This is not possible in some systems easily, because the productivity of each units measure whit different measures that adding them together is not possible. But in this system, each part was measured on efficiency in according to one scale and the amount of efficiency of each part causes the efficiency of organization and the productivity of it is earned.

4 The fuzzy systems
These systems are a knowledge-based or rules-based system. The rule based is the heart of system that contain of rules. A if-then rule is a if-then diction which many it's words are membership functions. The start point for made a fuzzy system is earning a set of fuzzy if-then rules earned by expert persons. The other stage is the combine of these rules. The fuzzy systems use different ways for combination of tem. Usually are used 3 kinds of systems:
1) Net fuzzy systems
2) TSK fuzzy systems
3) Fuzzification and difuzzification systems.

The main structure of net system shows below. The rule based shows the collection of fuzzy if-then rules. The fuzzy inference engine combines these rules of fuzzy set in internal environment to external environment in the base of fuzzy logical rules. TSK systems uses easy math relation instead using the rules that in net systems use descriptive expression whit linguistic values. In fact, TSK system is the average weight from numbers of ports. For using net fuzzy system one any way is adding one fuzzification at input and defuzzification at output. The result is showed below. This system covers the problems of net fuzzy system and TSK system. In this paper, aim is a system whit fuzzification and difuzzification.

4 The usage of fuzzy systems and fuzzy in PMCI model:

The productivity in hospital

<table>
<thead>
<tr>
<th>Item</th>
<th>Average fuzzy number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Management</td>
<td>(50.32,77.82,95.64)</td>
</tr>
<tr>
<td>2 Sentry and reception</td>
<td>(38.93,51.15,76.89)</td>
</tr>
<tr>
<td>3 Expert medico</td>
<td>(85.03,92.91,99.04)</td>
</tr>
<tr>
<td>4 Nurse</td>
<td>(71.96,85.94,95.66)</td>
</tr>
<tr>
<td>5 Para clinical services</td>
<td>(60.81,69.77,82.67)</td>
</tr>
<tr>
<td>6 Services and hygiene</td>
<td>(36.45,42.93,55.00)</td>
</tr>
<tr>
<td>7 Oficial sector</td>
<td>(30.81,39.23,50.43)</td>
</tr>
<tr>
<td>8 Drogstore</td>
<td>(33.95,59.30,84.25)</td>
</tr>
<tr>
<td>9 Medical equipment</td>
<td>(60.81,77.84,90.74)</td>
</tr>
<tr>
<td>10 Sport space</td>
<td>(28.55,44.18,49.39)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>extreme</th>
<th>middle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Management</td>
<td>11857</td>
<td>9656</td>
</tr>
<tr>
<td>2 Sentry and reception</td>
<td>9534</td>
<td>6343</td>
</tr>
<tr>
<td>3 Expert medico</td>
<td>12281</td>
<td>11522</td>
</tr>
<tr>
<td>4 Nurse</td>
<td>11862</td>
<td>10657</td>
</tr>
<tr>
<td>5 Para clinical services</td>
<td>10252</td>
<td>8652</td>
</tr>
<tr>
<td>6 Services and hygiene</td>
<td>6821</td>
<td>5324</td>
</tr>
<tr>
<td>7 Oficial sector</td>
<td>6254</td>
<td>4865</td>
</tr>
<tr>
<td>8 Drogstore</td>
<td>10534</td>
<td>7354</td>
</tr>
</tbody>
</table>
Table 1 data of the case study

This hospital's managers intend to improve the productivity of this hospital by PMCI model.

4.1 The first stage: determining the important outputs with the necessity particulars.

There is provided a questionnaire and ask form sick persons that ask 3 question for each case
1) From point of you, what is the measure of importance of item x? x₁
2) From point of you, what is the extreme measure of importance of item x? x₂
3) From point of you, what is the least measure of importance of item x? x₃

Which we have items in below
You consider the results as a (x₂, x₁, x₃) and then we obtain the result and introduce the important cone as important outputs.

The number of pattern is 124 persons. The result is showed below: We grasp that the doctor, nurses, tools, management and service have most important that its 70%. So, we can select them for PMCI model. We define the particular for these outputs now.

4.2 The second stage: determining the related particulars

<table>
<thead>
<tr>
<th>Item</th>
<th>least</th>
<th>defuzzification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Management</td>
<td>6240</td>
<td>75.42</td>
</tr>
<tr>
<td>Sentry and reception</td>
<td>4828</td>
<td>54.53</td>
</tr>
<tr>
<td>Expert medico</td>
<td>10544</td>
<td>92.48</td>
</tr>
<tr>
<td>Nurse</td>
<td>8923</td>
<td>84.88</td>
</tr>
<tr>
<td>Para clinical services</td>
<td>7541</td>
<td>70.77</td>
</tr>
<tr>
<td>Services and hygiene</td>
<td>4520</td>
<td>44.33</td>
</tr>
<tr>
<td>Oficial sector</td>
<td>3821</td>
<td>39.93</td>
</tr>
<tr>
<td>Drogstore</td>
<td>4211</td>
<td>56.38</td>
</tr>
<tr>
<td>Medical equipment</td>
<td>7541</td>
<td>76.81</td>
</tr>
<tr>
<td>Sport space</td>
<td>3541</td>
<td>41.57</td>
</tr>
</tbody>
</table>

Table 2 indexes of outputs

It is defined important output particular for each of the important outputs in according to defined particular.

4.3 The third stage: the contingency relationship

For earning it, is used fuzzy system for each important outputs. We said the fuzzy system include rule based and inference engine, fuzzification, difuzzification.

In this paper, there is the contingency relationship for tools of medical. The rule based includes:

- If efficiency is (65, 70, 75) then effectiveness is (-5, 0, 5)
- If efficiency is (70, 75, 80) then effectiveness is (5, 10, and 15)
- If efficiency is (75, 80, 85) then effectiveness is (15, 20, and 25)
- If efficiency is (80, 85, 90) then effectiveness is (25, 30, and 35)
- If efficiency is (85, 90, 95) then effectiveness is (35, 40, and 45)
- If efficiency is (90, 95, 100) then effectiveness is (45, 50, 55)
- If efficiency is (60, 65, 70) then effectiveness is (-10, 5, 0)
- If efficiency is (55, 60, 65) then effectiveness is (-15,10, -5)
- If efficiency is (50, 55, 60) then effectiveness is (-20, -15, -10)
If efficiency is (50, 55, 60) then effectiveness is (-25, -20, -15)

![Mamdani fuzzy system](image1)

It's certain that effectiveness doesn't have direct relationship with efficiency. So some times these have inverse relationship.

Now, by inserting these data, efficiency numbers for different manners can be obtained. As you see, this picture is the medium graphic of on fuzzy in matlab. We can obtain the total efficiency the numbers for different effectiveness by this fuzzy system that the result is below:

Design the fuzzy systems for all important outputs are necessary.

The forth stage: providing feedback reports.

As was said before, in this stage, firstly the information is collected in a specific period of time. Then in the base of the contingency relationship, the efficiency numbers is determined for each particular. Then total productivity is obtained by adding total productivity is obtained by adding total efficiency numbers of each output.

In according to calculated efficiency, the numbers which are needed for PMCI model are the related efficiency to medium and maximum effectiveness that is obtained them by using fuzzy systems from the contingency graph insert them to the productivity particular formula which will be point.

5. The productivity particular formula of whole organization

For earning it, we must multiply related efficiencies which average effectiveness of one by one of particular in importance of it and add them together, the total should divides the whole productivity with effectiveness maximum on its weight that is:

\[ W_i = \text{the measure of } i \text{ in the productivity} \]
\[ e_i = \text{the earning efficiency from the necessity table for } i \text{ in related to maximum effectiveness in a period of time} \]
\[ N = \text{the number of important factors in the productivity of whole of organization} \]

For this case study:

5. Conclusion

Because the measurement is necessary in improvement of productivity in each organization, so the way is important. PMCI technique is a model that has been used in several organizations, the fuzzy logic and fuzzy systems is used, because, they don't have certain outputs. In this paper, PMCI model at fuzzy environment is implied and can offer away that is useful in uncertain environment.

References
Study of soil erosion and sediment transport in small watersheds

E.B.M. BEILICCI, R.F. BEILICCI, C.A. GÎRBACIU
Faculty of Hydrotechnics
„Politehnica” University of Timisoara
G. Enescu Street, No. 1/A, postal code 300022
Romania
beilicci_erika@yahoo.com http://www.hidro.upt.ro

Abstract: Nowadays, the study of soil erosion and sediment transport due to overland flow in small watersheds, and of their own effects on the environment, is more and more important. The propagation area of soil erosion and its damage values are practically speaking hard to calculate. This paper presents a study method of soil erosion and resulted sediment transport in small catchments, a method in more steps, one of them being the soil erosion and sediment transport modelling with the use of WEPP program. On the basis of the obtained results, we can plan land management which consider the present situation and the perspectives of an area and harmonizes the interests of all its determining factors (the human factor, the environment, the economic factor etc.). At the same time, we can establish the economic and technical efficiency of different soil erosion and sediment transport control measures and works (agrotechnic, forest, hydrotechnical measures; structural and non – structural measures), necessary for the abatement of negative effects of soil erosion and sediment transport on environment.

Key words: Romania—Hydrographical basin—Soil erosion—Overland flow—Sediment transport—Modelling.

1 Introduction
The physical phenomena of soil erosion caused by water follow the next stages: the stage of solid particle dislocation and entrapment, under the action of rain spots and hydrodynamic forces (from hillslopes and from river beds), the stage of actual transport of sediments caused by water movement, and the sedimentation stage. In time, soil erosion and sediment transport brings changes and decreases in soil physical and chemical properties, favouring the occurrence of land movements. Also it produces clog of storage reservoirs, general erosion of the river bed downstream of reservoirs, ecological problems due to depreciation of the original qualities of the aquatic ecosystem etc.

The determinant factor of runoff and water erosion on the inclined land surfaces is represented by kinetic rain energy, directly related to the rain intensity. Besides this factor, an important role in the development of land erosion is played by: the surface runoff characteristics (liquid flow, runoff depth, water velocity, superficial runoff type) and the land characteristics (infiltrations' properties, soil texture and structure, declivity, relief, exposure, vegetation, etc.). Soil particles are being dislocated and entrapped by the rain spots force and by the overland flow. For this reason, the study on solid runoff on hillslopes cannot be separated from the study on liquid runoff, the liquid runoff determining the solid runoff.

The study of mechanisms and effects of soil erosion on the environment caused by water is necessary in two distinct situations:
- in the project stage of engineering works which are going to be executed on the hillslopes or the river bed water courses, intending to satisfy the economic necessities and to protect the environment against the negative effects of erosion processes. In this stage, we used mathematical models of these processes characteristics.
- in the operation stage of these works, the operation mode being based on the prognoses models of erosion processes. These will have to be exploited in an optimal way in order to satisfy, for a longer period of time, the economic and the environmental necessities.

Soil erosion and sediment transport (caused by overland flow) are a natural phenomenon’s, parts of the geological processes and climatic changes. These processes have increased during the last decades with the growth of the world population and its food needs, involving a larger land surfaces for agriculture purposes and an intensive practice of agriculture and animal husbandry. At the present time, over 2000 million hectares of land are affected by moderate and excessive erosion, over 55 % of these being caused by water erosion and almost 33 % by wind erosion. Each year, about 5 – 7 million hectares are lost due to
different kinds of erosion, and approximately 25 000 million tones of soil are lost due to soil erosion by overland flow. Soil degradation incurs a life-expectancy risk to approximately 1000 million people. In Romania, soil erosion caused by overland flow affects 6.3 million hectares, from which 2.3 million benefits from erosion control works (at present, many works are degraded). Soil erosion and landslides (present of about 0.7 million hectares) cause approximately 41.5 t/ha.year soil losses. Out of the entire agricultural land, we lose each year about 150 \cdot 10^6 tons of soil, which contain about 1.5 \cdot 10^6 tons of humus and 0.45 \cdot 10^6 tons of N, P, K. The direct economic damage is decreasing by 20 % each year in the agricultural production. [2]

The total soil losses caused by different kind of erosion (tons/ha.year), and the areas which are more affected by soil erosion from Romania are shown in figure 1 and figure 2 (after Motoc, 1983). [3] [4]

![Fig.1](image1)

**Fig.1**
Total soil loss due to the different kinds of erosion in Romania.

![Fig.2](image2)

**Fig.2**
Soil erosion affected areas in Romania.

The most affected areas by soil erosion are located on Moldova Plateau and Meridionali Carpathen Mountains, while the least affected are in the West Plain and in the south of Muntenia (The Baragan plain).

Due to the growth of world’s population and its food needs, floods, which are more and more frequent and affect many regions of the world, are causing significant material damage and human life losses. There is a need to elaborate thorough studies on the causes and the characteristics of sediment transport and on the measures aimed to reduce their negative effects.

The problems that need to be solved are the following:

- the control of sediment transport, especially in regions of high climatic aggression and erosion risk;
- the control of erosion and other hillslope degradation processes, especially where there is a desire to set up vine and tree plantations and on arable land with irrigation systems;
- the arrangement of the torrential hydrographic network and reshaping of small valleys;
- the management of inclined lands;
- the agricultural assurance use of degraded lands and improvement of their fertility;
- the elimination of water in excess from hillslopes and narrow alluvial plains;
- the limitation of landslides, the arrangement and differential capitalization of those affected by active landslides;
- the protection of reservoirs against silting;
- the embellishment of landscapes in areas with important erosion focuses.

### 2 Problem Formulation

For areas with a high risk occurrence of sediment transport processes with important potential negative effects, we need to carry out researches and studies that allow us to know which are the causes, the evolution and the effects of soil erosion and sediment transport on the environment.

Such a research supposes the elaboration or the updating of topographic, geomorphologic, climatic, soil (concerning the erosion and other degradation processes), hydrological, hydrogeological, geotechnical, vegetation and social – economic studies.

Based on these studies we can elaborate models for prognosis of soil erosion and sediment transport evolution and effects. The stages of such a model might be:

1. The carrying out the studies mentioned above.
2. The establishment of an initial stage of the hydrographical basins’ status, on which we shall report the later results.
3. The modelling of hillslope erosion processes with one of the forecasting programs currently existing in the world (WEPP, RUSLE, CREAMS etc.) with different land management hypotheses. The model will be chosen depending on the capabilities and the application possibilities of each specific situation.

4. The overland flow and sediment transport modelling from the hydrographical basins (Duflow, SMS, HEC etc.).

5. The interpretation of results and the determination of the necessary measures aimed to reduce the overland flow and sediment transport, respectively to reduce their negative effects on the environment.

2.1 Studies Required by this Method
The studies topographic, geomorphologic, climatic, soil, hydrological, hydrogeological, geotechnical, vegetation and social – economic are surveys. They have to be carried out by specialized institutions, and based on their results it needs to be established an initial stage by a multidisciplinary team. In order to exemplify the application of suggested method, we chose WEPP program for soil erosion and sediment transport modelling in the considered hydrographical basins, and DUFLOW program for the overland flow in the hydrographical basins.

2.2 WEPP (Water Erosion Prediction Project) Model Description
WEPP (Water Erosion Prediction Project) is software for the prediction of watershed soil erosion, developed by the USDA Forest Service, the Agricultural Research Service, the Natural Resources Conservation Service, the US Department of the Interior Bureau of Land Management and Geological Survey, frequently used in many countries. WEPP model may be used in both hillslope and watershed applications. The model is a distributed parameter, continuous simulation, and erosion prediction model, implemented as a set of computer programs for personal computers. The hillslope component of the WEPP erosion model requires a minimum of four input data files to run: climate file, slope file, soil file and plant/management file. The watershed component requires a minimum of seven input data files: each hillslope information file, structure file, slope file, soil file, management file, climate file and channel file. WEPP considers that the hillslope is divided in numerous parallel rills; the surface erosion occurs on interrill surfaces and the dislocated soil particles are transported downhill by rill flow (rill erosion is also considered in calculus). WEPP produces different kind of outputs, in various quantities, depend the user’s needs. The most basic output contain: the runoff and erosion summary information, which may be produced by a storm, on monthly, annual or average annual basis. The time – integrated estimates runoff, erosion, sediment delivery and sediment enrichment are contained in output, as well as the spatial distribution of erosion on the hillslope (figure 3). [5] [6]

2.3 DUFLOW Model Description
It is a one-dimensional program for quantitative and qualitative modelling of overland flow in open runways. It was developed by the International Institute for Hydraulic and Environmental Engineering (IHE) Delft, The Rijkswaterstaat (Public Works Department), the Tidal Water Division, The Hague, The Delft University of Technology, Holland. [7] DUFLOW is designed to cover a large range of applications, such as propagation of tidal waves in estuaries, flood waves in rivers, operation of irrigation and drainage systems, etc. Basically, free flow in open channel systems is simulated, where control structures like weirs, pumps, culverts and siphons can be included. As in many water management problems, the runoff from catchments areas is important; a simple precipitation-runoff relation is part of the
model set-up in DUFLow. The DUFLow software consists of the following parts: DUFLow water quantity (with this program one can perform unsteady flow computations in networks of open water courses) and DUFLow water quality (this program is useful in simulating the transportation of substances in free surface flow and can simulate more complex water quality processes).

DUFLow is based on the one-dimensional partial differential equation that describes non-stationary flow in open channels [8]. These equations, which are the mathematical translation of the laws of conservation of mass and of momentum, read:

\[ \frac{\partial Q}{\partial t} + \frac{\partial Q}{\partial x} = 0 \] (1)

\[ \frac{\partial Q}{\partial t} + gA \frac{\partial H}{\partial x} + \frac{\partial (\alpha Qv)}{\partial x} + \alpha \frac{\partial Q}{\partial x} = B \gamma w^2 \cos(\Phi - \phi) \] (2)

while the relation:

\[ Q = v \cdot A \] (3)

holds and where:

- \( t \) time (s);
- \( x \) distance as measured along the channel axis (m);
- \( H(x, t) \) water level with respect to reference level (m);
- \( v(x, t) \) mean velocity (averaged over the cross-sectional area) (m s\(^{-1}\));
- \( Q(x, t) \) discharge at location \( x \) and at time \( t \) (m\(^3\) s\(^{-1}\));
- \( R(x, H) \) hydraulic radius of cross-section (m);
- \( A(x, H) \) cross-sectional flow area (m\(^2\));
- \( b(x, H) \) cross-sectional flow width (m);
- \( B(x, H) \) cross-sectional storage area (m\(^2\));
- \( g \) acceleration due to gravity (m s\(^{-2}\));
- \( C(x, H) \) coefficient of De Chézy (m\(^{1/2}\) s\(^{-1}\));
- \( w(t) \) wind velocity (m s\(^{-1}\));
- \( \Phi(t) \) wind direction (degrees);
- \( \gamma(x) \) wind conversion coefficient (-);
- \( \alpha \) correction factor for non-uniformity of the velocity distribution in the advection term, defined as:

\[ \alpha = \frac{A}{Q^2} \int (\gamma y z)^2 dydz \] (4)

where the integral is taken over the cross-section \( A \) (m\(^2\)).

Equations (1) and (2) are discretized in space and time using the four-point implicit Preissmann scheme.

For a unique solution of the set of equations additional conditions have to be specified at the physical boundaries of the network and at the sections defined as hydraulic structures. The user-defined conditions at the physical boundaries may be specified as levels, discharges or a relation between both, for instance a (tidal) elevation \( H \), a discharge \( Q \), or a so-called \( QH \) relation. At internal junctions the (implicit) condition states that the water level is continuous over such a junction node, and that the flows towards the junction are in balance since continuity requires:

\[ \sum_{j=1}^{n} Q_{ji} + q_i = 0 \] (5)

where:

- \( i \) indication for the junction node;
- \( Q_{ji} \) discharge from node \( j \) to node \( i \) (m\(^3\) s\(^{-1}\));
- \( q_i \) additional or lateral flow to node \( i \) (m\(^3\) s\(^{-1}\)).

The above equations are solved at each time step. They are transformed into a system of (linear) equations for the water levels. After the water levels are computed using a standard solution method for systems of linear equations, the discharges are found by substituting the computed water levels into equations.

Equation (5) is not used in nodes where a water level is prescribed as boundary condition. In such a node no equation is needed because the water level is already known. Discharge boundary conditions are taken into account as the additional flow \( q_i \).

To start the computations, initial values for \( H \) and \( Q \) are required. These initial values must be provided by the user; they may be historical measurements, obtained from former computations or just a first reasonable guess. Additionally wind stress and rainfall conditions can be specified.

The application of this model supposes a plan of the study area for of the river network division and the hydrographical basin. The river network is divided into sectors of different lengths by nodes, in such a way that the linear sectors between two consecutive nodes, following the axe curves of the river bed. In each node, we need to show the bed level and the width of the water mirror on different levels. The area of hydrographical basin is delimited by the highest slope line, and, subsequently, successively, the associated flow areas which will connect in nodes.

The results from simulation with DUFLow are: the variation of water levels and water discharges in each node of network.

### 3 Problem Solution

The proposed method was applied to the hydrographical basin of Manastur, Timis County, Romania.

The studied territory – Manastur hydrographical basin – is situated in the depression between Zarand and Poiana Rusca Mountains, at the crossover line between Lipova Plateau and Occidental Field (figure 4). To model the solid runoff phenomena to it was chosen the Topla river sub-basin with a length of 11.56 km, a surface of 2357.40 ha; which contains 56 sectors of secondary valleys with a total length of 31.58 km; the area plan with the separation on sub-basins corresponding to each valley is represented in figure 5.
We have considered the following hypotheses [9]:

- the rain intensity constant on the entire surface of the hydrographical sub-basin
- the same soil type for the entire hydrographical sub-basin - Typic Hapludalfs (with more subtypes), with a middle and heavy texture (reduced permeability), being part of hydrological group C, with an initial humidity of 70% (table 1), according with the study carried out by OSPA [10] [11]
- soil use is the same all over the sub-basin surface
- there are no works for soil erosion and sediment transport control.

Table 1 Soil characteristics.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Depth (in)</th>
<th>Sand (%)</th>
<th>Clay (%)</th>
<th>Organic (%)</th>
<th>CEC (meq/100 g)</th>
<th>Rock (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.906</td>
<td>37.2</td>
<td>21.2</td>
<td>1.800</td>
<td>11.4</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>14.170</td>
<td>38.1</td>
<td>20.9</td>
<td>1.670</td>
<td>10.9</td>
<td>0.0</td>
</tr>
<tr>
<td>3</td>
<td>26.770</td>
<td>31.0</td>
<td>30.5</td>
<td>0.800</td>
<td>14.4</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>34.250</td>
<td>28.8</td>
<td>36.2</td>
<td>0.001</td>
<td>19.3</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>40.550</td>
<td>29.3</td>
<td>32.0</td>
<td>0.001</td>
<td>18.6</td>
<td>0.0</td>
</tr>
<tr>
<td>6</td>
<td>51.180</td>
<td>34.4</td>
<td>31.3</td>
<td>0.001</td>
<td>18.3</td>
<td>0.0</td>
</tr>
</tbody>
</table>

WEPP model was applied for each sectors of river bed (ditch), on the corresponding hydrographical sub-basin, and calculating the soil loses on each hillslope during the torrential rain, in four hypotheses (which have resulted from the torrential rain distribution charts with the 10% assurance and having a duration of 5, 15, 30, 60 minutes, elaborated on the basis of measurements and statistical processing of the meteorological data on the Bega hydrographical basin area (table 2). [12]

For the application of the DUFLOW model, the river network was divided into sectors of 74 nodes. In all the 74 nodes, there were operated cross sections, introducing the width of the water mirror on different levels (from 0.2 to 0.2 meters). In each node, there was introduced the water flow resulted from the overland flow on slopes (after establishing the area of the hydrographical basin of each node and its runoff coefficients).

Table 2 Characteristics of torrential rains.

<table>
<thead>
<tr>
<th>Storm Amount (mm)</th>
<th>Storm duration (hr)</th>
<th>Max intensity (mm hr^2)</th>
<th>% Duration to Peak Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.00</td>
<td>0.08</td>
<td>189.375</td>
<td>20</td>
</tr>
<tr>
<td>23.30</td>
<td>0.25</td>
<td>94.132</td>
<td>20</td>
</tr>
<tr>
<td>28.80</td>
<td>0.50</td>
<td>58.176</td>
<td>20</td>
</tr>
<tr>
<td>42.00</td>
<td>1.00</td>
<td>42.420</td>
<td>20</td>
</tr>
</tbody>
</table>

The runoff coefficients “k” is in accordance with the method taken over from the US Soil Conservation Service, where the soil is classified in one of the four hydrological groups (A, B, C, D).

The river network with the nodes can be seen in figure 6.

The simulation results are presented in the graphs from figure 7 to 10, and show the variation of the followings:

Fig.4 Studied hydrographical basin localization.

Fig.5 Area plan.

Fig.6 Hydrographical network nodes.
1. Specific soil losses on the entire hydrographical basin.
2. Surface runoff volume on the hillslopes.
3. Water volume and sediments, which go through the outlet section of the basin according to the rain duration.

The model simulation results show that we can establish relatively good correlations (2 order polynomial regression curve) between rain period and volumes, which were computed (soil losses, runoff volumes, water volumes which go out, sediment volumes). It was chose the 2 order polynomial regression because the R-squared (R is the regression coefficient) are better, in comparison with other regression types: exponential, logarithmic, linear, power or moving average. Also, can be observed that the difference between WEPP results and values of regression equation are significant in case of short rain duration, but are relative small in the case of long rain duration.

Fig.7 Soil losses variation on the entire hydrographical basin depending on rain duration.

<table>
<thead>
<tr>
<th>Rain duration (min)</th>
<th>5</th>
<th>15</th>
<th>30</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil loss (t/ha)</td>
<td>0.13</td>
<td>0.39</td>
<td>0.96</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Fig.8 Variation of runoff volume on whole hydrographical basin depending on rain duration.

<table>
<thead>
<tr>
<th>Rain duration (min)</th>
<th>5</th>
<th>15</th>
<th>30</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface runoff</td>
<td>2309.28</td>
<td>9959.01</td>
<td>16116.70</td>
<td>372444.70</td>
</tr>
</tbody>
</table>

Fig.9 Variation of water volume which go out from the hydrographical basin depending on rain duration.

<table>
<thead>
<tr>
<th>Rain duration (min)</th>
<th>5</th>
<th>15</th>
<th>30</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sediment volume</td>
<td>7.5</td>
<td>241.8</td>
<td>590.4</td>
<td>1742.9</td>
</tr>
</tbody>
</table>

Fig.10 Variation of sediment volume which go out from the hydrographical basin depending on rain duration.

WEPP program was run only in four hypotheses, considering on the entire studied surface a single type of soil and no soil erosion and sediment transport control measures.

The modelling results of DUPOWER program are flow hydrographs and water levels in different characteristic cross sections of river network. The maximum flow values and the time of their occurrence related to the rainfall start in node 18 (outgoing cross section) are shown in table 3.

The flood duration is around 12 hours in each of the four hypotheses. At the starting point of the simulation, we have considered that the water discharge in each node is zero (the riverbed is dry).

The water discharge values are not so high because the watershed is narrowed and occurs important delays in the flood wave propagation (we have considered the most unfavourable situation: the same rain intensity on all the watershed area and covering the entire watershed).
Table 3 DUFLOW model results.

<table>
<thead>
<tr>
<th>Rain duration and intensity</th>
<th>$Q_{\text{max}}$ (m$^3$s$^{-1}$)</th>
<th>T (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 min; 3 mm min$^{-1}$</td>
<td>5.30</td>
<td>8</td>
</tr>
<tr>
<td>15 min; 1.55 mm min$^{-1}$</td>
<td>6.20</td>
<td>7.8</td>
</tr>
<tr>
<td>30 min; 0.96 mm min$^{-1}$</td>
<td>6.90</td>
<td>8</td>
</tr>
<tr>
<td>60 min; 0.7 mm min$^{-1}$</td>
<td>7.10</td>
<td>9</td>
</tr>
</tbody>
</table>

The water level in node 18 (outlet sections) has no important variations, because the riverbed cross section is very wide.

This step of the modelling process serves in soil erosion and sediment transport control measures designing stage, for the calculation of the surface runoff volume and the soil losses, due to soil erosion.\[12\]

The next step contains: suggestions of different soil erosion and sediment transport control measures in the hydrographical sub-basins area, program simulation and computation of the surface runoff volume and soil losses. The final decision and placement of the efficient works of soil erosion and sediment transport control measures will be made according to economic calculations and cost/benefit analyses.

As a critical observation, WEPP and DUFLOW models can only be applied on small hydrographical basins (as shown in the precedent paragraphs), as the models can only use one file containing climatic characteristics, the same type of rain or rainfall on all the hydrographical basins (same intensity and period). For larger hydrographical basins, it is necessary to divide them into smaller basins with the same precipitation characteristics but different association times. In order to obtain flow hydrographs, level hydrographs and concentration variation of the suspension materials from the outgoing cross sections it is necessary to integrate the flow on all the hydrographical basins.\[13\]

WEPP model is also able to model hillslope erosion for more than one year. In order to do this, beside the topographic characteristics, we also need a good knowledge of climate area characteristics (annual medium temperature, annual precipitations, rainfall intensity, medium high of snow, wind direction etc.), and of land use prognoses during each year of the simulation.

In present time, in Romania exist only a single specialized institution for research in the field of soil erosion - Research and Development Centre for Soil Erosion – located in Perieni – Barlad, East Romania – Moldav region, Vaslui county. There exist experimentally plots for soil erosion phenomena research, but is not studied the resulted sediments transport. For presented model application, is study now the possibility of chose in Timis County in the Vest part of Romania, an one small hydrographical basin, that will be convert in an experimentally basin, equipped adequate for soil erosion and sediment transport study and measurements. Then, will be possible the comparison between model application results and data provided from measurements in situ. The next steps are the followings: WEPP model adaptation for Romanian conditions, model calibration, and model results validation.

4 Conclusions

On the ground of the obtained results, we can plan land management, which takes into consideration the present situation and the perspectives of the area, and harmonizes the interests of all the important involved factors (the human factor, the environment, the economic factor etc.). At the same time, we can establish economic and technical efficiency soil erosion and sediment transport control measures and works (agrotechnical, forest, hydrotechnical measures; structural and non – structural measures), necessary for the abatement of negative effects of soil erosion and sediment transport on the environment.

In the future, research on soil erosion and on sediment transport phenomena will have to concentrate on the following aspects:

1. The use of geo-informational systems (GIS) in soil erosion and sediment transport phenomena studies in hydrographical basins – it is necessary because of the possibilities it offers in the analysis of the numerous factors which interfere in the triggering and the development of erosion processes (particularly factors related to the land topography), the higher costs and the longer period of time necessary for their monitoring by other methods, especially when it comes to large surfaces. In the same time, GIS system allows the acquisition, the storage and the processing of data that characterize the status and evolution of land degradation caused by wind and water erosion, at any point in time. This allows policy makers to react in time and to adopt measures regarding:

   - the management of land affected by erosion
   - the decrease of erosion rate to the admissible limits
   - the permanent control of overland flow on hillslopes and of the concentrated ones in the torrential formation network
the reduction and control of watercourse pollution with chemicals coming from landslopes
the reduction of the warping and pollution rate of lakes by alluvia and chemicals washed off from the hillslopes
the prevention of flood occurrence on lands and other social – economic objectives, located at the basis of the hillslopes and along the water courses, due to the warping of flood waves mitigation hydro works. [14]

2. The diversification of methods aimed at acquiring and processing data field in order to increase the accuracy of the research on quantitative and qualitative changes caused by soil erosion and sediment transport of hydrographical basin elements.

3. The permanent development of models used to predict the erosion phenomena, the creation of forecasting models that satisfy all the requirements of a complex model.

4. The creation of a multidisciplinary research team for soil and hillslope erosion hydrological modelling, given by the development of a model for complex processes occurring on hillslopes implies a compromise between the processes we wish to integrate it in the model and the available data.

5. The need to develop a three-dimensional overland flow model as close as possible to the physical phenomena of hillslope overland flow, based on which we can elaborate the fundamental physical hydrological model of a hydrographical basin. [15]

6. The study of the physical - mechanical interactions between the water current alluvia, which influence the dynamics of certain pollutants. It is necessary to include these interactions in water quality models and in integrated water quality monitoring. [16]

7. The creation of representative experimental hydrographical basin (small hydrographical basin) where we may study the causes, manifestations and effects of soil erosion and sediment transport; also the possibility for us to test the effectiveness of different measures aimed at reducing the effects of soil erosion and sediment transport. These experimental hydrographical basin must be equipped with a sufficiently dense point network of tracking the erosion processes (hydrometric posts - in which we observe liquid flows and sediment transport, water levels and meteorological water quality, standard soil profiles, etc.) evolution.

References:
Indoor air Quality in apartment buildings of Estonia

ALO MIKOLA, TEET-ANDRUS KOIV
Department of Environmental Engineering,
Tallinn University of Technology,
Ehitajate tee 5, 19086 Tallinn,
ESTONIA
teet.koiv@ttu.ee

Abstract: - During the large-scale renovation of apartment buildings in Estonia there has risen the problem that the existing indoor air quality does not correspond to contemporary requirements. Some renovation methods, for example, replacing the old-type windows with new ones reduce the air change in apartments. Therefore the indoor air quality of apartment buildings should be studied carefully. The use of the metabolistic CO\textsubscript{2} method in determining the air change gives a good possibility for assessing the quality of the indoor air. To generalize the situation of indoor air quality in apartment buildings the large-scale measuring in different types of buildings has been carried out. The analysis of the measurement results shows that in many apartment buildings natural ventilation cannot guarantee the necessary quality of indoor air.

Key-Words: - Indoor air quality, CO\textsubscript{2} concentration in indoor air, air change rates, simulation of air change, apartment buildings, natural ventilation.

1 Introduction

Large-scale renovation of old apartment buildings in Estonia has raised the question of how to guarantee in those buildings an indoor climate that corresponds to contemporary requirements. So far renovation in apartment buildings has first of all been guided by saving energy, leaving the indoor climate in the background. It is necessary to find out the quality of the existing indoor air in Estonian apartment buildings.

An investigation into the existing situation of the air change and ventilation systems in apartment buildings requires large-scale measuring in apartments of different types of buildings. The methods used must guarantee the possibly most precise results [1], [2], [3], [4], [5], [6], [7]. The use of metabolistic CO\textsubscript{2} method in determining the air change gives a good possibility for assessing the quality of the indoor air. The theoretical base of the method and previous similar studies are described in [8]. This method has widely been used in public buildings [9], [10], [11], [12], [13], but the knowledge of its more exact application in apartment buildings in countries with a cold climate is not sufficient. Therefore in the given investigation a case study has been made in the course of which the methods used are specified and adapted to the conditions in apartment buildings. For better knowledge simulations of indoor climate have been carried out.

2 Criteria for assessing the CO\textsubscript{2} content in the indoor air

In accordance with the requirements [14] in Estonia living area must have natural or mechanical ventilation that guarantees the air change necessary for human activity. According to requirements in Estonia [15] air velocity in living spaces, the volume of the room per person and the content of harmful substances in the indoor air must not exceed the values permitted.

CO\textsubscript{2} content in the indoor air of residential buildings is determined by the standard [15] and the design criteria [15]. The CO\textsubscript{2} maximum concentrations determined by the EVS-EN 15251:2007 according to the classes of indoor climate for energy calculations and ventilation are presented in Table 1.
Table 1  Recommendable CO₂ levels over external air concentration and on the concentration 350 ppm (EVS-EN-15251) [16]

<table>
<thead>
<tr>
<th>Category</th>
<th>Indoor air CO₂ concentration above outdoor air level, ppm</th>
<th>Indoor air CO₂ concentration at outdoor air level 350 ppm, ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>350</td>
<td>700</td>
</tr>
<tr>
<td>II</td>
<td>500</td>
<td>850</td>
</tr>
<tr>
<td>III</td>
<td>800</td>
<td>1150</td>
</tr>
<tr>
<td>IV</td>
<td>&gt;800</td>
<td>&gt;1150</td>
</tr>
</tbody>
</table>

In analyzing the indoor climate in apartment buildings the use of the values given in Table 1 is not of much help, because mostly we have to do with natural ventilation without a possibility to regulate the air flow rate.

The Danish University of Technology has studied the standard in detail and gives the air flow rate of the classes of the indoor climate and the maximum norms of the CO₂ content in the indoor air [17]. These norms also correspond to the values that are given in the designing criteria of the indoor climate, Table 2. Due to that the present study uses the maximum values given in the designing criteria CR 1752 for accessing the CO₂ content in the indoor air, whereby the CO₂ content in the external air is considered to be 350 ppm. In new and already existing residential buildings it is essential to consider the classes B (II) and C (III) of thermal comfort. Values of class A (I) aim first of all at guaranteeing the high quality of the indoor climate that cannot be achieved in buildings with natural ventilation.

Tabel 2  Class of the indoor climate for rooms with human activity (CR 1752) [14]

<table>
<thead>
<tr>
<th>Category</th>
<th>Expected percentage dissatisfied, %</th>
<th>CO₂ concentration at outdoor air level 350 ppm, ppm</th>
<th>Indoor air CO₂ concentration, ppm</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>15</td>
<td>460</td>
<td>810</td>
</tr>
<tr>
<td>B</td>
<td>20</td>
<td>660</td>
<td>1010</td>
</tr>
<tr>
<td>C</td>
<td>30</td>
<td>1190</td>
<td>1540</td>
</tr>
</tbody>
</table>

The standard for determining the initial parameters of energy efficiency [15] allows short-time deviations from the required parameters of the indoor climate. The permitted deviation is up to 5%.

3 Results

3.1 Measuring of the CO₂ content in the indoor air in apartments

The CO₂ concentrations measured in the bedrooms of apartments in the summer period varied from 340 to 3200 ppm. According to the designing criteria of the indoor climate the maximum value of the CO₂ content for class B (II) is 1010 ppm and for class C (III) 1540 ppm (with the CO₂ content in the external air 350 ppm). In all the apartments investigated in summer during 88% of the measuring period the CO₂ content of the indoor air corresponded to class B and 97% to class C, Fig 1. Considering the fact that on an average people spend at home only about 14 hours a day, the CO₂ concentration corresponded to level B in 82% of the time people spent at home and to level C in 95% of the time respectively. Bearing in mind the fact that in determining the initial parameters of energy efficiency the standard permits CO₂ concentration to exceed the maximum values 5% in determining the classes B and C, in 75% of the investigated apartments the air quality corresponded to both class B and class C.

From the results it can be concluded that owing to frequent airing the rooms by opening windows in the summer time the exceeding of the CO₂ concentration is not a problem. The analysis of the results of measuring showed that the level of the maximum CO₂ concentration in the summer period is chiefly exceeded as a result of active daily human activity. In the night period the CO₂ level mostly corresponded to the permitted values of class C.

Fig.1 Cumulative distribution of the results of the measuring of the CO₂ concentration in the summer period

The results of the measuring in the winter period varied from 370 to 4000 ppm. In 13 % of the apartments the CO₂ level for a short time exceeded
4000 ppm – the maximum limit of the measuring unit. In the apartments investigated the CO₂ content in the indoor air corresponded to class B in 59% of the time of measuring and to class C in 87% of the time, Fig. 2. As on an average people spend at home only about 14 hours a day, the CO₂ concentration corresponded to level B in 38% of the time the apartments were used and to level C in 81% of the time. Bearing in mind the fact that in determining the initial parameters of energy efficiency the standard permits the CO₂ content to exceed the maximum values within 5% in the winter period. In 17% of the investigated apartments the CO₂ content in the indoor air corresponded to class B and 34% to class C.

![Fig. 2: Cumulative distribution of the results of the measuring of the CO₂ concentration in the winter period](image)

For assessing the impact of replacing the old-type windows with new ones on the quality of the room air an investigation has been made to separately study the CO₂ content in the indoor air in apartments with new and old windows, Fig.3. In apartments with old windows the CO₂ concentration in the indoor air corresponded to the maximum value of class B in 66% of the measuring period and in apartments with new windows to 56%. In apartments with old windows the CO₂ concentration corresponded to the maximum value of class C in 94% of the measuring period and in apartments with new windows to 84%. Considering the fact that on an average people spend at home only about 14 hours a day, the CO₂ concentration in the air of the apartments with old windows corresponded to level B in 49% of the time spent at home and in apartments with new windows in 34% of the time, respectively. In apartments with old windows the CO₂ concentration corresponded to level C in 91% of the time spent at home and in apartments with new windows in 77% of the time. Within 5% of the permitted exceeding the maximum value, apartments with old windows corresponded to class B in 22% of the apartments investigated and to class C in 55% of the apartments investigated. In apartments with new windows 15% of the apartments corresponded to class B and 25% to class C. Hence, the impact of changing the windows on the quality of the indoor air is apparent, and it turns out that replacing the old windows with new ones essentially deteriorates the air change in the apartments. It should be added that in apartments with old windows the rooms were aired much more frequently and therefore their impact on the quality of the indoor air was even bigger than are shown by the results of measuring.

![Fig. 3: Impact of changing windows on the CO₂ level during the period of measuring](image)

In Table 3 we can see calculated air changes in bedrooms and in Figure 4 average airflows in investigated bedrooms.

### Table 3 Air change rates in bedrooms calculated by indoor air CO₂ content measurements

<table>
<thead>
<tr>
<th>Code</th>
<th>Window</th>
<th>Ventilation type*</th>
<th>Door position</th>
<th>Air exchange rate, 1/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>4151</td>
<td>new</td>
<td>na+cook</td>
<td>open</td>
<td>0,09</td>
</tr>
<tr>
<td>2131</td>
<td>new</td>
<td>na</td>
<td>close</td>
<td>0,14</td>
</tr>
<tr>
<td>2121</td>
<td>new</td>
<td>na+cook</td>
<td>close</td>
<td>0,17</td>
</tr>
<tr>
<td>3121</td>
<td>new</td>
<td>na</td>
<td>close</td>
<td>0,21</td>
</tr>
<tr>
<td>3111</td>
<td>new</td>
<td>na+cook</td>
<td>open</td>
<td>0,23</td>
</tr>
<tr>
<td>1231</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0,24</td>
</tr>
<tr>
<td>1222</td>
<td>old</td>
<td>na+cook</td>
<td>open</td>
<td>0,31</td>
</tr>
<tr>
<td>4141</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0,33</td>
</tr>
<tr>
<td>1251</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0,40</td>
</tr>
<tr>
<td>1141</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0,48</td>
</tr>
<tr>
<td>1131</td>
<td>new</td>
<td>na+cook</td>
<td>close</td>
<td>0,55</td>
</tr>
<tr>
<td>3131</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0,60</td>
</tr>
</tbody>
</table>
3.2 Results of the simulation

A simulation of air change in typical apartment buildings with natural ventilation was carried out by software IDA ICE 4.0 on the basis of Estonian TRY. The results are given in Figure 4. The average difference of air change in months was more than 2 times and that of the maximum values was about 10 times.

### Table 1: Average air flows in investigated bedrooms

<table>
<thead>
<tr>
<th>No</th>
<th>Old/New</th>
<th>Method</th>
<th>State</th>
<th>Air Flow</th>
</tr>
</thead>
<tbody>
<tr>
<td>4161</td>
<td>old</td>
<td>na+cook</td>
<td>open</td>
<td>0.64</td>
</tr>
<tr>
<td>2141</td>
<td>old</td>
<td>na</td>
<td>close</td>
<td>0.67</td>
</tr>
<tr>
<td>2111</td>
<td>old</td>
<td>na</td>
<td>open</td>
<td>0.73</td>
</tr>
<tr>
<td>4131</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>0.74</td>
</tr>
<tr>
<td>1211</td>
<td>old</td>
<td>na+cook</td>
<td>open</td>
<td>0.85</td>
</tr>
<tr>
<td>1111</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>1.01</td>
</tr>
<tr>
<td>1202</td>
<td>new</td>
<td>na+cook</td>
<td>open</td>
<td>1.01</td>
</tr>
<tr>
<td>4121</td>
<td>new</td>
<td>na</td>
<td>open</td>
<td>1.11</td>
</tr>
<tr>
<td>1172</td>
<td>old</td>
<td>na+cook</td>
<td>open</td>
<td>1.15</td>
</tr>
<tr>
<td>1191</td>
<td>new</td>
<td>na+cook</td>
<td>open</td>
<td>1.18</td>
</tr>
<tr>
<td>2151</td>
<td>old</td>
<td>na+cook</td>
<td>open</td>
<td>1.34</td>
</tr>
<tr>
<td>1181</td>
<td>old</td>
<td>na</td>
<td>open</td>
<td>1.47</td>
</tr>
</tbody>
</table>

*na – natural ventilation; na+cook – natural ventilation and cooker hood

### Fig. 4

Air change rate in an apartment with an average air change value in the case of natural ventilation during a year (results of the simulation)

### 4 Conclusions

The purpose of the current study was to examine the situation of indoor air quality and air change in old-type Estonian apartment buildings. In order to achieve exact results an investigation into the CO₂ concentration in the indoor air has been carried out. In addition, indoor air quality simulations by special software were made. Air change in bedrooms was calculated by using the metabolic CO₂ method. The situation of indoor air quality in Estonian apartment buildings has been generalized applying the same method. As the analysis shows natural ventilation cannot guarantee the quality of indoor air in apartment buildings. Therefore the ventilation of old-type apartment buildings should be renovated.

CO₂ measurements were carried out in 31 brick apartment buildings in different Estonian towns. Air quality assessment based on the measurement results in the winter period varied from 370 - 4000 ppm. In the apartments studied the indoor air CO₂ concentration satisfied the thermal comfort level B during 59 % and level C during 78 % of the measurement period. According to the measurement results 17 % of the apartments corresponded to level B and 35 % to level C. The air change rates in bedrooms varied from 0.09 – 1.42 h⁻¹. The simulations of natural ventilation show that exhaust airflow changes greatly. Therefore it can be concluded that natural ventilation does not enable to assure sufficient air change in apartment buildings.

Replacing the old-type windows with new ones might cause several problems with the quality of the indoor climate. The study shows the indoor air CO₂ concentration increases when the new windows are mounted. From the measuring results we can conclude that replacing the old-type windows with new airtight
ones must be accompanied by the renovation of the ventilation system.

References:


Simplified design approach of rectangular spiral antenna for UHF RFID tag

MAKROUM EL MOSTAFA\textsuperscript{1,2}, RIFI MOUNIR\textsuperscript{1}, LATRACH MOHAMED\textsuperscript{3}, BENBASSOU ALI\textsuperscript{4}

\textsuperscript{1}Networks Laboratory, Computer, Telecom & Multimedia- EST
\textsuperscript{2}CED sciences de l'ingénieurs– ENSEM
Km 7, Route El Jadida, B.P. 8012 Oasis Casablanca Oasis Casablanca, Morocco
\textsuperscript{3}Radio & Microwave Group, ESEO, Graduate School of Engineering,
4 Rue Merlet de la Boulaye, PO 30926, 49009 Angers, France
\textsuperscript{4}Laboratory of Transmission and Processing of Information EMC / Telecom.
School of Technology (TSE) PO. 2427 FES MOROCCO
elmostafa1980@hotmail.com

Abstract: In this paper we present a method to simplify the calculation of spiral antennas for RFID tag settings without resorting to numerical analysis methods. It saves up to 99% of the time required by the simulation based on the method of moments. Thus we present a theoretical and experimental study for the design of the spiral antennas for RFID label in the UHF band. We present in this study the S11 parameter that enables us to evaluate the evolution of current distribution and therefore the resonance frequency of the spiral antennas. This parameter is calculated theoretically by applying the method of moments to wired antenna formed by the rectangular copper spiral printed on a dielectric substrate. The experimental validation of our theoretical models is performed using a network analyzer. The confrontation theory-experience allows us to draw some interesting conclusions concerning the number of loops of the spiral and the choice of dielectric substrate.

Key-Words: Antennas, Spiral antennas, RFID tags, UHF RFID tags, Method of moments, S11 parameter.

1. Introduction

During these last decades, the technologies of information and communications (ICT) have known unprecedented development. The identification technologies are part of these information technologies. Due to the recent development of microelectronics and wireless systems, new contactless identification technologies have emerged: the radio identification technology (or RFID for Radio-Frequency IDENTification). These new technologies, by their greater flexibility, make the exchange of information much faster and efficient.

RFID is a technology to recognize or identify with greater or lesser distance (contact tens of meters) in a minimum of time, an object, an animal or a person with an electronic tag. We can cite, for example, contactless smartcard systems, highway tolls without stopping, parking or building access control, etc. ...

It falls into the category of automatic identification technologies (AIDC, Automatic Identification and Data Capture), as the bar code character recognition, pattern recognition, or magnetic cards.

RFID systems use mainly four frequency bands [1], [2] and [3]: 125KHz (LF band, Low Frequency), 13.56 MHz (HF, High Frequency), 860-960MHz (UHF, Ultra High frequencies), 2.45 GHz (microwave). In recent years a growing interest in the field of industry and research focused on passive UHF RFID technology. It presents a low cost solution. It also helps to have a data rate higher (around 20 kbit / s) and achieve a reading range greater than other technologies called passive RFID. This interest has helped put in place in most parts of the world regulations and industry standards for market development of this technology.

In the first part of this document, a brief presentation of passive RFID technology is exposed. Then the modeling of a rectangular spiral RFID antenna using the moment method and the results of simulations and measurements are discussed and presented in a second part. Then we develop a method for estimating the peak current and thus the resonant frequency of spiral antennas.
2. The principle of passive RFID technology

An automatic identification application RFID, as shown in Fig. 1, consists of a base station that transmits a signal at a frequency determined to one or more RFID tags within its field of inquiry. When the tags are "awakened" by the base station, a dialogue is established according to a predefined communication protocol, and data are exchanged.

![Fig.1: Schematic illustration of an RFID system](image)

The tags are also called a transponder or tag, and consist of a microchip associated with an antenna. It is an equipment for receiving an interrogator radio signal and immediately return via radio and the information stored in the chip, such as the unique identification of a product.

Depending on the operating frequency of the coupling between the antenna of the base station and the tag may be an inductive coupling (transformer principle) or radiative (far-field operation). In both cases of coupling, the chip will be powered by a portion of the energy radiated by the base station.

To transmit the information it contains, it will create an amplitude modulation or phase modulation on the carrier frequency. The player receives this information and converts them into binary (0 or 1). In the sense reader to tag, the operation is symmetric, the reader transmits information by modulating the carrier. The modulations are analyzed by the chip and digitized.


To predict the resonant frequencies of the currents induced in the antenna structures forming tags, such as spiral antennas rectangular ICs, we have used initially the model based on the theory of diffraction by thin wires [4]. We arrive at an integro-differential equation, and whose resolution is based on the method of moments [5]. Although a very good result is obtained, the computation time required is a major drawback. It is quite high. In a second step, the study is to find a method to simplify the estimation of the resonance frequency of rectangular spiral antennas in various frequency ranges.

3.1 Method of Moments

It is an integral analysis method used to reduce a functional relationship in a matrix relationship which can be solved by conventional techniques. It allows a systematic study and can adapt to very complex geometric shapes.

This method is more rigorous and involves a more complicated formalism leading to heavy digital development. It applies in cases where the antenna can be decomposed into one or several environments: the electromagnetic field can then be expressed as an integral surface. It implicitly takes into account all modes of radiation.

Moreover, the decomposition of surface current to basis functions, greatly simplifies the solution of integral equations which makes the method simple to implement.

This procedure is based on the following four steps:

- Derivation of integral equation.
- Conversion of the integral equation into a matrix equation.
- Evaluation of the matrix system.
- And solving the matrix equation.

3.2 Formulation of the method of moments [5] [6]

We have chosen the configuration shown in Figure 2, a rectangular metal track, printed on an isolating substrate. It consists of length A, width B and thickness e.

![Fig.2: Dimensions of the track used in the simulation](image)

The theory of antennas for connecting the induced current in the metal track to the incident electromagnetic field (Ei, Hi) using integro-differential equation as follows:
The method used to solve such equations is the method of moments [4].

The problem thus is reduced to solving a linear system of the form:

\[ [V] = [Z_{mn}] [I] \]

With:

\([I]\) representing the currents on each element of the structure.
\([V.]\) representing the basic tension across each element \(m\) in length \(\Delta\) given by:

\[ V_m = E_i(m) \Delta \]

\([Z_{mn}]\) represents the generalized impedance matrix, reflecting the EM coupling between the different elements of the antenna.

The rectangular loop will be discretized into \(N\) identical segments of length \(\Delta\):

\[ \Delta = \frac{2(A + B)}{N} \]

The discretization step is chosen so as to ensure the convergence of the method of moments.

\[ \Delta = \frac{\lambda}{20} \]

We conclude from this relationship that the number of segments required for convergence of numerical results is:

\[ N = 40 \frac{(A + B)}{\lambda} \]

\(\lambda\) : Represents the smallest wavelength of EM field incident.

The simulations will focus on frequencies between:

\[ 100 \text{MHz} < f < 1.8 \text{GHz} \]

\(17 \text{cm} < \lambda < 3 \text{m}\)

The number of segments will be: \(N = 76\). The illumination of the loop is done by an plane EM wave, arriving in tangential impact such that the incident electric field \(E_z\) is parallel to the longest track.

The amplitude of the incident field is normalized 1V/m. we are interested in a loop short-circuited to highlight the resonance phenomena relating to the geometric characteristics of the loop.

\[ I(z) = A_c \sin \left( k \frac{A + B}{2} - |z| \right) + A_d \cos \left( k \frac{A + B}{2} - |z| \right) \]

With:

\(A_c\) : The amplitude of common mode current.
\(A_d\) : The amplitude of the differential mode current.
\(k = \omega \sqrt{\mu \varepsilon}\) : wave Vector.

Indeed, \(A_c\) and \(A_d\) can be written as:

\[ A_c = \frac{e_c(0)}{Z_0 \cos \left( k \frac{A + B}{2} \right)} \]
\[ A_d = -\frac{e_d(0)}{Z_0 \sin \left( k \frac{A + B}{2} \right)} \]

with \(Z_0\) is the characteristic impedance of a transmission line.
\(e_c(0)\) and \(e_d(0)\) : respectively represent the equivalent electromotive forces at the center of the tracks of common and differential modes.

The coefficients \(A_c\) and \(A_d\) take an infinite value only if:

\[ Z_0 \cos \left( k \frac{A + B}{2} \right) = 0 \]
\[ Z_0 \sin \left( k \frac{A + B}{2} \right) = 0 \]
for $Z_0 \cos \left( \frac{kA+B}{2} \right) = 0$

$$\cos \left( \frac{kA+B}{2} \right) = 0 \quad (12)$$

$$\Rightarrow \frac{kA+B}{2} = (2N + 1) \frac{\pi}{2} \quad (13)$$

with $N=0, 1, 2, \ldots$

$k = \frac{\omega \sqrt{\varepsilon r}}{c}$. We have considered the isolated loop, surrounded by air, so $\varepsilon_r = 1$ and $k = \frac{\omega \sqrt{\varepsilon_0 \mu_0}}{c}$

(13) becomes:

$$\frac{2nfA+B}{c} = (2N + 1) \frac{\pi}{2} \quad (14)$$

Hence the resonant frequency common mode:

$$F_{Re} = (2N + 1) \frac{c}{2(A+B)} \quad (15)$$

for $Z_0 \sin \left( \frac{kA+B}{2} \right) = 0$

$$\sin \left( \frac{kA+B}{2} \right) = 0$$

$$\Rightarrow \frac{kA+B}{2} = N, \pi \quad (17)$$

With $N=1, 2, 3, \ldots$

(17) Becomes:

$$\frac{2nfA+B}{c} = N, \pi \quad (18)$$

Therefore the resonant frequency of the differential mode is:

$$F_{Rd} = N \frac{c}{(A+B)} = 2N \frac{c}{2(A+B)} \quad (19)$$

We therefore find the result observed in fig.3, which was obtained by the method of moments.

The resonance frequency of the loop can be easily linked to the length $A$ and width $B$ of the loop by the following approximate relation:

$$F_R = N \frac{c}{2(A+B)} \quad (20)$$

With $c = 3108 \text{ m/s}$, speed of EM waves in vacuum and $N = 1, 2, 3, \ldots$

4. Simulations and measurements

The simulations were done under the MATLAB environment. The experimental validation was performed at the Laboratory LTPI / RUCI in Fez.

4.1 Achievement

We have made various prototypes of antennas as shown in Fig.4, using as the substrate, glass epoxy, type FR4 with relative permittivity $\varepsilon_r = 4.32$ and 1.53 mm thick.

Fig.4: Antenna loops made

4.2 Measures and results

The coefficient of reflection of antennas made, were measured with a vector network analyzer HP-type operating in the 100Hz-6000MHz band (Fig. 5).

Fig.5: Measurement of the reflection coefficient of the antenna using a vector network analyzer.

5. Evaluation of peaks and frequency of resonance of induced currents in function of geometric characteristics of printed loop

The assessment of the size and position of resonance peaks of currents distributed on the printed tracks is crucial for designers of antennas tags. Indeed the action of these peaks can completely change the normal operation of the transponder.

In this part the evolution of the amplitude of these peaks and their resonance frequency will be studied in function of geometric characteristics of loops.

5.1 Evaluation of the peaks as a function of the perimeter of the printed loop and the report ($B/A$)

In fig.6 we found the current to peak resonances for loops with different perimeters (from 40 cm to 120 cm) and reports Width / Length ($B/A = 1/4, 1/3$)
and 2/5). We find in this figure a linear variation of these peaks as a function of the perimeter of these loops to the same ratio (B / A) thereof. Note that we have made loops whose resonant frequencies are between 400 MHz and 3 GHz.

The peak current amplitude at resonance can be easily connected to the perimeter of the loop by the following equation:

\[ I_{\text{pic}} = \alpha \cdot P \]  

(21)

With:

- \( I_{\text{pic}} \): The amplitude of peak current at resonance.
- \( \alpha \): The slope of the line.
- \( P \): Scope of the loop.

- For the \( \frac{B}{A} = \frac{1}{4} \)
  \[ \alpha_1 = 1.735 \times 10^{-3} = \frac{1}{4} \times 6.94 \times 10^{-3} \]  

  (22)
- For the \( \frac{B}{A} = \frac{1}{3} \)
  \[ \alpha_2 = 2.353 \times 10^{-3} = \frac{1}{3} \times 6.94 \times 10^{-3} \]  

  (23)
- For the \( \frac{B}{A} = \frac{2}{5} \)
  \[ \alpha_3 = 2.776 \times 10^{-3} = \frac{2}{5} \times 6.94 \times 10^{-3} \]  

  (24)

From the above we can write \( \alpha \) as follows:

\[ \frac{\alpha}{\pi} = \frac{B}{A} \times K_t \]  

(25)

With:

\( K_t = 6.94 \times 10^{-3} \).

The equation can be written as follows:

\[ I_{\text{pic}B/A} = \frac{B}{A} \times K_t \times P \]  

(26)

5.2 Evolution of the resonance frequencies of the induced currents in function of numbers of loops of rectangular spiral antennas (fixed perimeter).

The evaluation of the resonance peaks of currents on printed circuit tracks is crucial for designers of printed antennas. Indeed the action of these peaks can completely change the normal operation of the circuit.

We calculated the resonance frequency of the induced currents for four rectangular spiral antennas, respectively having a loop, two loops, three loops and four loops and the same perimeter 64cm, shown in Fig.7

![Fig.7: Dimensions of the Ics tracks used in the simulation and measurements](image)

In fig.8 we observe that for the same scope the resonance frequencies of induced currents remain almost unchanged in function of the number of loops.

![Fig.8: Evolution of the resonance frequencies of the induced currents in function of numbers of loops](image)
Fig. 8: Reflection coefficient $S_{11}$ on the tracks of the ICs in fig.7-simulation by MOM and measurement

However, we notice a slight difference between the frequencies of resonance peaks of these antennas evaluated theoretically and those measured experimentally. This difference is surely due to the fact that in our theoretical model we do not take account of the dielectric permittivity of the substrate. Indeed, for the simulation we considered a spiral surrounded by air.

To account for the influence of the dielectric permittivity of the substrate on the positions of resonance frequencies, consider the case of a single loop. On Fig.9 we have plotted the $S_{11}$ parameter for different values of $\varepsilon_r$ substrate. We effectively note that this setting actually influence the position of the resonance frequency.

Fig. 9 Influence of $\varepsilon_r$ on peak resonances of spiral antennas

We can take advantage of this finding to try to design antennas that resonate at particular frequencies by playing on the nature of the dielectric substrates. This will allow the miniaturization of spiral antennas for RFID applications.

5.3 Evolution of the resonance frequencies of the induced currents in function of numbers of loops of rectangular spiral antennas (A and B fixed).

We have calculated the resonance frequencies of the induced currents for three rectangular spiral antennas, respectively having a loop, 2 loops and 3 loops and the same lengths ($A = 24\text{cm}$) and widths ($B = 8\text{cm}$): Fig.10

Fig. 10: Dimensions of tracks of printed antennas used in the simulation

On Fig.11 we see that for the same lengths ($A = 24\text{cm}$) and width ($B = 8\text{cm}$) resonance frequencies of the induced currents can be easily represented by the following approximate relation:
\[ F_{R(N\ boucles)} = \frac{F_{R(1\ boucle)}}{N} \] (27)

With: N=2, 3, 4,..........

Fig.11: Frequencies of resonance in function of numbers of loops on the tracks of the ICs of the fig.10

6. Conclusion

This paper presents the design of antennas for passive RFID tags. The first part concerned the quick introduction of this technology. It was followed by modeling of a spiral RFID UHF antenna using the theory of the antennas.

Finally we have presented a method of estimating the peak current and resonance frequency of rectangular spiral RFID antennas.

This study showed that the amplitudes of the resonance peaks of rectangular spiral antennas printed vary linearly as a function of geometrical characteristics thereof. This linearity can be used by designers of printed antennas to assess the amplitude of current peaks at resonance with simple graphs that can be drawn as a function of geometrical characteristics of loops. As well as frequencies of resonance of the induced currents are virtually unchanged in function of numbers of loops for the same perimeter.

Currently we are trying to establish a relationship between the resonance frequency of such antennas and the nature of the dielectric substrate on which it is printed. This will surely improve the performance of printed antennas and also to contribute to their miniaturization.

References:
Maximal Allocated Benefit and Minimal Allocated Cost for Interval Data

Sahand Daneshvar, Mozghan Mansouri Kaleibar

Islamic Azad University Tabriz Branch, Islamic Azad University Tabriz Branch and Young Researchers Club
Tabriz- Iran, Tabriz -Iran
sahanddaneshvar1@yahoo.com, mozghanmansouri953@gmail.com

Abstract:—This Paper deals with calculating the minimal and maximal shares of individuals or organizations based on different criteria. Suppose that players are selfish and the score for each criterion for a player is an interval. Each player makes any possible efforts to bring about his or her ideal condition. In this paper a new scheme to calculate the minimal allocated cost and the maximal allocated benefit for interval data is offered. In this scheme also the new models have been proposed for avoiding zero weight occurrences. Here ultimate allocated is achieved for each player with suitable coalition within several defined coalition.

Keywords:— Assurance Region Method, DEA, Game Theory, Minimum Game, Weight Selection, Zero Weight.

1 Introduction
Consider n players each have m criteria for evaluating their competency or ability, which is represented by interval. For example, consider a usual classroom examination, the higher score for a criterion is, the better player is judged to perform as regard to that criterion. Each person has a right to choose two sets of weights for the criteria that he or she supposes to be the score of player for each of them, as follows:

\[0,0\] It is assumed that the higher the interval score for a criterion is, the better player is judged to perform as regard to that criterion. Each person has a right to choose two sets of non negative weights \[w^d = (w^d_1, ..., w^d_m)\] to the criteria that are most preferable to the player. Using the weight \[w^d\], the relative scores of player \(k\) to the total score are defined as follows:

\[\sum_{i=1}^{m} w^d_i \cdot x^1_{ik}, \sum_{i=1}^{m} w^d_i \cdot (\sum_{j=1}^{m} (x^1_{ij} + x^u_{ij}))\] (1)

The denominator represents the total score of all players as measured by player k weight selection. The numerators indicate player k self-evaluation as a lower and upper bound lower bound. Therefore, the statement (1) represents player k lower and upper relative importance in accordance to the \(x^1_{ik}, x^u_{ik}\). We assume that the weighted scores are transferable. Player k wishes to maximize this ratio by selecting the most preferable weights \(w^d_i\), for each of \(x^1_{ik}, x^u_{ik}\) as follows:

\[\text{Max} \quad \sum_{i=1}^{m} w^d_i \cdot x^1_{ik} \quad \text{s.t} \quad w^d_i \geq 0 \quad (\forall i) \quad (2)\]

\[\text{Max} \quad \sum_{i=1}^{m} w^d_i \cdot (x^1_{ij} + x^u_{ij}) \quad \text{s.t} \quad w^d_i \geq 0 \quad (\forall i) \quad (3)\]

The motivation behind this problem is that the player k wishes to maximize his lower relative efficiency by (2), and upper relative efficiency by (3). As we can see in (2) the lower weighted sum of its record to the total weighted sum of all player records is maximized and in (3) the upper weighted sum of its record is maximized. We reformulate the problem, without losing generality. We normalize the data set X so that it is row-wise

2 Basic Models of The Game
We introduce the basic models and structures of the game.

2.1. Selfish behavior
Let \([x^l_{ij}, x^u_{ij}]\) be the score of player \(j\) in the criterion \(i\), for \(i = 1, ..., m\) and \(j = 1, ..., n\) and \(x^l_{ij} > 0, x^u_{ij} > 0\). It is assumed that the higher the interval score for a criterion is, the better player is judged to perform as regard to that criterion. Each person \(k\) has a right to choose two sets of weights for the criteria that he or she supposes to be the score of player for each of them, as follows:

\[0,0\] It is assumed that the higher the interval score for a criterion is, the better player is judged to perform as regard to that criterion. Each person \(k\) has a right to choose two sets of non negative weights \[w^d = (w^d_1, ..., w^d_m)\] to the criteria that are most preferable to the player. Using the weight \[w^d\], the relative scores of player \(k\) to the total score are defined as follows:

\[\sum_{i=1}^{m} w^d_i \cdot x^1_{ik}, \sum_{i=1}^{m} w^d_i \cdot (\sum_{j=1}^{m} (x^1_{ij} + x^u_{ij}))\] (1)

The denominator represents the total score of all players as measured by player k weight selection. The numerators indicate player k self-evaluation as a lower and upper bound lower bound. Therefore, the statement (1) represents player k lower and upper relative importance in accordance to the \(x^1_{ik}, x^u_{ik}\). We assume that the weighted scores are transferable. Player k wishes to maximize this ratio by selecting the most preferable weights \(w^d_i\), for each of \(x^1_{ik}, x^u_{ik}\) as follows:

\[\text{Max} \quad \sum_{i=1}^{m} w^d_i \cdot x^1_{ik} \quad \text{s.t} \quad w^d_i \geq 0 \quad (\forall i) \quad (2)\]

\[\text{Max} \quad \sum_{i=1}^{m} w^d_i \cdot (x^1_{ij} + x^u_{ij}) \quad \text{s.t} \quad w^d_i \geq 0 \quad (\forall i) \quad (3)\]

The motivation behind this problem is that the player k wishes to maximize his lower relative efficiency by (2), and upper relative efficiency by (3). As we can see in (2) the lower weighted sum of its record to the total weighted sum of all player records is maximized and in (3) the upper weighted sum of its record is maximized. We reformulate the problem, without losing generality. We normalize the data set X so that it is row-wise
normalized, i.e., \( \sum_{j=i}^{n} x_{ij} = 1 \) (\( \forall i \)). We divide the row \((x_{i1},...,x_{im})\) by the row-sum \( \sum_{j=1}^{m} (x_{ij} + x_{i}^{*}) = 1, i = 1,...,m \) (4)

Thus, using the Charnes-Cooper transformation scheme, the fractional programs (2) and (3) can be expressed using a linear programs follows:

\[
c^1(k) = \text{Max} \sum_{i=1}^{n} w_{i}^k x_{i}^1 \\
\text{s.t} \sum_{i=1}^{n} w_{i}^k = 1 \quad (5)
\]

\[
c^* (k) = \text{Max} \sum_{i=1}^{n} w_{i}^k x_{i}^n \\
\text{s.t} \sum_{i=1}^{n} w_{i}^k = 1 \quad (6)
\]

After solving problems (5) and (6), if the optimal value of the problem (5) and (6) are \(c^1(k)\) and \(c^*(k)\) respectively, then \(c^1(k) + c^*(k)\) may be considered as optimal value of the problem. Now the problem is to maximize the objectives (5) and (6) on the simplex \( \sum_{i=1}^{n} w_{i}^k = 1 \).

Apparently, the optimal solution is given by assigning 1 to \( w_{i(k)} \) and \( w_{i(k')} \) for the criterion \( i(k) \) and \( i(k') \) such that \( x_{i(k)} = \text{Max} \{x_{i}^1 | i = 1,...,m\} \) and \( x_{i(k')} = \text{Max} \{x_{i}^n | i = 1,....,m\} \) respectively and assigning 0 to the weight of remaining criteria. We denote this optimal value by \(c(k)\).

\(c^1(k) = x_{i}^1\) and \(c^*(k) = x_{i}^n\) for \(k=1,...,n \) (7)

Theorem 1. \( \sum_{k=1}^{n} c^1(k) + c^*(k) \geq 1 \) (8)

Proof. Let the optimal weight for player \( k \) be \( w_{k}^* = (w_{k}^{*1},...,w_{k}^{*n}) \) and \( w_{k}^* = (w_{k}^{1*},...,w_{k}^{n*}) \), \( w_{i(k)} = 1 \) and \( w_{i(k')} = 0 (\forall i \neq i(k)) \) and \( w_{i(k')} = 1 (\forall i \neq i(k')) \). Then we have

\[
\sum_{k=1}^{n} c^1(k) + c^*(k) = \sum_{k=1}^{n} c^1(k) + \sum_{k=1}^{n} c^*(k) = \sum_{k=1}^{n} \sum_{i=1}^{n} w_{i}^k x_{i}^1 + \sum_{k=1}^{n} \sum_{i=1}^{n} w_{i}^k x_{i}^n \\
\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i}^k x_{i}^1 + \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i}^k x_{i}^n \geq \sum_{i=1}^{n} x_{i}^1 + \sum_{i=1}^{n} x_{i}^n = 1
\]

The inequality above follows from \( x_{i(i)} \geq x_{i}^1 \) and \( x_{i(i')} \geq x_{i}^n \) and the last equality follows from the row-wise normalization.

This theorem assert that, if each player sticks to his egoistic sense of value and insists on getting the portion of the benefit as designated by \(c^1(k)\) and \(c^*(k)\), the sum of shares usually exceeds 1 and hence \(c^1(k) + c^*(k)\) cannot fulfill the role of division of the benefit. If eventually the sum of \(c^1(k)\) and \(c^*(k)\) turns out to be 1, all players will agree to accept the division \(c^1(k) + c^*(k)\), since this is obtained by the players most preferable weight selection. The latter case will occur when all players have the same and common optimal weight selection, we have the following theorem.

Theorem 2.

The equality \( \sum_{k=1}^{n} c^1(k) + c^*(k) = 1 \) holds if and only if our data satisfies the condition \( x_{1}^1 = x_{2}^1 = .... = x_{m}^1 \) and \( x_{1}^n = x_{2}^n = .... = x_{m}^n \), \( \forall k = 1,...,n \). That is, each player has the same score with respect to the \( m \) criteria.

Proof. The (if) part can be seen as follows. Since \( c^1(k) = x_{1}^1 \) and \( c^*(k) = x_{1}^n \) for all \( k \), we have:

\[
\sum_{k=1}^{n} c^1(k) + c^*(k) = \sum_{k=1}^{n} c^1(k) + \sum_{k=1}^{n} c^*(k) = \sum_{k=1}^{n} x_{1}^1 + \sum_{k=1}^{n} x_{1}^n = 1
\]

The (only if) part can be proved as follows. Suppose \( x_{i}^1 > x_{j}^1 \) and \( x_{i}^n > x_{j}^n \) then there must be column \( h \neq 1 \) and \( h' \neq 1 \) such that \( x_{1h} < x_{1h'} \) and \( x_{ih} < x_{ih'} \), otherwise the second row sum cannot attain 1. Thus we have

\( c^1(1) \geq x_{1}^1 \), \( c^* (1) \geq x_{1}^n \) and \( c^1(2) \geq x_{2}^1 \), \( c^* (2) \geq x_{2}^n \) \( c^1(h) \geq x_{1h} \), \( c^* (h) \geq x_{1h} \) \( c^1(h') \geq x_{1h'} \), \( c^* (h') \geq x_{1h'} \) \( c^1(j) \geq x_{jh} \), \( \forall j \neq 1 \), \( h \) \( c^* (j) \geq x_{jh} \), \( \forall j \neq 1 \), \( h' \) . Hence it holds

\[
\sum_{k=1}^{n} c^1(k) + c^*(k) = \sum_{k=1}^{n} c^1(k) + \sum_{k=1}^{n} c^*(k) \geq \sum_{j=1}^{n} x_{1}^1 + \sum_{j=1}^{n} x_{1}^n \\
\sum_{j=1}^{n} x_{1}^1 + \sum_{j=1}^{n} x_{1}^n > \sum_{j=1}^{n} x_{1}^1 + \sum_{j=1}^{n} x_{1}^n = 1
\]

This leads to a contradiction. Therefore player1 must have the same score in all criteria. The same relation must hold for the other players.

In the above case, only one criterion is needed for describing the game and the division proportional to this score is a fair division. However, such situation might occur only in rare instances. In the majority of cases, we have \( \sum_{k=1}^{n} c^1(k)+c^*(k) > 1 \).

2.2 Coalition with additive property

Let the coalition S be a subset of player set \( N = (1,....,n) \). The record for coalition S is defined by \( x_{j} (S) = \sum_{j \in S} x_{j}^1, x_{j}^n (S) = \sum_{j \in S} x_{j}^n (j = 1,...,m) \) (6)

These coalitions aim to maximize the outcome \( c^1(S), c^* (S) \)
Maximal Allocated Benefit and Minimal Allocated Cost for Interval Data

\[ c^1(S) = \text{Max} \sum_{i=1}^{n} w_i x^i(S) \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (7) \]
\[ w_i \geq 0 \forall i \]

\[ c^*(S) = \text{Max} \sum_{i=1}^{n} w_i x^*(S) \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (8) \]
\[ w_i \geq 0 \forall i \]

The \( c^1(S) \) and \( c^*(S) \) with \( c^1(\emptyset) = 0 \) and \( c^*(\emptyset) = 0 \) defines a characteristic function of the coalition \( S \). Thus this game is represented by \((N,e)\).

**Definition 1.** A function \( f \) is called sub-additive if for any \( S \subseteq N \) and \( T \subseteq N \) with \( S \cap T = \emptyset \) the following statement holds:

\[ f(S \cup T) \leq f(S) + f(T) \]

**Definition 2.** A function \( f \) called super – additive if for any \( S \subseteq N \) and \( T \subseteq N \) with \( S \cap T = \emptyset \) the following statement holds:

\[ f(S \cup T) \geq f(S) + f(T) \]

**Theorem 3.** The characteristic function \( c \) is sub – additive, for any \( S \subseteq N \) and \( T \subseteq N \) with \( S \cap T = \emptyset \) we have

\[ c^1(S \cap T) + c^*(S \cap T) \leq c^1(S) + c^*(S) + c^1(T) + c^*(T) \]

**Proof.** By renumbering the indexes, we can assume that \( S = \{1,...,h\}, T = \{h+1,...,k\} \) and \( S \cup T = \{1,...,k\} \). For these sets, it holds that

\[ c^1(S \cup T) + c^*(S \cup T) = \text{Max} \sum_{i=1}^{k} x^i \quad + \quad \text{Max} \sum_{j=h+1}^{k} x^j \quad + \quad \text{Max} \sum_{j=h+1}^{k} x^j \]

\[ = c^1(S) + c^1(T) + c^*(S) + c^*(T) \]

**Theorem 4.** \( c^1(N) + c^*(N) = 1 \).

**Proof.**

\[ c^1(N) + c^*(N) = \sum_{i=1}^{n} w_i \sum_{j=1}^{h} x^i + \sum_{i=1}^{n} w_i \]

\[ = \sum_{j=1}^{h} \left( \sum_{i=1}^{n} x^i + \sum_{i=1}^{n} w_i \right) = \sum_{i=1}^{n} w_i = 1 \]

**2.3 A DEA minimum game**

The opposite side of the game can be constructed by \((N,d)\) as follows:

\[ d^1(k) = \text{Min} \sum_{i=1}^{n} w_i x^i \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (9) \]
\[ w_i \geq 0 \forall i \]

\[ d^*(k) = \text{Min} \sum_{i=1}^{n} w_i x^* \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (10) \]
\[ w_i \geq 0 \forall i \]

The optimal value \( d^1(k) + d^*(k) \) assures the minimum division that player \( k \) can expect from the game.

**Theorem 5.** \[ \sum_{k=1}^{n} d^1(k) + d^*(k) \leq 1 \quad (11) \]

Analogously to the max game, for the coalition \( S \subseteq N \), we define

\[ d^1(S) = \text{Min} \sum_{i=1}^{n} w_i x^i \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (12) \]
\[ w_i \geq 0 \forall i \]

\[ d^*(S) = \text{Min} \sum_{i=1}^{n} w_i x^* \]
\[ s.t \sum_{i=1}^{n} w_i = 1 \quad (13) \]
\[ w_i \geq 0 \forall i \]

**Theorem 6.** The min game \((N,d)\) is super – additive we have \( d^1(S \cup T) + d^*(S \cup T) \geq d^1(S) + d^*(T) \) +\( d^*(S) + d^*(T) \) for each \( S, T \subseteq N \) with \( S \cap T = \emptyset \).

**Proof.** By renumbering the indexes, we have \( S = \{1,...,h\}, T = \{h+1,...,k\} \) and \( S \cup T = \{1,...,k\} \). For these sets it holds that

\[ d^1(S \cup T) + d^*(S \cup T) \]
\[ = \text{Min} \sum_{i=1}^{k} x^i + \text{Min} \sum_{j=h+1}^{k} x^j \]

\[ + \text{Min} \sum_{j=h+1}^{k} x^j \]

\[ = d^1(S) + d^1(T) + d^*(S) + d^*(T) \]

Thus these game start from \( d^1(k) > 0 \) and \( d^*(k) > 0 \) and enlarges the gains by the coalition until the grand coalition \( N \) with \( c^1(N) + c^*(N) = 1 \) is reached.

**Theorem 7.** \[ d^1(S) + d^*(S) + c^1(N \setminus S) + c^*(N \setminus S) = 1 \forall S \subseteq N \]

**Proof.** By renumbering the indexes, we can assume that \( S = \{1,...,h\}, N = \{1,...,n\} \) and \( N \setminus S = \{h+1,...,n\} \). For this sets, it holds that
\[ d^i(S) + d^c(S) + c^i(N \setminus S) + c^c(N \setminus S) \]

\[ = \text{Min} \sum_{j=1}^{n} x_j^1 + \text{Min} \sum_{i=1}^{\frac{n}{2}} x_i^n \]

\[ + \text{Max} \sum_{j=1}^{\frac{n}{2}} x_j^2 + \text{Max} \sum_{i=1}^{\frac{n}{2}} x_i^n \]

\[ = \text{Min} \left( \sum_{j=1}^{n} x_j^1 - \sum_{i=1}^{\frac{n}{2}} x_i^n \right) \]

\[ + \text{Max} \left( \sum_{j=1}^{\frac{n}{2}} x_j^2 - \sum_{i=1}^{\frac{n}{2}} x_i^n \right) + \text{Max} \sum_{j=1}^{\frac{n}{2}} x_j^1 \]

\[ = \text{Min} \left( \sum_{j=1}^{n} x_j^1 \right) \]

\[ + \text{Max} \sum_{j=1}^{\frac{n}{2}} x_j^2 = 1 \]

### 3 Extensions

In this section, we extend basic model to maximal allocated benefit and minimal allocated cost for interval data and discuss the zero weight.

#### 3.1 Maximal allocated benefit

Suppose that there are \( s \) criteria for representing benefits. Let \( \{y_j^1, y_j^2\}_{j=1}^{s} \) be the benefits of player \( j (j = 1,...,n) \) where \( u (u_1,...,u_n) \) is the virtual weights for benefits. Analogous to the expression (1) we define the relative score of player \( j \) to the total scores as:

\[
\frac{\sum_{i=1}^{n} u_i y_j^1}{\sum_{i=1}^{n} u_i (y_j^1 + y_j^2)} \quad (14), \quad \frac{\sum_{i=1}^{n} u_i y_j^2}{\sum_{i=1}^{n} u_i (y_j^1 + y_j^2)} \quad (15)
\]

Player \( j \) wishes to maximize his benefits. We can express this situation by linear programs below:

\[ \text{Max} \sum_{i=1}^{n} u_i y_j^1 \]

\[ \text{s.t} \sum_{i=1}^{n} u_i \left( \sum_{j=1}^{s} y_j^1 \right) = 1 \quad (16) \]

\[ \sum_{i=1}^{n} u_i y_j^1 \geq 0 \quad (j = 1,...,n), \quad u_i \geq 0 \quad \forall i \]

\[ \text{Max} \sum_{i=1}^{n} u_i y_j^2 \]

\[ \text{s.t} \sum_{i=1}^{n} u_i \left( \sum_{j=1}^{s} y_j^2 \right) = 1 \quad (17) \]

\[ \sum_{i=1}^{n} u_i y_j^2 \geq 0 \quad (j = 1,...,n), \quad u_i \geq 0 \quad \forall i \]

The weights of benefits are nonnegative. A characteristic function of the coalition \( S \) is defined by the linear program below:

\[ c^i(S) = \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \]

\[ \text{s.t} \sum_{i=1}^{n} u_i \left( \sum_{j \in S} y_j^1 \right) = 1 \quad (18) \]

\[ \sum_{i=1}^{n} u_i y_j^1 \geq 0 \quad (j = 1,...,n) \]

\[ u_i \geq 0 \quad \forall i \]

\[ c^c(S) = \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^2 \]

\[ \text{s.t} \sum_{i=1}^{n} u_i \left( \sum_{j \in S} y_j^2 \right) = 1 \quad (19) \]

\[ \sum_{i=1}^{n} u_i y_j^2 \geq 0 \quad (j = 1,...,n) \]

\[ u_i \geq 0 \quad \forall i \]

In the program (18), (19) the benefits of all players are nonnegative. Since the constraints of program (18) and (19) are the same for all coalitions, we have the following theorem.

**Theorem 8.** The maximal allocated benefits game satisfies a sub- additive property.

**Proof.** For any \( S \subset N \) and \( T \subset N \) with \( S \cap T = \phi \), we have:

\[ c^i(S \cup T) + c^c(S \cup T) = \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 + \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^2 \]

\[ = \text{Max} \sum_{i=1}^{n} u_i \left( \sum_{j \in S} y_j^1 + \sum_{j \in T} y_j^1 \right) \]

\[ + \text{Max} \sum_{i=1}^{n} u_i \left( \sum_{j \in S} y_j^2 + \sum_{j \in T} y_j^2 \right) \]

\[ = \text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \right) + \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in T} y_j^1 \]

\[ + \text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^2 \right) + \text{Max} \sum_{i=1}^{n} u_i \sum_{j \in T} y_j^2 \]

\[ \leq c^i(S) + c^c(S) + c^i(T) + c^c(T) \]

#### 3.2 Minimal allocated cost

Suppose that there are \( m \) criteria for representing costs. Let \( \{x_i^1, x_i^2\}_{i=1}^{m} \) be the costs of player \( j (j = 1,...,n) \) where \( v (v_1,...,v_m) \) is the virtual weights for costs. Player \( j \) wishes to minimize his costs then we have:
Maximal Allocated Benefit and Minimal Allocated Cost for Interval Data

\[ \begin{align*}
\text{Min} & \sum_{i=1}^{n} y_i x_{i o}^1 \\
\text{s.t} & \sum_{j=1}^{n} y_j \sum_{j=1}^{n} x_{j o}^1 = 1 \quad (20) \\
\sum_{j=1}^{n} v_j x_{i o}^1 & \geq 0 \quad (j = 1, \ldots, n), \quad v_i \geq 0 \quad \forall i
\end{align*} \]

\[ \begin{align*}
\text{Min} & \sum_{i=1}^{n} y_i x_{i o}^a \\
\text{s.t} & \sum_{j=1}^{n} y_j \sum_{j=1}^{n} x_{j o}^a = 1 \quad (21) \\
\sum_{j=1}^{n} v_j x_{i o}^a & \geq 0 \quad (j = 1, \ldots, n), \quad v_i \geq 0 \quad \forall i
\end{align*} \]

The weights of costs are non-negative. A characteristic function of the coalition S is defined by the linear program below:

\[ \begin{align*}
d(S) &= \text{Min} \sum_{i=1}^{n} y_i \sum_{j=1}^{n} x_{j o}^1 \\
\text{s.t} & \sum_{j=1}^{n} y_j \sum_{j=1}^{n} x_{j o}^1 = 1 \quad (22) \\
\sum_{j=1}^{n} v_j x_{i o}^1 & \geq 0 \quad (j = 1, \ldots, n), \quad v_i \geq 0 \quad \forall i
\end{align*} \]

\[ \begin{align*}
d(S) &= \text{Min} \sum_{i=1}^{n} y_i \sum_{j=1}^{n} x_{j o}^a \\
\text{s.t} & \sum_{j=1}^{n} y_j \sum_{j=1}^{n} x_{j o}^a = 1 \quad (23) \\
\sum_{j=1}^{n} v_j x_{i o}^a & \geq 0 \quad (j = 1, \ldots, n), \quad v_i \geq 0 \quad \forall i
\end{align*} \]

In the program (22), (23) the costs of all players are nonnegative. Minimal allocated costs game, satisfies a super-additive property.

**Theorem.** The maximal allocated cost game \((N,c)\) and min game \((N,d)\) are dual games, for any \(S \subseteq N\), we have \(d^1(S) + d^a(S) + c^1(N \setminus S) + c^a(N \setminus S) = 1\).

**Proof.**

\[
\begin{align*}
1(N \setminus S) + c^a(N \setminus S) = \\
\text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \right) + \text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^a \right) \\
= \text{Max} \left( \sum_{i=1}^{n} u_i \left( \sum_{j \in N} y_j^1 - \sum_{j \in S} y_j^1 \right) \right) + \\
\text{Max} \left( \sum_{i=1}^{n} u_i \left( \sum_{j \in N} y_j^a - \sum_{j \in S} y_j^a \right) \right) \\
= \text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in N} y_j^1 - \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \right)
\end{align*}
\]

\[
+ \text{Max} \left( \sum_{i=1}^{n} u_i \sum_{j \in N} y_j^a - \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^a \right) \\
= \text{Max} \left( 1 - \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \right) + \text{Max} \left( 1 - \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^a \right) \\
= 1 - \text{Min} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^1 \right) - \text{Min} \left( \sum_{i=1}^{n} u_i \sum_{j \in S} y_j^a \right) \\
= 1 - d^1(S) - d^a(S).
\]

### 3.3 Avoiding occurrence of zero weights and setting preference on weights

In \(A\) and \(B\), we presented a scheme for determining the weights through the program \((18), (19), (22)\) and \((23)\). Some weight may happen to be zero for all optimal solutions. This means that the corresponding criterion is not accounted for in the solution of the game at all. Let us suppose that all players agree to put preference on certain criteria. The zero weight issue can thus be solved in this way. If all players agree to incorporate preference regarding criteria, we can apply the following "assurance region method". For example, we set constraints on the ratio \(w_i, w_l\) \((i=2, \ldots, m)\) as: \(L_i = \frac{w_i}{w_l} \leq U_i, \ (i=2, \ldots, m)\).

where \(L_i\) and \(U_i\) denote lower and upper bounds of the ratio \(\frac{w_i}{w_l}\), respectively. These bounds must be set by agreement among all players. The program \((5)\) is now modified as:

\[ \begin{align*}
c(S) &= \text{Max} \sum_{i=1}^{n} w_i x_i^1(s) \\
\text{s.t} & \sum_{i=1}^{n} w_i x_i^1 = 1 \quad (24) \\
\sum_{i=1}^{n} w_i x_i^1 & \geq 0 \quad \forall i
\end{align*} \]

Similarly, we can avoid occurrence of zero weight in linear programs of maximal allocated benefits and minimal allocated costs. Then, we have:

\[ \begin{align*}
c^1(S) &= \text{Max} \sum_{i=1}^{n} u_i x_i^1 \\
\text{s.t} & \sum_{i=1}^{n} u_i x_i^1 = 1 \\
\sum_{i=1}^{n} u_i y_i^1 & \geq 0 \quad (i = 1, \ldots, n) \quad (25) \\
\sum_{i=1}^{n} u_i y_i^a & \leq \sum_{i=1}^{n} u_i y_i^a \quad (i = 2, \ldots, s), \quad u_i \geq 0 \quad \forall i
\end{align*} \]
In program (18),(19),(22) and (23) we presented a scheme for computing maximal allocated benefits, minimal allocated cost for coalitions. Also we can compute maximal allocated benefits and minimal allocated costs for members of coalition, using programs (16),(17),(20) and (21). These values can determine the players expected percentages of the total benefit, cost in the game. Each player can increase benefit allocation and decrease cost allocation, establishing coalition. In other words, the possible ultimate benefit allocated to the coalition can increase and minimal cost allocated can decrease once the best circumstance is provided. There is a question, how can player \( j \) establish coalition? Now, knowing this it’s easy for the player to examine which other players, she/he can establish coalition with so that he can reach the ultimate benefit ratio and minimal cost ratio. Each player can establish coalition in different ways, chance coalition, coalition concerning players of minimal allocated benefit (cost), coalition concerning players of maximal allocated benefit (cost), coalition with the player enjoying the ultimate benefit and the players with the minimal allocated benefit (cost), coalition with the player enjoying the minimal benefit (cost) and the players with maximal allocated benefit (cost). Clearly, coalition with the player having the ultimate allocated benefit would be better than the others. Coalition with the player having the minimal allocated benefit would be poorer in comparison to others. Now, it’s easy to understand that a player with the minimal benefit ratio establish a coalition with the player who has allocated the ultimate benefit for him/her self and player with the maximal benefit ratio establish a coalition with the player who has allocated the minimal benefit for him/her self. Having established the coalition the player cost ratio would be less or unchanged. Clearly, player with the minimal cost ratio establish a coalition with the player who has allocated the minimal cost for him/herself, and player with the maximal cost ratio establish a coalition with the player who has allocated the minimal cost for himself. These results represented by the example below.

4 Numerical Example

There are 10 players in this game. Each player uses 2 cost criteria and 4 benefit criteria for first of interval. We compute maximal allocated benefit and the minimal allocated cost for each player for first of interval. Table (2) shows this results (we set constraints in the ratio \( u_i \)). Similarly, we can compute maximal allocated benefit and minimal allocated cost for end of interval.

### TABLE 1

<table>
<thead>
<tr>
<th>Player</th>
<th>( x_{i1} )</th>
<th>( x_{i2} )</th>
<th>( y_{i1} )</th>
<th>( y_{i2} )</th>
<th>( y_{i3} )</th>
<th>( y_{i4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>20</td>
<td>800</td>
<td>200</td>
<td>350</td>
<td>340</td>
</tr>
<tr>
<td>2</td>
<td>70</td>
<td>18</td>
<td>900</td>
<td>160</td>
<td>320</td>
<td>470</td>
</tr>
<tr>
<td>3</td>
<td>80</td>
<td>22</td>
<td>1000</td>
<td>175</td>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>4</td>
<td>110</td>
<td>30</td>
<td>950</td>
<td>185</td>
<td>290</td>
<td>510</td>
</tr>
<tr>
<td>5</td>
<td>90</td>
<td>17</td>
<td>960</td>
<td>186</td>
<td>280</td>
<td>480</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>24</td>
<td>870</td>
<td>210</td>
<td>360</td>
<td>370</td>
</tr>
<tr>
<td>7</td>
<td>65</td>
<td>26</td>
<td>780</td>
<td>165</td>
<td>300</td>
<td>440</td>
</tr>
<tr>
<td>8</td>
<td>75</td>
<td>32</td>
<td>670</td>
<td>150</td>
<td>400</td>
<td>500</td>
</tr>
<tr>
<td>9</td>
<td>50</td>
<td>29</td>
<td>810</td>
<td>170</td>
<td>410</td>
<td>510</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>16</td>
<td>910</td>
<td>190</td>
<td>420</td>
<td>390</td>
</tr>
</tbody>
</table>

We now apply this approach to the data above.

### TABLE 2

<table>
<thead>
<tr>
<th>Player</th>
<th>maximal allocated benefit for first of interval</th>
<th>minimal allocated cost for first of interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11.17%</td>
<td>6.71%</td>
</tr>
<tr>
<td>2</td>
<td>10.66%</td>
<td>7.69%</td>
</tr>
<tr>
<td>3</td>
<td>11.56%</td>
<td>9.40%</td>
</tr>
<tr>
<td>4</td>
<td>11.56%</td>
<td>12.82%</td>
</tr>
<tr>
<td>5</td>
<td>11.10%</td>
<td>7.26%</td>
</tr>
<tr>
<td>6</td>
<td>11.73%</td>
<td>7.3%</td>
</tr>
<tr>
<td>7</td>
<td>9.98%</td>
<td>8.72%</td>
</tr>
<tr>
<td>8</td>
<td>11.35%</td>
<td>10.7%</td>
</tr>
<tr>
<td>9</td>
<td>11.63%</td>
<td>6.71%</td>
</tr>
<tr>
<td>10</td>
<td>11.91%</td>
<td>6.84%</td>
</tr>
</tbody>
</table>

7th player has the minimal allocated benefit and the 10th has the maximal allocated benefit. Also, the 1st has maximal allocated cost and 4th and 9th have the minimal allocated cost.
Maximal Allocated Benefit and Minimal Allocated Cost for Interval Data

Consider in Table 3(A), $S_1$ is chance coalition, $S_2$ is coalition concerning players of minimal allocated benefit, $S_3$ is coalition concerning players of maximal allocated benefit and in Table 3 (B), $S_4$ is chance coalition $S_7$ is coalition concerning players of minimal allocated cost, $S_8$ is coalition concerning players of maximal allocated cost, $S_9$ is coalition with the players enjoying the minimal cost and the players with maximal allocated cost, $S_{10}$ is coalition with the players enjoying the maximal cost and the players with minimal allocated cost.

Table (3) shows, establish the coalition the player benefit ratio are increased or unchanged. Also, establish the coalition the player cost ratio are decreased or unchanged. 7th player with the least benefit ratio has the most benefit in $S_1$. 10th player with the maximal benefit ratio has the most benefit in $S_5$. 4th player with the most cost ratio has the least cost in $S_9$.

<table>
<thead>
<tr>
<th>Player $j$</th>
<th>Modified maximal allocated benefit for first of interval</th>
<th>Modified minimal allocated cost for first of interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.22%</td>
<td>6.80%</td>
</tr>
<tr>
<td>2</td>
<td>10.33%</td>
<td>9.13%</td>
</tr>
<tr>
<td>3</td>
<td>11.12%</td>
<td>10.53%</td>
</tr>
<tr>
<td>4</td>
<td>10.90%</td>
<td>14.47%</td>
</tr>
<tr>
<td>5</td>
<td>10.78%</td>
<td>11.34%</td>
</tr>
<tr>
<td>6</td>
<td>9.91%</td>
<td>7.52%</td>
</tr>
<tr>
<td>7</td>
<td>9.22%</td>
<td>8.84%</td>
</tr>
<tr>
<td>8</td>
<td>9.11%</td>
<td>10.24%</td>
</tr>
<tr>
<td>9</td>
<td>10.20%</td>
<td>6.98%</td>
</tr>
<tr>
<td>10</td>
<td>10.52%</td>
<td>12.41%</td>
</tr>
</tbody>
</table>

Conclusion

In this paper, we have studied the common weight issues that connect the game solution with arbitrary weight selection behavior of the players. Regarding this subject, we have proposed a method for compute maximal allocated benefit and minimal allocated costs for interval data. We have introduced coalitions and the ways for finding the best coalitions. In this sense, avoided occurrence zero weight by assurance region method. Furthermore a numerical example, have been calculated with proposed ways, has been considered.

Acknowledgment

We are grateful to the anonymous referees and Professors Nakabayashi and Tone for their helpful comments and suggestions. This paper is supported by Islamic Azad University Tabriz Branch and Young Researchers Club.

References:

[6]. W.W. Cooper, L.M. Seiford, and K. Tone, Data envelopment analysis, a comprehensive text with models application references and DEA- solver software, Boston: Klawer Academic Publishers, 2000

<table>
<thead>
<tr>
<th>coalition</th>
<th>minimal allocated benefit for coalition</th>
<th>members of coalition</th>
<th>minimal allocated benefits of total game benefits</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>31.77%</td>
<td>10</td>
<td>12.25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>11.94%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>10.51%</td>
</tr>
<tr>
<td>$S_2$</td>
<td>33.76%</td>
<td>10</td>
<td>11.92%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>12.44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>13.56%</td>
</tr>
<tr>
<td>$S_3$</td>
<td>31.52%</td>
<td>5</td>
<td>11.47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>11.21%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>10.51%</td>
</tr>
<tr>
<td>$S_4$</td>
<td>31.55%</td>
<td>10</td>
<td>11.47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>12.27%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>11.73%</td>
</tr>
<tr>
<td>$S_5$</td>
<td>29.94%</td>
<td>10</td>
<td>12.09%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>10.12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>10.13%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>coalition</th>
<th>minimal allocated cost for coalition</th>
<th>members of coalition</th>
<th>minimal allocated cost of total game cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>28.86%</td>
<td>1</td>
<td>6.71%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>11.69%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>7.3%</td>
</tr>
<tr>
<td>$S_2$</td>
<td>13.42%</td>
<td>1</td>
<td>5.47%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>6.71%</td>
</tr>
<tr>
<td>$S_3$</td>
<td>35.17%</td>
<td>4</td>
<td>12.55%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>10.06%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>9.21%</td>
</tr>
<tr>
<td>$S_4$</td>
<td>31.54%</td>
<td>4</td>
<td>11.54%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>10.06%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>6.71%</td>
</tr>
<tr>
<td>$S_5$</td>
<td>28.19%</td>
<td>4</td>
<td>10.70%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9</td>
<td>6.71%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>6.71%</td>
</tr>
</tbody>
</table>

TABLE 3 (A)

TABLE 3 (B)
An Investigation of the Protection Performances for Current-Limiting Fuses Using Stainless Steel Wires

NARONG MUNGKUNG\textsuperscript{1}, SOMCHAI ARUNRUNGRUSMI AND TOSHIFUMI YUJI\textsuperscript{2}

\textsuperscript{1}Department of Electrical Technology Education
King Mongkut’s University of Technology Thonburi, Bangkok, THAILAND
\textsuperscript{2}Faculty of Education and Culture, Miyazaki University, JAPAN
e-mail: narong_kmutt@hotmail.com

Abstract: - Recently, owing to a high demand of and high dependence on electric power, the reliability of the distribution voltage of electric power systems has become more important. 22 kV high voltage cutout fuses have been used widely, but some melting accidents have occurred due to pollution/contamination of bushings. As the number of contaminants in bushings increases, the electric fields on the fuse element surface rise to a critical level, triggering corona discharge. A fuse element material that is more resistant to corona discharge than the Ag element is desired. A new type of current-limiting fuse made of stainless steel wires, which is low-cost and simple in design has been developed. Characteristics tests, including a blocking test with a high surge current when a voltage is applied, have been conducted. Using a model fuse, it was confirmed that this new type of fuse possesses similar performance characteristics to those with an Ag element.

Key-Words: - Large current breaking test, Stainless steel wire, 22kV high voltage cutout, Arc extinguishing tube

1 Introduction

In response to improvements in the reliability of electric supply as well as increases in short-circuit capacity, which accompany increases in electrical demand, there has been a demand for verification of improvement in the blocking capabilities of high currents in switching protection instruments used for electrical power distribution, as well as characteristics related to these blocking capabilities [1]. There has also been a gradual increase in the number of equipment with high-pressure cutout fuses installed.

In 1995, there were numerous melting accidents involving 22 kV high-voltage current-limiting cutout fuses in Japan's Kyushu coastal district [2]. Based on the authors' research results, such accidents are caused when, for example, the porcelain tube of a high-voltage cutout fuse is polluted or contaminated. In such cases, a highly electrical field is produced between the porcelain tube and fuse element, and a corona discharge is generated in the fuse element area, causing a melting accident to occur due to thermal degradation of the silver element (hereinafter referred to as “Ag”) used regularly in fuses [3,4]. As a result, improvements are necessary, and a fuse structure that prevents fuse-melting accidents caused by thermal degradation must be developed. High-voltage cutout fuses using stainless steel wires are being developed by Kakimoto [5] et al. However, since such fuses are of an emission-type structure, there is the problem of cracked gas [6] being emitted from the arc extinguishing tube, and a fusion sound being heard when the fuse melts. This in turn causes concern to the living environment.

The purpose of this research is to develop a low-cost 22 kV high-pressure cutout current-limiting fuse, by using stainless steel wires (SUS304) that are resistant to corona discharges as the element, and by comparing this to the presently used current-limiting fuse that uses Ag [7,8].

In this experiment, a 22 kV high-pressure cutout current-limiting fuse (hereinafter referred to as “model fuse”) with a newly structured element portion using stainless steel wires instead of the currently used Ag element was first experimentally created. Then, an oscillatory impulse high-current blocking characteristics test (hereinafter referred to as “blocking test”), which is one of the characteristics tests for fuses, was carried out. It was confirmed that the oscillatory impulse high-current blocking characteristics of the model fuse that is proposed here, a simulated fuse with a thin wire portion in the center, has equivalent blocking characteristics as fuses that use Ag. The results that were derived after examining the blocking characteristics of stainless steel wires, as substituted for Ag, are reported below.

2. Experimental Arrangement

For blocking tests, flashovers are prevented in the main wire area, where recovery work is difficult, and a lighting surge flashover is made to occur near the transformer. In such a state, the blocking capabilities of
the high-pressure cutout fuse to completely block the short-circuit current that starts flowing are examined [9].

![Experimental circuit for large current breaking test](image)

**Fig.1** Experimental circuit for large current breaking test.

Figure 1 shows the circuit for the blocking characteristic test. The power source and measuring apparatus used in this experiment are the power voltage of the arrester and a shock voltage current generator used for current characteristic tests, respectively. In the blocking test, the model fuse was substituted for the lightning protection component, and the terminal voltage and current characteristics were surveyed by measuring the terminal voltage and current waveform with an oscilloscope (Yokogawa; DL1540).

![Peculiar current waveform](image)

**Fig.2** Peculiar current waveform

Figure 2 shows the prospective current (ip) waveform when the current voltage (Vco) of the condenser is 15 kV. Two types of model fuses with different structures were used for the blocking test. Each of these is shown in Figure 3 and Figure 4. As shown in Figure 3, Type (1) represents the model fuse with a structure where one stainless steel wire is set horizontally linearly in the atmosphere. As shown in Figure 4, the Type (2) model fuse is composed of two stainless steel wires as the fuse element, but the central part of the element is composed of one wire. Since the model fuse will definitely melt in the central part, Type (2) is structured with a gap (d). The structure of this gap has been examined by giving sufficiently taking the heat conduction equation [10] into account. The purpose of the Type (2) model fuse is to extinguish arcs generated in the fuse's central element area when melting of the fuse occurs [11]. To do so, a transparent arc extinguishing tube (Poly Tetra Fluoro Ethylene; PTFE) with an inner diameter (hereinafter referred to as " φp ") of 0.5 mm was affixed, and the acrylic pipe was filled with arc-extinguishing sand.
3. Experiment results

3.1 Pre-arching process of fuse element
Figure 5 and Figure 6 are photographs of the Type (2) fuse melting as taken with the image converter camera when the applied voltage was 30kV, the element area consisted of Ag with a diameter (hereinafter referred to as "φs") of 0.1 mm and stainless steel wire with φs=0.2 mm, and an arc-extinguishing tube in the shape of φp=1.0 mm with a length (hereinafter referred to as "lp") of 100 mm was used. The value of resistance differs [12] for the material of the two different fuse elements, but characteristics tests were conducted such that the structure of the element part was the same for both model fuses.

In Figure 5, the emission of light is spread out overall, but in Figure 6, a strong emission is prominent in the center of the element part of the model fuse. This confirms that the gap in the center of the fuse element definitely contributes to the melting mechanism in the central area. It is assumed that for Ag, depending on the quality of the material, if an unexpected electric current passes through the Ag, an adiabatic increase in temperature occurs [13] and there is light emission overall.

Figure 7 shows the voltage and current waveforms measured at the same time the photographs were taken. As shown in (a), which represents the model fuse that uses Ag, melting occurs at the current waveform's peak value of approximately 4.5 kA, about 6.0 µs after power distribution begins. The arc current at approximately 40µs after the start of power distribution is about 300A. For the model fuse that uses stainless steel wires, Figure 7 (b) shows that 7.0 µs after power distribution begins, the current waveform's peak value is approximately 2.8 kA, at which fusion occurs. After a current distribution period of approximately 22 µs, the electric current subsides to nearly 0.0 A. No. 5 of Figure 6 (which corresponds to (5) of Figure 7 (b)) is a photograph taken approximately 8.0 µs after power distribution started, and is the first frame in which an emission can be clearly confirmed. This matches the peak value of the voltage and current waveform in Figure 7 (b), which confirms that melting occurred in the element portion of the fuse. The model fuse using stainless steel wires was confirmed as having the same blocking ability as the fuse using Ag.

3.2 Results of blocking test
Figure 8 shows the power voltage between fuse terminals and the temporal change during the blocking test that was performed using the Type (2) model fuse. In this case, the stainless steel wires were of φs = 0.1 mm, φp = 0.5 mm, and lp = 100 mm. In (a) and (b) of the
same figure, the charging voltage (V\text{co}) of the condenser is 22.5 kV and 37.5 kV, and the upper part of the waveform represents the terminal voltage, while the lower part represents the current waveform. When V\text{co}=30\text{kV}, the same waveform as the one for 22.5 kV was obtained. In a regular fuse melting process, after power distribution begins, the fuse element becomes heated by Joule heat, from the power distribution current \[14\]. The period from when power distribution starts and meltdown begins (hereinafter referred to as “Tf”) is the meltdown time (hereinafter referred to as “Tm”), and the period from when power distribution starts until the arc is generated is the fusion time (hereinafter referred to as “Tp”). The electric currents/voltage waveforms from the results of the blocking characteristics test are shown as the results from only one test. As shown in (b) of the same figure, an increase in applied voltage, or injected energy, results in a shortening of the fusion time. This is referred to as the “exploding wires phenomenon,” and a large amount of research is being carried out in regards to this phenomenon \[15,16\]. The breaking of copper wires by high currents is also reported on as a “fracture phenomenon” by Professor Arai of the Tokyo Denki University. The stainless steel wires used in this experiment, however, are composite materials and are not categorically the same as the ones discussed in reference material \[17\].

Figure 9 shows the power voltage and temporal change when the Type (2) model fuse is filled with arc-extinguishing sand. Compared to Figure 8, there is a prominent change in the power voltage and current waveform, depending on the applied voltage. Figure 9 (a) differs from Figure 8 in that the current value decreases to 0.0 A after surpassing Tp, due to the absorption of heat by the arc-extinguishing sand. A reascent in the current cannot be confirmed. Once the current is reduced to nearly 0.0 A, the current value rises again. In Figure 9 (b) and (c), the current is reduced to nearly 0.0 A, as an arc is generated at a point when the current waveform has passed its peak value, after starting to decrease due to a sudden increase in terminal voltage immediately after the Tf value. After fusion of the fuse element portion, the arc-extinguishing tube was visually confirmed as having slightly melted. This is because the structure of the gap between the arc-extinguishing tube and stainless steel wire is small, and fusion occurs when the central area of the stainless steel (one wire’s worth of element) is effectively heated by a high current. Generally, when a metal wire “explodes” through melting, there is (1) an increase in temperature while in the solid condition, (2) a melting process after reaching the melting point, (3) an increase in temperature while in the liquid state, and (4) vaporization at boiling point, after which an arc is generated, undergoes a four-stage change, and then the

\[
\frac{1}{R} \frac{dR}{dt} = K_{\text{gas}} (P_a - R \dot{i}) \]

\[1\]

Fig.7 Voltage and current waveforms.
metal is fused [19]. For the model fuses that were experimentally created, after power begins in the stainless steel element, the melting point is reached after an increase in temperature by Joule heat while in the solid state. Afterwards, the stainless steel element melts, transforms to a liquid state, and increases in temperature. It then reaches boiling point, vaporizes, undergoes a four-stage change of state, leading to the blocking of the fuse. This is based on the generalization of the authors’ past research results [20]. After the blocking of the fuse, there were also some residual substances, such as soot, inside the acrylic pipe, from after the fusion of the arc-extinguishing tube.

Figure 10 shows the relationship (hereinafter referred to as “Tm-V characteristic”) between Tm and applied voltage (hereinafter referred to as “V”) of the Type (2) fuse. Tm-V indicates that the heating characteristics inside the arc-extinguishing tube are influenced by the inner diameter of the arc-extinguishing tube. The figure indicates 3 ensemble average values; those marked with ▲ (when there is no arc-extinguishing sand in the acrylic pipe) and △ (when the acrylic pipe is filled with arc-extinguishing sand) are for when 1p=80 mm, and those marked with ■ (when there is no arc-extinguishing sand in the acrylic pipe) and □ (when the acrylic pipe is filled with arc-extinguishing sand) are for when 1p=100 mm. Figure 10 also shows that after the model fuse melted, the fusion time of the model fuse without arc-extinguishing sand in the acrylic pipe is shorter than for the model fuse with arc-extinguishing sand in the acrylic pipe. This is due to the heat dissipating more easily if the fuse element comes in contact with the arc-extinguishing sand.

3. The gap that was designed intentionally in the fuse element area definitely contributes to the fusing mechanism in the fuse’s central area.

4. Emissions confirm that fusion of the fuse element starts in the central area when an oscillatory impulse high-current is applied.

In the future, the application of stainless steel wires in actual 22 kV high-pressure cutout fuses will be considered after clarifying the characteristics of AC voltage superposition through experiments.
Acknowledgement
We would like to thank the National Research University Project of Thailand's Office of the Higher Education Commission for financial support

Reference

Fig.9 Voltage and current waveforms of Type② in 0.1 mm wire fuse with PTFE tube and arc extinguishing sand
Fig. 10 Relation between applied voltage and breaking time.


Narong Mungkung was born in Lopburi Province, Thailand, in 1965. He received the B.S.Ind.Ed. degree in electrical engineering from King Mongkut’s University of Technology, Thonburi, Thailand, in 1992, the M.S.Tech.Ed degree in electrical technology from King’s Institute of Technology, North Bangkok, Thailand, in 1998, the M.S. degree and D.Eng in electrical engineering from Nippon Institute of Technology, Saitama, Japan, in 2000 and 2003. Currently he is Assoc Prof. of Department Electrical Technology Education, Faculty of Industrial Education and technology, King Mongkut’s University of Technology Thonburi, Thailand. His main research interests are in engineering education, electrical discharge, plasma application, energy conversion system and high voltage.

Somchai Arunrungrasmi was born in Chonburi Province, Thailand, in 1972. He received the B.S.Ind.Ed. degree in electrical engineering from King Mongkut’s University of Technology, Thonburi, Thailand, in 1995, the M.S. degree in electrical engineering from King Mongkut’s University of Technology Thonburi, Thailand, in 2000. Currently he is a lecturer of Department Electrical Technology Education, Faculty of Industrial Education and technology, King Mongkut’s University of Technology Thonburi, Thailand. His main research interests are in engineering education, image processing, geographic information system and system design.

Toshifumi Yuji was born in Miyazaki, Japan, on June 2, 1976. He received the M.E. degree from the University of Miyazaki, Miyazaki, Japan, in 2003 and
the Ph.D. degree from the Tokyo Institute of Technology, Tokyo, Japan, in 2007. He became a Research Associate at Hiroshima National College of Maritime Technology in 2005 and Assistant Professor at Oita National College of Technology in 2006. He is currently a Lecturer in Faculty of Education and Culture, University of Miyazaki, Miyazaki, Japan. His research interests are plasma application and electric power installation. He is a member of the Institute of Electrical Engineers of Japan and the Japan Society of Applied Physics.
Assessment and forecast stability of slopes in the Carpathian and sub-Carpathian area of Dâmboviţa and Prahova. Classification and risk factors

ALEXANDRU ISTRATE, MADALINA FRINCULEASA, GICA PEHOIU
Department of Geography, Faculty of Humanities
“Valahia” University of Târgovişte
Str. Lt. Stancu Ion, no. 34-36, 130105, Târgoviste, Damboviţa
ROMANIA
alexandruistrate@yahoo.com; madalina_chitescu@yahoo.com; gpehoiu@yahoo.com; www.valahia.ro

Abstract: To determine risk factors in assessing and forecasting, the stability of slopes is a need for classification and ranking, so it is possible to establish a degree of urgency in the analysis of paddocks.

The paper is addressed to the risk factor of landslides, which affects the slopes of space flyisch and fore-syncline of the Carpathian space between rivers Argeş and Proviţa. Analysis of spatial distribution of landslides shows the role of geo-structural factor, that has the regional breakdown.

Within a geo-structural unit, lithologic factor can be considered as constant, no matter at what level scale geological and geo-chronological formations fall, favoring the occurrence of landslides.

Spatial distribution of landslides seems no coincidence, these focusing on certain areas that overlap alignment of tectonic fractures, usually crossing the general geological structure. Therefore, in the researched area, tectonic factor must have a large slope stability assessment and prognosis. In areas affected by tectonic fractures are found to accelerate the process of altering the rocks, which causes a reduction in the resistance forces on the one hand, and increasing gravitational instability, on the other. However, faults and their surrounding areas promote the movement of groundwater, contributing to increased gravitational instability.

Distribution analysis of landslides on the slopes is for every slope values that determine the maximum frequency reached for relatively small, up to 15°, of flyisch, and up to 30°, where is located Carpathian fore-syncline. The explanation of such sharing should be sought in the mineralogical composition of the deposits involved in the phenomenon of slip and geological structure of the perimeter analyzed. From here we can see the correlation between various factors which condition each other and form the whole trigger mechanism of landslides.

Key-words: geo-structural factors, spatial distribution, frequency of landslides, tectonic factor, illite-montmorillonite couple, fault.

1 Introduction
Slope instability is a constant presence in the of Carpathian flyisch and fore-syncline, being, on the one hand, a consequence of natural evolution of relief, and on the other, a result of human activity on environmental factors. Concerns for the limitation of the effects on human habitat appear to be insufficient, and that due to the volume of work that it entails, and thus their costs.

Therefore, it must be taken into account the problems of assessment and forecasting the behavior of slopes in time, approach that proves extremely difficult and complicated.

The assessment and forecast models of slope stability are considered eight factors of favorability, namely: arrangement of structural-tectonic, lithologic and petrographical nature, geomorphological configuration of the land surface, climate and hydrological regime, the movement of groundwater, seismicity region, arboreal vegetation coverage and impact of human activity.

Of the more than 150 landslides analyzed over the years include the most detailed elements of a number of 136, which covers the pleated and folded structures in the space of Carpathian fore-synclinal interfluve Argeş - Proviţa. Their monitoring was conducted over the period 1965-2000 years.

Preliminary data included their knowledge of their spatial and temporal location, information on gradient slopes and areas affected.

Just a few of these slides were taken from detailed data in respect of spatial geometry, physical and mechanical parameters of the deposits involved the phenomenon of slip and the geological sub-layers, hydrological and hydro-geological conditions, analysis of the status of efforts slopes etc.
2 Distribution of landslides

By their location in relation to geological structure has resulted in a statistical distribution, which allows some conclusions in the risk assessment and classification of the region. Correlation in distribution of landslides with flyisch units (Fig. 1) and fore-synclinal creates an overview of the extent and spatial distribution of the phenomenon. One can easily see that the highest frequency belongs to flyisch, which has values nearly double (86 landslides) in relation to fore-synclinal (50 landslides) (Fig. 2). This image allows some comments on favorability factors in landslide occurrence.

But within a region with high risk of instability phenomena in the development of the land, it is noted that the distribution and frequency of a certain structure, which should found significant weight factors in assessment and forecast slope instability.

For the regional analysis of the phenomena of land instability, lithologic factor has a major share.

On the canvas of Ceahlău, land slip processes are located on the Early Cretaceous formations, mainly on marly facies, the marly gritty schists, which are Sinaia layers and Comarnic layers.

Spectacular phenomena of instability are giving by post-tectonic deposits in the internal party of Ceahlău canvas, belonging Turonian - Senonian (Late Cretaceous), in the perimeter Dealul Frumos - Runcu. All upper Cretaceous, the period Vraconian - Turonian, in the predominantly shale facies, is due to landslides in the Macla and Teleajen canvases, in the line Măgura - Fieni - Vulcana de Sus.

The relative uniformity of lithologic facies in geological structures out lengthwise (from east to west) would suggest that the distribution of instability phenomena have the same spatial distribution, with variations depending on local conditions (slope, arboreal vegetation coverage, etc.). However, there is a concentration of landslides in some alignments that overlap large tectonic faults crossing.

Another area with relatively high frequency of instability is the process which overlaps post-tectonic structures of internal flyisch, consisting of Eocene deposits or Paleogene - Lower Miocene of synclinal Râul Alb - Malu cu Flori - Văleni Dâmboviţa and Talea - Runcu.

South of Tarcău canvas develops a fore-syncline where also the land sliding phenomena have a certain distribution and frequency.

Therefore, one can say that the regional factor in geo-structural has a large share of the risk assessment of landslides and it correctly reflects the distribution and frequency.
faulted. In this case, the lithologic factor favors the occurrence of landslides in clay-shale facies of Pucioasa layers. But, as in the cases listed above, there is a group of landslides after some alignments.

In the fore-syncline, lithologic factor remains the main favorability factor by shale facies of Badenian, Pontian and Dacian sequence of clay-sand. But some landslides are grouped, as with flysch of some areas, although the area of outcrop of geological formations that favor the occurrence of landslides is much higher.

Quaternary, especially the Piedmont area of Cândești developed interfluve Dâmbovița - Argeș, favoring instability phenomena through intercalations of clay and clay-shale of Cândești facies. In this case, should be considered also hydro-geological factor, which has a special weight because of the presence of aquifers under pressure.

3 Distribution of landslides in relation to geological formations

Distribution of landslides in geological formations allows some comments (Fig. 3). The analysis of this sharing shows the highest frequency observed in the area of outcrop of Paleogene of canvas Tarcău and post-tectonic structures of Cretaceous canvases (50 landslides). A lower frequency submit Cretaceous and Pliocene deposits (36, respectively 38 landslides), which would suggest that lithologic factor is a constant.

Regarding spatial distribution patterns of structures and lithologic complexes ought to landslides are located on the east-west directions. The situation would be consistent with the broad of geo-structural units, but the reality is different, they focusing on specific alignments, not incidentally overlapping large tectonic fractures.

Returning to Fig. 1, we can notice that the field of Cretaceous flysch, Lower Ialomicioara Valley is affected by a transverse fault of the geological structure in the sequence along which many landslides can be very deep. In this space, canvas of Ceahlău has a complex structure, given by a succession of syncline fold, some faulted overfolds, being affected across the fault of Ialomicioara.

To the south, the area of Teleajen canvas, landslides are located along faults of superposition at the head of tectonic structures, as in the perimeters of settlements Cucuteni and Fieni, or the canvas area in the territory of the localities Macla and Măgura, on the Bizău Valley.

Similar situations on can meet in the field of external flysch, respectively Tarcău canvas and post-tectonic structures. An obvious example is Cricovul Dulce Valley, where a series of landslides was produced along the fault line upper Vișinești, a transverse fault of general geological structure of Tarcău canvas. A cross-cutting aspect have also the perimeters of locality Vârfuri and Ulmetului fault, located slightly west.

In the same way can be seen sliding Vulcana Băi village perimeter, following, more or less, the fault of Ialomița.

Fig. 3. The distribution of the landslides according to the geological formations

The role of tectonic factor can be shown also in the fore-syncline, such as northern flank of syncline Valea Lungă, where cross faults is related to the extended Tarcău canvas. As can be evidenced Ştileb fault, along the valley with the same name, is extended to the south of Ulmetului fault. Some landslides from the same flank can be traced to the southward extension of the faults Vișinești and Vârfuri.

Also, in the Moreni town perimeter are some landslides that may be linked with fault of salt in the northern flank of anticline diapir with the same name. The same situation is encountered in Ocnita village area, where are found some slides related to faults of mounting cross Lăculețe - Ocnita - Colibași.

It is to be noticed a differentiation factor in the distribution of tectonic instability phenomena of the land, explained by the following processes:
- The existence of tectonic forces, for mobility or reactivation of faults of others, which has the effect of changing the status of efforts of slopes;
- A system of cracks of one side of the fault plane, which results in a decrease in rock strength and acceleration of deterioration processes;
- Hydro-geological faults, which also favors the acceleration of deterioration and hence a
4 The correlation between landslides and slope angle

Distribution analysis of landslides based on slope angle of the slope founds an inverse proportionality. Specifically, the frequency increases with decreasing slope (Fig. 4), which means that the slope is not a big major risk factor, as specified in the regulations for the development of risk maps to slip the slopes.

However, such a result must have an explanation. From the analysis pointed graph is observed number of 80 cases, the slopes up to $15^\circ$, which represents a rate of 59%, 28 cases of slopes corresponding to $30^\circ$, with a rate of 21%, and the remaining 20% returning more than $30^\circ$.

In the case of fore-syncline, an important one and subsidence have produced the slopes of up to $30^\circ$, representing 32%, similar to the previous (Fig. 6).

Therefore, regarding fore-syncline, slope inclination express in some measure proportionality between it and the risk of landslides.

Returning to Fig. 2, where the frequency of landslides is higher unity of flyisch, to landslide, might suggest that the slope would play an important role in triggering landslides. This correlation, regionally, is consistent with the role of slope in landslide occurrence, whereas, in fact, the formation unit slope has a high percentage compared with the fore-syncline. Therefore, at the regional level, slope factor can express the degree of risk of occurrence of landslides. But within a geo-structural unit with a high risk, it is noted that landslides have high frequency on sections of the slope with low gradient.

Factor explaining this behavior is the mineralogical, that the report is determined by illit - montmorillonit couple.
Numerous laboratory studies were performed on different types of slide show that illit is transforming in montmorillonit, which increases under clays fatty deposits surrounding sensitive clays, which have a high capacity to absorb water (Fig. 7). Water absorption may be so great that there is a gravity water forming an undercurrent. However, the occurrence of gravitational water greatly reduces internal friction and cohesion forces of clay particles.

Fig. 7. The corellation between flowing limit and plasticity index - landslide Vulcana Bai

In this situation occurs hydro-geological factor, the pore water pressure reducing the normal component of gravity and, thus, reducing the resistance forces.

The role of hydro-geological factor is accentuated when the groundwater is mineralized, especially the sulphate, as in the Paleogene formations of syncline Slănic - Bezdead - Buciumeni or Râul Alb - Malu cu Flori - Văleni Dâmboviţa. Furthermore, the contribution of groundwater mineral fraction accelerates the process of transition illit to montmorillonit.

Therefore, triggering landslides on slopes of less than 15° is explained by an accumulation of factors, including the hydro-geological and mineralogical ones, which have a high share.

Lower frequency of landslides on slopes steeper slope can be explained by several findings of the field and laboratory work.

In areas with high slopes, thickness of altered shell is lower, hence the lower value of tangential component of gravity.

In other cases, the size composition of deposits is somewhat rough, which gives greater permeability and allows faster drainage of the slope. Thus, contact time rock - water is relatively short, so the transition illit - montmorillonit does not occur or is significantly slowed.

The analysis in this case shows that the conversion to montmorillonit is produced from top to bottom, having an important role it rainfall infiltration.

5 Conclusions

Analysis of landslides in the Carpathian region of fore-syncline and flysch in the interfluve Argeş - Provîţa refer some matters which may result in assessment and forecast the behavior of slopes in time.

At regional level, geo-structural factor has a defining role in zoning a territory, increasing its share with a degree of geological and tectonic complexity.

Inside high-risk regions such as flysch and even Carpathian fore-syncline, lithologic factor can be considered a constant in achieving conditions of landslide occurrence.

In the area examined was noted that the slip phenomena have a certain distribution in space, which incidentally are associated with large tectonic fractures, with a regional or local aspect. On notice the important role of transverse faults deterioration of geological structure, which would suggest further efforts to introduce changes in steady-state of slopes. Therefore, tectonic factor becomes especially important in particular for forecasting the evolution of slopes.

Very good results were obtained by studying the frequency of landslides based on slopes gradient. Logically, the slope should have a high proportion in evaluating of slopes stability in a region. However, to obtain a high frequency of landslides on slopes of up to 15-16°, in the case of Carpathian flysch, and in proportions roughly equal to the slopes of up to 15° and from 15 to 30°, in the fore-syncline region.

The explanation of such behavior is given by the contribution of hydro-geological and mineralogical factors that have a role in triggering the instability phenomena of the land.

Therefore, the eight factors that determine slope instability should have a certain hierarchy, some of them with regional importance, while others introduce spatial differences and make us understand that the distribution of landslides is not entirely coincidental. On the other hand, it has to be introduced a factor to expressing the effects on the human habitat, so that the following classification and ranking of landslides in the degree of urgency in taking measures to limit property damage and human victims.

In connection with those stated, may draw attention to the fact that landslides are a major risk
factor in terms of population and human settlements. However, the more susceptible to such phenomena are especially localities in the geographical area of the rivers Dâmboviţa and Prahova.

References:
Numerical results for the blood micro-circulation modelled by a porous media

YVON MADAY and NOURA MORCOS
Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie (Paris 6)
Boite Courrier 187, 4, place Jussieu, 75252 Paris cedex 05
FRANCE
maday@ann.jussieu.fr; noura.mmm@hotmail.com; http://www.ann.jussieu.fr/

TONI SAYAH
Département de mathématiques, Faculté des sciences
Université Saint-Joseph
B.P 11-514 Riad El Solh, Beyrouth 1107 2050
LIBAN
tsayah@fs.usj.edu.lb; http://www.fs.usj.edu.lb/profs/tsayah.htm

abstract: We consider a non periodic homogenization model to simulate the blood flow at the level of the microvascularised tissues. We solve elliptic partial differential equations with two length-scales on the domain and we show numerical results and comparisons.

Key-Words Porous media, stochastic Homogenization, reduced basis method.

1 Introduction
Blood flow at the level of the microvascularised tissues can be considered as a flow through a porous media. This gives rise to models of homogenization types in which we solve elliptic partial differential equations (Darcy’s law) on structures having two length-scales. Due to inhomogeneity, the homogenization is not periodic and the simulations involve the resolution of a large number of parameterized cell problems. After solving the various cell problems, a global macroscopic homogenized problem can be set up and solved. The reduced-basis techniques of numerical simulations allow to speed up the computation of this large number of cells problems without any loss of precision. This reduced-basis method is based on the weak variational form and relies on a Galerkin method on an appropriate discrete space built from preliminary generic computations. The control of the approximation’s error on some output of interest is performed through a posteriori error estimation. An application of this is for simulation of the propagation of an image-enhancing tracer in the capillary bed. This will provide insight for the interpretation of the tracer behavior for the measurement and classification of cancer tumor. The article is organized as follows. In section 2, we present our problem and define the corresponding parameters. In section 3, we give a presentation of homogenization theories. We introduce in the section 4 the RB approach for a parameterized cell problem and notably derive a posteriori error bounds. In section 5, we propose a new scalar product to remedy to the problem of convergence and we end the paper with numerical results and comparisons.

For the existing numerical homogenization strategies, we can cite [9], [3], [5], [2], [11], [1], [6], [7], [8]. For the RB, we can cite [10], [4] ...

2 Description of the model
Let $\Omega$ be an open $C^1$ bounded domain of $\mathbb{R}^n$ with a boundary $\Gamma = \partial \Omega$. We denote by $\vec{n}$ the exterior unit normal to $\Gamma = \Gamma_D \cup \Gamma_N \cup \Gamma_{N1}$. We are interested in the behavior of a sequence of scalar functions $u^\varepsilon$ that satisfy

\[-\text{div}(A^\varepsilon(x) \nabla u^\varepsilon(x)) = f(x), \forall x \in \Omega,\]

\[u^\varepsilon|_{\Gamma_D} = 0; \quad A^\varepsilon \nabla u^\varepsilon \cdot \vec{n}|_{\Gamma_{N1}} = 0\]

and $A^\varepsilon \nabla u^\varepsilon \cdot \vec{n}|_{\Gamma_N} = -1.$ (1)
Numerical Results for the Blood Micro-Circulation Modelled by a Porous Media

for a sequence of scalars \( \varepsilon > 0 \). We suppose that \( f \in L^2(\Omega) \) and we are looking for the asymptotic limit of the sequence \( u^\varepsilon \) when \( \varepsilon \to 0 \). For every \( \varepsilon > 0 \), we suppose that \( A^\varepsilon \in L^\infty(\Omega, \mathcal{M}_{\alpha,A,\beta_A}) \), where \( \mathcal{M}_{\alpha,A,\beta_A} \) is the set of uniformly positive definite square matrices with uniformly positive definite inverse. The problem (1) is well posed and for every \( \varepsilon > 0 \), there exists a unique solution \( u^\varepsilon \in H^1_{\Gamma_D}(\Omega) = \{ u \in H^1(\Omega)/u|_{\Gamma_D} = 0 \} \) with the relation \( ||u^\varepsilon||_{H^1(\Omega)} \leq C(\Omega)||f||_{L^2(\Omega)} \).

3 Two-scale homogenization

First, we will introduce the two-scale homogenization theory to get explicit expressions for the homogenized problem. We assume that tensors \( A^\varepsilon \) are functions of two coupled variables on the set locally defined by a fast microscopic variable \( \varepsilon^{-1}x \) coupled with the slow macroscopic variable \( x \) in \( \Omega \).

\[
A^\varepsilon(x) = A(x, \frac{x}{\varepsilon}),
\]

where \( A(x,.) \) is \( Y = [0,1]^n \)-periodic, \( x \in \Omega \) and \( A \in L^\infty(\Omega, \mathcal{M}_{\alpha,A,\beta_A}) \). We denote by \( x \) the macroscopic variable and by \( y = \frac{x}{\varepsilon} \) the microscopic variable. As \( A(x, \frac{x}{\varepsilon}) \) is locally periodic, one possible manner to get explicit expression for the homogenized problem is to perform a formal two-scale analysis with the following Ansatz:

\[
u^\varepsilon(x) = \sum_{i=0}^{+\infty} \varepsilon^i u_i(x, \frac{x}{\varepsilon}) = u_0(x, \frac{x}{\varepsilon}) + \varepsilon u_1(x, \frac{x}{\varepsilon}) + \varepsilon^2 u_2(x, \frac{x}{\varepsilon}) + \ldots
\]

where, for any \( x \in \Omega \), the functions \( u_i(x,.) \) are \( Y \)-periodic. Inserting the Ansatz (3) into equation (1) allows us to write \( u^\varepsilon \) as:

\[
u^\varepsilon = u^* + \varepsilon \sum_{i=1}^{n} w_i \partial_i u^* + r_\varepsilon,
\]

where \( (w_i(x, .))_{1 \leq i \leq n} \) are \( n \) \( Y \)-periodic cell functions parameterized by their macroscopic position \( x \in \Omega \) and verified the following n cell problems,

\[
- \text{div}_y(A(x,y)[e_i + \nabla_y w_i(x,y)]) = 0, \forall y \in Y,
\]

and the function \( u_0 = u^* \) does not depend on the fast variable \( \varepsilon^{-1}x \) and verifies the homogenized problem:

\[
\begin{align*}
-\text{div}(A^*(x) \nabla u^*(x)) &= f, \forall x \in \Omega, \\
A^*(x) \nabla u^*|_{\Gamma_D} &= 0, \\
A^*(x) \nabla u^*|_{\Gamma_N} &= -1, \\
A^*(x) \nabla u^*|_{\Gamma_{N_1}} &= 0.
\end{align*}
\]

where the entries \( (A^*_{ij}(x))_{1 \leq i,j \leq n} \) of the homogenized matrix \( A^* \) can be explicitly computed with the cell functions \( w_i(x,.) \),

\[
A^*_{ij}(x) = \int_Y A(x,y)(e_i + \nabla_y w_i(x,y))e_j dy,
\]

Where, provided \( u^* \in W^{2,\infty}(\Omega) \), the correction error \( r_\varepsilon \) can be estimated to locally scale as \( \varepsilon \) (far enough from the boundary), and to globally scale as \( \sqrt{\varepsilon} \) (see [4]).

4 The reduced-basis method

To reduce the computation time, we use the RB method to improve the repeated numerical treatment of parameterized cell problems (5). Therefore, a computational procedure based on an offline/online approach should naturally allow for a reduction of the computation time, when in particular, a large number of parameterized cell problems occurs in the limit \( \varepsilon \to 0 \) of the homogenization strategies, in order to compute the homogenized problem (6) with non-periodic coefficients.

The RB approach for the parameterized cell problems (5) should significantly decrease the expense of computations in terms of CPU time for the homogenization problems where the offline stage is short compared to the online stage.

Let \( X \), be the quotient space \( H^1(Y)/\mathbb{R} \) of \( Y \)-periodic functions that belong to the Sobolev space \( H^1(Y) \) equipped with the norm \( ||u||_X = (\int_Y \nabla u . \nabla u)^{\frac{1}{2}} \) induced from the inner product \( (u, v)_X = \int_Y \nabla u . \nabla v \). In the dual space \( X' \) of \( X \), the dual norm is defined as \( ||g||_{X'} = \sup_{v \in X} \frac{g(v)}{||v||_X} \). For any integer \( 1 \leq i \leq n \), the i-th cell problem (5) for the cell functions \( w_i(x,.) \) rewrites in the following weak form:

\[
\begin{cases}
\text{Find } w_i(x,.) \in X \text{ solution for : } \\
\quad a(w_i(x, .), v; x) = f_i(v; x), \forall v \in X.
\end{cases}
\]
where \( x \in \Omega \) plays the role of a parameter, \( a \) is a continuous and coercive bilinear form in \( X \times X \) parameterized by \( x \in \Omega \) defined as 
\[
\int_Y A(x, y) \nabla u(y). \nabla v(y)dy
\]
and \( f_i \) is a linear and continuous form defined as 
\[
f_i(v; x) = -\int_Y A(x, y)e_i. \nabla v(y)dy, \quad 1 \leq i \leq n.
\]
We introduce the n-dimensional square symmetric matrix of elements \( (s_{ij})_{1 \leq i,j \leq n} \) defined by:
\[
s_{ij}(x) = -f_j(w_i(.); x) = \int_Y A(x, y)e_j. \nabla w_i(x, y)dy.
\]  
We consider the set \( M = \{w_i(.), x \in \Omega, 1 \leq i \leq n \} \) which can be approximated by \( X_N = \text{span}\{\xi_j, 1 \leq j \leq N \} \) of \( X \). To compute numerically the cell problem, we introduce the set \( M_p = \{w_i(x_k, .), 1 \leq k \leq p, 1 \leq i \leq n \} \) induced by a discrete sample \( D = \{x_k \in \Omega, 1 \leq k \leq p \} \) of parameter values distributed over the parameter space \( \Omega \). This is termed as the offline stage. \( M_p \) can be approximated by a finite dimensional subspace \( X_N = \text{span}\{\xi_j, 1 \leq j \leq N \} \) of \( X \). Then, the reduced basis method suggest to solve the variational problem (8) through a Galerkin projection method which consists to build the finite subspace \( X_N \subset X \) with an hilbertian basis adapted to the equation (8). \( \forall x \in \Omega \) et \( 1 \leq i \leq n \), the Galerkin approximation \( w_{iN} \in X_N \) of \( w_i(x, .) \) satisfy :
\[
a(w_{iN}(x, .), v; x) = f_i(v; x), \quad \forall v \in X_N
\]  
The approximation of \( w_i(x, y), i = 1,...,n \) for some given \( x_k, k = 1,...,p \), is calculated with a criteria of a posteriori error estimation which will be presented later. To recall the estimation errors showed in [4], we begin by defining the quantities:
\[
A^*_{ij}(x) = \int_Y A(x, y)(e_i + \nabla_y w_{iN}(x, y)). e_j dy
\]
\[
A_{ij}(x) = \int_Y A(x, y)e_i.e_j dy
\]
\[
A^*_{ij}(x) = \int_Y A(x, y)e_i.e_j dy
\]  
Since the matrix \( s \) is the part which interests us in our calculation of the homogenized matrix, one will state two theorems which enable us to define the a posteriori bounds error of this matrix.  

**Theorem 0.1.** [4] For \( 1 \leq i \leq n \), we define the a posteriori error estimation \( \Delta_N(w_i(x, .)) \) by:
\[
\Delta_N(w_i(x, .)) = (\| a(w_i(.,.) - w_{iN}(x, .), ; x) \|_X)/\alpha_A,
\]
and the corresponding affectivities \( \eta_i(x) \) by \( \eta_i(x) = \Delta_N(w_i(x, .))/\| w_i(x, .) - w_{iN}(x, .) \|_X \) which satisfy the N-independent inequalities: \( 1 \leq \eta_i(x) \leq \gamma_A^i \).  

**Theorem 0.2.** [4] Let \( w_i(.) \) and \( w_{iN}(x, .) \) be two solutions of the problem (8) and \( w_{iN}(x, .) \) two solutions of the problem (10) with \( i \neq j \), we have \( |s_{ij}(x) - s_{ij,N}(x)| \leq \Delta_{ij,N}(x) \) with 
\[
\Delta_{ij,N}(x) = \frac{\Delta^*_i\Delta^*_j}{\alpha_A}.
\]
Finally, in the online stage, for any \( x \in \Omega \), \( w_i(x, y) \) is to be approximated, using the Reduced Basis method, by some vector \( w_{iN}(x, y) \) in the Galerkin projection space \( X_N \) of size \( N \) that writes:
\[
w_i(x, y) = \sum_{j=1}^{N} w_{iN}(x)\xi_j(y),
\]
where \( w_{iN}(x) \) is computed by \( N \times N \) problems.  

5 Application  
In this section we simulate the blood flow in the capillary as in the (FIGURE 1). The purpose of this work is to model a porous media, for which we will use a permeability tensor which vanishes on some region of the domain. This fact gives a very small values of the constant of coercivity \( \alpha_A \) and follows a non pertinent a posteriori error. To solve this problem, we will introduce a new scalar product which gives better numerical results.  

In this section, we take \( n = 2 \). Let \( \hat{\lambda} = [0, 1]^2 \) be the reference cell where the permeability is defined as follow:
\[
K = \begin{pmatrix} K_{11}(\hat{y}_1, \hat{y}_2) & 0 \\ 0 & K_{22}(\hat{y}_1, \hat{y}_2) \end{pmatrix},
\]
with 
\[
K_{11}(\hat{y}_1, \hat{y}_2) = K_{22}(\hat{y}_1, \hat{y}_2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\hat{y}_1 - \mu}{\sigma} \right)^2} + \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{\hat{y}_2 - \mu}{\sigma} \right)^2} + \varsigma,
\]
where \( (\hat{y}_1, \hat{y}_2) \in [0, 1]^2, \sigma = 0.07, \mu = 0.5, 0 < \varsigma < 1\). We denote by \( \hat{\lambda} = H^2(\hat{\lambda})/\mathbb{R} \).  
To build a random porous media, first we introduce the function \( \phi_{\theta}(\hat{y}) = -2\theta \hat{y}^3 + 3\theta\hat{y}^2 + (1 - \theta)\hat{y} \)
which is stable on $[0,1]$ and have a reciprocal function for every random $\theta \in [-2,1]$. Then we will modify $K$ in every cell by replacing $y_1$ by $\phi_0(y_1)$ and $y_2$ by $\phi_0(y_2)$, and we denote by $\psi_0(y_1, y_2) = (\phi_0(y_1), \phi_0(y_2))$.

With the chosen permeability tensor, we will solve first the cell problem in order to compute $u^c$ solution of (1), by applying the reduced bases method: For all $i \in \{1, 2\}$

$$
\left\{ \begin{array}{ll}
-\text{div}(K_\theta(y)(c_i + \nabla_y w_i(\theta,y))) = 0 & \forall y \in Y = [0,1]^2 \\
 w_i \in H^1_0(Y)/\mathbb{R},
\end{array} \right.
$$

(13)

with $K_\theta(y) = K \circ \psi_0(y)$ and

$$(K_\theta)_{11}(y) = K_{11} \circ \psi_0(y) = (K_\theta)_{22}(y) = K_{22} \circ \psi_0(y) = \frac{1}{\sigma^2} \left( e^{-\frac{|\phi_0(y_1) - \mu_1|^2}{2\sigma^2}} + e^{-\frac{|\phi_0(y_2) - \mu_2|^2}{2\sigma^2}} \right) + c.$$

From the last expression of $K$, we can deduce that the coercivity coefficient $\alpha_A = \xi$ of the bilinear form $a$ is very small. Since, the theorem 0.2 gives a non pertinent results because we divide by $\alpha_A$. To remedy to this problem, we will introduce a new inner product $[[\hat{u}, \hat{v}]]_{\hat{X}} = \int_{\hat{X}} (K\nabla \hat{u} \nabla \hat{v})$ For all $\hat{u}, \hat{v} \in \hat{X}$ and we denote by $[[\cdot]]_{\hat{X}}$ the associated norm.

**Theorem 0.3.** With this new inner product, we define the a posteriori estimate

$$\Delta_\beta(\theta) = (\| a(w_i(\theta, \cdot) - w_iN(\theta, \cdot), \cdot; \theta) \|_{\hat{X}}^2)/(\beta^A)^d$$

and the effectivities

$$\eta_{i\beta}(\theta) = \Delta_\beta(w_i(\theta, \cdot)) / \| w_i(\theta, \cdot) - w_iN(\theta, \cdot) \|_{\hat{X}}.$$

We have the inequality $1 \leq \eta_{i\beta}(\theta) \leq \frac{(\zeta_A)^d}{(\beta^A)^d}$ with

$$\zeta_A = \max_{y_1 \in [0,1]} \phi_{\mu}^{-1}(y_1)$$

**Theorem 0.4.** With this new inner product, the theorem 0.2 becomes $\| s_{ij}(\theta) - s_{ij}^N(\theta) \| \leq \Delta_{i,j,N}^\beta(\theta)$ with

$$\Delta_{i,j,N}^\beta(\theta) = \frac{\Delta_i^\beta \Delta_j^\beta}{(\beta^A)^d}.$$

Where $\Delta_i^\beta = \| a(w_i(\theta, \cdot) - w_iN(\theta, \cdot), \cdot; \theta) \|_{\hat{X}}$.

### 6 Numerical Results

In this section, we show numerical results for $n = 2$, $\Omega = [0,1]^2$, $\theta \in [-0.4,0.4]$ and $f = 0$. We choose $\Gamma_N = \{1\} \times [0.35,0.65]$, $\Gamma_D = \{0\} \times [0.35,0.65]$ and $\Gamma_{N1}$ is the rest of the boundary. In this work, these numerical computations are done with the Freefem++ software. More precisely, we use classical P1 Lagrange finite elements. The repartition of the permeability is given by the left figure of (FIGURE 2). The right color of (FIGURE 2) represents the flow repartition in the porous media where $\varepsilon = 0.01$.

The (FIGURE 3) represents the variation in logarithmic scale of:

$$A_\beta = \max_{1 \leq i \leq 2, \theta_k \in [-1,2]} \| w_i(\theta_k, \cdot) \|_{\hat{X}}^2,$$

with respect of $N$, where $\zeta = 10^{-14}$ and $10^{-10}$.

We conclude that with the new inner product values, $A_\beta$ is not affected by the variation of $\zeta$.

Next, in (FIGURE 4), we compare the error values in logarithmic scale, of the a posteriori estimate of the solution with the new scalar product $A_\beta$ and a posteriori estimate of the solution with the old inner product $A_N = \max_{1 \leq i \leq 2, x_k \in D} \| w_i(x_k, \cdot) \|_{\hat{X}}$.

We conclude that the errors of the solutions computed with the new product scalar is much better than the other one. Hence the advantage of this method.

The next table shows the CPU time (in seconds) given by the Freefem++ code to approximate the FE matrix for the homogenized problem either with a direct FE approach or with a RB method for a porous medium and for a small $\varepsilon$, where $NB$ is the bases number, $\mathcal{N}$ is the number of the degrees of freedom of the cell, $\mathcal{N}$ is the number of the degrees of freedom of $\Omega$, **offline part** is the time of the bases construction, **BRA** is the time of the calculation of $A^*$ using the reduced basis method and **EFA** is the time of the calculation of $A^*$ using the finite element method.

<table>
<thead>
<tr>
<th>NB</th>
<th>$\mathcal{N}$</th>
<th>offline part</th>
<th>$\mathcal{N}$</th>
<th>BRA</th>
<th>EFA</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>100</td>
<td>1.17 s</td>
<td>121</td>
<td>0.017s</td>
<td>0.764s</td>
</tr>
<tr>
<td>5</td>
<td>400</td>
<td>5.62 s</td>
<td>121</td>
<td>0.018s</td>
<td>3.011s</td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>1.476 s</td>
<td>10201</td>
<td>1.810s</td>
<td>64.604s</td>
</tr>
<tr>
<td>5</td>
<td>400</td>
<td>5.93 s</td>
<td>10201</td>
<td>1.816s</td>
<td>263.945s</td>
</tr>
</tbody>
</table>

As we can see, the RB method is faster than the FE method. We can notice also that the offline part depends only on the microscopic mesh while the online part depends on the macroscopic part.
Conclusion: In this paper, we use the RB method to solve and speed up the resolution of the Darcy equation in the porous media using the locally homogenization model. We introduce a new scalar product to obtain pertinent results. In a next publication, we will show three dimensional results.

References

Figure 3. $A_\beta$ with respect of $N$, with $\zeta = 1e-10$ (left) and $\zeta = 1e-14$ (right).

Figure 4. $\zeta = 1e-14$, $A_\beta$ with respect of $N$ (left), $A_N$ with respect of $N$ (right).
## Author Index

### A
- Abdullah N. L. 36
- Abu Hanifah F. 86
- Agamalov O. N. 176
- Akma Ibrahim N. 92
- Alam Md. S. 50
- Ali B. 262
- Amir Abas M. 102
- Antoun Simao R. 154
- Arema S. 92
- Arshad A. M. 102
- Arunrungrusmi S. 276
- Azizpour M. J. 204
- Azrilah A. A. 23, 30

### B.
- Bakar R. A. 107
- Basri H. 23, 30
- Bauk S. 50
- Beilicci E. B. M. 249
- Beilicci R. F. 249

### C.
- Capotescu V. 159
- Castellanos C. G. 130
- Che-Ani A. I. 6
- Cipriana M. A. 60
- Coelho B. 154
- Coelho Oliveira A. 154
- Comaniciu C. 185
- Cook C. W. 66
- Croitoru A. 124

### D.
- Dahlui M. 102
- Daneshvar S. 269
- Daud Md. N. 6
- Daud N. 224
- Decaro F. P. 66
- Decaro N. E. 66
- Delgado E. 130

### E.
- Ebrahimi G. H. 243
- El Mostafa M. 262

### F.
- Filimon S. 60
- Frinculeasa M. 284

### G.
- Gavrilot A. 118, 124
- Girbaciuc C. A. 249
- Gupta N. 190

### H.
- Habibah Arshad N. 86
- Hafez Nawi A. 201
- Haripersad R. 228
- Hilmi Fadzil M. 102

### I.
- Idrus R. 36, 43
- Istrate A. 284

### J.
- Jamil M. 18
- Jenn-Tzey Y. 114

### K.
- Kadirgama K. 107
- Kaleibar M. M. 269
- Khoshnevis Yazdi S. 53
- Khusairy Azim A. 102
- Kobasko N. I. 136, 142, 148
- Kofli N. T. 18
- Koiv T. A. 257

### M.
- Maday Y. 290
- Mahmud Z. 78
<table>
<thead>
<tr>
<th>Author</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malek H. Ab.</td>
<td>224</td>
</tr>
<tr>
<td>Maleque M. A.</td>
<td>107</td>
</tr>
<tr>
<td>Mares C.</td>
<td>237</td>
</tr>
<tr>
<td>Masodi Md. S.</td>
<td>78</td>
</tr>
<tr>
<td>Mastorakis N. E.</td>
<td>124</td>
</tr>
<tr>
<td>Mat K.</td>
<td>18</td>
</tr>
<tr>
<td>Mazurenko E. A.</td>
<td>136</td>
</tr>
<tr>
<td>Medvedev A. M.</td>
<td>136</td>
</tr>
<tr>
<td>Meng Chun K.</td>
<td>43</td>
</tr>
<tr>
<td>Mikola A.</td>
<td>257</td>
</tr>
<tr>
<td>Mirzazadeh A.</td>
<td>243</td>
</tr>
<tr>
<td>Mohamad Yusof M. Y.</td>
<td>201</td>
</tr>
<tr>
<td>Mohamed A.</td>
<td>12</td>
</tr>
<tr>
<td>Mohamed L.</td>
<td>262</td>
</tr>
<tr>
<td>Mohammadi Majd H.</td>
<td>204</td>
</tr>
<tr>
<td>Mohd-Tawil N.</td>
<td>6</td>
</tr>
<tr>
<td>Morcos N.</td>
<td>290</td>
</tr>
<tr>
<td>Moskalenko A. A.</td>
<td>136</td>
</tr>
<tr>
<td>Mounir R.</td>
<td>262</td>
</tr>
<tr>
<td>Moya Chaves F. D.</td>
<td>209</td>
</tr>
<tr>
<td>Muhamad N.</td>
<td>12</td>
</tr>
<tr>
<td>Mungkung N.</td>
<td>276</td>
</tr>
<tr>
<td>Mustapha R.</td>
<td>12</td>
</tr>
<tr>
<td>Rahman M. N. A.</td>
<td>18</td>
</tr>
<tr>
<td>Ramly A.</td>
<td>6</td>
</tr>
<tr>
<td>Rios-Bolivar A.</td>
<td>165</td>
</tr>
<tr>
<td>Rodzo’an N. A.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Rothlisberger S.</td>
<td>130</td>
</tr>
<tr>
<td>Saidfudin M.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Salmalian K.</td>
<td>215</td>
</tr>
<tr>
<td>Savin P.</td>
<td>1</td>
</tr>
<tr>
<td>Sayah T.</td>
<td>290</td>
</tr>
<tr>
<td>Sepulveda F. A.</td>
<td>130</td>
</tr>
<tr>
<td>Shao-Kang H.</td>
<td>114</td>
</tr>
<tr>
<td>Shiang-Yen T.</td>
<td>36, 43</td>
</tr>
<tr>
<td>Skakouri B.</td>
<td>53</td>
</tr>
<tr>
<td>Sobhanallahi M. A.</td>
<td>243</td>
</tr>
<tr>
<td>Soleimani M.</td>
<td>215</td>
</tr>
<tr>
<td>Soo Mee J. W.</td>
<td>195</td>
</tr>
<tr>
<td>Tsitsiashvili G.</td>
<td>171</td>
</tr>
<tr>
<td>Wee Nee C.</td>
<td>43</td>
</tr>
<tr>
<td>Omar M. Z.</td>
<td>12, 18, 23, 30</td>
</tr>
<tr>
<td>Osman S. A.</td>
<td>18</td>
</tr>
<tr>
<td>Othman I. R.</td>
<td>78</td>
</tr>
<tr>
<td>Othman J.</td>
<td>78</td>
</tr>
<tr>
<td>Palai S. S.</td>
<td>201</td>
</tr>
<tr>
<td>Panchal V. K.</td>
<td>190</td>
</tr>
<tr>
<td>Paven Ionela G.</td>
<td>60</td>
</tr>
<tr>
<td>Pehoiu G.</td>
<td>284</td>
</tr>
<tr>
<td>Peik See K.</td>
<td>36</td>
</tr>
<tr>
<td>Perez Teixeira R. L.</td>
<td>154</td>
</tr>
<tr>
<td>Peh Wen L.</td>
<td>43</td>
</tr>
<tr>
<td>Ping-Jung T.</td>
<td>114</td>
</tr>
<tr>
<td>Prasad D.</td>
<td>201</td>
</tr>
<tr>
<td>Prasad D.</td>
<td>201</td>
</tr>
<tr>
<td>Rahman Khan Md. A.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. M.</td>
<td>107</td>
</tr>
<tr>
<td>Saidfudin M.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Salmalian K.</td>
<td>215</td>
</tr>
<tr>
<td>Savin P.</td>
<td>1</td>
</tr>
<tr>
<td>Sayah T.</td>
<td>290</td>
</tr>
<tr>
<td>Sepulveda F. A.</td>
<td>130</td>
</tr>
<tr>
<td>Shao-Kang H.</td>
<td>114</td>
</tr>
<tr>
<td>Shiang-Yen T.</td>
<td>36, 43</td>
</tr>
<tr>
<td>Skakouri B.</td>
<td>53</td>
</tr>
<tr>
<td>Sobhanallahi M. A.</td>
<td>243</td>
</tr>
<tr>
<td>Soleimani M.</td>
<td>215</td>
</tr>
<tr>
<td>Soo Mee J. W.</td>
<td>195</td>
</tr>
<tr>
<td>Tsitsiashvili G.</td>
<td>171</td>
</tr>
<tr>
<td>Wee Nee C.</td>
<td>43</td>
</tr>
<tr>
<td>Omar M. Z.</td>
<td>12, 18, 23, 30</td>
</tr>
<tr>
<td>Osman S. A.</td>
<td>18</td>
</tr>
<tr>
<td>Othman I. R.</td>
<td>78</td>
</tr>
<tr>
<td>Othman J.</td>
<td>78</td>
</tr>
<tr>
<td>Palai S. S.</td>
<td>201</td>
</tr>
<tr>
<td>Panchal V. K.</td>
<td>190</td>
</tr>
<tr>
<td>Paven Ionela G.</td>
<td>60</td>
</tr>
<tr>
<td>Pehoiu G.</td>
<td>284</td>
</tr>
<tr>
<td>Peik See K.</td>
<td>36</td>
</tr>
<tr>
<td>Perez Teixeira R. L.</td>
<td>154</td>
</tr>
<tr>
<td>Peh Wen L.</td>
<td>43</td>
</tr>
<tr>
<td>Ping-Jung T.</td>
<td>114</td>
</tr>
<tr>
<td>Prasad D.</td>
<td>201</td>
</tr>
<tr>
<td>Rahman Khan Md. A.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. M.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. N. A.</td>
<td>18</td>
</tr>
<tr>
<td>Ramly A.</td>
<td>6</td>
</tr>
<tr>
<td>Rios-Bolivar A.</td>
<td>165</td>
</tr>
<tr>
<td>Rodzo’an N. A.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Rothlisberger S.</td>
<td>130</td>
</tr>
<tr>
<td>Saidfudin M.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Salmalian K.</td>
<td>215</td>
</tr>
<tr>
<td>Savin P.</td>
<td>1</td>
</tr>
<tr>
<td>Sayah T.</td>
<td>290</td>
</tr>
<tr>
<td>Sepulveda F. A.</td>
<td>130</td>
</tr>
<tr>
<td>Shao-Kang H.</td>
<td>114</td>
</tr>
<tr>
<td>Shiang-Yen T.</td>
<td>36, 43</td>
</tr>
<tr>
<td>Skakouri B.</td>
<td>53</td>
</tr>
<tr>
<td>Sobhanallahi M. A.</td>
<td>243</td>
</tr>
<tr>
<td>Soleimani M.</td>
<td>215</td>
</tr>
<tr>
<td>Soo Mee J. W.</td>
<td>195</td>
</tr>
<tr>
<td>Tsitsiashvili G.</td>
<td>171</td>
</tr>
<tr>
<td>Wee Nee C.</td>
<td>43</td>
</tr>
<tr>
<td>Omar M. Z.</td>
<td>12, 18, 23, 30</td>
</tr>
<tr>
<td>Osman S. A.</td>
<td>18</td>
</tr>
<tr>
<td>Othman I. R.</td>
<td>78</td>
</tr>
<tr>
<td>Othman J.</td>
<td>78</td>
</tr>
<tr>
<td>Palai S. S.</td>
<td>201</td>
</tr>
<tr>
<td>Panchal V. K.</td>
<td>190</td>
</tr>
<tr>
<td>Paven Ionela G.</td>
<td>60</td>
</tr>
<tr>
<td>Pehoiu G.</td>
<td>284</td>
</tr>
<tr>
<td>Peik See K.</td>
<td>36</td>
</tr>
<tr>
<td>Perez Teixeira R. L.</td>
<td>154</td>
</tr>
<tr>
<td>Peh Wen L.</td>
<td>43</td>
</tr>
<tr>
<td>Ping-Jung T.</td>
<td>114</td>
</tr>
<tr>
<td>Prasad D.</td>
<td>201</td>
</tr>
<tr>
<td>Rahman Khan Md. A.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. M.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. N. A.</td>
<td>18</td>
</tr>
<tr>
<td>Ramly A.</td>
<td>6</td>
</tr>
<tr>
<td>Rios-Bolivar A.</td>
<td>165</td>
</tr>
<tr>
<td>Rodzo’an N. A.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Rothlisberger S.</td>
<td>130</td>
</tr>
<tr>
<td>Saidfudin M.</td>
<td>23, 30</td>
</tr>
<tr>
<td>Salmalian K.</td>
<td>215</td>
</tr>
<tr>
<td>Savin P.</td>
<td>1</td>
</tr>
<tr>
<td>Sayah T.</td>
<td>290</td>
</tr>
<tr>
<td>Sepulveda F. A.</td>
<td>130</td>
</tr>
<tr>
<td>Shao-Kang H.</td>
<td>114</td>
</tr>
<tr>
<td>Shiang-Yen T.</td>
<td>36, 43</td>
</tr>
<tr>
<td>Skakouri B.</td>
<td>53</td>
</tr>
<tr>
<td>Sobhanallahi M. A.</td>
<td>243</td>
</tr>
<tr>
<td>Soleimani M.</td>
<td>215</td>
</tr>
<tr>
<td>Soo Mee J. W.</td>
<td>195</td>
</tr>
<tr>
<td>Tsitsiashvili G.</td>
<td>171</td>
</tr>
<tr>
<td>Wee Nee C.</td>
<td>43</td>
</tr>
<tr>
<td>Omar M. Z.</td>
<td>12, 18, 23, 30</td>
</tr>
<tr>
<td>Osman S. A.</td>
<td>18</td>
</tr>
<tr>
<td>Othman I. R.</td>
<td>78</td>
</tr>
<tr>
<td>Othman J.</td>
<td>78</td>
</tr>
<tr>
<td>Palai S. S.</td>
<td>201</td>
</tr>
<tr>
<td>Panchal V. K.</td>
<td>190</td>
</tr>
<tr>
<td>Paven Ionela G.</td>
<td>60</td>
</tr>
<tr>
<td>Pehoiu G.</td>
<td>284</td>
</tr>
<tr>
<td>Peik See K.</td>
<td>36</td>
</tr>
<tr>
<td>Perez Teixeira R. L.</td>
<td>154</td>
</tr>
<tr>
<td>Peh Wen L.</td>
<td>43</td>
</tr>
<tr>
<td>Ping-Jung T.</td>
<td>114</td>
</tr>
<tr>
<td>Prasad D.</td>
<td>201</td>
</tr>
<tr>
<td>Rahman Khan Md. A.</td>
<td>107</td>
</tr>
<tr>
<td>Rahman M. M.</td>
<td>107</td>
</tr>
</tbody>
</table>