ESTIMATION OF LOAD DIAGRAMS IN MV/LV SUBSTATIONS

J. N. FIDALGO
Department of Electric Engineering and Computers
Faculty of Engineering of Porto University
Rua Dr. Roberto Frias, nº 378, 4200-465, Porto
PORTUGAL
jfidalgo@inescporto.pt http://www.fe.up.pt http://www.inescporto.pt

Abstract: - Efficient power systems planning and exploration require the estimation of load diagrams at the different levels of the distribution networks. In particular, the planning of new MV/LV substations requires the assessment of their expected load curves under different exploration scenarios. The deregulation process and environment preservation constraints also compel the need of higher efficiency levels in network investments.
This paper describes the methodologies adopted for the estimation of load curve diagrams in MV/LV substations. The assessment process is based on billing data (monthly energy consumption, hired power contracts, activity codes and weekday type), which is the unique information generally available at this level of the network. The most common approaches use measurements in typical classes of consumers defined by experts to construct inference engines that, most of the times, only estimate peak loads. In this paper, two different approaches are tested. The first one is based on the definition of classes using a clustering algorithm and uses Artificial Neural Networks (ANN) for the estimation of the MV/LV substation load curve. In the second one, an ANN is trained to output directly the load diagram estimated for each individual consumer. This article describes the adopted methodologies and presents some representative results. Performance attained is discussed as well as a method to achieve confidence intervals.

Key-Words: - Load estimation, Clustering, Artificial neural networks, Distribution networks

1 Introduction

During last decade, one has witnessed not only the growth of transmission and distribution power networks but especially how they got more and more complex. Particularly, distribution systems have been continuously spreading and its complexity is also increasing not only because of new operation alternatives but also in consequence of the advent of deregulation.
The competition in an open energy market also demands a higher energetic efficiency due to the will to explore the existing infra-structures to the limits and postpone investments as much as possible. At the same time, environment preservation and the need for energy efficiency have also become more demanding. In this framework, load curve estimation is becoming each time more fundamental to an efficient management and planning of electric distribution systems.
Consumptions' estimation studies have been carried out by several investigators [1-3]. Some distribution utilities performed these studies, modeling consumers' behavior for planning purposes, and using inference processes usually based on linear regression. The main purpose is frequently the estimation of peak power [12] or a rough estimation of the whole annual consumption.
In the present project, the network planning needs are accessed by a software application named Load Curve Management Module that estimates hourly load diagrams for LV and MV clients. The main goal of is to provide 24-hour diagram estimates at medium voltage /low voltage (MV/LV) substations.
The project included measurement campaigns, model development and testing. A short description of the measurement campaigns can be found in section 3.
Two different approaches were considered to perform load estimation. The first one includes three main procedures: a) definition of consumers' classes (clustering); b) inference of load diagrams of MV/LV public stations; c) estimation of error bars providing a image of hourly consumption's variability. In the second approach, an ANN is trained to output directly the load diagram estimated for each individual consumer.
This paper is organized as follows. Section 2 clarifies the main objective of the paper. Section 3 describes the measurement campaigns Section 4 describes the project phases, in particular the ones related to the approach considered. Section 5 presents the results obtained in terms of clustering process. Sections 6 and 7 describe the ANN and the results in the first approach. Section 8 describes the second approach and presents the related results.

2 Load curve estimation

In the proposed load curve estimation approach, ANN inputs consist mainly on consumers billing data (hired power, activity codes, monthly energy consumption). On the other hand, the estimation period can be extended to several months or years, provided that some forecasts exists for monthly energy consumption. It is particularly adapted for planning of distribution network evolution. As an example, suppose that power distribution company aims at installing a new MV/LV station that will fed a given set of consumers; the question is: what is the load diagram expected for this station on a given year period?

3 Measurement campaigns

Initial modeling contributed to define the scope of the measurement campaigns. These campaigns were implemented taking into account the need for hourly base diagrams, considering two year periods, Winter and Summer, and also two groups, one for weekends and holidays and another for working days. Samples were taken accordingly in order to cover a large spectrum of possibilities.

The measurement campaigns result in a collection of consumptions' evolution data, in order to implicitly characterize consumers' behavior. A large spectrum of possible load curves is quintessential to represent the whole universe of consumers. To accomplish this purpose, two measurement campaigns were carried out (one in the Summer and another in the Winter). Load curve recorders (LCR) were installed in a variety of consumers located in the neighborhoods of 5 different Portuguese cities. Observation areas include urban, semi-urban and rural types. The power of a consumer or group of consumers was registered every 15 minutes during a period of two or three weeks. The peak loads as well as the date and hour were also registered.

At the end of the observation period, the LCR transferred the information gathered directly to a PC. This information was completed with the information available in EDP databases, such as commercial characteristics (hired power, monthly energy consumption, peak loads, activity codes).

The basic idea consists of implementing a mechanism for load curve estimation of MV/LV public stations and MV individual consumers. This device will aggregate consumptions estimates following the network tree-structure, and assessing load curves at secondary substation, primary substation and distribution centers.

4 Project phases

In this project, all phases were addressed in a innovative way, based on the use of artificial neural networks (ANN).

4.1 Clustering

The definition of classes was performed by clustering collected load diagrams, in order to avoid biasing introduced by preconceived ideas about the way consumers behave. Each clustering training pattern contains 24 elements - the registered power consumptions a each hour of a given day. This training set was presented to a Kohonen clustering tool, in order to obtain different load evolution classes. Results were compared with other classification tools, namely with fuzzy clustering algorithm which provided quite similar results.

4.2 Load Curve Estimation

When dealing with MV networks, the main issue is to get load diagrams in any point of the network, to be used later for planning purposes (Fig. 1). The available information from utility data base consists mainly of commercial data and energy consumptions.

One intends to estimate load diagrams essentially for MV/LV public stations and MV individual clients. It was also decided to aggregate LV consumers, dependent from public stations, in order to evaluate their accumulated load diagram. This will avoid the need for the characterization of each LV individual consumer, reducing the size of the data base needed for future studies. Furthermore, there are no imperative knowledge requirements for a single LV client.

The data obtained from the measurements campaigns was divided following the season (Summer or Winter), the weekday (workday or weekend) and the type of consumer (LV or MV). Other available data of LV consumers are the monthly energy consumption, the activity code and hired power. All those consumers are fed by MV/LV public substations. For MV consumers, the accessible activity code, peak power, hired power and energy measures for different tariffs are known.

These curves and values must be available in several points of the network, for instance in a MV/LV public station or in a feeder. Additional parameters are evaluated for each load curve (peak power use, load factor, loss factor, etc.), which help the characterization of single or aggregated consumer behavior.

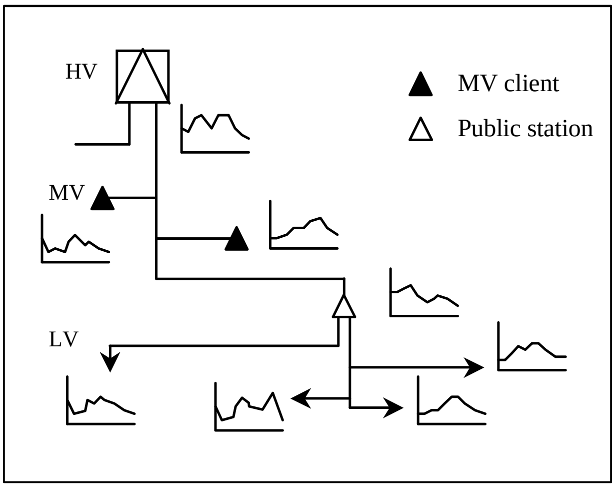

Fig. 1 – Load diagram aggregation in a typical MV network

Estimated load values fed the network analysis tool, in the network planning department. Information is transferred according to the planner needs. He may choose to study a line load, a substation or group of substations load, Distribution Centers loads, etc.

4.3 Estimation of confidence intervals

Load consumption is always characterized by a considerable variability. For similar conditions (season of the year, workday or weekend, and so on), a given consumer or a set of consumers might present two quite different diagrams for two consecutive days. Fortunately, in general, one can observe some kind of behavior pattern, and load curves obtained for similar circumstances define a kind of fuzzy diagram.

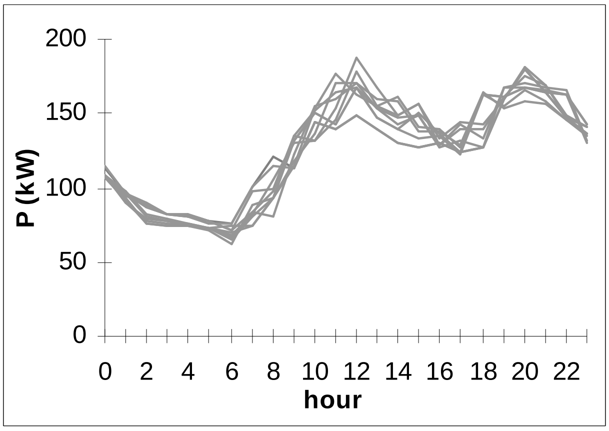

Fig. 2 - Load curves of a MV/LV public station for different workdays

Instead of a simple estimated load curve, it would be interesting to obtain a given bandwidth in such way that the probability that a real load curve be inside that band is, let's say, 0.9. The width of that band will somehow represent a measure of dispersion of loads curves. Fig. 2 presents aggregated load diagrams in a given public station on different week days.

On the present work, we propose to train an auxiliary ANN to learn load curves dispersion (error bounds distribution). These error bars will depend not only on the type and number of consumers aggregated in a given public station, but also on the hour of the day. This way, more complete information on load diagrams dispersion is obtained. Results in section 7.2 show this feature with some detail.

5 Natural and operation classes

MV consumers' modeling needed two different types of analysis: MV clients and MV/LV public stations. In fact, a lot of public stations have no load measurement at all, and it is not possible to infer directly its hourly consumption.

Moreover, as a result of the different behavior according to time and season of the year, the analysis was divided into Winter/Summer and week/weekend day cases. The establishment of the partial models for all the mentioned cases was followed by the development of a integration procedure to cope adequately with intermediate situations.

The modeling process and the subsequent data handling are applied to each referred cases.

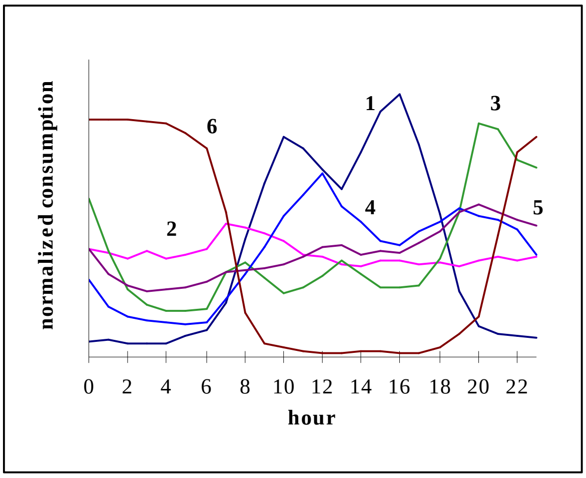

Fig. 3 - Classes diagrams

5.1 Natural classes

One of the fundamental steps of this approach was the identification of natural classes from registered diagrams instead of defining a priori the classes. For that purpose, two different methods were tried: fuzzy clustering [4] and self-organized neural networks (Kohonen maps [5]).

In the LV case study, the best clustering performance was obtained when load diagrams were separated into six classes both with Kohonen and fuzzy clustering.

Kohonen prototypes obtained are shown in Fig. 3. The fuzzy prototypes achieved are similar to Kohonen's. This leads to an independent confirmation of the results obtained before.

5.2 Operational classes

Although valuable, the natural classes are not useful for operational purposes. In fact, they are defined directly from diagrams and not from consumers' characteristics available in commercial data base.

After obtaining the cluster's prototypes one must to induce the relation between classifications and commercial data (tariff class, hired power and monthly energy consumption), in order to generalize the classification of consumers for which only commercial data is known, (what constitutes the real future operational conditions).

Some experiences were carried out for determining a good combination of clustering (number of classes based on load evolution) and inference of classification rules (based on commercial data). This inference process was based on the observation of the distribution of classes' members on the 3D space of commercial data (tariff class, hired power and monthly energy consumption).

The best classification performance was obtained with four classes (Figure 4) and the following classification rules:

- D Night consumers
- A Domestic consumers (Tc=0), low hired power (Pc≤6.6kW) and low energy consumption (E<600kWh)
- B Industrial consumers (Tc=4)
- C Other consumers

The comparison Fig. 4 to Fig. 3 show the match of class D to 6, B to 1, A to 3 to and C to the rest. Operational classes were effectively used in the inference process described in next section.

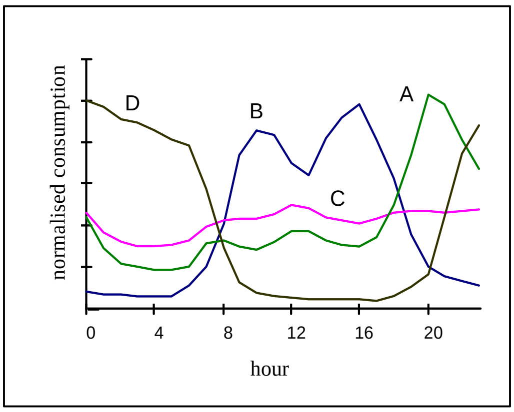

Fig. 4 - Operational classes

6 ANN implementation

ANN is the basic tool used in this work for inference purposes. All ANN were trained with the Adaptive Backpropagation (ABP) training algorithm [6]. The ABP is based on the classic backpropagation but uses an individual adaptive learning rate for each weight, which provides a much faster learning process. The stop training criterion was based in the well known cross validation principle, which fights against overfitting.

For the MV clients case, an ANN is used to estimate their consumption curves directly from commercial data. For the Public Stations case, the available data for each one of these stations only comprises:

- Number of LV consumers of each operational class;
- Total energy consumption of each class.

The daily diagram estimation in an hourly base (p0, p1, ..., p23) is made using a back-propagation neural network, as the one represented on Figure 5.

ni and Ei are respectively the number of consumers and the total energy (monthly) of class i (the indices' i=0..3 relate to the 4 classes, A a D previously described).

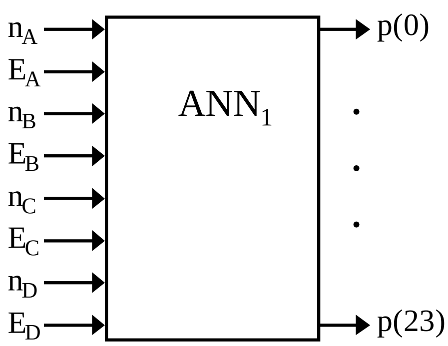

Fig. 5 - Inputs/Outputs from ANN

To train this ANN, 2000 patterns were generated from the data file derived from measurement campaigns. Each pattern was generated to include from 80 to 160 LV consumers, randomly selected from basic samples.

For each pattern, the 8 ANN values were settled from the classification of prototype elements, followed by energy counting and sum for each class. The values from the outputs have equivalence in the 24 time intervals from the aggregated diagram of the consumers belonging to each pattern. From the 2000 generated patterns, 1500 were taken out to train the ANN and the remaining were used for testing.

7 Results (Approach 1)

This section begins with the presentation of some illustrative examples where one's compare inferred load diagrams with the real thing. After, we propose a method for assessing error bandwidth around the predicted load curve. The idea is to obtain a measure of confidence intervals.

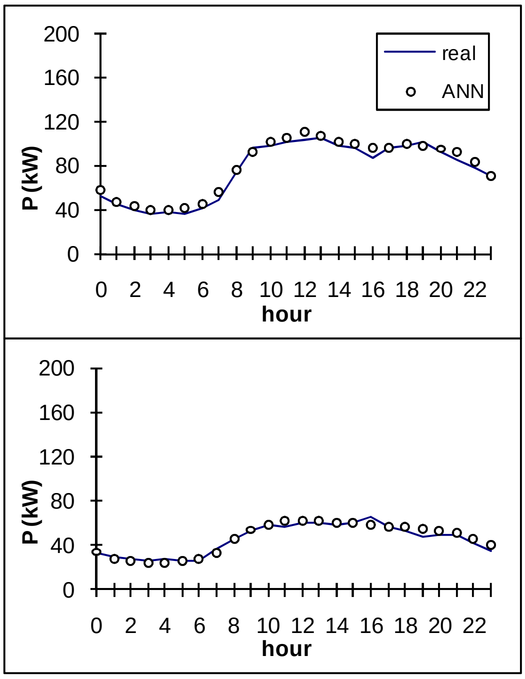

Fig. 6 - Some inference tests (LV Summer workday)

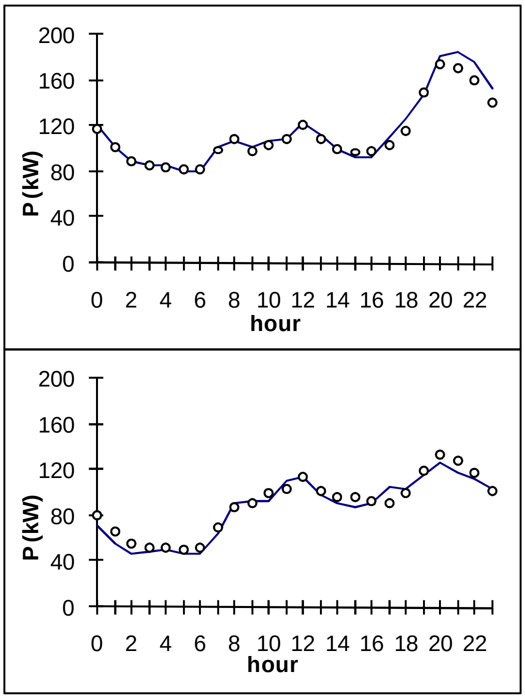

Fig. 7 - Some inference tests (LV Winter workday)

7.1 Load diagrams inference

Figure 6 presents pattern examples of the test set, comparing the real diagram (real) and ANN outputs (NN). Examples shown referred to LV consumers, summer and workdays.
Figure 7 shows similar results but for winter workdays. Global results show that the ANN is capable of estimating the test set diagrams presenting a mean absolute percentage error around 10%. This may be considered a good result, especially if we take into account the arbitrariness inherent to loads behavior.

7.2 Confidence intervals

Despite the good quality of approximation achieved (Fig. 6 and 7), it is always desirable to access the confidence intervals in order to provide a characterization of the accuracy of such estimates. It would be interesting to obtain a given bandwidth around ANN estimated load diagram in such way that the probability of a real load curve be inside that error bounds is, let's say, 0.9.
There has been some interesting work in the area of confidence interval prediction for ANN [7-10]. In most of those studies, authors assume Gaussian or t-student distributions and estimate output variance as a function of inputs variance and of input/output transferring function, using Bayes rule.
Here, we used an auxiliary ANN (called ANNd) to learn load curves' dispersion (error bounds distribution) depending not only on the type and number of consumers aggregated in a given public station, but also on the hour of the day, as shown in Fig. 8. This figure presents aggregated load diagrams of a given public station on different week days (lines) as well as ANN estimation (circles).
The analysis of a large amount of figures like this one has shown that there is a pattern on the errors spreading. If there is not, we can only evaluate average errors. We can observe that consumptions' dispersion is not homogeneous, that is, the same consumer or group of consumers does not present the same uncertainty around a medium load curve for all the hours of the days. For instance, the dispersion before 7 a.m. is usually smaller than at (e.g.)11:00.
Inputs of ANNd are the same of ANN1. Its outputs are the absolute values of the differences between ANN1 outputs and load consumption curves. This way, ANNd produces an error dispersion measure of diagrams estimated by ANN1.
It must be stressed that there are two kinds of errors: a) errors arising from ANN implementation limitations and b) errors (called dispersion errors) related with the nature of predicted values. We

shouldn’t expect that ANNd learn the approximation errors that ANN1 couldn’t learn, but we hope that learn something about the way total error distributes itself over ANN1 outputs and as a function of its inputs.

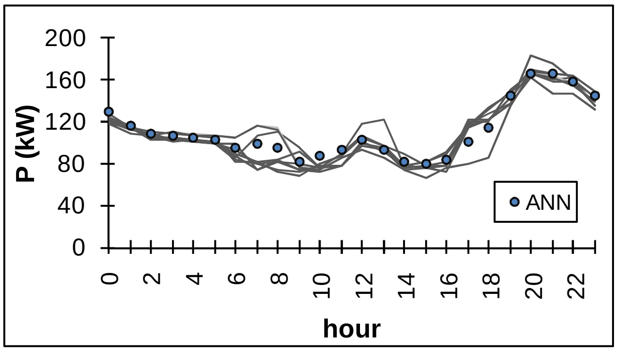

Fig. 8 - Public station consumption for different workdays

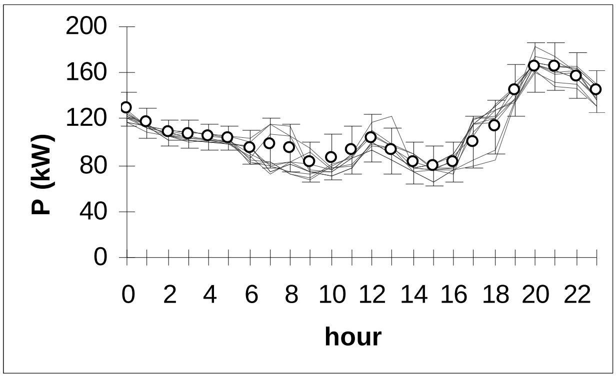

Fig. 9 - The 90% confidence intervals

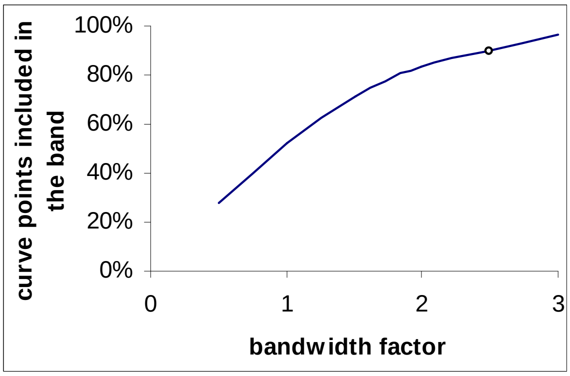

Fig. 10 - Inclusion factor versus bandwidth

Fig. 9 is a superposition of Fig. 8 with the 90% confidence intervals (that corresponds to 2.5 times the output of ANNd, as one can see in Fig. 10. This figure characterizes the relation between what we have called bandwidth factor and the percentage of hourly consumptions within the bands.

These bands represent a measure of the confidence interval of the estimated load curve. Presented results contribute to confirm that adopted tools are the most suitable for the proposed objectives.

8 Alternative approach

A second alternative approach for load curve estimation was tested within the scope of a second project developed under the framework of a contract with the Portuguese distribution company (EDP). This assignment was motivated by the opening of MIBEL (open electricity market for Portugal and Spain). The main objectives of this project include:

1. Consumers’ characterization, including load profiling for LV consumers;
2. Loss allocation and derivation of loss factors;
3. Load research.

Consumers’ characterization phase comprises two main tasks:

- To attain typical mean diagrams of the several types of consumers (low, medium and high voltage types);
- Load profiling (in this project only for LV consumers). Load profiling is an essential tool for open electricity markets [19-22].

The second objective of the project aims at estimate typical losses in LV, MV and HV networks, and loss allocation within each subtype of consumers. Finally, the third goal is study on the relations among consumers use of electrical devices and their diagrams.

These tasks were completed after the conclusion of a measurement campaign to supply the required portrayal of LV consumers’ behavior. Similarly to Approach 1, in addition to the data collected from the measurement operation, the information sources include data attained from consumers inquiring on energy utilization and company’s data base, containing information on consumers’ type, hired power, annual consumption and so on.

Consumers were divided by voltage level, type (Residential, Industrial, Commercial, Hotels and Restaurants, and Others), hired power, type of region (urban, semi-urban and rural) and by consumption strata. These stratum boundaries within each group were settled by Kohonen organizing maps. Sampling resources are distributed within each group according to its number of consumers and its variance (or pattern deviation) following the Neyman stratified sampling methodology [18].

The number of samples was settled by assuming a Gaussian distribution of consumptions and appointing the confidence interval required and the maximum error allowed within each stratum considered. Within each stratum a random selection was performed to choose the consumers to be analyzed. The following step consists of the installation of the hourly meters in the selected consumers. The next phase includes data gathering (15 min measurements on a permanent basis) and collection, and data processing (outliers filtering, data organization, etc).

Fig. 11 shows examples of individual diagrams collected for a number of consumers. In this illustration the top graphic shows examples of residential consumers, the middle graphic contains commercial type examples and the bottom graphic shows industry diagrams patterns.

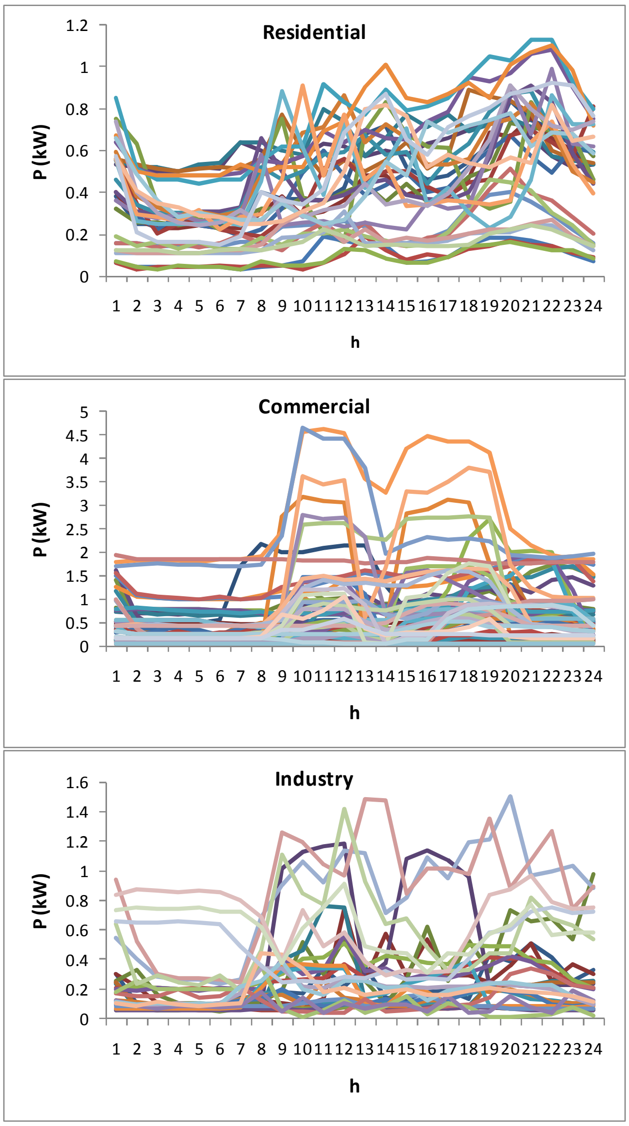

Fig. 11 – Real diagram examples for different types of consumers

This figure shows a large variety of real load diagrams. As we can see, quite different diagram shapes may occur even within the same type of costumers. However, despite the natural fluctuation of loads, it is possible to derive a characteristic behavior for each type.

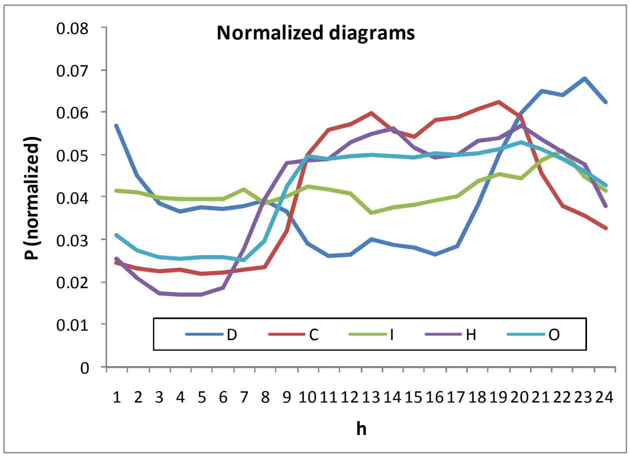

Fig. 12 – Normalized mean diagrams for the five types of consumers considered (January, workdays)

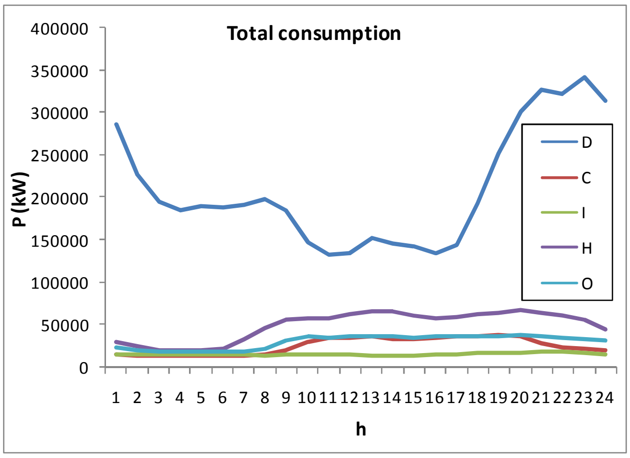

Fig. 13 – Total consumption for the five types of consumers considered (January , workdays)

Fig. 12 shows the normalized mean diagrams for the five LV types of consumers considered (domestic, commercial, industry, hotel or restaurant, other) for the month of January and for workdays.

Fig. 13 shows the total consumption for each of the types considered in the workdays of January. As we can see, the domestic (*i.e.* residential) consumers have a significant contribution to the total system load. Although their energy consumption is usually smaller than the other types, the number of domestic consumers is considerably larger than the number of industry, commerce or hotel type of consumers.

This methodology for load estimation used here comprises the following steps:

1. Generation of data patterns;
2. Training an ANN where the inputs are: weekday type (workday, Saturday, Sunday), consumer type (domestic, commercial, industry, hotel or restaurant, other), contracted hired power, month energy consumption and month. These inputs were selected based on their availability but also because of their discrimination power [14]. The ANN output is the load diagram to be estimated;
3. Aggregation of the consumers' load diagrams fed by each example MV/LV substation. Note that is this case the ANN provides an individual estimate for each consumer;
4. Tests and performance evaluation.

Fig. 14 shows the input/output structure of ANN_2. The mean absolute percentage error obtained in this case was similar to the obtained in the first case (10%). Given the usual load fluctuation of LV consumers' diagrams, this performance is satisfactory.

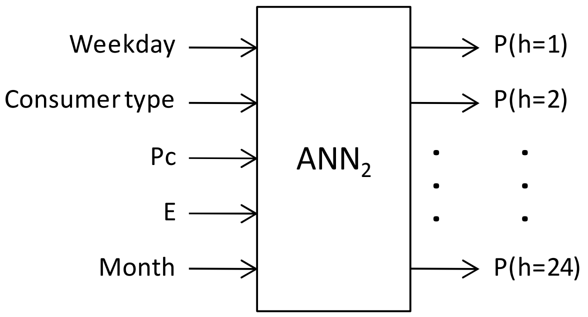

Fig. 14 – ANN_2 input/output scheme

In the examples shown in Fig 15, it is possible to confirm that ANN2 is also capable of good estimates, despite the variety of diagram shapes. The bottom example in this figure shows the diagram for a MV/LV substation whose consumers are mainly domestic type, presenting the peak late in the afternoon. On the other hand, the two top examples show two MV/LV substations where the influence of industry and commerce types of consumers are evident. In these cases, there are two similar peaks of consumption: one late in the morning and the other in the middle of the afternoon.

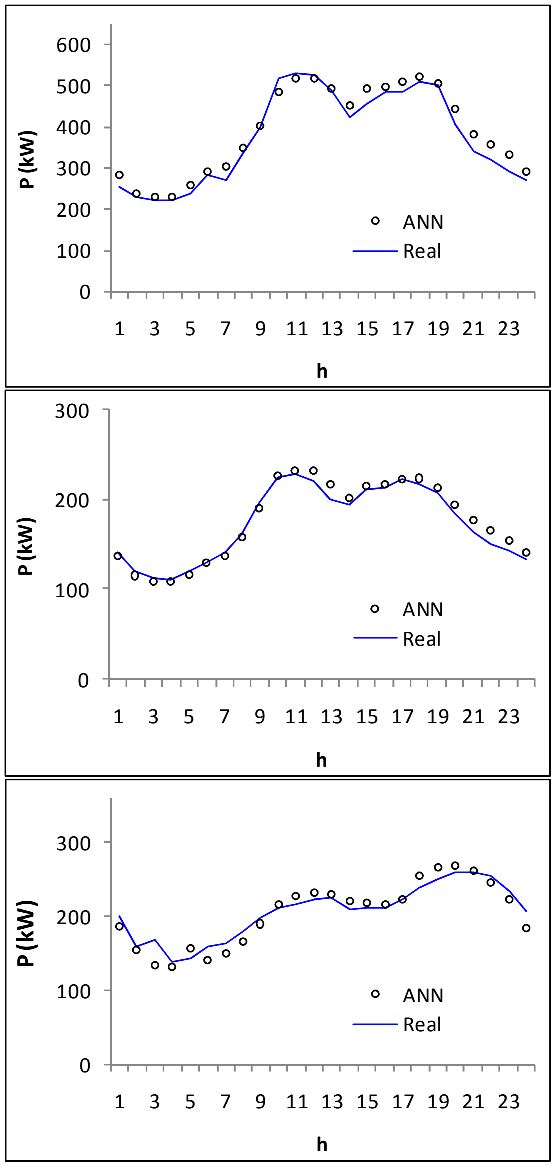

Fig. 15 – Load estimation examples (ANN_2)

The first approach includes the determination of confidence intervals. The same methodology may be applied here, using an auxiliary ANN to learn load curves' dispersion depending on the type and number of consumers aggregated in a given public station, the hour of the day, the contracted hired power weekday type (workday, Saturday, Sunday), consumer type , contracted hired power, month energy consumption and the month of the year.

9 Further developments

The next step of the current development is to test the load estimation based on the profiles approved by the Portuguese regulatory entity. Another interesting issue that may be included is weather influence, specially the temperature. This well known effect is used in most forecast techniques, and its inclusion in this work may lead to better results also in estimated load curves.

10 Conclusions

This article proposes two inference mechanisms to estimate MV/LV load curve estimation. As an additional result, confidence intervals were derived, using a complementary ANN to access the error of the first one. This approach has the capital advantage of including all kind of errors inherent to the load estimator ANN. Confidence intervals are useful to represent the uncertainty of the estimated diagrams.
The results obtained support the adopted approach, showing that this methodology constitutes a powerful tool especially for distribution planning.

References:

[1] P. Juuti, E. Lakervi, J. Partanen, "The Use of Customer Load Profiles in Distribution Network Design and Operation Planning", Proceedings ISEDEM, Singapore, 1988

[2] J. Rutten, L. Heygele, "Consumption figures as a basis for the determination of load", Proceedings CIRED, Liège, 1991

[3] K. Livik et al., "Estimation of annual coincident peak demand and load curves based on statistical analysis and typical load data", Proceedings CIRED, Brighton, 1993

[4] J.C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York, 1981.

[5] T. Kohonen, "Self Organization and Associative Memory", Springer-Verlag, Berlin, 1984

[6] F. M. Silva, L. B. Almeida, "Acceleration Techniques For The Backpropagation Algorithm", In "Neural Networks", Springer-Verlag, 1990

[7] David A. Nix, and Andreas S. Weigend, "Learning Local Error Bars for Nonlinear Regression", In Advances in Neural Information Processing Systems, MIT Pess, 1995

[8] G. Chryssolouris, "Confidence Interval Prediction for Neural Networks", in Trans. on Neural Networks, Vol. nº. 1, Jan 1996

[9] T. Heskes "Practical Confidence and Prediction Intervals", In Advances in Neural Information Processing Systems, MIT Press, 1997

[10] J. N. Fidalgo, M.T. Ponce de Leão. M.A. Matos, "Assessing Error Bars in Distribution Load Curve Estimation", Lecture Notes in Computer Science, Springer-Verlag, New-York, 1997

[11] M.A. Matos, M.T. Ponce de Leão, "Electric Distribution System Planning with Fuzzy Loads", International Transactions in Operational Research, Vol.2, No.1, Elsevier,1995

[12] Shamsuddin Ahmed, "Peak Electric Load Estimation: Al-Ain City", WSEAS Transactions On Systems, Issue 2, Volume 3, ISSN 1109-2777, April 2004

[13] Marta Marmiroli, Yoshitaka Ohta, Junjiro Sugimoto, Ryuichi Yokoyama, "Neural Networks Method To Forecast Electricity Price For Markets With High Volatility", WSEAS Transactions On Power Systems, Issue 8, Volume 1, ISSN 1790-5060, August 2006

[14] J. N. Fidalgo, "Feature Selection Based On ANN Sensitivity Analysis - A Practical Study", 2001 International Conference On Neural Networks And Applications, WSEAS 2001, Tenerife, Fev 2001

[15] Han-Ching Kuo, Yuan-Yih Hsu, "Distribution system load estimation and service restoration using a fuzzy set approach", IEEE Transactions on Power Delivery, Volume 8, Issue 4, Oct. 1993 Page(s):1950 – 1957

[16] Nazarko, J, Zalewski, W., "The fuzzy regression approach to peak load estimation in power distribution systems", IEEE Transactions on Power Systems, Volume 14, Issue 3, Aug. 1999 Page(s):809 - 814

[17] Amjady, N., "Short-term hourly load forecasting using time-series modeling with peak load estimation capability", IEEE Transactions on Power Systems, Volume 16, Issue 3, Aug. 2001 Page(s):498 – 505

[18] Cochran, W.G. (1977). Sampling Techniques, 3rd ed., New York: Wiley

[19] Gianfranco Chicco, Roberto Napoli, " Electric Energy Costumer Characterisation for Developing Dedicated Market Strategies", Porto Power Tech, PPT 2001, Porto, 2001

[20] ADEME, e EDF, "Demand-Side Management End-Use Metering Campaign in the Residential Sector", Programa SAVE, Contrato nº 4.1031/93.58, ADEME, e EDF, 1998

[21] José R. Saenz Javier Bilbao, José Luis Berasategui, Arantxa Tapia, Esther Torres, "Load Curves of Domestic Users: An Approach",

[22] José Antonio Jardini et. al., "Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers", IEEE Trans. on Power Delivery, Vol.15, No. 1, Jan. 2000