

## Improving the accuracy of effort estimation through Fuzzy set combination of size and cost drivers

CH. SATYANANDA REDDY, KVSVN RAJU

Department of Computer Science and Systems Engineering  
College of Engineering, Andhra University  
Visakhapatnam  
INDIA

[satyanandau@yahoo.com](mailto:satyanandau@yahoo.com), [kvsvn.raju@gmail.com](mailto:kvsvn.raju@gmail.com)

**Abstract:-** In this research, it is investigated the precision of size and cost drivers in the estimation of effort using Constructive Cost Model (COCOMO). It is imperative to stress that uncertainty at the input level of the COCOMO yields uncertainty at the output, which leads to gross estimation error in the effort estimation. Instead of using a single number to represent the size, it can be characterized as a fuzzy value. Cost drivers also expressed through an unclear category which needs subjective assessment. Fuzzy logic has been applied to the COCOMO using the symmetrical triangles and trapezoidal membership functions to represent the cost drivers and size. Using trapezoidal membership function for the size and cost drivers, a few attributes are assigned the maximum degree of compatibility when they should be assigned lower degrees. To overcome the above limitations, in this work, it is concentrated to use Gaussian membership function for the COCOMO parameters. In addition, this paper proposes to incorporate both size and cost drivers together, with a fuzzy set using Gaussian membership function. The present work is based on COCOMO dataset and the experimental part of the study illustrates the approach and compares it with the standard version of the COCOMO. It has been found that the proposed method is performing better than ordinal COCOMO and the achieved results were closer to the actual effort.

**Key-Words:-** Constructive Cost Model, COCOMO, Fuzzy based Effort Estimation, Gaussian Membership Function, Software Cost Estimation, Software Effort Estimation, Software Size and Project Management.

## 1 Introduction

Accurate and timely prediction of the development effort and schedule required to develop software system is one of the most critical activities in managing software projects. The precision and reliability of the effort estimation is very important for software industry because both overestimates and underestimates of the software effort are harmful to software companies. Nevertheless, accurate estimation of software development effort in reality has major implications for the management of software development. If a manager's estimate is too low, then the software development team will be under considerable pressure to finish the product quickly. On the other hand, if a manager's estimate is too high, then too many resources will be committed to the project. In point of fact, estimating software development effort remains a complex problem attracting considerable research attention. It is very important to investigate novel methods for improving the accuracy of such estimates. As a result, many models for estimating software development effort have been proposed and are in use.

This paper proposed to extend the Constructive Cost Model (COCOMO) [4] by incorporating the concept of fuzziness into the measurements of software size. The effort multipliers and scale factors of the COCOMO were described in natural language as very low, low, nominal, high, very high and extra high and these were represented by fixed numerical values. Furthermore, these values are represented with fuzzy interval values. The

advantages of this over quantization are that they are more natural and they mimic the way in which humans interpret linguistic values. Fuzzy logic-based cost estimation models are more appropriate when vague and imprecise information is to be accounted for.

Though, many membership functions were used in the literature [12] to represent the size and cost drivers, many of them are not appropriate to clear the vagueness in the size and cost drivers. The triangular, trapezoidal membership functions are being used in COCOMO to replace the conventional quantization by using fuzzy interval values [10]. With these membership functions, the transition from one interval to an adjacent interval is abrupt rather than gradual. Therefore after studying the behavior of the software size and cost drivers [5], to get emphasize, a way of propagation of uncertainty and to attain smoother transition in the Membership Function (MF), this work attempts to achieve a fuzzy based effort estimation by using fuzzy sets for its size and cost drivers together.

Hence, in this study, it has been proposed and validated empirically, that the size of a software project can be specified by distribution of its possible values and the uses of Gaussian MF to represent the size and cost drivers in the COCOMO. Consequently in this paper, the combination of size and cost drivers have been fuzzified together to get accuracy in the estimation. It has been found that the use of Gaussian MF for its size and cost drivers are performing better, as these membership functions demonstrates a smoother transition in its intervals, and the achieved results were closer to the actual effort.

The rest of this paper is organized as follows: Section 2 briefly describes the related work done for estimating the effort through different fuzzy logic approaches. Section 3 gives an overview of a proposed fuzzy effort estimation model and reveals the methodology used in this research. Section 4 presents experimental design and application of membership functions to COCOMO using fuzzy logic tool box. Section 5 summarizes the experimental results. The final section concludes that the accuracy of effort estimation can be improved through the proposed model and the estimated effort can be very close to the actual effort.

## 2 Related Work

Papers were reviewed regarding aspects related to research on software development effort estimation based on a fuzzy logic model. The fuzzy logic model uses the fuzzy logic concepts introduced by L.A.Zadeh [15]. Study showed that fuzzy logic model has a place in software effort estimation. Attempts have been made to fuzzify some of the existing models in order to handle uncertainties and imprecision problems. Using real project data, Gray and MacDonell [9] compared Function Point Analysis, Regression techniques, feed forward neural network and fuzzy logic in software effort estimation. Their results showed that fuzzy logic model achieved good performance, being outperformed in terms of accuracy only by neural network model with considerably more input

variables. In their fuzzy logic model, triangular membership functions were defined for the small, medium, large intervals of size.

Fuzzy logic had also been applied to algorithmic models to cater for the need of fuzziness in the input. The first realization of the fuzziness of several aspects of COCOMO was that of Fei and Liu [8]. The authors observed that an accurate estimate of delivered source instruction (KDSI) cannot be made before starting a project, and it is unreasonable to assign a determinate number for it. Ryder [13] researched on the application of fuzzy logic to COCOMO and Function Points models. Musflek et al. [12] worked on fuzzifying basic COCOMO model without considering the adjustment factor. On the other hand, Idri et al., [2] proposed fuzzy intermediate COCOMO with the fuzzification of cost drivers. The effort multiplier for each cost driver is obtained from fuzzy set, enabling its gradual transition from one interval to a contiguous interval. Validation results showed that the fuzzy intermediate COCOMO can tolerate imprecision in its input (cost drivers) and generate more gradual outputs.

Ahmed and Saliu [1] geared up further by fuzzifying the two different portions of the COCOMO model i.e. nominal effort estimation and the adjustment factor. They proposed a fuzzy logic framework for effort prediction by integrating the fuzzified nominal effort and the fuzzified effort multipliers of the intermediate COCOMO model. The mainstream of the work is concentrated on fuzzifying either size or cost drivers with the representation of triangular or trapezoidal membership functions. Hence, in this work,

it is proposed to incorporate both size and cost drivers together with a fuzzy set interval values using Gaussian MF.

### 3 Research Methodology

#### 3.1 Formulation of the problem

In COCOMO, effort is expressed in Person Months (PM). It determines the effort required for a project based on software project's size in Kilo Source Line Of Code (KSLOC) as well as other cost factors known as scale factors and effort multipliers by as shown in (1).

$$PM = A.(Size)^{1.01 + \sum_{i=1}^5 SF_i} \cdot \prod_{i=1}^{17} EM_i \quad (1)$$

where A is a multiplicative constant, and the set of Scale Factors (SF) and Effort Multipliers (EM) are defined the model [6]. It contains 17 effort multipliers and 5 scale factors. The standard numeric values of the cost drivers are given in Appendix.

Traditionally, the problem of software cost estimation relies on a single (numeric) value of size of given software project to predict the effort. However, the size of the project is, based on some previously completed projects that resemble the current one (especially at the beginning of the project). Obviously, correctness and precision of such estimates are limited. It is of principal importance to recognize this situation and come up

with a technology using which we can evaluate the associated imprecision residing within the final results of cost estimation. The technology endorsed here deals with fuzzy sets. Using fuzzy sets, size of a software project can be specified by distribution of its possible values.

Commonly, this form of distribution is represented in the form of a fuzzy set. It is important to stress that uncertainty at the input level of the COCOMO model yields uncertainty at the output [12]. This becomes obvious and, more importantly, bears a substantial significance in any practical endeavor. By varying the size using fuzzy set, (that reflects a level of designer's confidence as to the estimate), we can easily model the effort that impacts the estimation accuracy. Obviously, a certain monotonicity property holds, which is less precise estimates of size gives rise to less detailed effort estimates.

Cost drivers are often expressed through an unclear category which needs subjective assessment. The cost drivers and scale factors of the COCOMO were described in natural language as very low, low, nominal, high, very high and extra high and these were represented by fixed numerical values [6]. But it is not appropriate to give a fixed numerical number to each of the scales. Instead of using fixed numbers to characterize the cost drivers, interval values were used and these were represented using various membership functions triangular, trapezoidal etc. [2]. However still there was some linearity by using these functions. Overlapped symmetrical triangles or trapezoids reduce fuzzy systems to precise linear systems [3]. Furthermore

there is a possibility when using a trapezoidal function that some attributes are assigned the maximum degree of compatibility when they should be assigned lower degrees. In order to avoid this linearity it is proposed a more continuous Gaussian function to represent the cost drivers.

### 3.2 Proposed Research Method

In this investigation, it is proposed to use Gaussian MF for the size and cost drivers. Furthermore, in this paper, it is concentrated to incorporate both size and cost drivers with a fuzzy set-based generalization of the COCOMO model. For example, a small software project can be described by a fuzzy set  $K$  in the form shown in Fig.1. The grades of membership capture a notion of partial membership of an element to the concept (fuzzy set). In general, a fuzzy set  $K$  is described by its membership function  $K(x)$  which expresses the degree of membership of  $x$  to the fuzzy set  $K$  describing a certain concept (say, small project, high reliability, etc.).

In this work, it is projected to use Gaussian MF to represent the linguistic values of cost drivers and size. Gaussian MF gives more continuous transition from one interval to another [11]. A typical representation of cost driver using Gaussian MF is shown in Fig.1. Gaussian Bell curve sets give richer fuzzy system with simple learning laws that tune the bell curve variance which is represented in (2).

$$\mu_{Ai}(x) = \text{Gaussian}(x, c_i, \sigma_i) = e^{-1/2(x-c_i/\sigma_i)^2} \quad (2)$$

Where  $c_i$  is the center of the  $i^{\text{th}}$  fuzzy set and  $\sigma_i$  is the width of the  $i^{\text{th}}$  fuzzy set. We have defined the fuzzy sets corresponding to the various associated linguistic values for each cost driver.

In this research, a new fuzzy effort estimation model is proposed by augmenting the technology of fuzzy sets for both size and cost drivers to deal with linguistic data, and to generate fuzzy membership functions and rules for size and cost drivers obtained from (3). In the next step, we evaluate the COCOMO model using (1), and size and cost drivers obtained from fuzzy sets ( $F\_Size\_EM_{ij}$ ) rather than from the classical size and  $EM_{ij}$ .  $F\_Size\_EM_{ij}$  is calculated from (4), the classical size and  $EM_{ij}$  and the membership functions  $\mu$  defined for the various fuzzy sets associated with the size and cost drivers.

$$F\_Size\_EM_{ij} = F(\mu_{A1}^{V_i}(P) \dots \mu_{Aj}^{V_i}(P), \mu_A(P), EM_{i1} \dots EM_{ij}, \text{Size}) \quad (3)$$

For ease,  $F$  is taken as a linear function, where the  $\mu_{Aj}^{V_i}$  is the membership function of the fuzzy set  $A_j$  associated with the cost driver  $V_i$  and  $\mu_A$  is the membership function associated with the size.

$$F\_Size\_EM_{ij} = \sum_{j=1}^{k_i} \mu_{Aj}^{V_i}(P) \times EM_{ij} \times \text{Size} \quad (4)$$

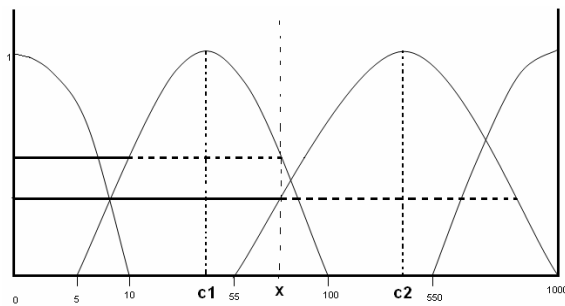


Fig.1. Representation of size and cost drivers Using Gaussian Membership Function

## 4 Design Methodology

The proposed cost estimation model was implemented using fuzzy logic tool box of MATLAB software. The fuzzy inference system (FIS) is used to implement the various processing steps. Options were provided for creating and editing FIS with fuzzy logic tool box software using graphical tools or command line functions. This GUI tool allows us to edit the higher level features such as number of input and output variables of the FIS. Membership functions were added for software size and for each cost driver using 'addmf' command. Each cost driver in fuzzy COCOMO can be defined with membership function. The membership function editor 'mfedit' that allows us to inspect and modify all the membership functions. For each membership function we can change the name, type and parameters. The size and cost drivers are defined and customized to the Gaussian MF using the command 'gaussmf' (x, [sig c]).

The size and cost drivers are measured using a rating scale of six linguistic values: 'very low', 'low', 'nominal', 'high', 'very high' and 'extra high'. The assignment of linguistic values to the cost drivers (or project attributes) uses conventional quantification where the values are intervals. For example, in the case of the DATA cost driver, we have defined a fuzzy set for each linguistic value with a Gaussian-shaped MF shown in Fig.1. We note that the fuzzy sets associated with the DATA cost driver satisfy the normal condition. The evaluation consists in comparing the accuracy of the estimated effort with the actual effort. A common criterion for the evaluation of cost estimation models is the Magnitude of Relative Error (MRE) [7] which is defined in (5).

$$\text{MRE} = \frac{|\text{Actual Effort} - \text{Predicted Effort}|}{\text{Actual Effort}} \times 100 \quad (5)$$

The Gaussian MF that has been proposed in this work gives accurate effort than by using any other membership functions. When it uses trapezoidal function the peak value is linear but in Gaussian function it touches the peak at only one point. Hence, Gaussian function is performing better than trapezoidal function, as it demonstrates a smoother transition between its intervals. The results clearly indicate that such fuzzy set modeling approach affects significantly the estimation outcomes.

## 5 Experimental Results

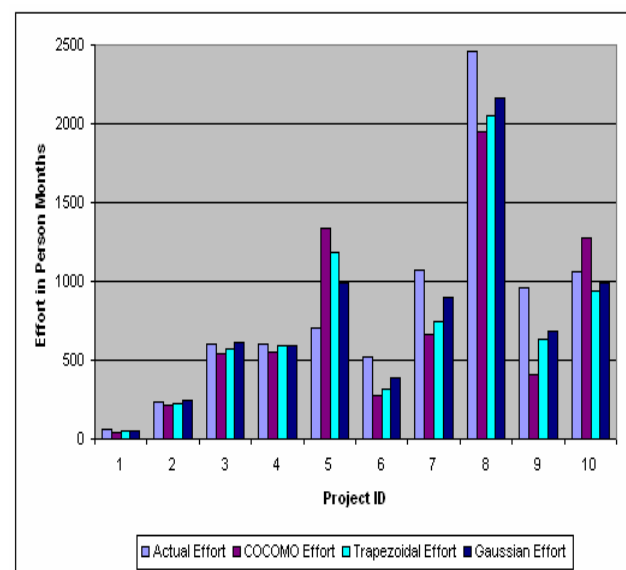
Experiments were done by taking original data from COCOMO dataset [14]. The software development efforts obtained when using COCOMO and other membership functions were observed. After analyzing the results attained by means of applying COCOMO, trapezoidal MF for cost drivers, and Gaussian MF for both size and cost drivers together, it is observed that the effort estimation of the proposed model is giving more precise results than the other models.

**Table 1. Results and Comparison of Effort Estimation in Person Months**

Project ID	Actual Effort	Effort in Person Months (PM)		
		COCOMO	Trapezoidal MF for the cost drivers alone	Gaussian MF for the size & cost drivers together
1	61	45.63	48.49	54.48
2	237	214.10	227.52	241.92
3	599	539.60	573.44	611.67
4	603	553.43	588.12	595.72
5	702	1335.1	1179.5	990.64
6	523	278.86	311.28	387.59
7	1075	661.3	749.95	902.38
8	2455	1945.4	2046.8	2163.3
9	958	408.33	628.64	687.84
10	1063	1275.9	943.75	993.06

The effort estimated by means of fuzzifying size and cost drivers together and using Gaussian MF is yielding better estimate which is very nearer to the actual effort. Therefore, using fuzzy sets, size and cost drivers of a software project can be specified by distribution of its possible values, by means of which we can evaluate the associated imprecision residing within the final results of cost estimation.

Table 1 shows the results obtained for some of the data sets taken from COCOMO dataset, which includes the effort estimated using Constructive Cost Model and the effort obtained using trapezoidal MF for the cost drivers alone, and the effort achieved using Gaussian MF for both size and cost drivers together i.e. the proposed fuzzified model.



**Fig.2. Chart representing the comparison of effort estimation.**

It has been found that proposed model is performing better than ordinal COCOMO and Gaussian function is performing better than trapezoidal function, as it demonstrates a smoother transition in its intervals, and the achieved results were closer to the actual effort.

Figure 2 shows the chart representing comparative analysis of the actual effort with that of the effort estimated using COCOMO, trapezoidal MF for cost drivers, and Gaussian MF for size and cost drivers. Effort in person months is scaled along with y-axis. Actual effort, COCOMO effort, and effort obtained using trapezoidal MF for cost drivers alone, and effort obtained using Gaussian MF for both size and cost drivers, were represented for each sample projects, which were taken along with x-axis.

The magnitude of relative errors was calculated using (5). For example, the relative error calculated for project 1 for COCOMO, trapezoidal and for the proposed model is 25.20, 20.51 and 10.69 respectively. In the case of second project it is 9.66, 4.00 and 2.08. The Mean Magnitude of Relative Error (MMRE) is 32.65, 23.22 and 14.58 respectively. Figure 3 shows the chart representing relative errors which are represented along with y-axis against each project, which is taken along with x-axis. This clearly shows that there is a decrement in the relative error, so that the proposed model is more suitable for effort estimation.

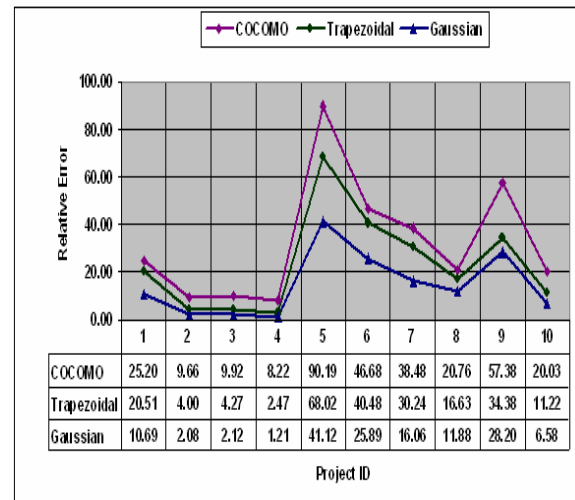


Fig.3. Assessments of Magnitude of Relative Errors

## 6 Conclusions and Future Research

The use of fuzzy modeling techniques offers an attractive alternative in the software industry. In this paper it has been proposed and examined the use of fuzzy sets rather than classical intervals in the COCOMO. Using fuzzy sets, size of a software project can be specified by distribution of its possible values and these fuzzy sets were represented by Gaussian MF. For each cost driver and size, its associated linguistic values are represented by Gaussian shaped MF. The relative error for COCOMO using Gaussian function is lower than that of the error obtained using trapezoidal MF.

From the experimental results, it is concluded that, by fuzzifying the size and cost drivers of the project, it can be proved that the resulting estimate impacts the effort. The effort generated using the proposed model gives better result than that of using ordinal COCOMO. This illustrates that by fuzzifying size and cost drivers by using Gaussian MF, the accuracy of effort estimation can be improved and the estimated effort is very close to the actual effort. Moreover, by capturing the uncertainty of the initial data (estimates), one can monitor the behavior (quality) of the cost estimates over the course of the software project. This facet adds up anew conceptual dimension to the models of software cost estimation by raising awareness of the decision making with regard to the quality of the initial data needed by the model.

This work can be extended by integrating with neural networks. By using this extended approach with the standard COCOMO models, we can take advantage of the features of neural network, such as learning ability and good interpretability. Therefore, a promising line of future work is to extend our model to the neuro-fuzzy approach.

## 7 Appendix

### COCOMO COST DRIVERS

Cost Drivers	Range	Description
RELY	0.82-1.26	Required Software Reliability
DATA	0.90-1.28	Database Size
CPLX	0.73-1.74	Product Complexity
RUSE	0.95-1.24	Developed for Reusability
DOCU	0.81-1.23	Documentation Match to Life-Cycle Needs
TIME	1.00-1.63	Execution Time Constraint
STOR	1.00-1.46	Main Storage Constraint
PVOL	0.87-1.30	Platform Volatility
ACAP	1.42-0.71	Analyst Capability
PCAP	1.34-0.76	Programmer Capability
PCON	1.29-0.81	Personnel Continuity
APEX	1.22-0.81	Applications Experience
PLEX	1.19-0.85	Platform Experience
LTEX	1.20-0.84	Language and Tool Experience
TOOL	1.17-0.78	Use of Software Tools
SITE	1.22-0.80	Multi site Development
SCED	1.43-1.00	Required Development Schedule

## References:

- [1] Ahmed, M. A., Omolade Saliu, M., Alghamdi, J., Adaptive fuzzy logic-based framework for software development effort prediction, *Elsevier Science, Information and Software Technology*; Volume 47, Issue 1, 1 January 2005, Pages 31-48.
- [2] Ali Idri and Alain Abran, COCOMO Cost Model Using Fuzzy Logic, *Proceedings of 7th International Conference on Fuzzy Theory and Technology*, Atlantic City, New Jersey, Feb 27-Mar 03, 2000.
- [3] Ali Idri, Alain Abran, La Logique Floue Appliquee Aus Modeles d'Estimation d'Effort de Development de Logiciels Cas Du Modele COCOMO, 2004, <http://www.gelog.etsmtl.ca/publications/pdf/560.pdf>.
- [4] Boehm, B.W., Software Engineering Economics, *Prentice-Hall*, Englewood Cliffs, NJ, 1981.
- [5] Boehm, B.W., Royce, W.W., Le COCOMO Ada, *Genie logiciel & Systemes experts*, 1989.
- [6] Boehm, B.W., et al., Cost Models for Future Software Life Cycle Processes: COCOMO2.0, *Annals of Software Engineering on Software Process and Product Measurement*, Amsterdam, 1995.
- [7] Briand L.C., Emam K.E. Surmann D. And Wieczorek I., An Assessment and Comparison of Common Software Cost Estimation Modeling Techniques, *ISERN-1998-27*.
- [8] Fei. Z, X. Liu., f-COCOMO: Fuzzy Constructive Cost Model in Software Engineering, *Proceedings of IEEE International Conference on Fuzzy System*, IEEE, 1992, pp. 331-337.
- [9] Gray A. R., S. G. MacDonell, Applications of Fuzzy Logic to Software Metric Models for Development Effort Estimation, *Fuzzy Information Processing Society 1997 NAFIPS' 97*, Annual Meeting of the North American, 21 – 24 September 1997, pp. 394 – 399.
- [10] Leonard J. Jowers A, James J. Buckley B, Kevin D. Reilly A, Estimation of f-COCOMO Model Parameters Using Optimization Techniques, 2006, <http://sunset.usc.edu/events/2006/CIIForum/pages/presentations/2006SEWorldJowers-Buckley-Reilly-c.pdf>.
- [11] Mohammad Azzeh, Daniel Neagu, Peter Cowling, Improving Analogy Software Effort Estimation using Fuzzy Feature Subset Selection Algorithm, *Proceedings of the 4th international workshop on Predictor models in software engineering*, PROMISE'08, May 12–13, 2008, Leipzig, Germany, Pages 71-78.

- [12]Musflek P, Pedrycz W., Succi G, Reformat M,  
Software Cost Estimation with Fuzzy Models,  
*Applied Computing Review*, 2000, Vol. 8, No. 2,  
pages 24-29.
  
- [13]Ryder. J., Fuzzy Modeling of Software Effort  
Prediction, *Proceedings of IEEE Information  
Technology Conference*, Syracuse, NY, 1-3 Sept  
1998, pp: 53-56.
  
- [14]Tim Menzies, Promise Software Engineering  
Repository,[http://promise.site.uottawa.ca/SERep  
ository/datasets-page.html](http://promise.site.uottawa.ca/SERepository/datasets-page.html)
  
- [15]Zadeh. L. A., Fuzzy Sets, Information and  
Control, Vol. 8, 1965, pp. 338-353.