International Journal on Advanced Science Engineering Information Technology

Vol.10 (2020) No. 2
ISSN: 2088-5334

Adopting the Appropriate Performance Measures for Soft Computing-based Estimation by Analogy

Muhammad Arif Shah[a,b], Dayang N. A. Jawawi[a,c*], Mohd Adham Isa[a], Muhammad Younas[a,c], Ahmad Mustafa[a]

[a] *Software Engineering Department, School of Computing, Faculty of Engineering, Universit Teknologi Malaysia, Johor Bahru, Malaysia*

[b] *City University of Science and Information Technology Peshawar, Pakistan*

[c] *Government College University, Faisalabad, Pakistan*
*E-mail: * arif.websol@gmail.com*

***Abstract*— Soft Computing based estimation by analogy is a lucrative research domain for the software engineering research community. There are a considerable number of models proposed in this research area. Therefore, researchers are of interest to compare the models to identify the best one for software development effort estimation. This research showed that most of the studies used mean magnitude of relative error (MMRE) and percentage of prediction (PRED) for the comparison of their estimation models. Still, it was also found in this study that there are quite a number of criticisms done on accuracy statistics like MMRE and PRED by renowned authors. It was found that MMRE is an unbalanced, biased, and inappropriate performance measure for identifying the best among competing estimation models. The accuracy statistics, e.g., MMRE and PRED, are still adopted in the evaluation criteria by the domain researchers, stating the reason for "widely used," which is not a valid reason. This research study identified that, since there is no practical solution provided so far, which could replace MMRE and PRED, the researchers are adopting these measures. The approach of partitioning the large dataset into subsamples was tried in this paper using estimation by analogy (EBA) model. One small and one large dataset were considered for it, such as Desharnais and ISBSG release 11. The ISBSG dataset is a large dataset concerning Desharnais. The ISBSG dataset was partitioned into subsamples. The results suggested that when the large datasets are partitioned, the MMRE produces the same or nearly the same results, which it produces for the small dataset. It is observed that the MMRE can be trusted as a performance metric if the large datasets are partitioned into subsamples.**

***Keywords*—estimation by analogy; performance metrics; MMRE; PRED; software development effort.**

I. INTRODUCTION

Software development effort estimation is one of the essential activities in the software engineering process in general and specifically in software project management. It becomes nearly impossible to plan and control a project without the accurately estimated figures. There are a considerable number of models presented to predict the software development effort, but unfortunately, none could show absolute success to estimate accurately in all the cases. The estimation models can be categorized into parametric models and nonparametric models [1]. The parametric models deal with historical projects with the help of numerical or statistical analysis [2, 3]. The non-parametric models follow soft computing based models such as artificial neural networks, fuzzy logic, genetic algorithm, estimation by analogy, or case-based reasoning [4-9]. The performance of each model is compared with other related models to show the degree of improvement. There are different performance measures used for validating the model performance, such as, relative error (RE), mean relative error (MRE), the mean magnitude of relative error (MMRE), a median of the magnitude of relative error (MdMRE), percentage of prediction (PRED), the balanced mean magnitude of relative error (BMMRE) and mean absolute residual (MAR). The most adopted performance measure for software development effort estimation is MMRE [10, 11]. The researchers want to keep the value of MMRE less than (0.25) for their estimation models as the acceptable range of MMRE is equal to or less than (0.25) [10]. The MMRE is usually used as the cross-validation method, which is the standard evaluation process [6]. One use of MMRE is to compare and select the best model among the available pool of competing estimation models. The model with the lowest MMRE is indicated as the best model, such as Shepperd and Schofield [8], and Myrtveit and Stensrud [12]. The MMRE can be calculated, as shown

in Equation 1. Where xi describes the variable of interest or estimated value, and x denotes the actual value.

$$MMRE = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|x_i - x|}{x_i}\right) \quad (1)$$

The soft computing based estimation by analogy was initiated by Idri and Abran [13], and it is still the interest of many researchers. Most of the studies in this research domain adopted MMRE and PRED (.25) as the performance evaluation metric for comparing and ranking the estimation models even though these performance measures are much criticized by the domain researchers [Foss, Stensrud [14] [15-21].

This study (1) highlights the criticisms done by different researchers on PRED, MMRE, and various versions of it and the reasons behind the adoption of these performance measures even after proven and published researches against them. (2) This research also indicates to the performance measure to be adopted for soft computing-based estimation by analogy.

There are overwhelming related works on soft computing-based estimation by analogy. The very first study in this domain was published by Idri and Abran [13], and the number of publications is still growing. There could be found many soft computing techniques used with estimation by analogy and each technique, which makes it difficult to identify the best in a typical situation or circumstances. The performance of each technique evaluated, and the weaknesses and strengths need to be identified for its appropriateness. There have been published many studies to deal with the comparison of techniques and models of soft computing-based estimation by analogy. These studies used MRE, MMRE, MdMRE, MAR, BMMRE, and PRED as performance metrics. There are also a considerable number of studies performed on these performance metrics, questioning their reliability and validity for unbiased model comparison. These studies have claimed and proved that the discussed performance metrics does not guarantee that an inferior estimation model would be avoided, and it may wrongly mark the inferior model as the superior model.

Côté, Bourque [22], brought forward MMRE and other performance metrics. There are a few studies that seem to be related to this research. Miyazaki, Terakado [23], criticized MMRE that it even produces lower values for models that calculate underestimated predication and believed that summary statistics could perform better. Still, the implications of MMRE were not investigated. Kitchenham, Pickard [24], attempted to find the necessary measurement output of MMRE and related metrics. According to them, PRED relates to kurtosis, and MMRE is a metric to find the variance of variable y=y^. They suggested observing the box plot of y=y^ both. Stensrud, Foss [25], indicated the strength of MRE by showing that the size and MRE are virtually uncorrelated. Foss, Myrtveit [26], suggested that MRE is dependent on the scale while indicating a consequence of the project manager may falsely have faith in high accuracy for small ERP projects concerning MMRE. Foss, Stensrud [14], performed a simulation and stated that MMRE does not select the best model all the time and called it a biased performance measure for model comparison. The authors suggested to use an amalgam of the models' theoretical justifications. Stensrud, Foss [27], presented an empirical validation for showing the relation between MMRE and project size. At the same time, they criticised MMRE for not being an appropriate metric of MRE for large and small projects. They suggested to partition the large datasets in subsamples for better use of MMRE.

Shepperd and Kadoda [15], suggested considering the context of prediction in the evaluation of estimation models. Myrtveit, Stensrud [16], stated that MMRE is an unreliable validation procedure and invalid for selecting between the competing estimation models. They strongly urge the development of reliable measures to have confidence in comparing the prediction models. Menzies, Port [28], discussed that MMRE is conventionally biased alongside overestimates, although PRED categorizes accurate prediction systems. Port and Korte [29], anticipated that increase or decrease in PRED and MMRE values depend up the size of data. Stensrud, Foss [25], showed that the causes of inferior model selection and conclusion instability are the models based on accuracy statistics MBRE, MIBRE, MRE, and MMRE. They argued that metrics should represent the functional form of a prediction model to avoid inferior model selection. Shepperd and MacDonell [19], proposed a new framework for predictive system comparison based on their standardized accuracy, guessing the random predictions using Monte Carlo and computing the effect sizes. Langdon, Dolado [20] identified the limitation in the framework (MARP0) proposed by Shepperd and MacDonell [19]. They argued that MARPO causes overestimation due to standardized accuracy measures. They further stated that calculating and unbiased MARPO is practical in software engineering datasets:

II. Materials and Method

In soft computing-based estimation by analogy, the results are evaluated by different performance measures. Over time, different studies used various performance measures. There were a total of 62 studies found, published until January 2018, related to the study domain, which used the performance metrics MRE, MMRE, MAR, MdMRE, PRED, BMMRE, MEMRE, MIBRE, MBRE, SD, LSD, RSD, and RMSE. Some studies used more than one metric. The frequencies of these performance measures are shown in Figure 1, which shows MRE, MMRE, and PRED are the most used performance metrics by these studies.

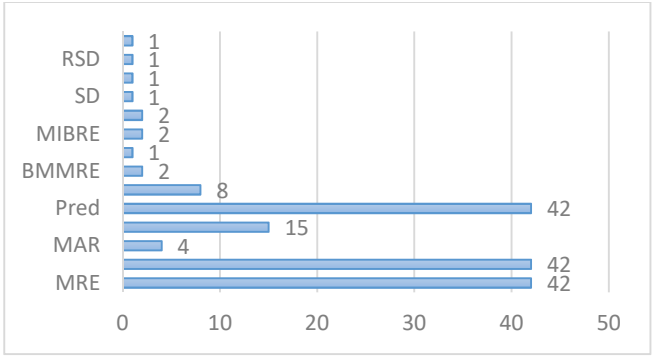

Fig. 1 Frequencies of the performance measures used by different studies.

The percentage of the studies can be seen in Figure 2, which shows that MRE appeared in 28% of the total domain studies, MMRE appeared in 27% of the total studies and PRED was also used by 27% soft computing-based estimation by analogy studies. In contrast, MdMRE was the interest of 10% of the studies. Figure 2 shows that MRE, MMRE, and PRED are the most widely adopted performance measures for soft computing-based estimation by analogy study domain. Though these are the widely adopted performance measures, many authors, as discussed in the next section, criticized the validity of these measures.

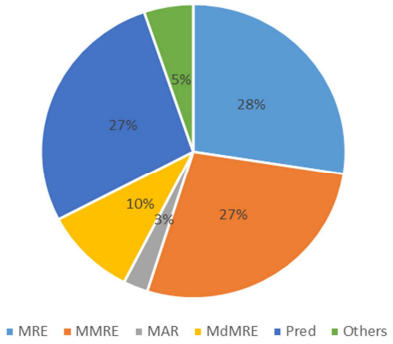

Fig. 2 The percentage of performance measures adopted by domain studies

A. *Criticism on Accuracy Statistics*

The literature revealed that MMRE is the most widely adopted performance measure for soft computing based estimation by analogy, but at the same instant, it is criticized by many studies, e.g., Kitchenham, Pickard [24], Stensrud, Foss [27], Myrtveit, Stensrud [16], and Port and Korte [29], etc. The accuracy statistics such as MMRE were criticized as unreliable, inappropriate for model comparison, unbalanced, unable to classify superior and inferior for large datasets, etc. accurately. Some of the other with their one-liner criticism on the accuracy statistics are provided in the next paragraph.

Jorgensen [30] and Briand, Langley [31] indicated that a few excessively high MRE values affect MMRE values. According to Kitchenham, Pickard [24] MMRE produces different assessment values for different estimation models. Shepperd and MacDonell [19], Stensrud, Foss [27], Shepperd and Kadoda [15] and Menzies, Port [28] termed accuracy statistics such as MMRE as a biased performance measure for prediction model comparison. Myrtveit and Stensrud [32], criticized that MER, MMRE, MIBRE, and MBRE lead to inferior model selection and conclusion instability. The details are presented in Table 1.

TABLE I
THE ACCURACY STATISTICS CRITICIZED BY DIFFERENT STUDIES

Study Reference	**Criticism**
Jorgensen [29]	A few excessively high MRE values affect the MMRE value
Briand, Langley [30]	A few excessively high MRE values affect the MMRE value
Kitchenham, Pickard [23]	Different assessment values for different estimation models
Stensrud, Foss [24]	MMRE is an inappropriate measure due to its dependence on project size
Foss, Stensrud [13]	MMRE is not reliable for comparison of estimation models
Foss, Myrtveit [25]	MMRE is an inappropriate measure due to its dependence on project size
Stensrud, Foss [26]	MMRE is biased as compared to Absolute Residuals in model comparison due to its dependence on the number of features, characteristics size and distribution type
Shepperd and Kadoda [15]	Validation procedure of MMRE is unreliable due to biasness in competing estimation models
Menzies, Port [27]	MMRE is biased as compared to PRED
Port and Korte [28]	The accuracy of MMRE and PRED are affected by dataset size
Li, Xie, and Goh [32]	MMRE is an unbalanced measure which gives better for underestimation rather overestimation
Myrtveit and Stensrud [31]	MER, MMRE, MIBRE, and MBRE leads to inferior model selection and conclusion instability
Shepperd and MacDonell [19]	MMRE is biased accuracy statistic
Langdon, Dolado [19]	$MARP_0$ leads to an overestimation

Table 1 shows the authors and their criticism of the accuracy statistics used as the performance measure for soft computing-based estimation by analogy. Most of the studies criticized MMRE out of all the available performance measures. Most of the critiques indicated that MMRE is unable to make a comparison of models when different datasets are followed or when the size of a dataset is bigger. According to them, it is highly likely that MMRE would select an inferior model as a superior model in the estimation or prediction model comparison.

B. *Solutions Proposed by the Existing Studies*

There were a few solutions suggested and provided by some studies, as shown in Table 2. Though the studies did not provide any applicable metric that would replace the MMRE at once, however, there were some handy suggestions that would help the researchers to develop or come up with an unbiased solution. Few of the solution provided by the studies are discussed as:

Jorgensen [30], Briand, Langley and Wieczorek [31] and Li, Xie and Goh [33] Included MdMRE in the process of evaluation with MMRE, Stensrud, Foss, Kitchenham and Myrtveit [25] suggested that the dataset should be partitioned into subsamples, and evaluation should be performed for each subsample, Port and Korte [29] stated that considering Standard Error is important for accuracy measure to avoid biased results. Kitchenham, Pickard, MacDonell, and Shepperd [24] indicated that the MMRE as the measure of spread and PRED as the kurtosis of *z* variable when *z* equals estimated divided by actual. Kurtosis and measure of spread, as well as the skewness of *z* and the measure of the central location, is necessary.

Solutions suggested by these studies can be summed up with two significant highlights. Firstly, the short-term solution: large the dataset should be partitioned into

subsamples, and evaluation should be performed on each subsample. Secondly, to come up with the composite of existing or a novel, balanced, appropriate and unbiased performance measures for estimation, which will be capable of dealing with large or small and complex or simple datasets. In this paper, the short-term solution is validated on Estimation by Analogy (EBA).

TABLE II
SOLUTIONS PROVIDED FOR THE CRITICIZED ACCURACY STATISTICS BY DIFFERENT STUDIES

Study Reference	Solution
Jorgensen [29]	Included MdMRE in the process of evaluation with **MMRE**
Briand, Langley [30]	Included MdMRE in evaluation with **MMRE** due to less sensitive to extreme values
Kitchenham, Pickard [23]	The MMRE as the measure of spread and PRED as the kurtosis of z variable when z equals *estimated divided by actual.* Kurtosis and measure of spread, as well as the skewness of z and the measure of the central location, is necessary
Stensrud, Foss [24]	The dataset should be partitioned into subsamples and evaluation should be performed for each subsample
Foss, Stensrud [13]	Suggested to look for existing other existing statistical analysis methods to adopt for prediction models in software engineering instead of reinventing new ones
Port and Korte [28]	Considering Standard Error is important for accuracy measure to avoid biased results
Li, Xie, and Goh [32]	Included MdMRE in the process of evaluation with **MMRE**
Myrtveit and Stensrud [31]	The functional form of estimation models may avoid biases in model comparison and to integrate multiple accuracy metrics such as MMRE, MAR, and MMER
Shepperd and MacDonell [19]	Proposed MARP0 framework for model comparison on standardized accuracy, guessing and calculation of sizes effect

C. The Short-Term Solution

In the short-term solution, there are taken two different datasets of relatively large and small scales such as International Software Benchmarking Standard Group (ISBSG) Release 2011 [33] and Desharnais to check the effects of MMRE evaluation criteria. The ISBSG dataset is a much bigger dataset with the total number of 5052 project data, concerning Desharnais, which has the data of 81 projects only. The ISBSG dataset, which is relatively bigger than Desharnais, is also partitioned into several subsamples so that the effect of MMRE on the small, large, and partitioned dataset is observed estimation by analogy model was utilized to check the validity MMRE as a performance evaluation metric.

1) Estimation by Analogy (EBA): Shepperd (1997), introduced Estimation by Analogy (EBA) as a replacement of algorithmic models. In EBA, the effort of the target project is estimated by making an analogy with similar projects completed in the past. The EBA has widely been adopted due to its simplicity and resemblance to human behavior [1]. There are four main steps of EBA, such as,

- Past dataset
- Similarity function
- The retrieval rules
- Solution function

EBA executes the estimation process concerning the following steps

- Producing dataset from the gathered data of past projects
- Adopting the proper features or attributes for comparisons such as Lines of Code (LOC) and Functional Points (FP)
- Retrieving the past project and finding the similarity between past and targeted project
- Estimating the targeted project's effort

2) Similarity Function: In EBA, a comparison between the features of two projects is made by finding the similarity between them using a similarity function. The most prominent similarity functions are Manhattan Similarity (MS) and Euclidean Similarity (ES) [34]. The ES is shown in Equation 1.

$$Sim(p, p') = \frac{1}{\left[\sqrt{\sum_{i=1}^{n} w_i Dis(f_i, f_i')} + \delta\right]} \quad (2)$$

$$Dis(f_i, f_i') = \begin{cases} (f_i, f_i')^2, \text{if } f_i \text{ and } f_i' \text{are numeric or ordinal} \\ 1 \quad \text{if } f_i \text{ and } f_i' \text{are nominal and } f_i = f_i' \\ 0 \quad \text{if } f_i \text{ and } f_i' \text{are nominal and } f_i \neq f_i' \end{cases}$$

Where, w_i represents the assigned weight to each attribute, p, and p' represents the target and past project. The i^{th} feature of past and target project is represented by f and f'. δ is introduced to retrieve non-zero result while n determines the total number of attributes. The formula of MS is almost similar to that of ES, but it also calculates the absolute differences between the features, as shown in Equation 3.

$$Sim(p, p') = \frac{1}{\left[\sqrt{\sum_{i=1}^{n} w_i Dis(f_i, f_i')} + \delta\right]} \quad (3)$$

$$Dis(f_i, f_i') = \begin{cases} |f_i, f_i'|^2, \text{if } f_i \text{ and } f_i' \text{are numeric or ordinal} \\ 1 \quad \text{if } f_i \text{ and } f_i' \text{are nominal and } f_i = f_i' \\ 0 \quad \text{if } f_i \text{ and } f_i' \text{are nominal and } f_i \neq f_i' \end{cases}$$

3) Solution Function: The similar projects found through similarity functions are utilized by the solution function to estimate the development effort. The closest

analogy, mean of most similar projects, an average of most related projects, the median of the related projects, and the inverse distance weighted mean of the projects through Equation 4.

$$C_p = \sum_{K=1}^{K} \frac{Sim(p, p_k)}{\sum_{i=1}^{k} Sim(p, p_i)} Cp_k \quad (4)$$

Where the ***k*th** most related projects are denoted by p_k, p denoted the new project, $Sim\ (p, p_k)$ shows the similarity between project p and p_k, Cp_k shows the value of effort of the ***k*th** most related project. k itself represents the total number of most related projects.

III. Results and Discussion

EBA estimation model was utilized to assess the performance of MMRE on the large, small, and partitioned datasets. The MMRE values of EBA on these datasets are shown in Table 3. The value of MMRE was calculated as 0.1736 for Desharnais (relatively small dataset); it proves the finding of previous studies, which indicated that MMRE has better values for small projects. The value of MMRE for the large dataset (ISBSG) was calculated as 1.0998, which ironically is indicating the poor performance of EBA. The same model was indicated as a better model by MMRE with a small dataset, but it was indicated less effective (as the value of MMRE is maximum). When the large dataset (ISBSG) was partitioned into subsamples, the MMRE value was produced as 0.3471, which also shows better results for the EBA estimation model. The results of small and partitioned datasets are nearly similar as compared to a large dataset.

TABLE III
MMRE values of EBA on small, large and partitioned datasets

Dataset	MMRE
Small (Desharnais)	0.1736
Large (ISBSG)	1.0998
Partitioned (ISBSG)	0.3471

Though the study domain of this research is to identify the appropriate performance measure for soft computing-based estimation by analogy, the criticism done on the existing performance measure applies to all the estimation and prediction models. The studies reviewed for accuracy statistics, e.g., MMRE, MdMRE, etc. were strictly from the domain of soft computing-based estimation by analogy to highlight the need for appropriate performance measures for the selected research domain.

It seems that most of the studies criticized MMRE, which is the most prominent accuracy measure. The summary of the criticisms done on the MMRE indicates to its bias in comparing the estimation models. Most of the studies called it unbalanced, biased, or inappropriate measures for competing estimation models. According to these studies, the performance of MMRE is inversely proportional to the size and nature of the project or dataset. The MMRE identifies the superior model as superior until the data is not complex and large, but the case is changed for complex and large projects or datasets. It may classify the superior model as an inferior or inferior model as superior. It raises the question of what makes the researchers use these performance measures even after valid, critical, and justified criticism done by renowned authors as MMRE is still widely adopted performance measure. It was found in the review process of this study that from the origination of soft computing-based estimation by analogy research domain until recently (January 2018) the studies adopted MMRE in their evaluation criteria, such as [35,36,37,38]. Most of the studies did not provide any specific reason, and some provided the only reason "MMRE is the most widely adopted performance measure." This might be due to the lack of novel and replaceable solutions since the solutions provided by each of the critiques were only suggestions and not some practical replacements for MMRE. There was an attempt made by Shepperd and MacDonell [20] who proposed a new framework for measuring the performance, but that was also claimed as biased by Langdon, Dolado [20].

There were many solutions provided or suggested by the critiques, but for the short term, the most applicable and expedient could be dividing the large project or dataset into subsamples, MMRE could better utilize these subsamples as it works best for small project data. The other most feasible solution seems to use a composite of the existing statistical techniques for predicting the inferior model as inferior and superior as superior, e.g., the composite of MdMRE and MAR or any other capable of integration accuracy statistics could lead to a better evaluation method as compared to MMRE and its individual flavors.

The approach of the short-term was tried in this paper using the EBA model. Two datasets small and large with specific projects were considered for it, such as Desharnais and ISBSG release 11. The ISBSG dataset is a large dataset with respect to Desharnais. The ISBSG dataset was partitioned into subsamples. The results suggested that when the large datasets are partitioned, the MMRE produces the same or nearly the same results, which it produces for the small dataset. It is observed the MMRE can be trusted as a performance metric if the large dataset is partitioned into subsamples.

IV. Conclusion

Soft computing-based estimation by analogy is an important and lucrative domain for researchers. There are a considerable number of models proposed in this research area. Therefore, researchers are of interest to compare the models to identify the best one for software development effort estimation. This research showed that most of the studies used MMRE for the comparison of their estimation models. Still, it was also found in this study that there are quite a number of criticisms done on accuracy statistics like MMRE and PRED by renowned authors. It was found that MMRE is an unbalanced, biased, and inappropriate performance measure for identifying the best among competing estimation models. The accuracy statistics, e.g., MMRE and PRED, are still adopted in the evaluation criteria by the domain researcher stating the reason of "widely used," which is not a valid reason. This research study identified that, since there is no practical solution provided so far, which could replace MMRE and PRED, the

researchers are adopting these measures. This research indicated future research directions for a performance measure, such as trying a different composition of the existing accuracy measures or develop new ones to check against all the estimation models proposed to date. The importance of performance measurement is highlighted by the scope of future studies in this area. It is apprehended that all the estimation models proposed to date will have to be re-evaluated to have unbiased estimation techniques, approaches, models, and frameworks. The approach of the short-term was tried in this paper using the EBA model. Two datasets small and large were considered for it, such as Desharnais and ISBSG release 11. The ISBSG dataset is a large dataset with respect to Desharnais. The ISBSG dataset was partitioned into subsamples. The results suggested that when the large datasets are partitioned, MMRE produces the same or nearly the same results, which it produces for the small dataset. It is observed the MMRE can be trusted as a performance metric if the large dataset is partitioned into subsamples.

ACKNOWLEDGMENT

The authors express their deepest gratitude to Universiti Teknologi Malaysia (UTM) for supporting the publication of this research work under UTM-TDR grant vot no. 06G23 and Ministry of Education (MOE) Malaysia for FRGS grant vot no. 5F117

REFERENCES

[1] M. Jørgensen, *Forecasting of software development work effort: Evidence on expert judgement and formal models.* International Journal of Forecasting, **23**(3): p. 449-462, 2007.

[2] B. Boehm, B. Clark, Horowitz and Brown, *Software cost estimation with Cocomo II with Cdrom*, Prentice Hall PTR, 2000.

[3] E. Mendes, *The use of Bayesian networks for web effort estimation: further investigation*. in *Web Engineering, 2008. ICWE'08. Eighth International Conference on*, IEEE, 2008.

[4] K. V. Kumar, et al., *Software development cost estimation using wavelet neural networks.* Journal of Systems and Software, **81**(11): p. 1853-1867, 2008.

[5] S. J. Huang, N. H. Chiu and L. W. Chen, *Integration of the grey relational analysis with genetic algorithm for software effort estimation.* European Journal of Operational Research, **188**(3): p. 898-909, 2008.

[6] M. O. Elish, *Improved estimation of software project effort using multiple additive regression trees.* Expert Systems with Applications, **36**(7): p. 10774-10778, 2009.

[7] J. Stefanowski, *An empirical study of using rule induction and rough sets to software cost estimation*. Fundamenta Informaticae, **71**(1): p. 63-82, 2006.

[8] M. Shepperd and C. Schofield, *Estimating software project effort using analogies.* IEEE Transactions on software engineering, **23**(11): p. 736-743, 1997.

[9] M. A. Ahmed and Z. Muzaffar, *Handling imprecision and uncertainty in software development effort prediction: A type-2 fuzzy logic based framework.* Information and Software Technology, **51**(3): p. 640-654, 2009.

[10] S. D. Conte, H. E. Dunsmore, and V. Y. Shen, *Software engineering metrics and models*. 1986: Benjamin-Cummings Publishing Co., Inc.

[11] Idri, Ali, Fatima azzahra Amazal, and A. Abran. "Analogy-based software development effort estimation: A systematic mapping and review." Information and Software Technology58, 206-230, 2015.

[12] I. Myrtveit and E. Stensrud, *A controlled experiment to assess the benefits of estimating with analogy and regression models.* IEEE transactions on software engineering, **25**(4): p. 510-525, 1999.

[13] A. Idri and A. Abran. *Towards a fuzzy logic based measures for software projects similarity*. in *Proc 6th MCSEAI'2000 Maghrebian Conference on Computer Sciences*. 2000.

[14] T. Foss, et al., *A simulation study of the model evaluation criterion MMRE.* IEEE Transactions on Software Engineering, **29**(11): p. 985-995, 2003.

[15] M. Shepperd, and G. Kadoda. *Using simulation to evaluate prediction techniques [for software]*. in *Software Metrics Symposium, 2001. METRICS 2001. Proceedings. Seventh International*. IEEE, 2001.

[16] I. Myrtveit, E. Stensrud, and M. Shepperd, *Reliability and validity in comparative studies of software prediction models.* IEEE Transactions on Software Engineering, **31**(5): p. 380-391, 2005.

[17] A. R. Gray and S. G. Macdonell, *Software metrics data analysis—exploring the relative performance of some commonly used modeling techniques.* Empirical Software Engineering, **4**(4): p. 297-316, 1999.

[18] M. Shepperd and G. Kadoda, *Comparing software prediction techniques using simulation.* IEEE Transactions on Software Engineering, **27**(11): p. 1014-1022, 2001.

[19] M. Shepperd, and S. MacDonell, *Evaluating prediction systems in software project estimation.* Information and Software Technology, **54**(8): p. 820-827, 2012.

[20] W. B. Langdon, et al., *Exact mean absolute error of baseline predictor, MARP0.* Information and Software Technology, **73**: p. 16-18, 2016.

[21] M. A. Shah, D. N. A. Jawawi, M. A. Isa, K. Wakil, M. Younas and A. Mustafa, *"MINN: A Missing Data Imputation Technique for Analogy-based Effort Estimation".* International Journal of Advanced Computer Science and Applications(IJACSA), **10**((2)), 2019.

[22] V. Côté, et al., *Software metrics: an overview of recent results.* Journal of Systems and Software, **8**(2): p. 121-131, 1988.

[23] Y. Miyazaki , et al., *Robust regression for developing software estimation models.* Journal of Systems and Software, **27**(1): p. 3-16, 1994.

[24] B. A. Kitchenham, et al., *What accuracy statistics really measure.* IEE Proceedings-Software, **148**(3): p. 81-85, 2001.

[25] E. Stensrud, et al. *An empirical validation of the relationship between the magnitude of relative error and project size*. in *Software Metrics, 2002. Proceedings. Eighth IEEE Symposium on*. IEEE, 2002.

[26] T. Foss, I. Myrtveit, and E. Stensrud. *MRE and heteroscedasticity: An empirical validation of the assumption of homoscedasticity of the magnitude of relative error*. in *Proc. ESCOM, 12th European software control and metrics conference. The Netherlands*. 2001.

[27] E. Stensrud, et al., *A further empirical investigation of the relationship between MRE and project size.* Empirical software engineering, **8**(2): p. 139-161, 2003.

[28] T. Menzies, et al. *Validation methods for calibrating software effort models*. in *Proceedings of the 27th international conference on Software engineering*. ACM, 2005.

[29] D. Port, and M. Korte. *Comparative studies of the model evaluation criterions mmre and pred in software cost estimation research*. in *Proceedings of the Second ACM-IEEE international symposium on Empirical software engineering and measurement*. ACM, 2008.

[30] M. Jorgensen, *Experience with the accuracy of software maintenance task effort prediction models.* IEEE Transactions on software engineering, **21**(8): p. 674-681, 1995.

[31] L. C. Briand , T. Langley, and I. Wieczorek. *A replicated assessment and comparison of common software cost modeling techniques*. in *Proceedings of the 22nd international conference on Software engineering*. ACM, 2000.

[32] I. Myrtveit, and E. Stensrud, *Validity and reliability of evaluation procedures in comparative studies of effort prediction models.* Empirical Software Engineering, **17**(1-2): p. 23-33, 2012.

[33] ISBSG (2011) International Software Benchmarking Standard Group from www.isbsg.org

[34] M. Shepperd and C. Schofield, Estimating software project effort using analogies. *IEEE transactions on software engineering, 23*(11), 736-743, 1997.

[35] I. Thamarai and S. Murugavalli. "Model for improving the accuracy of relevant project selection in analogy using differential evolution algorithm." Sādhanā 42.1, 23-31, 2017.