WILEY
Publishers Since 1807

JOHN WILEY & SONS, LTD., THE ATRIUM, SOUTHERN GATE, CHICHESTER P019 8SQ, UK

*** PROOF OF YOUR ARTICLE ATTACHED, PLEASE READ CAREFULLY ***

After receipt of your corrections your article will be published initially within the online version of the journal.

PLEASE NOTE THAT THE PROMPT RETURN OF YOUR PROOF CORRECTIONS WILL ENSURE THAT THERE ARE NO UNNECESSARY DELAYS IN THE PUBLICATION OF YOUR ARTICLE

☐ **READ PROOFS CAREFULLY**

ONCE PUBLISHED ONLINE OR IN PRINT IT IS NOT POSSIBLE TO MAKE ANY FURTHER CORRECTIONS TO YOUR ARTICLE

- This will be your only chance to correct your proof
- Please note that the volume and page numbers shown on the proofs are for position only

☐ **ANSWER ALL QUERIES ON PROOFS** (Queries are attached as the last page of your proof.)

- List all corrections and send back via e-mail to the production contact as detailed in the covering e-mail, or mark all corrections directly on the proofs and send the scanned copy via e-mail. Please do not send corrections by fax or post

☐ **CHECK FIGURES AND TABLES CAREFULLY**

- Check sizes, numbering, and orientation of figures
- All images in the PDF are downsampled (reduced to lower resolution and file size) to facilitate Internet delivery. These images will appear at higher resolution and sharpness in the printed article
- Review figure legends to ensure that they are complete
- Check all tables. Review layout, titles, and footnotes

☐ **COMPLETE COPYRIGHT TRANSFER AGREEMENT (CTA) if you have not already signed one**

- Please send a scanned signed copy with your proofs by e-mail. **Your article cannot be published unless we have received the signed CTA**

☐ **OFFPRINTS**

- 25 complimentary offprints of your article will be dispatched on publication. Please ensure that the correspondence address on your proofs is correct for dispatch of the offprints. If your delivery address has changed, please inform the production contact for the journal – details in the covering e-mail. Please allow six weeks for delivery.

Additional reprint and journal issue purchases

- Should you wish to purchase a minimum of 100 copies of your article, please visit http://www3.interscience.wiley.com/aboutus/contact_reprint_sales.html
- To acquire the PDF file of your article or to purchase reprints in smaller quantities, please visit http://www3.interscience.wiley.com/aboutus/ppv-articleselect.html. Restrictions apply to the use of reprints and PDF files – if you have a specific query, please contact permreq@wiley.co.uk. Corresponding authors are invited to inform their co-authors of the reprint options available
- To purchase a copy of the issue in which your article appears, please contact cs-journals@wiley.co.uk upon publication, quoting the article and volume/issue details
- Please note that regardless of the form in which they are acquired, reprints should not be resold, nor further disseminated in electronic or print form, nor deployed in part or in whole in any marketing, promotional or educational contexts without authorization from Wiley. Permissions requests should be directed to mailto: permreq@wiley.co.uk

STVR 419
pp: 1–34 (col.fig.: Nil)
PROD. TYPE: COM
ED: Senthil
PAGN: Babu – SCAN:

SOFTWARE TESTING, VERIFICATION AND RELIABILITY
Softw. Test. Verif. Reliab. (2009)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/stvr.419

On the selection of software defect estimation techniques

João W. Cangussu[1,*,†], Syed W. Haider[1], Kendra Cooper[1] and Michael Baron[2]

[1]*Department of Computer Science, The University of Texas at Dallas, Richardson, TX, U.S.A.*
[2]*Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX, U.S.A.*

SUMMARY

Estimating the number of defects in a software product is an important and challenging problem. A multitude of estimation techniques have been proposed for defect prediction. However, not all techniques are applicable in all cases. The selection of the proper approach to use depends on multiple factors: the features of the approach; the availability of resources; and the goals for using the estimated defect data. In this paper a survey of existing estimation techniques and a decision support approach for selecting the most suitable defect estimation technique for a project, with specific goals, is proposed. The results of the ranking are a clear indication that no estimation technique provides a single, comprehensive solution; the selection must be done according to a given scenario. Copyright © 2009 John Wiley & Sons, Ltd.

Received 4 August 2008; Revised 19 August 2009; Accepted 31 August 2009

KEY WORDS: defect estimation; software testing; estimation theory

1. INTRODUCTION

Estimating the number of defects in a software product is an important and challenging problem. **Q1**
The defect estimate is a fundamental metric used in managing the schedules and resources for testing, defining product release schedules, estimating the amount of customer support needed after a release, and predicting the level of customer satisfaction for a product. There are three main dimensions when addressing the defect estimation problem: (i) the model definition, (ii) the goal of the estimator, and (iii) the estimation technique used. The first dimension refers to how the behaviour of the problem is modeled. For example, some models rely on a single exponential decay for the decrease in the number of defects [1–3], while others present a double exponential solution [4].

*Correspondence to: João W. Cangussu, Department of Computer Science, The University of Texas at Dallas, P.O. Box 830688 M/S EC31, Richardson, TX 75083, U.S.A.
†E-mail: cangussu@utdallas.edu

Also, some models are based upon the capture–recapture of defects [5,6] and others are fully based on the defect decay behaviour [1]. The second dimension regards the goal of the estimator. In some cases one is interested in computing the number of remaining defects in a product [7], while others have the goal of identifying the most defect prone modules [8–10]. Once a model is available and the goal is defined the third dimension can be considered, that is, which estimator should be used. A large variety of estimators are available such as the maximum likelihood estimator (MLE), least-square estimator (LSE), Bayesian estimator, among many others. Although a correlation exists between models and estimators, the focus here is on the third dimension: the selection of a proper estimator taking into consideration both technical and practical aspects, such as cost, applicability, latency, etc.

A number of surveys of estimation techniques exist [11,12]. For example, Fenton's work [11] provides a broad critique of existing techniques; however, the analysis does not consider additional factors that occur in practice. The work of Briand *et al.* [12] presents a comprehensive survey that is focused on capture–recapture models.

The direction of this work is to provide a more comprehensive analysis, which recognizes that the 'best' technique available cannot always be selected due to resource and technical constraints. In practice, the 'right' choice, which better fits the current situation, needs to be made. That is, the goal is to classify existing defect estimation approaches and devise a mechanism for the selection of the best technique based on alternative user-defined scenarios. Figure 1 presents an overview of the process, which consists of three main parts: the project-independent characterization of estimation approaches, the project-dependent scenarios, and a multi-criteria decision making (MCDM) approach that is used to rank the estimation approaches [13,14].

As a part of the project-independent characterization of estimation approaches, the first step is to classify existing defect estimation techniques based on their requirements and technical aspects.

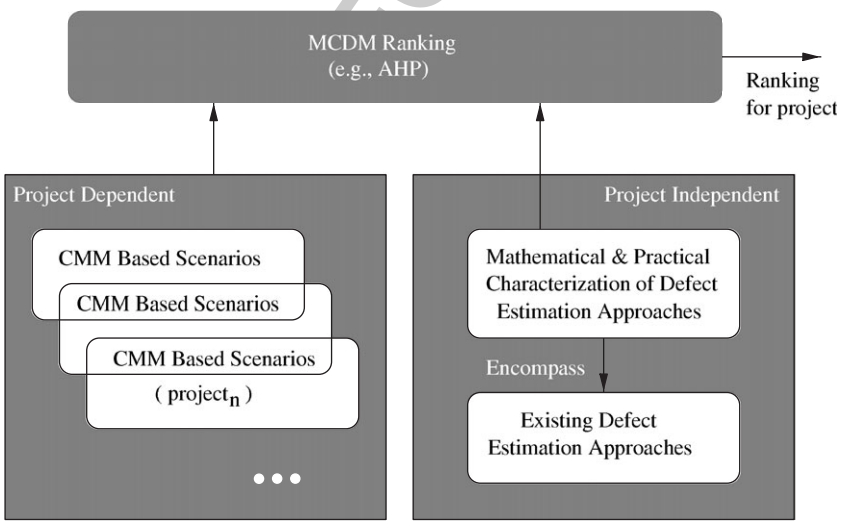


Figure 1. Overview of the classification and selection process of defect estimation techniques based on user-defined scenarios.

DOI: 10.1002/stvr

This classification is presented in Section 2. Next, a set of practical features that are general enough to properly characterize defect estimation techniques must be delineated. These features are presented in Section 3. In addition, at least one estimation approach representing each estimation class is selected to be considered for the ranking. The approaches are compared in Section 4 using these chosen features.

The selection of the best estimation technique is based on multiple features and needs to be tailored for given scenarios. This is an optimization problem and the use of fully discretized approaches (such as flowcharts and decision trees) could lead to extremely large solutions, which would be difficult to maintain and use. MCDM algorithms are better suited, as described in Section 5, to deal with such optimization problems and various techniques are available. Although, in theory, any of them could be used, here the Analytic Hierarchy Process (AHP) as described in Section 5 has been applied.

Finally, project-dependent example scenarios are created based on Capability Maturity model (CMM) levels. These are presented in Section 6 and the approaches are ranked according to them. CMM is used here to create scenarios that vary according to organization and process maturity. Conclusions are drawn in Section 7.

2. MATHEMATICAL CHARACTERIZATION OF EXISTING ESTIMATION TECHNIQUES: AN OVERVIEW

The goal of any estimator is to minimize the error between the actual value of the parameter and the estimated value. As described later in Section 4 a myriad of estimators are available that have been applied to software defect estimation problems. These estimators are based on different techniques and therefore have distinct mathematical requirements for their application. Here, an overview of the classes of estimators most commonly used for defect estimation is presented [3,11,15–18]. These classes include Bayesian Belief Networks (BBNs), Bayesian Parametric Approaches (BPAs), non-parametric Models, and Classical Parametric Approaches. The mathematical foundations of specific estimators within these classes are presented in the Appendix.

As can be seen from Figure 2, the first question to be asked is about the availability of proper prior knowledge. Prior knowledge refers to actual information/knowledge about the parameter of interest. For example, based on historical data of past similar projects the number of defects for a new project is expected to be in the range from X to Y. Another example could be that error-prone modules present a certain characterization or a certain probability density function (PDF) with respect to specific software metrics (size, complexity, etc.).

The answer to this question divides the choice of classes into non-Bayesian and Bayesian approaches. More specifically, if proper prior knowledge is not available, then either a Classical Parametric approach (Appendix A.3) or a non-parametric approach (Appendix A.4) needs to be used. Otherwise, either a BBN (Appendix A.1) or a BPA (Appendix A.2) can be used.

When designing an estimator to predict the total number of defects (T_D) without any prior knowledge, one could specify that the estimator should produce results within the interval $0<\hat{T_D}<\infty$. However, for example, with the use of prior knowledge such as μ_{T_D} the estimator could be designed

DOI: 10.1002/stvr

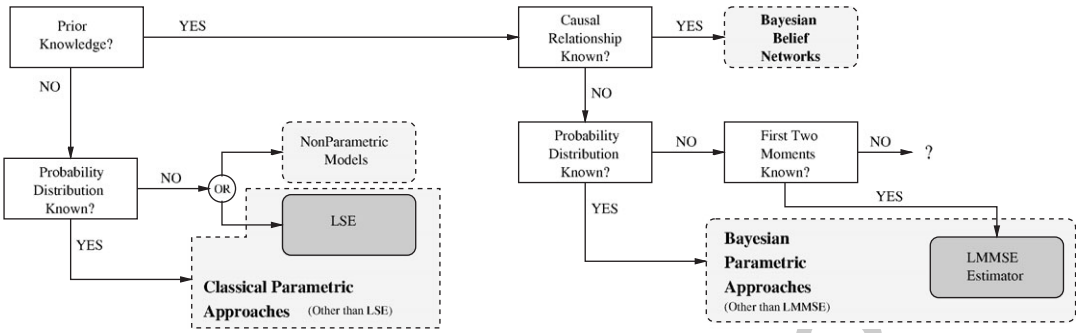


Figure 2. Classification of estimation approaches. Part of this figure has been extracted from the work of Kay [19].

to produce results within an expected interval $\mu_{T_D} - C < \hat{T_D} < \mu_{T_D} + C$ for a given constant C. This increases the accuracy of the estimator.

Figure 2 shows the decision tree for the application of a class of estimators; a more detailed tree with specific techniques is provided in the Appendix. The tree in Figure 2 is not intended to be comprehensive but includes most commonly used techniques for defect estimation of a software product. Brief descriptions of each of the four classes of estimators are presented next.

The BBN [20] is a powerful technique that has been extensively used to model, understand, and simulate complex systems. The BBNs have been successfully used in many distinct domains such as medical diagnostics [21], social interactions [22], and defect prediction [23,24].

The process of creating a BBN is iterative in nature; it involves using data, expert opinion, or observation of the phenomenon being modeled, to derive the probability distributions. In turn, more data are generated. Daniel *et al.* [25] have pointed out problems with the BBN approach. Among other issues they state '*Constructing a realistic and consistent graph often requires collaboration between knowledge engineers and subject matters experts*, *which in most cases is hard to establish*' [25]. Despite these problems the BBN approach is extremely helpful in modeling volatile systems such as most software development organizations. A brief description of the BBN approach is available in Appendix A.1.

In summary, the use of a BBN approach requires, in general, not only the availability of some prior knowledge but also domain expert knowledge in terms of causal relationships between elements of the development process and software product artifacts [25]. If such resources (Prior knowledge and causal relationship known) are not available, then a BPA is an alternative solution.

Before describing and comparing Bayesian and Classical Parametric approaches, some discussion on the mean square error (*MSE*) and bias of estimators are needed. This is because the goal of an estimator is to minimize the *MSE* between a parameter and its estimator $MSE(\hat{\theta}) = E[(\theta - \hat{\theta})^2]$.

In Figure 3 it can be seen that the PDF of a Classical estimator is defined over the range $-\infty < \theta < \infty$, whereas the PDF of a Bayesian estimator is truncated based on the prior knowledge that $-A < \theta < A$ for some given value A. When prior knowledge $p(\theta)$ of θ is available, the MSE can be rewritten as Equation (1). This is referred to as the Bayesian mean square error (*BMSE*).

DOI: 10.1002/stvr

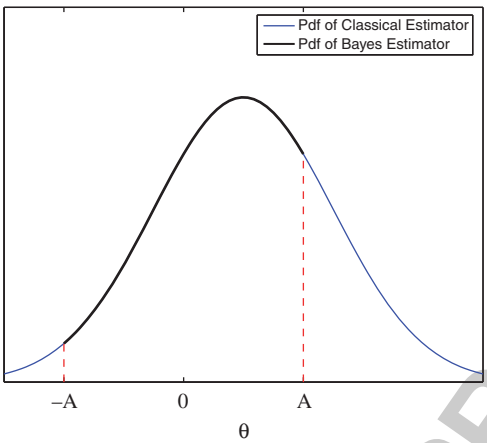


Figure 3. Classical vs Bayesian.

Given this knowledge about the range, it would be undesirable to estimate the values of θ outside of it. Hence, a Bayesian estimator can provide a more accurate estimate.

$$BMSE(\hat{\theta}) = \int \left[\int (\theta - \hat{\theta})^2 p(\theta|\mathbf{x}) \, d\theta \right] p(\mathbf{x}) \, d\mathbf{x} \tag{1}$$

In Equation (1) $p(\theta|\mathbf{x})$ is the posterior probability, which uses Baye's rule given by Equation (2). Note that the denominator is merely a normalizing factor.

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta) p(\theta)}{p(\mathbf{x})} \tag{2}$$

The posterior probability provides the probability of the parameter and takes into account the knowledge presented by the data set. The prior probability, on the other hand, is only based on the initial information about the parameter. The knowledge in the data in addition to the initial information about the parameter further refines the estimate. A description of three Bayesian estimators (Minimum Mean Square Error—MMSE, Maximum A Posteriori—MAP, and Linear Minimum Mean Square Error—LMMSE) from Figure 2 is given in Appendix A.2.

BPAs are better suited than Classical Parametric approaches, discussed later, in two situations. First, when some kind of prior information or knowledge is available about the parameter θ to be estimated (θ could be the number of defects in a software product, the most error-prone modules, etc.). Second, when an efficient minimum variance unbiased (MVU) estimator (refer to Appendix A.3.1), which has the lowest variance among all other estimators, cannot be found. Then, for a prior PDF a Bayesian estimator can be found that has the lowest variance among all the other estimators for a range of θ. Note that a Bayesian estimator is a biased (not with respect to the joint distribution) estimator as compared with Classical Parametric estimators. This is an advantage as the probable values for the parameter θ are in a smaller domain as compared with unbiased estimators.

The lack of prior knowledge limits the estimation process to Classical Parametric and Non-parametric approaches. Although information about the PDF could also be considered prior

DOI: 10.1002/stvr

knowledge, the distinction made here is that prior knowledge is more specific (for example, actual values of the parameters of a PDF), while a known PDF only provides information about the shape of the data with no specific information about its parameters.

If the PDF of the data is known or assumed, then the problem of finding an estimator $\hat{\theta}$ is simply finding a function of data that gives estimates of the parameter. Many defect estimators [3,12,18,26–29] assume a PDF for the data. For example, a Normal or Gaussian random variable with mean m and variance σ^2 ($N[m, \sigma^2]$) is often assumed. When the PDF is known or assumed, some estimators can be used: MVU, MLE, and Methods of Moments (refer to Appendix A.3). Otherwise, an LSE or a non-parametric model is needed (Appendix A.4).

In general, parametric models capture unimodal densities well, while Non-parametric methods may provide good fit for multimodal densities. Non-parametric methods are used when neither the PDF of the data is known nor reasonable assumptions regarding it can be made; they are also less sensitive to outliers and model deviations than Parametric approaches. Non-parametric methods are completely dependent on the data and therefore require more of it to provide reasonable results. That is, Non-parametric estimators need a larger data set in order to draw meaningful conclusions with the same confidence level than the application of a Parametric approach.

3. CHARACTERIZATION OF EXISTING ESTIMATION TECHNIQUES

In order to rank existing defect estimation techniques for the software testing process, a set of characteristics needs to be defined. As presented below, the set proposed here is composed of the following features: subjectivity, cost, applicability, latency, and expressiveness. This set is not assumed to be complete, but is intended to capture the major aspects required for the classification of the techniques.

- F_1 Subjectivity: the more subjective an approach is, the more reliant it is on the availability of experts and the less repeatable the process. For example, consider two different persons are going to use the same technique for the same data set/project. An approach with very low subjectivity will lead to the same (or similar) results, while an approach with high subjectivity may lead to considerably more distinct results due to the different levels of experience and expertise of each person. Subjectivity is an inversely proportional feature: the lower the subjectivity level associated with an specific estimation technique, the better the technique.
- F_2 Cost: the costs associated with a defect estimator plays an important role in their selection. Depending on the availability of resources, a very accurate technique may not be considered due to its high cost. Here, costs include the cost of collecting and maintaining project data and the cost of hiring experts to, for example, prepare a customized solution. Similar to subjectivity, cost is also an inversely proportional feature.
- F_3 Applicability: the less restricted and the smaller the number of constraints an approach has, such as the availability of historical data, availability of experts, etc., the more widely applicable the approach is. Assume an approach requires a certain number of data sets from past similar projects to be used. Since not all companies record such data and cannot rely on publicly available data, the applicability of such approach would be considered low. However, if an approach only uses data from the current project, then any company can apply

DOI: 10.1002/stvr

the approach. Another aspect to be considered in terms of applicability is the difficulty in satisfying the constraints. A small number of constraints that are very difficult to satisfy is not better than having a large number of more easily satisfied constraints. Applicability is a directly proportional feature: the higher the applicability level associated with an approach the better the approach.

F_4 Latency: the earlier the estimates can be made, the sooner decisions can be made in the testing process. An approach is not very useful if it can reach reasonable levels of accuracy only towards the end of the project. Also, latency here is not associated with the time required to collect historical data; it is computed only for the time required to compute an accurate estimation for the current project. If the time to acquired historical data is considered, then the approach would be penalized twice for the same feature: the high cost to collect/acquire the data and the low latency. Latency is inversely proportional: the lower the latency the better the approach.

F_5 Expressiveness: the more suitable information can be extracted from a technique, the more useful it is. For example, some techniques provide only the estimation of defects, while others are able to determine the number of defects per module or per file. Others may also be used to identify the source of problems (bad design, bad inspection, etc.) for a given module or project. Also, a technique is considered more expressive if it provides the required information; some techniques may provide lots of information but may not be very useful because they do not provide the specific information that is required. Expressiveness is directly proportional: the more useful information a technique provides, the better it is.

One could argue that accuracy is missing from the list of features presented above. However, estimation techniques presented in the literature [3,18,28] have all achieved levels of accuracy within a $\pm 10\%$ range. Accuracy is considered, in this paper, after a training period, when one is needed. Therefore, the question to be asked is not how accurate an approach is but how long does it take to achieve that level of accuracy. Latency, listed above, seems to include both factors and the inclusion of accuracy would be redundant.

Also, the features above should not be analysed in isolation. One could argue that there is, in general, a correlation between low subjectivity and high latency and low accuracy. However, the correlation is also impacted by the context of the project/application. For example, depending on a given scenario, the lack of historical data and expert knowledge could require the use of a technique with low subjectivity. This type of correlation and context analysis is a multi-criteria decision problem that is described in Section 5. The correlation may have an impact on the comparison of the performance between two alternative solutions but does not have a direct impact on the ranking of the estimation techniques presented here.

4. EXISTING APPROACHES

Existing defect estimation approaches are described and classified according to the schema defined in Section 2. The goal here is not to present a comprehensive list of existing approaches but to select representatives of each estimation class defined in Section 2 and individually characterize them according to the features defined in Section 3.

DOI: 10.1002/stvr

The BBN approach category is represented here by the work of Fenton and Neil [11]. Among many Bayesian Parametric estimators [15,30,31], the Bayes ED^3M[32] is selected as a representative for this class as it also presents a process to collect pertinent historical data and it is a more contemporary approach. Many reliability models have been proposed that predict the total number of defects in the software. Two of these models are presented: (i) Padberg's approach [16,28] based on Hypergeometric Distribution Software Reliability Growth Model (HGDM) and (ii) Musa–Okumoto Logarithmic Poisson model [3,33]. In addition, ED^3M[18] is discussed as the approach seems more applicable and has lower latency. All of these approaches are based on the MLE, which requires assumption about the probability distribution. To the best of our knowledge, a detailed study does not exist on the PDFs for software testing; this may be due to the lack of a large body of data sets available to individual researchers. In general, each assumed distribution is based on some assumptions, while ignoring other facts. Hence, it can be safely deduced that no model will work in all situations. Finally, the work of Barghout *et al.* [17] has been chosen as a representative of Non-parametric approaches. Indeed Non-parametric models are not commonly used for defect estimation. The reason for that may rest on the fact that latency for such approaches is considerably higher and not justifiable when compared with other existing estimators. Non-parametric approaches have been successfully applied with different goals [34] but not defect estimation.

4.1. BBN approaches

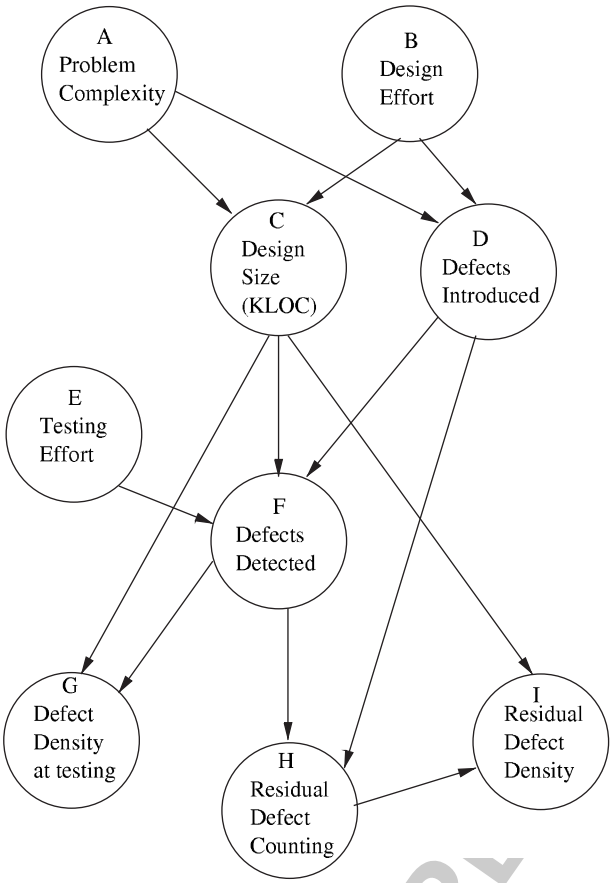
4.1.1. BBN^{AP_1} : Fenton's BBN estimators

The example presented by Fenton and Neil [11] is described here to exemplify the use and the requirements of a BBN for defect estimation (a brief description of BBN is available in Appendix A.1). Their BBN prototype is presented in Figure 4. As it can be seen, the BBN can compute the defect density during testing as well as after testing, i.e. the residual defect density. Assuming the probabilities given in their paper the probability of the defect density at testing can be computed given that the problem complexity is 'Very High (VH)', the design effort is 'High (H)', the testing effort is 'Low (L)', and the design Size is between '1 and 2 KLOC(1–2)'. The predictions can be computed using conditional probability. The results are shown in Figure 4.

Results for defect prediction using a BBN approach have been reported to be accurate [23,24,35]. Also, in case of unexpected results in the computation of the probabilities, the problem can be traced to its causal relationship. While the BBN approaches are very powerful technique, they require much more effort and data than other approaches. That is, the BBNs can be created purely from empirical data, purely from expert knowledge, or a combination of both.

Software development processes differ considerably from organization to organization and even within the same organization. This makes the construction of a general BBN capturing the causal relationships improbable. The expectation is that a BBN needs to be customized for each organization/group. Also, a substantial amount of data need to be collected to compute the probabilities and train the BBN. It should be noticed that large, reliable sources of data may be very hard to acquire. The use of expert knowledge can decrease or even eliminate the data dependency problem. However, such an approach may not be directly applied to companies where the development process is almost *ad hoc*. Also, the costs can increase considerably if an expert needs to be hired to develop the BBN.

DOI: 10.1002/stvr



Results for the computation of defect density

Percentage	Defect density at testing
89.20	0 — 0.001
09.49	0.001 — 0.002
01.20	0.002 — 0.003
00.11	0.003 — 0.004
00.00	0.004 — 0.005
00.00	> 0.005

Percentage	Residual defect density
40.85	0 — 0.001
50.39	0.001 — 0.002
08.29	0.002 — 0.003
00.47	0.003 — 0.004
00.00	0.004 — 0.005
00.00	> 0.005

Figure 4. Bayesian Belief Network for defect prediction [11].

4.2. Classical Parametric approaches

4.2.1. CPA^{AP_2} : Padberg's MLE

Padberg developed a defect estimation technique by assuming that software testing is a random experiment with Hypergeometric distribution. It should be noticed that Padberg's model is preceded by an apparently equivalent model; the m-Stage Testing model developed by Cai [36]. Padberg showed that when a growth quotient $Q(m)$ of the likelihood function $L(m)$ is greater than 1, it indicates that the likelihood function is indeed increasing and provides MLEs.

$$Q(m) = \frac{L(m)}{L(m-1)} = \frac{(m-w_1)\ldots(m-w_n)}{m^{n-1}\cdot(m-c_n)} \tag{3}$$

DOI: 10.1002/stvr

In Equation (3) m is the initial number of faults in the software, w_n is the number of newly discovered and rediscovered faults for nth test, and c_n is the cumulative number of faults discovered in n tests. The algorithm for finding the MLE is briefly presented here; more details can be found elsewhere [16,28]. For given data c_n first find $x = c_n + 1$ then $Q(x)$. If $Q(x) > 1$ then set $x = x + 1$ and find $Q(x)$ again. Keep repeating the steps until $Q(x) \leqslant 1$, then MLE will be $\hat{m} = x - 1$. The statistical performance of $Q(m)$ is not discussed in [16,28]. It is not known if the variance of $L(m)$ is asymptotically bounded by Cramer–Rao Lower Bound (CRLB) [19]; in other words it is not known if $Q(m)$ is asymptotically an efficient MVU estimator. Even though the underlying data model is not known, it can be observed from $Q(m)$ that the model is nonlinear and the MLE that is based on a nonlinear data model for finite data records does not achieve CRLB.

4.2.2. CPA^{AP_3} : *Musa–Okumoto model*

Another example of a defect estimator is the Musa–Okumoto logarithmic Poisson model, which is briefly described here. Refer to the work of Musa *et al.* [3,33] for more details. The model discussed is the execution-time model but there is another one available for calendar-time. Musa and Okumoto proposed a reliability model based on the assumption that the expected number of faults $\mu(t)$ by time t are Poisson distributed as given in Equation (4). The parameters to be estimated are λ_0 the initial failure intensity and θ the rate of reduction in the normalized failure intensity per failure. The data model given by Equation (5) is a nonlinear function of λ_0 and θ. Hence, the MLE will not achieve the CRLB for finite data set.

$$Pr[M(t) = m] = \frac{[\mu(t)]^m}{m!} \mathrm{e}^{\mu(t)} \tag{4}$$

$$\mu(t) = \frac{1}{\theta} \ln(\lambda_0 \theta t + 1) \tag{5}$$

Moreover, a closed-form solution of the MLE could not be found for Equations (4) and (5). Therefore, a numerical approximation of the MLE is needed. Hence, the approximation of the MLE is not guaranteed to be (asymptotically) an efficient MVU estimator.

4.2.3. CPA^{AP_4} : ED^3M

The data model of ED^3M [18] is given in Equation (6) where $\mathbf{D}$ is the defect data vector, $\mathbf{h}$ is the observation vector, and $\mathbf{w}$ is noise vector. Vectors are of dimensions $N \times 1$. It is assumed that $\mathbf{D}$ is normally distributed and the PDF of $\mathbf{D}$ is given by Equation (8). The initial number of defects in the software is given by R_0. An MLE estimator is developed for $\hat{R}_0$ in Equation (9).

$$\mathbf{D} = R_0 \mathbf{h} + \mathbf{w} \tag{6}$$

$$\text{where } h(n) = 1 - \frac{\lambda_2}{\lambda_2 - \lambda_1} \mathrm{e}^{-\lambda_1 n} + \frac{\lambda_1}{\lambda_2 - \lambda_1} \mathrm{e}^{-\lambda_2 n} \tag{7}$$

$$p(\mathbf{D}; R_0) = \frac{1}{(2\pi\sigma^2)^{N/2}} \mathrm{e}^{\left[-(1/(2\sigma^2))(\mathbf{D} - R_0\mathbf{h})^{\mathrm{T}}(\mathbf{D} - R_0\mathbf{h})\right]} \tag{8}$$

$$\hat{R}_0 = (\mathbf{h}^{\mathrm{T}}\mathbf{h})^{-1}\mathbf{h}^{\mathrm{T}}\mathbf{D} \tag{9}$$

DOI: 10.1002/stvr

As seen in Equation (6) the data model is linear; therefore, the MLE estimator $\hat{R}_0$ in Equation (9) can achieve the CRLB for finite data set and will be an efficient MVU estimator. With respect to latency, ED^3M performs slightly better than Padberg's approach and clearly outperforms Musa–Okumoto Model [18].

There is no subjectivity for ED^3M as the only input for the model is the number of defects per time unit. The only source of input is also the reason for lower cost as no other data and no expert opinions are required. Defects per time unit seems to be collected by any software company making ED^3M widely applicable. ED^3M presents latency that is very similar to Padberg's approach.

4.3. Bayesian parametric approaches

4.3.1. BPA^{AP_5} : *Bayes*ED^3M.

If R_0 is considered to be deterministic or in other words no prior knowledge about R_0 is available, then this leads to the MLE version of $\hat{R}_0$ as given by Equation (9) and is referred as the ED^3M.

On the other hand if some prior knowledge about R_0 is available, then a Bayesian version of $\hat{R}_0$, referred here as the BayesED^3M, is defined. Let the two pieces of prior knowledge be the prior mean μ_{R_0} and prior variance $\sigma^2_{R_0}$. The assumption that R_0 is Gaussian distributed is also made; more formally $R_0 \sim \mathcal{N}[\mu_{R_0}, \sigma^2_{R_0}]$. The Bayesian Estimator of $\hat{R}_0$ is given by Equation (10)

$$\hat{R}_0 = \mu_{R_0} + \sigma^2_{R_0}\mathbf{h}^{\mathrm{T}}(\sigma^2_{R_0}\mathbf{h}\mathbf{h}^{\mathrm{T}} + \sigma^2_n\mathbf{I})^{-1}(\mathbf{D} - \mu_{R_0}\mathbf{h}) \tag{10}$$

where $\sigma^2_n = \sigma^2/n$.

The BayesED^3M presents a lower latency with similar accuracy when compared with the ED^3M. There is no subjectivity associated with this approach as it depends on collected data. However, data from past similar projects need to be properly collected, which increases its cost. Since not all companies keep records of past projects, the applicability of the BayesED^3M is considered lower than the ED^3M.

4.4. Non-parametric approaches

4.4.1. NPA^{AP_6} : *Order Statistics non-parametric model*

The Parzen window approach has been used by Barghout *et al.* [17,37] in estimating inter-failure times for a software reliability model. The technique is called 'a non-parametric Order Statistics software reliability model' and referred hereafter as the Order Statistics non-parametric model.

Inter-failure times are, in general, assumed to be exponentially distributed. The conditional distribution of inter-failure times is expressed in terms of independent and identically distributed (i.i.d.) non-parametric distribution of the cumulative time of a failure and the unknown parameter N, which is the total number of failures. For the estimation of the non-parametric distribution of the cumulative time of a failure, the Parzen window method is used. Three different kernel functions viz., Gaussian, double exponential, and log normal are evaluated. The evaluation of each type of kernel function is further combined with two different techniques viz., likelihood cross-validation method and prequential likelihood method for the estimation of h (length of the side of the

DOI: 10.1002/stvr

hypercube). Hence, six different estimators are evaluated and compared against existing parametric models for software reliability for three different data sets. It is concluded that all six estimators did not perform as well as their parametric counterparts. However, after recalibration each of them performed consistently better than the parametric models for three of the data sets. Latency and applicability appear to be the two major concerns regarding the Non-parametric models.

5. RANKING METHODOLOGY

The problem of ranking and selecting alternatives to support decision making has been under investigation for centuries [38]. There are numerous solutions available, which can be categorized at a high level into single objective optimization (i.e. rank on one criterion) or multiple objective optimization (i.e. rank on two or more criterion). Within each of these categories, additional characteristics such as the number of alternatives available (finite or infinite), dimension, and scale used to describe the performance of an alternative with respect to a criterion, etc. distinguish among the approaches.

The main goal of single objective optimization is to find the best solution, which corresponds to the minimum or maximum value of a single objective function. This takes the form of a classic optimization problem: the objective function is the single criterion and the constraints are requirements that the alternatives need to meet. Different optimization techniques can be used, such as linear programming, discrete optimization, etc. [39]. This type of optimization is useful in providing decision makers with insights into the problem, but usually cannot provide a set of alternative solutions that consider conflicting objectives. An example of a single objective optimization problem is to choose the lowest cost alternative.

Many decision making problems, including the selection of an estimation technique, need to achieve several objectives: minimize cost, maximize the convergence rate, maximize applicability, etc. A multi-objective optimization problem with conflicting objectives does not have a single optimal solution. Instead, the interaction among different objectives results in a set of compromised solutions, largely known as the trade-off, non-dominated, non-inferior, or Pareto-optimal solutions. Different multi-objective optimization techniques can be used, such as multi-objective genetic or evolutionary algorithms, multi-attribute utility theory-based approaches, multi-attribute value theory based, outranking methods, and techniques based on the AHP. These techniques are often referred to as MCDM approaches.

The literature in MCDM is extensive. Studies are available that (i) characterize the suitability of the MCDM approaches for different kinds of problems [40]; (ii) present the results of using two or more MCDM approaches to the same problem [41,42]; and (iii) present the strengths and limitations of a MCDM approach, often using case studies [43,44].

A widely used MCDM algorithm is the AHP. The AHP uses the subjective judgment of a decision maker as input; the quantified weight of each alternative is the output. Key advantages of the AHP are that it is simple for decision makers to use, has been successfully applied on a wide variety of problems, tool support is available, and the approach can use qualitative information provided by the decision maker. However, there are well-documented problems with AHP. For example, AHP does not scale well to problems with a large number of alternatives. First, AHP becomes more difficult and time consuming for the decision makers to provide the pairwise comparisons (PCs)

DOI: 10.1002/stvr

between a large number of alternatives; this will lead to large comparison matrices as defined in Section 5.1. Second, larger matrices are more likely to become inconsistent and/or contradictory, which reduces the accuracy of the ranking. For the problem under investigation in this work, the number of alternatives is relatively small, making AHP a suitable choice.

The purpose of this section is to provide an overview of the AHP approach and describe how it can be applied to rank a set of defect estimation techniques with respect to a project with specific goals and constraints. Once the techniques are ranked the test manager, for example, can select the technique that is 'right' for the project. The application of the AHP approach is illustrated in Section 6.

5.1. Analytic hierarchy process

The AHP method has three main steps [45] (refer to Figure 5). The first step is to structure the decision making problem as a hierarchical decomposition, in which the objective or goal is at the top level, the criteria used in the evaluation are in the middle levels, and the alternatives are at the lowest level. The simplest form used to structure a decision problem consists of three levels: the goal at the top level, criteria used for evaluation at the second level, and the alternatives at the third level. This organization is used in the discussion on the second and third steps.

The second step is to create decision tables at each level of the hierarchical decomposition. The matrices capture a series of *PC* matrices using relative data. The comparison can be made using a nine point scale or real data if available. The nine point scale includes: $[9, 8, 7, \ldots, \frac{1}{7}, \frac{1}{8}, \frac{1}{9}]$, where 9 means extreme preference, 7 means very strong preference, 5 means strong preference, and continues down to 1, which means no preference. The reciprocals of the above levels are also available. For example, if a comparison between 'a' and 'b' is evaluated as 7, then the comparison between 'b' and 'a' is $\frac{1}{7}$.

For the top level of the decomposition, a single importance matrix (IM) is defined. In AHP, the values captured in the IM reflect the relative importance of criteria, i.e. the requirements, from the decision maker's perspective. For example, if the evaluation criteria are a, b, c, and d, then the decision maker compares their importance, two at a time in a four by four matrix. For example, a decision maker may determine that a is strongly preferred to b. In this case the PC of a/b has the value 5; b/a has the inverse value $\frac{1}{5}$. Evaluation criteria in the defect estimation selection problem include factors such as applicability, subjectivity, cost, latency, and expressiveness as described in Section 3.

The IM does not capture additional relationships among the criteria, such as synergistic or conflicting relationships. The criteria are assumed to be independent. This is likely to be the case in many decision making situations, in which the conflicting or synergistic relationships among the requirements are not well understood.

The IM for the top level provides a means to prioritize the requirements for the decision making process. Using the comparison data in the matrix a priority vector can be calculated, for example, with an eigenvector formulation. This priority vector is used in the third step, when the priorities are aggregated.

The decision tables for the middle level capture a PC matrix for the alternative solutions for each evaluation criterion. Consequently, each evaluation criterion has a PC^{C_i}. For example, if there are four evaluation criteria and five alternatives to choose from, then this results in four PCs that are five

DOI: 10.1002/stvr

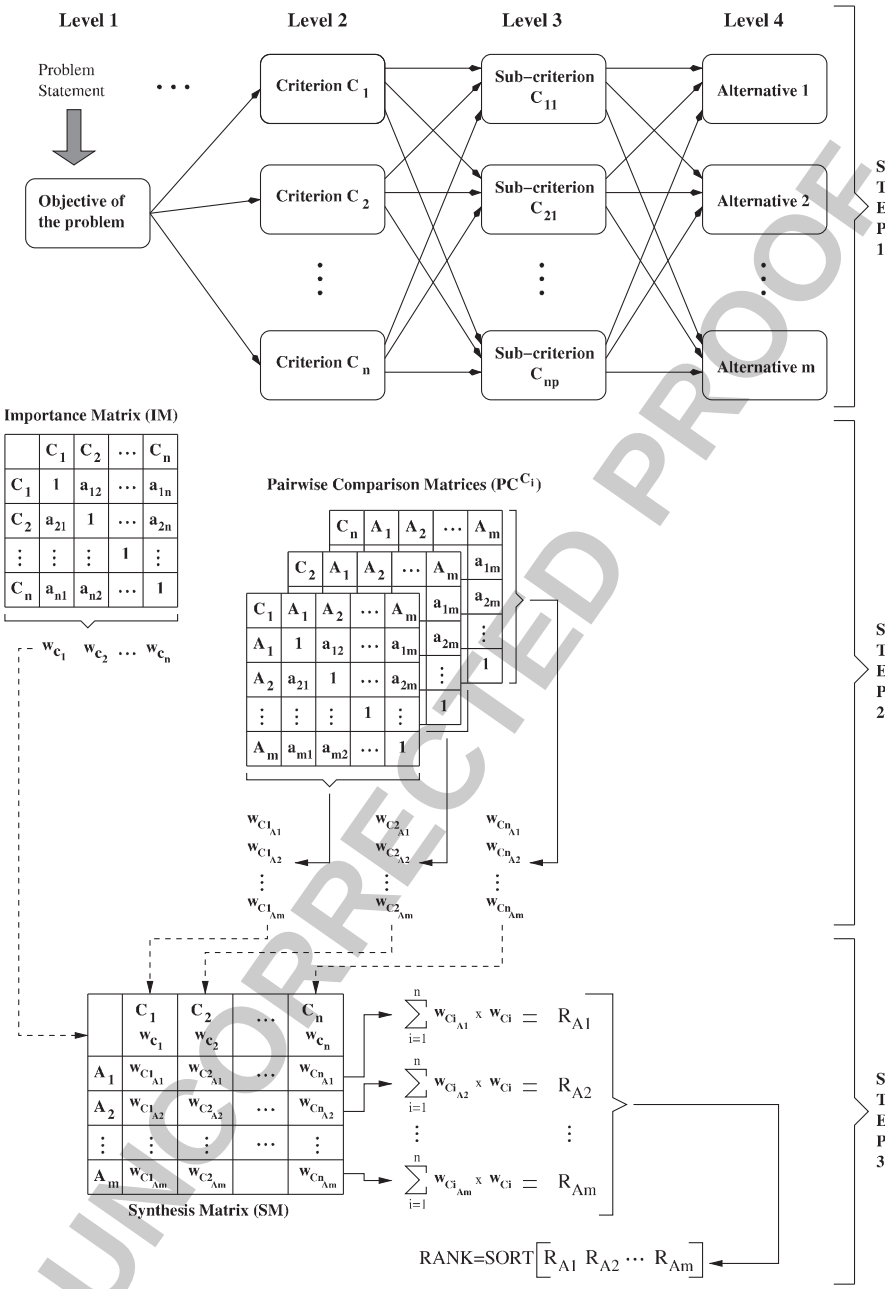


Figure 5. General structure of the AHP approach.

DOI: 10.1002/stvr

by five. The alternatives in the defect estimation selection problem are the estimation techniques. The priority vector for each of these decision tables is calculated. Here again, an eigenvector calculation can be used. The priority vector can be represented in normalized form, such that the sum of the elements is equal to one. This vector is used in the third step, when the priorities are aggregated. In AHP, the values captured in a PC matrix do not explicitly capture other relationships such as synergistic or conflicting relationships that are present in the alternative solutions.

The third step aggregates, or synthesizes, the priorities so that the 'best' alternative can be chosen. Options for the aggregation include a distributive mode and an ideal mode. As these two options have been shown to give the same result 92% of the time [14], the distributive mode is presented here. The reader is referred to the work of Saaty [14] for a discussion of the ideal mode. The distributive mode calculation uses the priority vector calculated in step one as a weight vector, W, and the normalized priority vectors calculated in step two to define a new matrix A. Each element A_{ij} represents the normalized priority calculated in step two for alternative i and evaluation criteria j. To calculate an alternative's overall priority, each row in the matrix is multiplied by the weight vector and the multiplied elements are added together:

$$\text{Alternative priority } i = \sum_{j=1}^{4} (w_j * A_{ij}) \tag{11}$$

The alternative with the largest value represents the best alternative.

6. RANKING OF EXISTING APPROACHES

A set of scenarios are defined here to evaluate the existing defect estimation approaches described in Section 4. The evaluation is based on the use of AHP [14,45] described in Section 5.

6.1. Defining the PC matrices

According to AHP, PC matrices are needed to evaluate the approaches with respect to the five criteria defined in Section 3. Each matrix refers to one criterion and reflects the PC between each of the approaches. It facilitates the process as one needs to concentrate only on the two approaches being compared. The overall comparison is achieved by computing the eigenvector associated with the largest eigenvalue of the matrix. Saaty's scale is used here to construct the PC matrices and the IMs. This approach is subjective as someone else may view the comparison between two approaches from a different perspective leading consequently to different matrices. This subjectivity does not diminish the application of AHP. Indeed it shows that the approach can be customized for different views and for different scenarios. It is clear that a quantitative and less subjective comparison could lead to more conclusive results. However, except for latency, all the features described in Section 3 are subjective defying a quantitative analysis. Latency could have been compared quantitatively if access to a large data set that could be used for all approaches was available. As this is not the case, published results are used here to make a less quantitative comparison for latency. Next, a description of how each matrix is created is presented.

The BBN approaches are considered more subjective than the others as they rely on expert opinion and the customization of the network. Therefore, the comparison of BBN^{AP_1} with the Classical

DOI: 10.1002/stvr

Parametric approaches and the Non-parametric approach shows a higher level of subjectivity. The level is lowered when compared with classical BPAs. BPA^{AP_5} is more subjective than the Classical Parametric approaches as it depends on past data and how the data are collected and used. It also presents a high level of subjectivity when compared with the Non-parametric approach. The Classical Parametric approaches present similar levels of subjectivity. However, a higher level is assigned to CPA^{AP_2} as it depends on the estimation of other parameters or the collection of proper data. CPA^{AP_4} is assigned the lowest level of subjectivity for the parametric approaches as it does not depend on any other estimation; even initial parameters are computed automatically. Finally, the Non-parametric approach has the lowest level of subjectivity when compared with others as it relies solely on data from current observations. Not even assumptions about the probability distribution are necessary.

Table I shows the matrix PC_{subj} regarding the subjectivity feature. The quantification of the comparison is given by the eigenvector associated with the largest eigenvalue of matrix PC_{subj}^{T}. A transposed matrix is used here to represent the fact that the lower the value of the subjectivity, the better the approach. This leads to a vector where the higher the value, the lower the subjectivity. Here, the results are normalized according to their summation to simplify the comparison. The results, represented by w_{subj}, are given as follows:

$$w_{subj} = [0.0216 \;\; 0.1321 \;\; 0.1724 \;\; 0.2385 \;\; 0.0369 \;\; 0.3985]^{\mathrm{T}} \tag{12}$$

As it can be seen, the Non-parametric approach NPA^{AP_6} has the lowest subjectivity (highest value) followed by the three Classical Parametric approaches. A small advantage is given to CPA^{AP_4} when compared with the other Classical Parametric approaches. As expected, the Bayesian approaches are more subjective and present the two lowest values in the vector.

With respect to cost, the classical Parametric and the Non-parametric approaches present the lowest cost level as they depend mostly on data from the current project (refer to Table II). The classical Bayesian approach BPA^{AP_5} depends also on data from previous projects and a consequent increase in cost. Finally, BBNs are more costly as they not only depend on data from past projects to train the network but they also depend on expertise opinion to construct the network. As for subjectivity, the values are computed based on the transpose of PW_{cost} to represent the inverse values (the lower the cost the better). As before, final results are normalized and presented in the

Table I. Pairwise comparison matrix (PC_{subj}) for the approaches described in Section 4 with respect to subjectivity (F_1).

F_1: Subjectivity	BBN^{AP_1}	CPA^{AP_2}	CPA^{AP_3}	CPA^{AP_4}	BPA^{AP_5}	NPA^{AP_6}
BBN^{AP_1}	1	9	9	9	3	9
CPA^{AP_2}	$\frac{1}{9}$	1	2	3	$\frac{1}{6}$	3
CPA^{AP_3}	$\frac{1}{9}$	$\frac{1}{2}$	1	2	$\frac{1}{6}$	3
CPA^{AP_4}	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{2}$	1	$\frac{1}{6}$	3
BPA^{AP_5}	$\frac{1}{3}$	6	6	6	1	9
NPA^{AP_6}	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	1

DOI: 10.1002/stvr

Table II. Pairwise comparison matrix (PC_{cost}) for the approaches described in Section 4 with respect to cost (F_2).

F_2: Cost	BBN^{AP_1}	CPA^{AP_2}	CPA^{AP_3}	CPA^{AP_4}	BPA^{AP_5}	NPA^{AP_6}
BBN^{AP_1}	1	9	9	9	6	9
CPA^{AP_2}	$\frac{1}{9}$	1	1	1	$\frac{1}{4}$	$\frac{1}{4}$
CPA^{AP_3}	$\frac{1}{9}$	1	1	1	$\frac{1}{4}$	$\frac{1}{4}$
CPA^{AP_4}	$\frac{1}{9}$	1	1	1	$\frac{1}{4}$	$\frac{1}{4}$
BPA^{AP_5}	$\frac{1}{6}$	4	4	4	1	1
NPA^{AP_6}	$\frac{1}{9}$	4	4	4	1	1

direct order: the higher the value the better (the lower the cost). The results of computing the eigenvector associated with the largest eigenvalue follow:

$$w_{cost} = [0.0220 \;\; 0.2269 \;\; 0.2269 \;\; 0.2269 \;\; 0.0702 \;\; 0.2269]^{\mathrm{T}} \tag{13}$$

As expected, the three Classical Parametric and the Non-parametric approaches present the same level of cost followed by the classical Bayesian and by the BBN approach.

As described in Section 3, applicability refers to the potential number of companies that can use the estimation technique. The smaller the number of restrictions and requirements to use, the better the approach. The requirements about the Classical Parametric approaches are mainly focused on assumptions about the distribution and, therefore, are assigned a higher value for applicability. The comparisons among the Classical Parametric approaches present some advantages for CPA^{AP_4} as no requirements about initial values are necessary. One would think that the Non-parametric approach would be more applicable because it does not rely on any assumption about the distribution. However, it does require a considerable amount of data to decrease variance leading to a lower applicability when compared with the classical approaches. The need for historical data decreases the applicability of Bayesian parametric and BBN approaches. The latter suffers even more due to the need of expert opinion for the customization of the network according to project and company's need. The higher the applicability the better. Therefore, there is no need to use the transpose of PC_{appl} (see Table III). Computing the eigenvector leads to the following values:

$$w_{appl} = [0.0241 \;\; 0.2385 \;\; 0.2385 \;\; 0.3435 \;\; 0.0473 \;\; 0.1081]^{\mathrm{T}} \tag{14}$$

Classical Parametric approaches present better applicability with CPA^{AP_4} showing a slightly higher value. They are followed by the Non-parametric and the BPA. The applicability of BBN presents the lowest level, as expected.

Latency here refers to how long it takes to get an accurate estimation for the project under consideration. It does not account for the time required to collect historical data. Non-parametric approaches depend on a large amount of data from the current project to produce a reasonable estimate and therefore is assigned the highest latency level. Classical Parametric approaches also depend on data from the current project but can produce good estimates at early stages. CPA^{AP_4} and CPA^{AP_2} present a better latency than CPA^{AP_3} and, therefore, are assigned a lower latency level.

DOI: 10.1002/stvr

Table III. Pairwise comparison matrix (PC_{appl}) for the approaches described in Section 4 with respect to applicability (F_3).

F_3: Applicability	BBN^{AP_1}	CPA^{AP_2}	CPA^{AP_3}	CPA^{AP_4}	BPA^{AP_5}	NPA^{AP_6}
BBN^{AP_1}	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{6}$
CPA^{AP_2}	9	1	1	$\frac{1}{2}$	6	3
CPA^{AP_3}	9	1	1	$\frac{1}{2}$	6	3
CPA^{AP_4}	9	2	2	1	6	3
BPA^{AP_5}	3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1	$\frac{1}{3}$
NPA^{AP_6}	6	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	3	1

Table IV. Pairwise comparison matrix (PC_{laten}) for the approaches described in Section 4 with respect to latency (F_4).

F_4: Latency	BBN^{AP_1}	CPA^{AP_2}	CPA^{AP_3}	CPA^{AP_4}	BPA^{AP_5}	NPA^{AP_6}
BBN^{AP_1}	1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{9}$
CPA^{AP_2}	6	1	$\frac{1}{2}$	1	3	$\frac{1}{3}$
CPA^{AP_3}	6	2	1	2	3	$\frac{1}{3}$
CPA^{AP_4}	6	1	$\frac{1}{2}$	1	3	$\frac{1}{3}$
BPA^{AP_5}	3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1	$\frac{1}{6}$
NPA^{AP_6}	9	3	3	3	6	1

The BPA improves on that as it has information from past similar projects. After being trained, a BBN can provide accurate estimations very early in the development process and therefore presents the lowest latency level.

Latency is inversionally proportional, that is, the lower the latency the better. A transpose of the matrix PW_{laten}, from Table IV, is used to compute the eigenvector.

$$w_{laten} = [0.4873 \;\; 0.0947 \;\; 0.0682 \;\; 0.0947 \;\; 0.2209 \;\; 0.0342]^{\mathrm{T}} \tag{15}$$

The values of the eigenvector show that the BBN has the highest value and therefore the lowest latency followed by the BPA. CPA^{AP_2} and CPA^{AP_4} are next with the same latency, while CPA^{AP_3} and NPA^{AP_6} present the lowest values (large latency).

Expressiveness is the last feature to be considered. The causal-relationship provided by a BBN contributes to its high level of usefulness. The availability of data also makes BPAs to have a higher expressiveness than the others as historical data can be used not only to estimate remaining defects but also to find correlation between metrics and, for example, to identify error-prone modules. The remaining approaches present a similar level of expressiveness. However, the outcome of CPA^{AP_4} is restricted to the number of remaining defects and a lower level is assigned to it. Using the values from Table V leads to the eigenvector in Equation (16).

$$w_{dou} = [0.6317 \;\; 0.0698 \;\; 0.0698 \;\; 0.0307 \;\; 0.1283 \;\; 0.0698]^{\mathrm{T}} \tag{16}$$

DOI: 10.1002/stvr

Table V. Pairwise comparison matrix (PC_{use}) for the approaches described in Section 4 with respect to expressiveness (F_5).

F_5: Expressiveness	BBN^{AP_1}	CPA^{AP_2}	CPA^{AP_3}	CPA^{AP_4}	BPA^{AP_5}	NPA^{AP_6}
BBN^{AP_1}	1	9	9	9	9	9
CPA^{AP_2}	$\frac{1}{9}$	1	1	3	$\frac{1}{2}$	1
CPA^{AP_3}	$\frac{1}{9}$	1	1	3	$\frac{1}{2}$	1
CPA^{AP_4}	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{3}$	1	$\frac{1}{6}$	$\frac{1}{3}$
BPA^{AP_5}	1.9	2	2	6	1	2
NPA^{AP_6}	$\frac{1}{9}$	1	1	3	$\frac{1}{2}$	1

The values of Equation (16) show that BBN^{AP_1} presents a much higher level of expressiveness followed, from a certain distance, by the BPA. The other approaches present similar levels with some disadvantages noticed for CPA^{AP_4}.

6.2. Defining the IMs

The IMs define the user perception of the PC between two criteria. Therefore, IMs are defined by each user. In order to simulate different users four scenarios have been selected. The scenarios are based on the maturity level of an organization as defined by the CMM [46]. Although CMM presents five levels, levels 4 and 5 are combined here as their major difference is the optimization goal through continuous process improvement in level 5. It is true that not all organizations at the same maturity level have the same requirements or resources. However, not all possible scenarios can be evaluated here; the four scenarios presented below are a good representation of common development environments. As needed, additional alternative scenarios can be easily evaluated by tuning the IM matrix.

Scenario I: the process at CMM level I is considered to be *ad hoc* and sometimes chaotic. However, even on this type of scenario an estimate of the number of defects can help to lessen the side effects of such environment. The *ad hoc* process is an indication that little data are available and expert knowledge can be considered non-existent. Therefore, applicability and cost seem to be the dominant factors in determining the estimation approach to be used. They are assigned equal importance when compared with each other and very high importance when compared with the other factors. Any organization would like to use accurate estimations as early as possible during the development process. Here latency is considered to have medium preference when compared with subjectivity and expressiveness. Organizations at this level do not have a repeatable process and subjectivity as well as expressiveness is of less importance to them. They are assigned here equal importance. The IM for Scenario I is shown in Table VI and the final weights are given in Equation (17). The results show that cost and applicability are the most important factors for Scenario I. A considerable gap exists between these two and the next factor latency. Finally, subjectivity and expressiveness are the factors with least impact and therefore the two lowest weights.

$$w_{sc1} = [0.0348 \;\; 0.4222 \;\; 0.4222 \;\; 0.0861 \;\; 0.0348] \tag{17}$$

DOI: 10.1002/stvr

Table VI. Importance Matrix for Scenario I.

	F_1	F_2	F_3	F_4	F_5
F_1	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{4}$	1
F_2	9	1	1	9	9
F_3	9	1	1	9	9
F_4	4	$\frac{1}{9}$	$\frac{1}{9}$	1	4
F_5	1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{4}$	1

Table VII. Importance Matrix for Scenario II.

	F_1	F_2	F_3	F_4	F_5
F_1	1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	1
F_2	6	1	2	6	6
F_3	6	$\frac{1}{2}$	1	6	6
F_4	3	$\frac{1}{6}$	$\frac{1}{6}$	1	3
F_5	1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$	1

Scenario II: at CMM level 2, the process is achieving some repeatability and some data start to become available. This fact decreases the importance of applicability and cost when compared with the other factors; they are assigned a medium importance. When compared with each other, cost becomes more important than applicability due to the availability of some data and maybe some expert knowledge; a low importance level is assigned in this case. The same comments regarding the relative importance of the remaining factors from Scenario I can be applied here. The only difference is that, in general, a lower level of importance is assigned. Table VII shows the importance matrix associated with Scenario II. Similar to Scenario I, cost and applicability are the most important. However, as can be observed from Equation (18), cost presents a higher weight than applicability. Also, the gap between applicability and latency has decreased. As before, the last two factors present equally low weights.

$$w_{sc2} = [0.0512 \;\; 0.4512 \;\; 0.3430 \;\; 0.1033 \;\; 0.0512] \tag{18}$$

Scenario III: as an organization achieves maturity level 3, a well-defined process is in place. The development process is consistent and mostly repeatable. Software quality is monitored, which implies the availability of more data when compared with CMM level 2. Also, expert knowledge is not a major issue. These facts decrease the importance of cost and applicability. The earlier an accurate estimation is achieved, the more effectively it can be used and the more repeatable is the process. Also, the less the subjectivity involved, the more repeatable the process. Therefore, a medium importance level is assigned to these factors when compared with cost, applicability, and expressiveness. Subjectivity seems to have a higher impact on repeatability than latency and it is therefore assigned a higher relative importance. As at level 3, organizations are interested not only in defect estimation but also in other information such as error-prone modules, the expressiveness presents more importance than cost and applicability. Also, applicability does not seem to be a

DOI: 10.1002/stvr

major issue at level 3 and a higher relative importance is assigned to cost when compared with applicability. Table VIII shows the IM for Scenario 3 and the weights are presented in Equation (19). For this level, cost and applicability are of less importance presenting the lowest two weights. The impact of subjectivity has increased as repeatability becomes more important. Latency and expressiveness are now the top two weights. This is a desired behaviour as more is expected to be done from the estimation techniques and the earlier they are available, the more can be accomplished and the higher the benefits.

$$w_{sc3} = [0.1701 \;\; 0.0718 \;\; 0.0552 \;\; 0.3861 \;\; 0.3168] \tag{19}$$

Scenario IV: at CMM level 4 quantification and predictability are added to the repeatability of level 3. Variations on the projects' outcomes are due to mostly normal variations but are still within the acceptable control limits. Measurement is a very important factor at level 4 leading to a higher importance to expressiveness when compared with other factors. The more you can extract from an estimation technique, the more useful it will be to manage the process. Repeatability is still very important at this level and the same comments regarding subjectivity and latency for Scenario III are applicable here. The same applies to cost and applicability. The IM matrix for Scenario 4 is presented in Table IX. The associated weights are given in Equation (20). At this level much more is expected to be accomplished by the techniques leading to the highest weight associated with expressiveness. Latency has the second highest weight.

$$w_{sc4} = [0.0989 \;\; 0.0375 \;\; 0.0291 \;\; 0.1777 \;\; 0.6568] \tag{20}$$

Table VIII. Importance matrix for Scenario III.

	F_1	F_2	F_3	F_4	F_5
F_1	1	4	4	$\frac{1}{2}$	$\frac{1}{5}$
F_2	$\frac{1}{4}$	1	2	$\frac{1}{6}$	$\frac{1}{3}$
F_3	$\frac{1}{4}$	$\frac{1}{2}$	1	$\frac{1}{6}$	$\frac{1}{3}$
F_4	2	6	6	1	2
F_5	5	3	$\frac{1}{2}$	$\frac{1}{2}$	1

Table IX. Importance matrix for Scenario IV.

	F_1	F_2	F_3	F_4	F_5
F_1	1	5	5	$\frac{1}{3}$	$\frac{1}{9}$
F_2	$\frac{1}{5}$	1	2	$\frac{1}{7}$	$\frac{1}{9}$
F_3	$\frac{1}{5}$	$\frac{1}{2}$	1	$\frac{1}{7}$	$\frac{1}{9}$
F_4	3	7	7	1	$\frac{1}{9}$
F_5	9	9	9	9	1

DOI: 10.1002/stvr

6.3. Ranking

Synthesis matrices (SM) are used to compute the final rank for each scenario. The weights obtained from the IMs are used to populate the first line of each matrix. That is, w_{sc1} is used to create SM^{sc_1}, w_{sc2} is associated with SM^{sc_2}, and so on. The rest of the SM matrix is populated using the weights computed from the PC matrices. The first column of SM^{sc_i} is filled with w_{subj}, the second column with w_{cost}, the third column with w_{appl}, the fourth column with w_{late}, and the fifth column with w_{dou}. The SM for Scenario 1 is given in Table X. The matrices for the other three scenarios can be obtained by replacing the first row w_{sc1} by w_{sc2}, w_{sc3}, and w_{sc4}.

The rank R of each approach AP_j for a given scenario sc_k is computed according to Equation (21).

$$R_{ap_j}^{sc_k} = \sum_{i=1}^{5} w_{sc_k}(i) \times w_{F_i} \tag{21}$$

where F_i, $i = 1, \ldots, 5$ represent the five features described in Section 3.

The results of Equation (21) for the four scenarios are presented in Table XI. As it can be seen, the importance of low cost and high applicability favour the Classical Parametric approaches for the first two scenarios. CPA^{AP_4} is ranked (1) for both scenarios. CPA^{AP_2} and CPA^{AP_3} are ranked (2) and (3) for the first scenario; their ranks are switched for scenario 2. The Non-parametric approach

Table X. Synthesis matrix (SM^{sc_1}) for Scenario I for ranking the selected defect estimation approaches.

	F_1	F_2	F_3	F_4	F_5
	$w_{sc1}^{F_1} = 0.034$	$w_{sc1}^{F_2} = 0.422$	$w_{sc1}^{F_3} = 0.422$	$w_{sc1}^{F_4} = 0.086$	$w_{sc1}^{F_5} = 0.034$
BBN^{AP_1}	0.0246	0.0220	0.0241	0.4873	0.6317
CPA^{AP_2}	0.1984	0.2269	0.2385	0.0947	0.0698
CPA^{AP_3}	0.2585	0.2269	0.2385	0.0682	0.0698
CPA^{AP_4}	0.3572	0.2269	0.3435	0.0947	0.0307
BPA^{AP_5}	0.0631	0.0702	0.0473	0.2209	0.1283
NPA^{AP_6}	0.0982	0.2269	0.1081	0.0342	0.0698

Table XI. Ranking of existing defect Estimation techniques for CMM-based scenarios.

	Scenario I		Scenario II		Scenario III		Scenario IV	
Existing approaches	Value	Rank	Value	Rank	Value	Rank	Value	Rank
BBN^{AP_1}	0.0843	(5)	0.1022	(5)	0.3954	(1)	0.5054	(1)
CPA^{AP_2}	0.2140	(2)	0.2077	(3)	0.1219	(4)	0.0977	(4)
CPA^{AP_3}	0.2138	(3)	0.2081	(2)	0.1219	(4)	0.0990	(3)
CPA^{AP_4}	0.2624	(1)	0.2499	(1)	0.1423	(3)	0.0909	(5)
BPA^{AP_5}	0.0753	(6)	0.0805	(6)	0.1443	(2)	0.1338	(2)
NPA^{AP_6}	0.1502	(4)	0.1516	(4)	0.0743	(6)	0.0733	(6)

DOI: 10.1002/stvr

occupies the fourth position for both scenarios. The Bayesian approaches have the two lowest rank positions for scenarios 1 and 2.

One question that arises is regarding the sensitiveness of the ranking to close values. For example, for Scenario I the value for CPA^{AP_2} is 0.2140 and the value for CPA^{AP_3} is 0.2138 presenting a small difference only in the third decimal case. As the summation of the values are normalized to 1, it represents only a 0.18% difference that is not significant enough to choose either of the approaches. The same is true for the other three scenarios where the values are either the same or differ from a small amount. In general, when considering only the Classical Parametric approaches, CPA^{AP_4} has a small advantage when compared with CPA^{AP_2} and CPA^{AP_3}. The difference for the first two scenarios are around 4.8 and 4.2% and drops to 2.04 and 0.13% for scenarios III and IV. The decrease in the difference is mainly due to the reduction of importance in the applicability and cost, which are the major advantages of CPA^{AP_4}.

Moving from Scenarios I and II to Scenarios III and IV leads to the importance of factors to switch from low cost and high applicability to high level of expressiveness and low latency. These changes result in the Bayesian approaches occupying the top of the rank. BBN^{AP_1} is ranked (1) and BPA^{AP_5} is ranked (2) for both scenarios. The Classical Parametric approaches are next in the rank. CPA^{AP_4} is ranked (3) while CPA^{AP_2} and CPA^{AP_3} are both ranked (4) for Scenario III. Their positions are exchanged for Scenario IV. Owing mainly to high latency and low expressiveness, the Non-parametric approach NPA^{AP_6} occupies the last position in the ranks for Scenarios III and IV.

In summary, Classical Parametric approaches are better suited for low cost and high applicability scenarios, whereas Bayesian approaches are a better fit for scenarios where low latency and high levels of expressiveness present higher precedence over cost and applicability. The non-parametric approaches do not seem to perform well in any of the studied scenarios. As stated earlier non-parametric approaches do not make any assumption about the parameter under estimation and therefore have to rely solely on the data from the current project to get the estimation; this leads to poor performance in terms of accuracy and latency. Classical Parametric approaches perform better due to the fact that a PDF that captures the dominant behaviour of the parameter is assumed. Bayesian approaches go even further by also relying on specific data from past projects and sometimes on expert knowledge.

Another question to be answered refers to the validity of the ranking. Is the ranking presented in Table XI definitive? The answer is no, as other scenarios could lead to a different ranking. That is, the ranking is based on the importance values given in Tables VI, VII, VIII, and IX. Any person interested in evaluating the impact of using a defect estimation technique can tune the values in the IM matrix to better fit the scenario for a specific company or project. The use of the new IM matrix could lead to a new ranking. This does not mean that the ranking presented here is incorrect, rather that not all possible scenarios have been evaluated. Despite that, our conjecture is that the features presented in Section 3 are the ones that have a major impact on the use of defect estimation techniques and combined with the scenarios based on the CMM provide a reasonable ranking for most common scenarios. The validity of the ranking for a given IM matrix refers to the validity of the AHP approach, which is not in the scope of this paper. Furthermore, AHP has shown to be very adequate to problems where the number of features is not very large as the one addressed in this paper. A potential scenario where the ranking could be wrong is if the IM matrix is not properly defined. However, the IM matrix is, in general, defined by the decision maker, who must have a clear understanding of the company or project under consideration.

DOI: 10.1002/stvr

A final question to be addressed refers to the need to use an MCDM approach, such as AHP, instead of the manual evaluation of a given scenario. Clearly, one could obtain the same results with a manual evaluation but at a higher cost. That is, using the proposed approach, a decision maker just has to plug in values in the IM matrix and the rank can be automatically computed while a manual evaluation could be error prone and more expensive due to the time to reevaluate the existing defect estimation techniques. In addition the use of an MCDM approach allows a novice to select the technique without a detailed knowledge of estimation techniques. Another important aspect of the work proposed here is that it can be easily expanded or modified. New features can be added or removed and the ranking can be easily recomputed under a new characterization.

7. CONCLUDING REMARKS

In this paper a decision support approach for selecting the most suitable defect estimation technique for a project, with specific goals, is proposed and validated.

The new approach consists of three main parts. The first part provides a project-independent characterization of defect estimation approaches, using mathematical and practical aspects. From a mathematical perspective, the existing estimation techniques are first broadly classified as Bayesian (require prior knowledge) or non-Bayesian (do not require prior knowledge) approaches. The Bayesian approaches are further classified into BBN and Parametric Bayesian approaches. The non-Bayesian approaches are further classified into Classical Parametric and Non-parametric approaches. From a practical perspective, additional factors are used to characterize the approaches. These factors include the degree of subjectivity, cost, applicability, latency, and expressiveness.

The second part provides a project-dependent characterization of a project, based on user-defined scenarios. As examples, scenarios are presented that are based on the CMM levels 1–4 of the organization. For example, the first scenario is based on a project to be developed in a CMM Level 1 organization, the second scenario for a CMM Level 2 organization, and so on. The scenarios can be customized according to the specific needs of any user.

The third part ranks the defect estimation to determine the 'best' approach to use for a project. Given the numerous, competing factors that determine the 'best' approach, an MCDM approach is needed. Owing to its simplicity and high applicability, AHP has been used here for the ranking of the estimation techniques; however, other MCDM approaches could have been used.

Although a limited number of scenarios are listed in the study presented in Section 6.2, they appear to provide good representation for many projects and companies. In any case, the proposed approach allows for the specification and evaluation of any other scenario; the ranking can be even expanded to include additional features.

The results of the ranking are a clear indication that no estimation technique provides a single, comprehensive solution; the selection must be done according to a given scenario. The results indicate that Classical Parametric approaches are better suited for low cost and high applicability scenarios, while Bayesian approaches are more appropriate to scenarios where cost is not a major problem and more importance is given to latency and the level of expressiveness for the approach. Non-parametric approaches suffer from high latency and do not seem to be well suited for the estimation of defects.

DOI: 10.1002/stvr

Several interesting directions for future work on this topic could be considered. First, the validation of the approach could be extended, by defining additional, alternative example scenario sets. The current collection of characterization features will be further explored using case studies in the industry to determine if (i) factors are missing or (ii) factors currently defined are not essential. Second, given the extensive literature on estimation approaches, additional representative approaches could be considered in the four estimation classes to refine the selection approach. Third, the tool support could be refined to improve its usability. For example, the steps in the approach could be realized in a GUI-based tool, rather than the current command line version.

APPENDIX A

Here a brief description of some estimators, as seen in Figure A1, for the four classes of estimators defined in Section 2 is presented. Figure A1 is an expanded version of Figure 2 presented in Section 2. The description of the estimators presented here are mostly based on the work of Kay [19].

A.1. Bayesian Belief Networks

A BBN has two components. The first component is a Directed Acyclic Graph (DAG), whose nodes represent random variables and whose arcs represent the dependencies between the variables; a child node depends on the parent node(s). If there is an Arc from variable A to variable B, then A is called a parent of B, and B is a child of A. The second component is the Conditional Probability Distributions (CPDs). Each variable has a CPD that defines the prior probability of the variable given the value of its parents.

Formally, a BBN is denoted as $B = \langle G, \theta \rangle$; where $G = \langle V, A \rangle$, $V = \{X_1, X_2, X_3, \ldots, X_n\}$ is the set of variables presented by the BBN, and A is the set of arcs in the DAG G. The graph G encodes independence assumptions, by which each variable X_i is independent of its non-descendants given its parents in G. θ represents the CPDs of the BBN; it denotes a set of parameters. Each parameter in the set is the CPD of Variable X_i denoted as $\theta_i = P(X_i|Pa(X_i))$ where $Pa(X_i)$ are parents of variable X_i. As given in Equation (A1) the BBN B defines a unique joint probability distribution over V.

$$P(X_1, X_2, X_3, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|Pa(X_i)) \tag{A1}$$

If X_i has no parents, its CPD is said to be unconditional; otherwise, it is conditional. If the variable represented by a node is observed, then the node is said to be an evidence node; otherwise, the node is said to be hidden or latent.

An example of a BBN is given in Figure A2 (extracted from Ben-Gal's work [47]). The BBN can be used to infer the source of a backache (variable A) due to some back injury (variable B). The inference is based on the practice or not of a wrong sport (given by variable S), the use of an uncomfortable chair (variable C), and the existence of co-workers with backache problems (variable W). The CPD associated with each node is also listed.

The true power of BBN lies in the inference that estimates the values of hidden nodes, given the values of the observed nodes. For example, Equation (A2) gives the probability of variable W

DOI: 10.1002/stvr

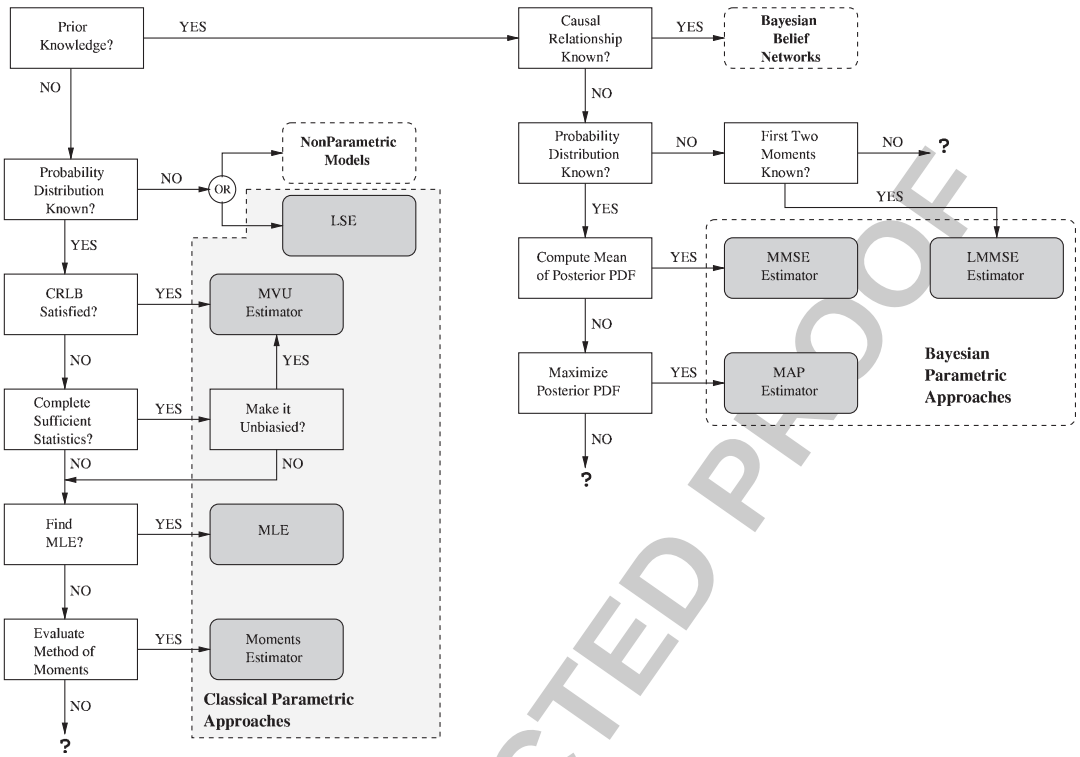


Figure A1. Expanded version of Figure 2 for the classification of estimation approaches. Part of this figure has been extracted from the work of Kay [19].

given that $C = \text{`}T\text{'}$ and $S = \text{`}T\text{'}$ by looking up the CPT and Equation (A3) gives the probability of variable A given that $C = \text{`}T\text{'}$ and $S = \text{`}T\text{'}$ by BBN inference.

$$P(W = \text{`}T\text{'} | C = \text{`}T\text{'}, S = \text{`}T\text{'}) = 0.9 \quad \text{and} \quad P(W = \text{`}F\text{'} | C = \text{`}T\text{'}, S = \text{`}T\text{'}) = 0.1 \tag{A2}$$

$$P(A = \text{`}T\text{'} | C = \text{`}T\text{'}, S = \text{`}T\text{'}) = 0.64 \quad \text{and} \quad P(A = \text{`}F\text{'} | C = \text{`}T\text{'}, S = \text{`}T\text{'}) = 0.36 \tag{A3}$$

A.2. Bayesian parametric approaches

A.2.1. Minimum mean square error (MMSE) estimator

In Equation (1) the BMSE can be minimized by minimizing the inner integral for $\hat{\theta}$, which will result in the estimator given by Equation (A4). In Equation (A4) it can be seen that the Bayesian

DOI: 10.1002/stvr

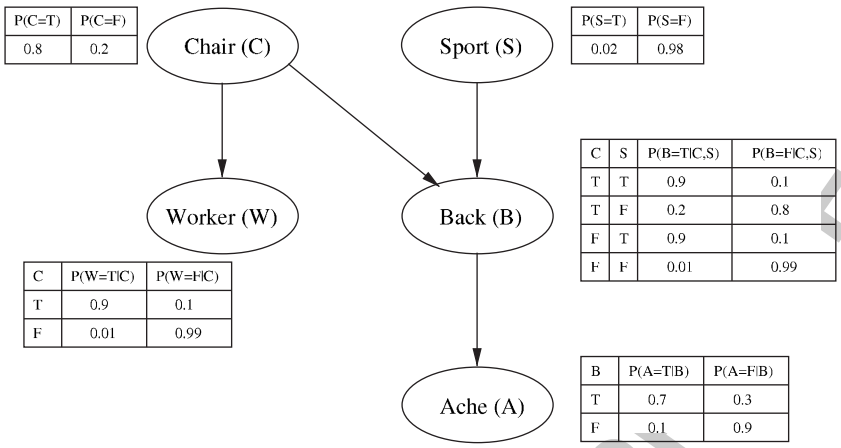


Figure A2. Example of a BBN for a backache problem [47].

estimator is the mean of the posterior PDF $p(\theta|\mathbf{x})$. Intuitively, an optimal estimator is the one that minimizes the BMSE when averaged over the range of θ given by the prior knowledge.

$$\hat{\theta} = E(\theta|\mathbf{x})] \tag{A4}$$

A closed-form solution for Equation (1) exists when both $\mathbf{x}$ and θ are jointly Gaussian. In this particular case the MMSE estimator is given by the following equation:

$$\hat{\theta} = \alpha\overline{\mathbf{x}} + (1-\alpha)\mu_\theta \quad \text{for} \quad \alpha = \frac{\text{prior variance}}{\text{prior variance} + \text{data variance}} \tag{A5}$$

where $\alpha : 0 < \alpha < 1$. Initially, the data set is smaller in size; hence, the weight is towards prior mean μ_θ. As more data points become available, α grows larger and the weight is shifted towards the sample mean. α as described in Equation (A5) is the ratio of variances. When the data set is smaller in size, its variance is greater than the prior variance. However, once the data set grows its variance decreases and it concentrates on the PDF $p(\theta)$ around the actual value of θ. For a sufficiently large data set, the variance of the data will become negligibly small and α will become almost 1, which in turn result in an estimator based on the data set.

A.2.2. Maximum a posteriori (MAP) estimator

Another criterion for finding the estimator is called the Bayes risk with an associated cost function, which is minimized in order to minimize the Bayes risk. One alternative to minimize the Bayes risk is to maximize the posterior probability. The estimator that results from this condition is called a MAP estimator. An MMSE is also categorized among Bayes risk functions. The cost function

DOI: 10.1002/stvr

associated with an MMSE is $(\theta-\hat{\theta})^2$ and it results in the estimator given by Equation (A4). Another type of cost function is called Hit or Miss function, which is given by the following equation:

$$C(\varepsilon)=\begin{cases}0 & |\varepsilon|<\delta\\ 1 & |\varepsilon|>\delta\end{cases} \tag{A6}$$

The Bayes risk for this cost function is given by the following equation:

$$Bayesrisk=E[C(\varepsilon)]=\int\left[\int C(\theta-\hat{\theta})\,p(\theta|\mathbf{x})\,\mathrm{d}\theta\right]p(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{A7}$$

In order to minimize the Bayes risk, the inner integral is minimized according to $\hat{\theta}$ as given by the following equation:

$$\frac{\partial g(\hat{\theta})}{\partial\hat{\theta}}=\frac{\partial}{\partial\hat{\theta}}\left[\int_{-\infty}^{\hat{\theta}}C(\hat{\theta}-\theta)\,p(\theta|\mathbf{x})\,\mathrm{d}\theta+\int_{\hat{\theta}}^{\infty}C(\theta-\hat{\theta})\,p(\theta|\mathbf{x})\,\mathrm{d}\theta\right] \tag{A8}$$

Solving Equation (A8) further results in the following equation, where minimizing the Bayes risk is maximizing the posterior probability; this results in the MAP estimator.

$$g(\hat{\theta})=1-\int_{\hat{\theta}-\delta}^{\hat{\theta}+\delta}p(\theta|\mathbf{x})\,\mathrm{d}\theta \tag{A9}$$

The MAP estimator is derived using the Baye's rule, where the posterior PDF is given by Equation (2). Therefore, a MAP estimator is given by either of the following equations:

$$\hat{\theta}=argmax\ \text{of}\ \theta\ [p(\mathbf{x}|\theta)\,p(\theta)] \tag{A10}$$

$$\hat{\theta}=argmax\ \text{of}\ \theta\ ln[p(\mathbf{x}|\theta)\,p(\theta)] \tag{A11}$$

It is easier to find a closed-form solution for a MAP than for an MMSE estimator for a given distribution because no integration is involved. In fact MAP estimator is intuitively similar to classical MLE and therefore experiences the same problem as classical MLE.

A.2.3. Linear minimum mean square error (LMMSE) estimator

The integration required in the MMSE estimator and maximization in the MAP estimator could become complex and involved, and may not result in closed-form solution. When both data and prior PDF are either Gaussian or jointly Gaussian, a closed-form solution is more likely to be found. Another BPA that can avoid this difficulty assumes that only the first two moments of the joint PDF $p(\mathbf{x},\theta)$ are known as given by Equations (A12) and (A13), and does not make any assumption regarding the distribution of the joint PDF. Another critical assumption is that the estimator is linearly related to data as given by Equation (A14). The name of the estimator linear minimum mean square error (LMMSE) estimator originates from the fact that the estimator is found using

DOI: 10.1002/stvr

the same criterion as an MMSE estimator and the relationship between the estimator and data is linear.

$$First\ moment = \begin{bmatrix} E[\theta] \\ E[\mathbf{x}] \end{bmatrix} \tag{A12}$$

$$Second\ moment = \begin{bmatrix} \mathbf{COV}_{\theta\theta} & \mathbf{COV}_{\theta x} \\ \mathbf{COV}_{x\theta} & \mathbf{COV}_{xx} \end{bmatrix} \tag{A13}$$

$$\hat{\theta} = \mathbf{a}^{\mathrm{T}}\mathbf{x} + x[N] \tag{A14}$$

where $E[\theta]$ is the expectation of the parameter θ (same applies to x) and $\mathbf{COV}_{ij}$ is the covariance between variables i and j. The coefficients a are found by minimizing the BMSE given in Equation (A15), which results in Equation (A16).

$$BMSE = E[(\theta - \hat{\theta})^2] \tag{A15}$$

$$\mathbf{a} = \mathbf{COV}_{xx}^{-1}\mathbf{COV}_{x\theta} \tag{A16}$$

Finally, the estimator is given by the following equation:

$$\hat{\theta} = E[\theta] + \mathbf{COV}_{\theta x}\mathbf{COV}_{xx}^{-1}(\mathbf{x} - E[\mathbf{x}]) \tag{A17}$$

A.3. Classical Parametric approaches

A.3.1. Minimum variance unbiased (MVU) estimator

The variability of estimates determines the efficiency of the estimator. The higher the variance of the estimator, the less effective (or reliable) are the estimates. Hence, various (non-constant) estimators can be found for the data, but the one with the lowest variance is the best unbiased estimator. Another important characteristic of the estimator is that it must be unbiased, i.e. $E[\hat{\theta}] = \theta$.

There are various methods available to determine the lower bound on the variance of the estimators, e.g. [48,49], but CRLB [19] is the most frequently used.

An estimator that is unbiased, satisfies the CRLB theorem, and is based on a linear model is called an efficient MVU estimator [19]. In other words it efficiently uses the data to find the estimates. An estimator with a variance that is always minimum when compared with other estimators but is not less than CRLB is simply called an MVU estimator. An estimator based on a linear data model is more likely to attain the CRLB and hence provide an efficient MVU estimator. The efficient MVU estimator for $\hat{\theta}$ and its variance (VAR) are given in the following equations:

$$\frac{\partial ln\ p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)(g(\mathbf{x}) - \theta) \tag{A18}$$

$$\hat{\theta} = g(\mathbf{x}) \tag{A19}$$

$$VAR(\hat{\theta}) = \frac{1}{I(\theta)} \tag{A20}$$

DOI: 10.1002/stvr

A defect estimator that is an efficient MVU estimator will provide very accurate estimates. In other fields such as signal processing and communication systems where the system or model under investigation is well defined in terms of physical constraints, it is possible to find an efficient MVU estimator. However, in software engineering no single model completely captures all the aspects of a software testing process. Different models [3,12,18,26–29] are based on different assumptions and this lack of consistency hints towards the absence of a mature testing model. Therefore, it is unlikely to find an efficient MVU estimator.

It is possible that for a given data model the CRLB cannot be achieved. In other words, a solution similar to the one given by Equation (A18) may not exist. Another approach to find an MVU estimator is based on finding the sufficient statistic such that minimal data are required to make the PDF of the data $p(\mathbf{x};\theta)$ independent of the unknown parameter θ. As discussed earlier $p(x[n];\theta)$ is dependent of both data $x[n]$ and θ. The existence of the sufficient statistic leads to the following equation; this is according to Neyman–Fisher Factorization Theorem [19]:

$$p(\mathbf{x};\theta) = g(T(\mathbf{x}),\theta)h(\mathbf{x}) \tag{A21}$$

where $T(\mathbf{x})$ is the sufficient statistic. If $p(\mathbf{x};\theta)$ can be factorized as given by Equation (A21), then the sufficient statistic $T(\mathbf{x})$ exists. Then, a function of $T(\mathbf{x})$ needs to be found, which is the estimator $\hat{\theta}$ such that $\hat{\theta}$ is unbiased $E[\hat{\theta}]=\theta$. Then we have to find a function of the estimator $\hat{\theta}$ as given by $\hat{\theta}=f(T(\mathbf{x}))$.

A.3.2. Maximum likelihood estimator (MLE)

If neither the sufficient statistic $T(\mathbf{x})$ for θ nor an unbiased function $f(T(\mathbf{x}))$ can be found, then an MLE estimator can be used. Many practical estimators proposed in the field for defects estimation are based on an MLE, e.g. [18,28,33]. An MLE is asymptotically (as $N\to\infty$) an efficient MVU estimator. For a linear data model the MLE achieves the CRLB for a finite data set [19]. Another important property is that if an efficient estimator exists, then MLE will produce it [19]. The basic idea is to find the value of θ that maximizes $ln\ p(\mathbf{x};\theta)$ the log-likelihood function for a given $\mathbf{x}$. If a closed-form solution does not exist, then a numerical method such as Newton–Raphson can be used to approximate the solution. However, the approximation may not necessarily converge to maximization of $\ln p(\mathbf{x};\theta)$ to produce an MLE. An example of numerical approximation of an MLE is given by John and Musa [3].

A.3.3. Methods of moments

Another method to find estimator if either MLE cannot be found or is computationally intensive is called the Method of Moments. The method of moments estimator is generally consistent [19]. Given $p(\mathbf{x};\theta)$, if we know the kth moment of $x[n]$, then μ_k is a function of θ as given by the following equation:

$$\mu_k = E(x[n]^k) = f(\theta) \tag{A22}$$

$$\theta = f^{-1}(\mu_k) \tag{A23}$$

DOI: 10.1002/stvr

$$\hat{\mu_k} = \frac{1}{N} \sum_{n=0}^{N-1} x^k[n] \tag{A24}$$

$$\hat{\theta} = f^{-1}\left(\frac{1}{N} \sum_{n=0}^{N-1} x^k[n]\right) \tag{A25}$$

The kth moment of the data $\mathbf{x}$ is approximated $\hat{\mu_k}$ by taking the average of $\mathbf{x}^{(k)}$ as given by Equation (A24). If f is an invertible function as given by Equation (A23), then substituting the approximation $\hat{\mu_k}$ into Equation (A23) results in the estimator $\hat{\theta}$ as given by Equation (A25).

A.3.4. Least-square estimator (LSE)

The concern regarding the estimators previously described in this Appendix from A.3.1 to A.3.3 was in finding an unbiased estimator with minimum variance when either the PDF is known or some assumption is made about it. When this is not the case, an LSE can be defined to minimize the square difference between the available data x and the chosen model M. That is, the parameter θ to be estimated is chosen to make M as close as possible to x, leading to the following equation:

$$LSE(\theta) = \sum_{n=0}^{N-1} (x[n] - M[n]^2 \tag{A26}$$

where N is the number of data points and the model M is a function of θ.

The use of the LSE does not depend on the PDF and, in theory, the estimator is valid for any PDF. This clearly improves the applicability of the estimator as it depends only on the data points for the current projects. However, the accuracy of the estimator depends on how good the model M is and the noise in the data x. As noise is a commonplace in the software process, the expectation is that an LSE would have poor accuracy when compared with other estimators. During the testing process, for example, noise can be characterized as changes in the personnel, variations in the complexity of the application under test, schedule pressures, etc.

A.4. Non-parametric approaches

Two well-known Non-parametric methods are the Parzen window and the kth-nearest neighbor. The Parzen window method estimates the conditional density function $p(\mathbf{x}|\theta)$ from the given data set. The kth-nearest neighbor estimates the posteriori probability $p(\theta|\mathbf{x})$. The goal is to find a smoothed version of the PDF of data as given by Equation (A27) over a region R; the underlying assumption is that data samples are i.i.d. Note that $\mathbf{x}$ is a single d-dimensional point and so data samples are represented by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

$$P = \int_R p(\mathbf{x}')\, \mathrm{d}\mathbf{x}' \tag{A27}$$

If k of these samples fall inside R, then according to binomial distribution the probability of these k samples is given by Equation (A28). The mean of choosing k points is given by Equation (A29).

DOI: 10.1002/stvr

Similarly the mean fraction k/n of the points is given by Equation (A30).

$$P_k = \binom{N}{k} P^k (1-P)^{n-k} \tag{A28}$$

$$E[k] = NP \tag{A29}$$

$$E[k/N] = P \tag{A30}$$

Note that the distribution of k/N is sharply peaked around the mean $E[k/N]$ as $VAR[k/N] \to 0$ as $n \to \infty$. Another assumption is that if $p(\mathbf{x})$ is continuous and R is so small that $p(\mathbf{x})$ is constant over R then $p(\mathbf{x})$ is given by Equation (A32) for a point $\mathbf{x}$.

$$P = \int_R p(\mathbf{x}')\,\mathrm{d}\mathbf{x}' \simeq p(\mathbf{x})V \tag{A31}$$

$$p(\mathbf{x}) = \frac{P}{V} = \frac{k/N}{V} \tag{A32}$$

A tradeoff needs to be considered in the size of R. When R is smaller, the averaging of $p(\mathbf{x})$ is avoided and a true value of $p(\mathbf{x})$ is obtained. However, the size of R must be large enough so that P is large, which results in a small value of $VAR[k/N]$. In summary there are two competing goals: (1) Decrease $VAR[k/N]$; (2) Minimize averaging of $p(\mathbf{x})$. There are two approaches to these goals. The first approach leads to the Parzen window method and the second approach leads to the kth-nearest neighbor method.

A.4.1. Parzen windows

The Parzen window technique is based on the approach that for a given $VAR[k/N]$ and, therefore, a fixed value of $k = k_N$, find a value of $V = V_N$ such that the average of $p(\mathbf{x})$ is minimized. For N data points, let R_N be a hypercube in d-dimensional space with each side of length h_N and volume V_N. $\phi((\mathbf{x}-\mathbf{x}_i)/h_N)$ is a function centred at $\mathbf{x}$; it is equal to 1 if the sample point $\mathbf{x}_i$ falls in the hypercube R_N. Therefore, k_N sample points falling inside R_N are given by Equation (A33).

$$k_N = \sum_{i=1}^{N} \phi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_N}\right) \tag{A33}$$

The estimate of $p(\mathbf{x})$ is obtained by substituting the value of k_N in Equation (A34).

$$p_N(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{N} \frac{1}{N}\phi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h_N}\right) \tag{A34}$$

Note that $1/V_n \int \phi((\mathbf{x}-\mathbf{x}_i)/h_n)\,\mathrm{d}\mathbf{x} = 1$ when centred at $\mathbf{x}_i$. The choice of ϕ and h_N determines the accuracy of the estimate.

DOI: 10.1002/stvr

A.4.2. K th-nearest neighbour

The Kth-nearest neighbour technique is based on an alternative goal. For a given value of averaging of $p(\mathbf{x})$ and, therefore, a fixed value of $V = V_N$, find a value of $k = k_N$ such that $VAR[k/N]$ is minimized. The idea is to extend the size of the cell centred at $\mathbf{x}$ so that k_N-nearest neighbor are captured. This will result in high resolution given that as $N \to \infty$, $k_N \to \infty$ as well so that $VAR[k/N] \to 0$. This fact refers back to Equation (A32), which can be rewritten in the form of the following equation:

$$p_N(\mathbf{x}) = \frac{P}{V} = \frac{k_N/N}{V_N} \tag{A35}$$

Let $k = \sqrt{N}$ rather than choosing some arbitrary function of data set. If $p_N(\mathbf{x})$ is also a reasonable approximation of $p(\mathbf{x})$, then from Equation (A35) V_N can be found by the following equation, where V_1 can be a function of data rather than some arbitrary function:

$$V_N \simeq \frac{1}{\sqrt{N}\, p_N(\mathbf{x})} = \frac{V_1}{\sqrt{N}} \tag{A36}$$

REFERENCES

1. Satoh D, Yamada S. Discrete equations and software reliability growth models. *12th International Symposium on Software Reliability Engineering*, (*ISSRE*). IEEE: New York, 2001; 176–184.
2. Yamada S, Ohba M, Osaki S. S-shaped reliability growth modeling for software error detection. *IEEE Transactions on Reliability* 1983; **R-32**(5):475–478.
3. John D, Musa KO. A logarithmic poisson execution time model for software reliability measurement. *Proceedings of the Seventh International Conference on Software Engineering*. IEEE, ACM, IEEE Press: New York, 1984; 230–238.
4. Cangussu JW, DeCarlo RA, Mathur AP. A formal model for the software test process. *IEEE Transactions on Software Engineering* 2002; **28**(8):782–796.
5. Runeson P, Wohlin C. An experimental evaluation of an experience-based capture–recapture method in software code inspections. *Empirical Software Engineering* 1998; **3**(3):381–406.
6. Briand L, Emam KE, Freimut B. A comparison and integration of capture–recapture models and the detection profile method. *Proceeding of Ninth International Symposium on Software Reliability Engineering* (*ISSRE 1998*), 1998; 32–41. **Q2**
7. Gaffney JE Jr. Metrics in software quality assurance, 1981.
8. Ostrand TJ, Weyuker EJ, Bell RM. Locating where faults will be. *Richard Tapia Celebration of Diversity in Computing Conference*. IEEE, IEEE Press: New York, 2005; 48–50.
9. Ostrand TJ, Weyuker EJ, Bell RM. Predicting the location and number of faults in large software systems. *IEEE Transactions on Software Engineering* 2005; **31**(4). **Q3**
10. Ohlsson N, Alberg H. Predicting error-prone software modules in telephone switches. *IEEE Transactions on Software Engineering* 1996; **22**(12):886–894.
11. Fenton NE, Neil M. A critique of software defect prediction models. *IEEE Transactions on Software Engineering* 1999; **25**(5):675–689.
12. Briand LC, Emam KE, Freimut BG, Laitenberger O. A comprehensive evaluation of capture–recapture models for estimating software defect content. *IEEE Transactions on Software Engineering* 2000; **26**(6):518–540.
13. Mollaghasemi M, Pet-Edwards J. *Making Multiple-Objective Decisions*. IEEE Computer Society Press: Silver Spring, MD, 1997.
14. Saaty T. *Fundamentals of Decision Making and Priority Theory*. RWS Publications, 1994. **Q4**
15. Littlewood B, Sofer A. A Bayesian modification to the Jelinski–Moranda software reliability model. *Software Engineering Journal* 1987; **2**(2):30–41.
16. Padberg F. A fast algorithm to compute maximum likelihood estimates for the hypergeometric software reliability model. *Second Asia–Pacific Conference on Quality Software*. IEEE: New York, 2001; 40–49.
17. Barghout M, Littlewood B, Abdel-Ghaly A. A non-parametric order statistics software reliability model. *Software Testing, Verification and Reliability* 1999; **8**(3):113–132.

DOI: 10.1002/stvr

18. Haider SW, Cangussu JW, Cooper K, Dantu R. Estimating defects based on defect decay model: ED^3M. *IEEE Transactions on Software Engineering* 2008. **Q5**
19. Kay SM. *Fundamentals of Statistical Signal Processing*: *Estimation Theory*. Prentice-Hall PTR: Englewood Cliffs, NJ, 1993.
20. Mittal A, Kassim A (eds.). *Bayesian Network Technologies*: *Applications and Graphical Models*. IGI Publications, 2007.
21. Pradhan M, Provan G, Middleton B, Henrion M. Knowledge engineering for large belief networks. *Proceeding of the Tenth Conference on Uncertainty in Artificial Intelligence*, 1994; 484–490.
22. Daniel BK, Zapata-Rivera JD, McCalla GI. A Bayesian computational model of social capital in virtual communities. *Communities and Technologies*. Kluwer: London, 2003; 287–305.
23. Fentona N, Neila M, Marsha W, Heartya P, Marqueza D, Krausec P, Mishrad R. Predicting software defects in varying development lifecycles using bayesian nets. *Information and Software Technology* 2007; **49**(1):32–43.
24. Fentona N, Neila M, Marsha W, Hearty P, Radlinski L, Krausec P. Project data incorporating qualitative factors for improved software defect prediction. *Proceedings of the Third international Workshop on Predictor Models in Software Engineering* (*International Conference on Software Engineering—ICSE*). IEEE Computer Society: Washington, DC, 2007.
25. Daniel BK, Zapata-Rivera JD, McCalla GI. A Bayesian belief network approach for modeling complex domains, chapter II. IGI Publications, 2007; 13–41. **Q6**
26. Goel AL, Okumoto K. Time-dependent error-detection rate model for software and other performance measures. *IEEE Transactions on Reliability* 1979; **R-28**(3):206–211.
27. Moranda P, Jelinski Z. Final report on software reliability study. *Technical Report 63921*, McDonnell Douglas Astronautics Company, MADC, 1972.
28. Padberg F. Maximum likelihood estimates for the hypergeometric software reliability model. *International Journal of Reliability*, *Quality and Safety Engineering* 2002.
29. Zhang P, Mockus A. On measurement and understanding of software development processes. *Technical Report ALR-2002-048*, Avaya Research Labs, November 2002.
30. Littlewood B, Verrall JL. A Bayesian reliability growth model for computer software. *Applied Statistics* 1973; **22**(3): 332–346.
31. Mazzuchi T, Soyer R. A Bayes empirical-Bayes model for software reliability. *IEEE Transactions on Reliability* 1988; **37**(2):248–254. DOI: 10.1109/24.3749.
32. Haider SW, Cangussu JW. Bayesian estimation of defects based on the defect decay model: ED^3M. 2006.
33. Iannino MO. *Software Reliability Measurement*, *Prediction*, *Application*. McGraw-Hill: New York, 1987.
34. Kenett RS, Polak M. Semi-parametric approach to testing for reliability growth, with application to software systems. *IEEE Transactions on Reliability* 1986; **35**(3):304–311.
35. Fenton NE, Neil M. Software metrics: Roadmap. *ICSE '00*: *Proceedings of the Conference on the Future of Software Engineering*. ACM: New York, NY, U.S.A., 2000; 357–370. DOI: http://doi.acm.org/10.1145/336512.336588.
36. Cai KY. *Software Defect and Operational Profile Modeling*. Kluwer: Dordrecht, 1998.
37. Barghout M, Littlewood B, Abdel-Ghaly A. A non-parametric approach to software reliability prediction. *Proceedings of the Eighth International Symposium on Software Reliability Engineering*, 2–5 November 1997; 366–377.
38. Hammond JS, Keeney RL, Raiffa H. *Smart Choices A Practical Guide to Making Better Decisions*. Harvard Business School Press: Cambridge, MA, 1999.
39. Nemhauser G, Rinnooy K, Todd M. *Handbooks in Operations Research and Management Science*, *Optimization*. North-Holland: Amsterdam, 1989.
40. Slinesi C, Kornyshova E. Choosing a prioritization method. Case of is security improvement. *Proceedings of 18th Conference on Advanced Information Systems Engineering*, Luxembourg, 2006; CEUR–WS/Vol–231.
41. Akhavi F, Hayes C. A comparison of two multi-criteria decision-making techniques. *Proceedings of the IEEE International Conference on Systems*, *Man and Cybernetics*, 2003; 956–961.
42. Scholl A, Manthey L, Helm R, Steiner M. Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: an empirical comparison. *European Journal of Operational Research* 2005; **164**:760–777.
43. Mahmoud M, Garcia L. Comparison of different multicriteria evaluation methods for the red bluff diversion dam. *Journal of Environmental Modelling and Software* 2000; **15**:471–478.
44. Kwiesielewicz M, van Uden E. Inconsistent and contradictory judgements in pairwise comparison method in the AHP. *Journal of Computers and Operations Research* 2004; **31**:713–719.
45. Saaty T, Vargas L. *Methods*, *Concepts and Applications of the Analytic Hierarchy Process*. Kluwer: Dordrecht, 2001.
46. SEI. CMMI. Available at: www.sei.cmu.edu/cmmi, online edn, November 2005. **Q7**
47. Ben-Gal I. Bayesian networks. In *Encyclopedia of Statistics in Quality and Reliability*, Ruggeri F, Kenett R, Faltin F (eds.). Wiley: New York, 2007.
48. Kendall SM, Stuart A. *The Advanced Theory of Statistics*, vol. 2. Macmillan: New York, 1979.
49. McaAulay RJ, Hofstetter EM. Barankin bounds on parameter estimation. *IEEE Transactions on Information Theory* 1971; **17**:669–676.

Author Queries Form

John Wiley

JOURNAL TITLE: **STVR** **30/9/2009**

ARTICLE NO: 419

Queries and / or remarks

Query No.	Details required	Author's response
Q1	Please provide the table of contents with the following details for this article: text, one figure (diagram or illustration selected from the manuscript or an additional eye-catching figure). The table of contents entry must contain the paper title and the authors' names (with the corresponding author indicated by an asterisk) and should contain the figure and no more than 80 words or 3 sentences of text summarizing the key findings presented in the paper.	
Q2	Please provide the place where the proceeding was held for References [6,21,37,41].	
Q3	Please provide page number for Reference [9].	
Q4	Please provide the location of the publisher for References [14,20,25].	
Q5	Please provide further details for References [18,28,32].	
Q6	Please provide book title for Reference [25].	
Q7	Please provide access date for Reference [46].	

COPYRIGHT TRANSFER AGREEMENT

Date: _____ Contributor name: _____

Contributor address: _____

Manuscript number (Editorial office only): _____

Re: Manuscript entitled _____

_____ (the "Contribution")

for publication in _____ (the "Journal")

published by _____ ("Wiley-Blackwell").

Dear Contributor(s):
Thank you for submitting your Contribution for publication. In order to expedite the editing and publishing process and enable Wiley-Blackwell to disseminate your Contribution to the fullest extent, we need to have this Copyright Transfer Agreement signed and returned as directed in the Journal's instructions for authors as soon as possible. If the Contribution is not accepted for publication, or if the Contribution is subsequently rejected, this Agreement shall be null and void. **Publication cannot proceed without a signed copy of this Agreement.**

A. COPYRIGHT

1. The Contributor assigns to Wiley-Blackwell, during the full term of copyright and any extensions or renewals, all copyright in and to the Contribution, and all rights therein, including but not limited to the right to publish, republish, transmit, sell, distribute and otherwise use the Contribution in whole or in part in electronic and print editions of the Journal and in derivative works throughout the world, in all languages and in all media of expression now known or later developed, and to license or permit others to do so.

2. Reproduction, posting, transmission or other distribution or use of the final Contribution in whole or in part in any medium by the Contributor as permitted by this Agreement requires a citation to the Journal and an appropriate credit to Wiley-Blackwell as Publisher, and/or the Society if applicable, suitable in form and content as follows: (Title of Article, Author, Journal Title and Volume/Issue, Copyright © [year], copyright owner as specified in the Journal). Links to the final article on Wiley-Blackwell's website are encouraged where appropriate.

B. RETAINED RIGHTS

Notwithstanding the above, the Contributor or, if applicable, the Contributor's Employer, retains all proprietary rights other than copyright, such as patent rights, in any process, procedure or article of manufacture described in the Contribution.

C. PERMITTED USES BY CONTRIBUTOR

1. Submitted Version. Wiley-Blackwell licenses back the following rights to the Contributor in the version of the Contribution as originally submitted for publication:

a. After publication of the final article, the right to self-archive on the Contributor's personal website or in the Contributor's institution's/employer's institutional repository or archive. This right extends to both intranets and the Internet. The Contributor may not update the submission version or replace it with the published Contribution. The version posted must contain a legend as follows: This is the pre-peer reviewed version of the following article: FULL CITE, which has been published in final form at [Link to final article].

b. The right to transmit, print and share copies with colleagues.

2. Accepted Version. Re-use of the accepted and peer-reviewed (but not final) version of the Contribution shall be by separate agreement with Wiley-Blackwell. Wiley-Blackwell has agreements with certain funding agencies governing reuse of this version. The details of those relationships, and other offerings allowing open web use, are set forth at the following website: http://www.wiley.com/go/funderstatement. NIH grantees should check the box at the bottom of this document.

3. Final Published Version. Wiley-Blackwell hereby licenses back to the Contributor the following rights with respect to the final published version of the Contribution:

a. Copies for colleagues. The personal right of the Contributor only to send or transmit individual copies of the final published version in any format to colleagues upon their specific request provided no fee is charged, and further-provided that there is no systematic distribution of the Contribution, e.g. posting on a listserve, website or automated delivery.

b. Re-use in other publications. The right to re-use the final Contribution or parts thereof for any publication authored or edited by the Contributor (excluding journal articles) where such re-used material constitutes less than half of the total material in such publication. In such case, any modifications should be accurately noted.

c. Teaching duties. The right to include the Contribution in teaching or training duties at the Contributor's institution/place of employment including in course packs, e-reserves, presentation at professional conferences, in-house training, or distance learning. The Contribution may not be used in seminars outside of normal teaching obligations (e.g. commercial seminars). Electronic posting of the final published version in connection with teaching/training at the Contributor's institution/place of employment is permitted subject to the implementation of reasonable access control mechanisms, such as user name and password. Posting the final published version on the open Internet is not permitted.

d. Oral presentations. The right to make oral presentations based on the Contribution.

4. Article Abstracts, Figures, Tables, Data Sets, Artwork and Selected Text (up to 250 words).

a. Contributors may re-use unmodified abstracts for any non-commercial purpose. For on-line uses of the abstracts, Wiley-Blackwell encourages but does not require linking back to the final published versions.

b. Contributors may re-use figures, tables, data sets, artwork, and selected text up to 250 words from their Contributions, provided the following conditions are met:

(i) Full and accurate credit must be given to the Contribution.
(ii) Modifications to the figures, tables and data must be noted. Otherwise, no changes may be made.
(iii) The reuse may not be made for direct commercial purposes, or for financial consideration to the Contributor.
(iv) Nothing herein shall permit dual publication in violation of journal ethical practices.

D. CONTRIBUTIONS OWNED BY EMPLOYER

1. If the Contribution was written by the Contributor in the course of the Contributor's employment (as a "work-made-for-hire" in the course of employment), the Contribution is owned by the company/employer which must sign this Agreement (in addition to the Contributor's signature) in the space provided below. In such case, the company/employer hereby assigns to Wiley-Blackwell, during the full term of copyright, all copyright in and to the Contribution for the full term of copyright throughout the world as specified in paragraph A above.

2. In addition to the rights specified as retained in paragraph B above and the rights granted back to the Contributor pursuant to paragraph C above, Wiley-Blackwell hereby grants back, without charge, to such company/employer, its subsidiaries and divisions, the right to make copies of and distribute the final published Contribution internally in print format or electronically on the Company's internal network. Copies so used may not be resold or distributed externally. However the company/employer may include information and text from the Contribution as part of an information package included with software or other products offered for sale or license or included in patent applications. Posting of the final published Contribution by the institution on a public access website may only be done with Wiley-Blackwell's written permission, and payment of any applicable fee(s). Also, upon payment of Wiley-Blackwell's reprint fee, the institution may distribute print copies of the published Contribution externally.

E. GOVERNMENT CONTRACTS

In the case of a Contribution prepared under U.S. Government contract or grant, the U.S. Government may reproduce, without charge, all or portions of the Contribution and may authorize others to do so, for official U.S. Government purposes only, if the U.S. Government contract or grant so requires. (U.S. Government, U.K. Government, and other government employees: see notes at end)

F. COPYRIGHT NOTICE

The Contributor and the company/employer agree that any and all copies of the final published version of the Contribution or any part thereof distributed or posted by them in print or electronic format as permitted herein will include the notice of copyright as stipulated in the Journal and a full citation to the Journal as published by Wiley-Blackwell.

G. CONTRIBUTOR'S REPRESENTATIONS

The Contributor represents that the Contribution is the Contributor's original work, all individuals identified as Contributors actually contributed to the Contribution, and all individuals who contributed are included. If the Contribution was prepared jointly, the Contributor agrees to inform the co-Contributors of the terms of this Agreement and to obtain their signature to this Agreement or their written permission to sign on their behalf. The Contribution is submitted only to this Journal and has not been published before. (If excerpts from copyrighted works owned by third parties are included, the Contributor will obtain written permission from the copyright owners for all uses as set forth in Wiley-Blackwell's permissions form or in the Journal's Instructions for Contributors, and show credit to the sources in the Contribution.) The Contributor also warrants that the Contribution contains no libelous or unlawful statements, does not infringe upon the rights (including without limitation the copyright, patent or trademark rights) or the privacy of others, or contain material or instructions that might cause harm or injury.

CHECK ONE BOX:

☐ Contributor-owned work

ATTACH ADDITIONAL SIGNATURE PAGES AS NECESSARY

Contributor's signature _____ Date _____

Type or print name and title _____

Co-contributor's signature _____ Date _____

Type or print name and title _____

☐ Company/Institution-owned work (made-for-hire in the course of employment)

Company or Institution (Employer-for-Hire) _____ Date _____

Authorized signature of Employer _____ Date _____

☐ U.S. Government work

Note to U.S. Government Employees
A contribution prepared by a U.S. federal government employee as part of the employee's official duties, or which is an official U.S. Government publication, is called a "U.S. Government work," and is in the public domain in the United States. In such case, the employee may cross out Paragraph A.1 but must sign (in the Contributor's signature line) and return this Agreement. If the Contribution was not prepared as part of the employee's duties or is not an official U.S. Government publication, it is not a U.S. Government work.

☐ U.K. Government work (Crown Copyright)

Note to U.K. Government Employees
The rights in a Contribution prepared by an employee of a U.K. government department, agency or other Crown body as part of his/her official duties, or which is an official government publication, belong to the Crown. U.K. government authors should submit a signed declaration form together with this Agreement. The form can be obtained via http://www.opsi.gov.uk/advice/crown-copyright/copyright-guidance/publication-of-articles-written-by-ministers-and-civil-servants.htm

☐ Other Government work

Note to Non-U.S., Non-U.K. Government Employees
If your status as a government employee legally prevents you from signing this Agreement, please contact the editorial office.

☐ NIH Grantees

Note to NIH Grantees
Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of Contributions authored by NIH grant-holders to PubMed Central upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, see www.wiley.com/go/nihmandate.

WILEY AUTHOR DISCOUNT CARD

As a highly valued contributor to Wiley's publications, we would like to show our appreciation to you by offering a **unique 25% discount** off the published price of any of our books*.

To take advantage of this offer, all you need to do is apply for the **Wiley Author Discount Card** by completing the attached form and returning it to us at the following address:

The Database Group
John Wiley & Sons Ltd
The Atrium
Southern Gate
Chichester
West Sussex PO19 8SQ
UK

In the meantime, whenever you order books direct from us, simply quote promotional code **S001W** to take advantage of the 25% discount.

The newest and quickest way to order your books from us is via our new European website at:

http://www.wileyeurope.com

Key benefits to using the site and ordering online include:

- Real-time SECURE on-line ordering
- The most up-to-date search functionality to make browsing the catalogue easier
- Dedicated Author resource centre
- E-mail a friend
- Easy to use navigation
- Regular special offers
- Sign up for subject orientated e-mail alerts

So take advantage of this great offer, return your completed form today to receive your discount card.

Yours sincerely,

Verity Leaver
E-marketing and Database Manager

*TERMS AND CONDITIONS

This offer is exclusive to Wiley Authors, Editors, Contributors and Editorial Board Members in acquiring books (excluding encyclopaedias and major reference works) for their personal use. There must be no resale through any channel. The offer is subject to stock availability and cannot be applied retrospectively. This entitlement cannot be used in conjunction with any other special offer. Wiley reserves the right to amend the terms of the offer at any time.

REGISTRATION FORM
FOR 25% BOOK DISCOUNT CARD

To enjoy your special discount, tell us your areas of interest and you will receive relevant catalogues or leaflets from which to select your books. Please indicate your specific subject areas below.

Accounting [ ]
- Public [ ]
- Corporate [ ]

Chemistry [ ]
- Analytical [ ]
- Industrial/Safety [ ]
- Organic [ ]
- Inorganic [ ]
- Polymer [ ]
- Spectroscopy [ ]

Encyclopedia/Reference [ ]
- Business/Finance [ ]
- Life Sciences [ ]
- Medical Sciences [ ]
- Physical Sciences [ ]
- Technology [ ]

Earth & Environmental Science [ ]

Hospitality [ ]

Genetics [ ]
- Bioinformatics/Computational [ ]
  Biology [ ]
- Proteomics [ ]
- Genomics [ ]
- Gene Mapping [ ]
- Clinical Genetics [ ]

Medical Science [ ]
- Cardiovascular [ ]
- Diabetes [ ]
- Endocrinology [ ]
- Imaging [ ]
- Obstetrics/Gynaecology [ ]
- Oncology [ ]
- Pharmacology [ ]
- Psychiatry [ ]

Non-Profit [ ]

Architecture [ ]

Business/Management [ ]

Computer Science [ ]
- Database/Data Warehouse [ ]
- Internet Business [ ]
- Networking [ ]
- Programming/Software [ ]
  Development [ ]
- Object Technology [ ]

Engineering [ ]
- Civil [ ]
- Communications Technology [ ]
- Electronic [ ]
- Environmental [ ]
- Industrial [ ]
- Mechanical [ ]

Finance/Investing [ ]
- Economics [ ]
- Institutional [ ]
- Personal Finance [ ]

Life Science [ ]

Landscape Architecture [ ]

Mathematics/Statistics [ ]

Manufacturing [ ]

Material Science [ ]

Psychology [ ]
- Clinical [ ]
- Forensic [ ]
- Social & Personality [ ]
- Health & Sport [ ]
- Cognitive [ ]
- Organizational [ ]
- Developmental and Special Ed [ ]
- Child Welfare [ ]
- Self-Help [ ]

Physics/Physical Science [ ]

[ ] I confirm that I am a Wiley Author/Editor/Contributor/Editorial Board Member of the following publications:

SIGNATURE: .....

PLEASE COMPLETE THE FOLLOWING DETAILS IN BLOCK CAPITALS:

TITLE AND NAME: (e.g. Mr, Mrs, Dr) .....

JOB TITLE: .....

DEPARTMENT: .....

COMPANY/INSTITUTION: .....

ADDRESS: .....

.....

.....

.....

TOWN/CITY: .....

COUNTY/STATE: .....

COUNTRY: .....

POSTCODE/ZIP CODE: .....

DAYTIME TEL: .....

FAX: .....

E-MAIL: .....

YOUR PERSONAL DATA
We, John Wiley & Sons Ltd, will use the information you have provided to fulfil your request. In addition, we would like to:

1. Use your information to keep you informed by post, e-mail or telephone of titles and offers of interest to you and available from us or other Wiley Group companies worldwide, and may supply your details to members of the Wiley Group for this purpose.

   [ ] Please tick the box if you do not wish to receive this information

2. Share your information with other carefully selected companies so that they may contact you by post, fax or e-mail with details of titles and offers that may be of interest to you.

   [ ] Please tick the box if you do not wish to receive this information.

If, at any time, you wish to stop receiving information, please contact the Database Group (databasegroup@wiley.co.uk) at John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, UK.

E-MAIL ALERTING SERVICE
We offer an information service on our product ranges via e-mail. If you do not wish to receive information and offers from John Wiley companies worldwide via e-mail, please tick the box [ ].

This offer is exclusive to Wiley Authors, Editors, Contributors and Editorial Board Members in acquiring books (excluding encyclopaedias and major reference works) for their personal use. There should be no resale through any channel. The offer is subject to stock availability and may not be applied retrospectively. This entitlement cannot be used in conjunction with any other special offer. Wiley reserves the right to vary the terms of the offer at any time.

Ref: S001W