

Research Article

An Improved Approach for Reduction of Defect Density Using Optimal Module Sizes

Dinesh Verma and Shishir Kumar

Department of Computer Science & Engineering, Jaypee University of Engineering and Technology, Guna, India

Correspondence should be addressed to Shishir Kumar; dr.shishir@yahoo.com

Received 4 April 2014; Revised 25 July 2014; Accepted 8 August 2014; Published 24 August 2014

Academic Editor: Robert J. Walker

Copyright © 2014 D. Verma and S. Kumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, software developers are facing challenges in minimizing the number of defects during the software development. Using defect density parameter, developers can identify the possibilities of improvements in the product. Since the total number of defects depends on module size, so there is need to calculate the optimal size of the module to minimize the defect density. In this paper, an improved model has been formulated that indicates the relationship between defect density and variable size of modules. This relationship could be used for optimization of overall defect density using an effective distribution of modules sizes. Three available data sets related to concern aspect have been examined with the proposed model by taking the distinct values of variables and parameter by putting some constraint on parameters. Curve fitting method has been used to obtain the size of module with minimum defect density. Goodness of fit measures has been performed to validate the proposed model for data sets. The defect density can be optimized by effective distribution of size of modules. The larger modules can be broken into smaller modules and smaller modules can be merged to minimize the overall defect density.

1. Introduction

The reliability of a software product is an important factor, being considered by the developer before the formal release of a product. Defect density (DD) is an important attribute that affects software reliability. Defect prediction is very important for estimation of defect density. There are many methods available that can be used to predict the number of defects in software during testing phases [1]. Estimation of this attribute is one of the approaches used to establish readiness for release. There are several methods that can be used to predict or estimate the value of this parameter. Six sigma methods are used by Fehlmann [2] for advanced prediction of defect density for software that has been developed and its development process is moving further for production. Mark [3] has presented a DevCOP method that has been used for estimation of defect density using verification and validation certificate technique. There are a few crucial factors that have an effect on the initial defect density. Analysis of these factors is important because it provides a quantitative method of identifying possible techniques for reducing the insertion

rate of defects. Further, this analysis can be used to estimate initial defect density that may be used later while planning the required effort for testing. In present scenario many models are available for calculating the defect density. Most studies used multiple numbers of factors in modelling for defect density estimation. The performance of these models increases when a number of factors are added for estimation of defect density [4]. The examples of such type of models are RADC model [5, 6] and ROBUST model [7, 8] that uses this multiplicativity of factors for performance enhancement.

Most of models earlier in the software reliability engineering focused on the complexity of code, skill or experience of developers, and development process. But the overall effect has been affected, with consideration of only these factors, when requirement changes [9]. Earlier studies show that defects that are distributed in a software system are random in nature. So the size of particular module may affect the defect density. Ferdinand [10] study argued that a number of defects are increased with number n of the code segment. Specifically, this theory asserts that the number of defects is proportional to a power of n . This consideration has been used in the

proposed approach to relate the defect density with module size.

2. Related Work

The software design for module of small size is easy as compared to larger module size. So the design of the software may be improved by minimizing the size of the module. But Basili and Perricone [11] mentioned that larger modules had a lower defect density. This observation was true even for the more complex module. This analysis done by the Basili and Perricone with a data set where most of the modules are of small size.

Alsmadi and Najadat [12] proposed techniques to generate a large number of randomly selected modules from tested data sets and make a comparison in every cycle between two modules where one is faulty and the other is not. This comparison can be used to estimate the defect density due to identified faulty modules.

The study by Withrow [13] shows that the defect density is decreasing with increasing size of modules up to a certain point and after that it increases with an increase in the size of the modules. These results of Withrow analysis support the hypothesis given by Banker and Kemerer [14].

Hatton [15] in his approach proposed logarithmic growth in the number of defects for software product of the size up to 200 lines of code with reference to Withrow's data. Malaiya and Denton [16] have evaluated the argument of Rosenberg [17] about observed decreases in defect density with respect to rising in module sizes. They have observed that it is misleading and proposed a model that confirms such observations.

Malaiya and Denton [16] proposed a model that takes into account both trends, that is, decreasing and rising trend of defect density. They have given two different regions for defect density. They had also applied it to the actual data to obtain parameter values.

In the next section a model proposed by Malaiya and Denton [16] for defect density has been considered and a modified model, with two additional parameters, has been proposed, which can be used to examine the effect of module size distribution of the overall defect density.

3. Proposed Model

In this paper two mechanisms have been considered which are responsible for categorizing the faults in the software system. In the present scenario most of the software system is comprised of modules and modules themselves are comprised of instructions [16]. Using these two considerations a composite model has been proposed by Malaiya and Denton [16] with using some parameters to control the defect density. In this model only the relation of defect density with the module size is established, but this model is not able to give the parameters for updating the module size in terms to reduce the defect density. So there is a need to add some parameters in the model that give idea for changes in the size of modules for minimizing the defect density. In this paper

the model proposed by Malaiya is modified by considering the argument given by Ferdinand [10] and adding two other parameters to increase the reduction in the defect density rate of software systems. The first category of faults may be related to those modules that occur due to interaction with others modules. The second category of faults may be an instruction-related fault that occurs due to incorrect implementation of individual module. In the proposed model defect density for module-related fault will be denoted by D_m and for instruction-related fault the defect density will be denoted by D_i and the overall defect density will be denoted by D .

First Category of Faults. These types of faults are associated with parameter passing among the modules. Some of the faults of first category may be related to global data as well as assumptions made by modules to one another. For such types of faults, uniform distribution among the modules has been considered. In terms of defect density, defects may be defined as the overheads that decline proportionately with increase in module size.

Defect density D_m for module-related faults with size s is given by

$$D_m(s) \propto \frac{1}{s^l}, \quad (1)$$

$$D_m(s) = \frac{\alpha}{s^l}, \quad (2)$$

where l is any nonzero integer that is added in the model to exponentially reduce the defect density and α is proportionality constant. In (2) the minimum possible value of s is one and α is a suitable parameter. The model presented with (2) seems consistent with the model proposed by Shen et al. [18].

A Second Category of Faults. This category represents the faults that occurred due incorrect implementation of instruction or module. So these faults can be treated as instruction faults. For obtaining the defect density with respect to these types of fault, first the probability of an incorrect instruction, which has two components, should be considered. The first component has been assumed as a constant that is β and the other component depends on the interaction among instructions. Thus, the second component is proportional to the module size s^n . So the defect density D_i due to instruction-related defects can be expressed as follows:

$$D_i = \beta + \gamma s^n, \quad (3)$$

where γ is another proportional parameter and n is any nonzero integer added to exponentially reduce the defect density. From (2) and (3), the total defect density can be given as follows:

$$\begin{aligned} D(s) &= D_m(s) + D_i(s) \\ &= \frac{\alpha}{s^l} + \beta + \gamma s^n. \end{aligned} \quad (4)$$

Equation (4) represents the total amount of defect density that occurs due to module-related faults and instruction-related faults, where $D(s)$ is dependent variable and s is independent variable.

From the above equation, it is obvious that total defect density can be optimized with respect to module size s , and for minimum defect density the concept of maxima and minima (based on calculus) may be used under the assumption that the given equation is well defined and differentiable within its domain.

Now for a finding the minimum value of defect density, differentiating (4), with respect to s , gives

$$\frac{d}{ds}D(s) = \frac{-l\alpha}{s^{l+1}} + n\gamma s^{n-1}. \quad (5)$$

For maxima and minima equate (5) to zero and put $(d/ds)D(s) = 0$. Consider the following:

$$\frac{-l\alpha}{s^{l+1}} + n\gamma s^{n-1} = 0, \quad (6)$$

$$s = \left(\frac{l\alpha}{n\gamma} \right)^{1/(l+n)}. \quad (7)$$

Equation (7) gives the minimum value of size s . To calculate the minimum defect density, with respect to this minimum size of modules, takes the second derivative of (5). Consider the following:

$$\frac{d^2}{ds^2}D(s) = \frac{l(l+1)\alpha}{s^{l+2}} + n(n-1)\gamma s^{n-2}. \quad (8)$$

Substituting value of s from (7) in (8), it gives

$$\frac{d^2}{ds^2}D(s) = \frac{l(l+1)\alpha}{(l\alpha/n\gamma)^{(l+2)/(l+n)} + n(n-1)\gamma \left(\frac{l\alpha}{n\gamma} \right)^{(l-2)/(l+n)}}. \quad (9)$$

This is only positive when constants α, β are positive and l, n are nonzero positive integers. So, at the point $s = (l\alpha/n\gamma)^{1/(l+n)}$ the defect density will be minimum under some restriction.

So the minimum defect density is

$$D_{\min}(s) = \frac{\alpha}{(l\alpha/n\gamma)^{l/(l+n)}} + \gamma \left(\frac{l\alpha}{n\gamma} \right)^{n/(l+n)} + \beta. \quad (10)$$

It is also possible to formulate a model for defect density verses module size with more variational parameters but it may become complicated to provide justification for the assumptions made for these parameters.

Special Cases. (i) If the value of $l = 1$ and $n = 1$ is taken, then it gives $D_m(s) = \alpha/s$ and $D_i(s) = \beta + \gamma s$, and total defect density is given by

$$\begin{aligned} D(s) &= D_m(s) + D_i(s) \\ &= \frac{\alpha}{s} + \beta + \gamma s. \end{aligned} \quad (11)$$

Differentiating (11), with respect to s , gives

$$\frac{-\alpha}{s^2} + \gamma = 0. \quad (12)$$

So, the module size s_{\min} for minimum defect density is

$$s_{\min} = \sqrt{\frac{\alpha}{\gamma}}, \quad (13)$$

and the minimum defect density is given by

$$D_{\min} = (2\sqrt{\alpha\gamma} + \beta). \quad (14)$$

(ii) If the value of $l = 2$ and $n = 2$ is taken, then it gives $D_m(s) = \alpha/s^2$ and $D_i(s) = \beta + \gamma s^2$, and total defect density is given by

$$\begin{aligned} D(s) &= D_m(s) + D_i(s) \\ &= \frac{\alpha}{s^2} + \beta + \gamma s^2. \end{aligned} \quad (15)$$

So, the module size s_{\min} for minimum defect density is given by

$$s_{\min} = \left(\frac{\alpha}{\gamma} \right)^{1/4}, \quad (16)$$

and the minimum defect density is given by

$$D_{\min}(s) = \frac{\alpha}{(\alpha/\gamma)^{1/2}} + \gamma \left(\frac{\alpha}{\gamma} \right)^{1/2} + \beta. \quad (17)$$

(iii) If the value of $l = 2$ and $n = 1$ is taken, then it gives $D_m(s) = \alpha/s^2$ and $D_i(s) = \beta + \gamma s$, and total defect density is given by

$$\begin{aligned} D(s) &= D_m(s) + D_i(s) \\ &= \frac{\alpha}{s^2} + \beta + \gamma s, \end{aligned} \quad (18)$$

$(d/ds)D(s) = 0 \Rightarrow (-2\alpha/s^3) + \gamma = 0 \Rightarrow (-2\alpha/s^3) = -\gamma \Rightarrow s^3 = (2\alpha/\gamma) \Rightarrow s = (2\alpha/\gamma)^{1/3}$, and under some suitable constraints over α, β , and γ . The point $s_{\min} = (2\alpha/\gamma)^{1/3}$ will give minimum point. Consider the following:

$$\Rightarrow D_{\min}(s) = \frac{\alpha}{(2\alpha/\gamma)^{2/3}} + \beta + \gamma \left(\frac{2\alpha}{\gamma} \right)^{1/3}. \quad (19)$$

(iv) If the value of $l = 1$ and $n = 2$ is taken, then it gives $D_m(s) = \alpha/s$ and $D_i(s) = \beta + \gamma s^2$, and total defect density is given by

$$\begin{aligned} D(s) &= D_m(s) + D_i(s) \\ &= \frac{\alpha}{s} + \beta + \gamma s^2, \end{aligned} \quad (20)$$

$(d/ds)D(s) = 0 \Rightarrow (-\alpha/s^2) + 2s\gamma = 0 \Rightarrow s^3 = (\alpha/2\gamma) \Rightarrow s = (\alpha/2\gamma)^{1/3}$, and under some suitable constraints over

TABLE 1: Parameter values with special case (i).

Data	Parameter values			Optimal size of module	Minimum defect density
	α	β	γ		
Basili	-6.136	18.384	-0.0489	—	—
Withrow	36.41	4.13	0.0009	203.388	4.486
PDR	704.71	0.125	0.00005	3760.904	0.496

TABLE 2: Parameter values with special case (ii).

Data	Parameter values			Optimal size of module	Minimum defect density
	α	β	γ		
Basili	6960.4	13.6	-0.0001	—	—
Withrow	2253.3	4.544	0.000000154	332.461334	4.574212
PDR	98450	0.5499	7E - 10	3443.73259	0.566503

α , β , and γ . So, $s_{\min} = (\alpha/2\gamma)^{1/3}$ is a point of minima and the minimum value is given by

$$D_{\min}(s) = \frac{\alpha}{(\alpha/2\gamma)^{1/3}} + \beta + \gamma \left(\frac{\alpha}{2\gamma} \right)^{2/3}. \quad (21)$$

Under some restricted conditions of constants α , β , and γ this type of model can be classified into two distinct regions [16]. In the first region defect density is decreasing with increasing rate of module size, and in another region defect density increasing with module size.

4. Analysis of Proposed Model

As discussed in the previous section the model has been formulated to show a relationship between software defect density and size of the module. In this section we have discussed the data collection for the analysis of the proposed model, method used to calculate the parameter value of the two cases, calculation of defect density using the parametric values, and finally optimal size of module for minimum defect density.

4.1. Data Collection. The proposed model has been analyzed with respect to the two distinct values for $l = 1$, $n = 1$ and $l = 2$, $n = 2$. The proposed model can also be analyzed with other distinct values for l and n , but in this paper only two special cases as case (i) and case (ii) have been used to validate the model. In this paper three different data sets have been used to analyze the model. The data made available by Basili and Perricone [11] has been considered as the first data set for analysis. This data set only comprises small sizes of modules. There are very few number of modules that are larger than 200 in size. The data set is given in Table 1. The second data set, given by Withrow [13], has been used for analysis. This data set comprises larger module sizes and is given in the Table 2. The third data set has been taken from the declining defect density that is noticed only in the small size of modules. The third data set has been collected from the PDR (Promise data repository) [19]. This data set has been collected from the 23 closed sources projects in which number of modules,

TABLE 3: Basili data [11].

S. number	Size (LOC)	Modules	DD		
			Observed	Case (i)	Case (ii)
1	0-50	258	16	15.817	16.084
2	51-100	70	12.6	13.433	13.096
3	101-150	26	12.4	11.008	11.209
4	151-200	13	7.6	8.574	8.794
5	>200	3	6.4	6.135	6.211

size, and defect density are available. This data set is given in Table 3.

4.2. Parameter Calculation. In the previous section we have discussed the three data sets that have been used to analyze the proposed model. To calculate the parameter values for α , β , and γ , curve fitting method has been used. The parameter values are obtained for all three data sets for two distinct values of l and n as in the case (i) and case (ii). The first data set comprises only small size of modules, so according to the model, parameter γ plays a very little role. This data set does not show any values in the second region where the growth of defect density is seen with increasing size of modules. Only the first region is seen where the defect density decreases with increasing size of the module. So the optimal size of modules and minimum defect density field in the Tables 1 and 2 of the first data set are kept blank. The parameter values are given in Table 1 in case (i) and in Table 2 for case (ii). The optimal values for size of module in case (i) have been calculated using (13); the minimum defect density with respect to this optimal size was calculated using (14) and is given in Table 1. Similarly, the optimal values for size of module and minimum defect density with respect to this size for case (ii) are given in Table 2.

4.3. Calculation of Defect Density. After calculating the parameter value for two cases, defect densities with respect to proposed model have been calculated. Finally, calculated values of defect density for two special cases are compared with the observed values of defect density using the graph.

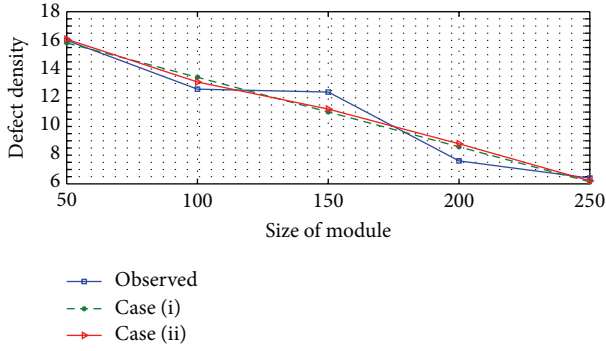


FIGURE 1: Defect density variation of Basili data with case (i) and case (ii).

TABLE 4: Withrow data [13].

S. number	Size (LOC)	Modules	DD		
			Observed	Case (i)	Case (ii)
1	4–62	93	5.4	4.765	5.112
2	64–97	39	4.9	4.584	4.771
3	103–154	52	3.4	4.503	4.638
4	161–250	53	1.8	4.501	4.589
5	251–397	46	5.2	4.580	4.583
6	402–625	31	5.6	4.755	4.611
7	651–949	22	6.8	5.066	4.700
8	1050–5160	26	8.3	5.240	4.767

The calculated values of defect density for the data made available by Basili and Perricone [11] are presented in Table 3 and plotted in Figure 1.

The analysis of decreasing defect density trend with increasing size of modules has been done on the data made available by Basili and Perricone [11]. It indicates the relationship between observed data and fitted values for case (i) and case (ii).

From the graph it can be inferred that there is no region where the growth of defect density with increasing size of module possible has been seen. Only the first region is seen where the defect density decreases with increasing size of the module. The case (i) and case (ii) give more linear relation between defect density and module size than the observed data. In case (ii), geometric progressive sizes are used that give minimum defect density with respect to the size.

In the data set given by Basili and Perricone [11], no larger modules are used, so another data set given by Withrow [13] that comprises most of larger size modules has been taken for analysis. Due to these larger size modules both regions can be seen. These data are mentioned in Table 4, for Ada modules with calculated values of defect density and plotted in Figure 2.

The analysis of growth in defect density trend with increasing size of larger modules has been done on the data made available by Withrow [13]. It shows both trends indicated by the proposed model. These trends have been seen in both special cases.

TABLE 5: PDR data set [19].

S. number	Size (LOC)	Modules	DD		
			Observed	Case (i)	Case (ii)
1	1–132	30197	5.6	5.470	6.200
2	133–465	5209	1.39	1.664	1.005
3	466–1263	1065	0.77	0.746	0.613
4	1266–2927	222	0.66	0.512	0.567
5	2928–9878	40	0.61	0.690	0.619

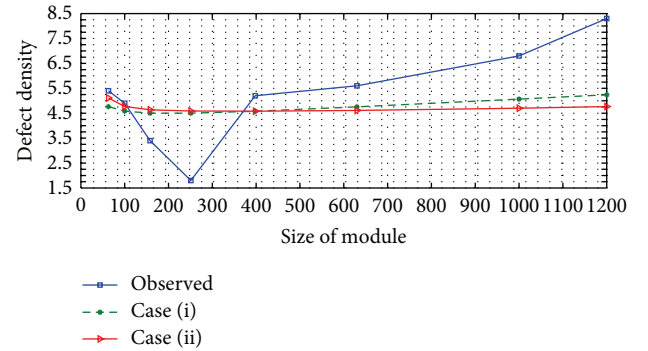


FIGURE 2: Defect density variation of Withrow data with case (i) and case (ii).

This increasing trend in defect density for larger modules has been identified. It may happen due to possibilities that large modules may not be tested too thoroughly in comparison to the smaller modules, resulting in relatively higher defect density in larger models. From the graph it is observed that the case (ii) gives better results in terms of minimization of defect density with increasing size of modules.

The third data set has been collected from the PDR (Promise data repository) [19]. These data with calculated values of defect densities in case (i) and case (ii) are presented in Table 5 and plotted in Figure 3.

This data set is also showing the two regions for decreasing and increasing defect density trends. The analysis for these two trends has been done on the data collected from the PDR and calculated values for two special cases.

From Figure 3, it can be inferred that case (ii) gives improved results in comparison to case (i). Both regions are seen in the graph. The analysis results for this data set show very small variation in defect density, because of the very large size of modules. It infers that once the size of module increases with geometric progression, the increasing rate of defect density reduces with size of the module.

4.4. Goodness of Fit. In the previous section, defect densities for case (i) and case (ii) have been calculated using the proposed model. After obtaining the values of defect density, the fitted values must be compared in terms of fitness of good measure. To perform goodness of fit measure, SPSS statistical analysis tool [20] has been used for calculating the different statistics measures. The sum of squares due to error (SSE) statistics measures the total deviation of the response values from the fit to the response values. A value of this

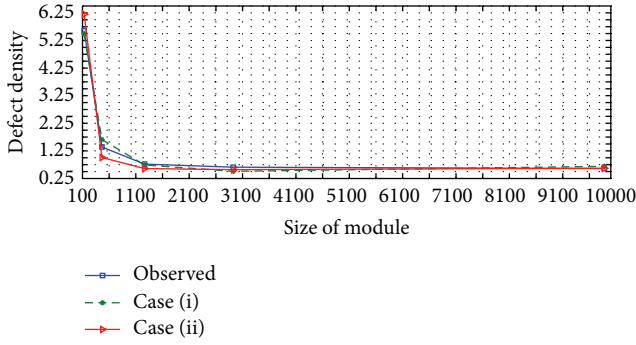


FIGURE 3: Defect density variation of PDR data with case (i) and case (ii).

TABLE 6: Goodness of fit measure statistics.

Data set	Case	Measure statistics		
		MSE	R^2	P value
Basilli	case (i)	0.238	0.697	0.058
	case (ii)	0.227	0.712	0.043
Withrow	case (i)	0.22	0.437	0.056
	case (ii)	0.029	0.529	0.041
PDR	case (i)	0.939	0.273	0.005
	case (ii)	0.992	0.102	0.000

measure closer to 0 indicates that the model has a smaller random error component, and the fit will be more useful for prediction. The R -square (R^2) statistics measure indicates how successful the fit is in explaining the variation of the data. This measure can take on any value between 0 and 1, with closer to 1 indicating that most of the proportion of the variance is accounted for the model. For example, an R^2 value 0.59 means that the predictors explain the 59% variation in the dependent variable. The P value is the estimated probability of rejecting the null hypothesis of the study question. If the P value is less than 0.05, then the null hypothesis can be rejected. In this study, the null hypothesis can be considered as “there is no goodness of fit between the observed value of defect density and fitted values of defect density.” Table 6 shows these measures for the all three data sets using case (i) and case (ii).

From the Table 6, it is inferred that, in all three data sets, case (ii) gives more goodness of fit measured with the observed data set.

5. Performance Evaluation

The model proposed by Malaiya and Denton [16] suggested a relation of module size with defect density in SDLC. In this model the modules are sized with small ranges, whereas in the model proposed through this paper, the ranges of module size are increased by using two additional parameters with integer power of module size in comparison with Malaiya’s model. This proposed approach has been mentioned in (4). On the basis of analysis using proposed model with the data sets (two data sets as used by Malaiya and Denton [16] and data set

from PDR [19]), improvement in terms of minimizing the defect density with optimum module size has been observed compared to Malaiya’s model. The graph shown in Figure 1 indicates closer results with the observed defect density when used with integer 2 power for sizes of modules. Similarly, graph in Figure 2 indicates improvised optimization in the defect density when used with the power of 2 of sizes of modules. The more improvised optimization can be seen in Figure 3 with data sets collected from the PDR with a power of 2 of sizes of modules. Furthermore, when the power of module size will be increased, the proposed model will likely to provide better results.

6. Conclusions and Future Work

Through this paper a model has been proposed that depicts the impact of module size on the defect density. Defect density reduces when the size of the module is reduced and gradually rises after an optimal size of the module. In this paper the module size has been considered in geometric progression. Two additional parameters have been added to the contemporary model proposed by Malaiya et al. Analysis of the proposed model has been performed with the data set as used by the Malaiya et al. and data set collected from the PDR. Analysis with two special cases reflects an optimization of defect density with respect to original observed defect density. Through proposed model an optimal size of the module has also been identified that may be used for effective distribution of module sizes during SDLC. A condition for optimal distribution of module sizes has also been identified in this paper. With the improvised approach presented through this paper, it may also be concluded that the optimization of defect density may be achieved by effective distribution of module sizes. The larger modules can be broken into smaller modules and very small modules can be merged together to optimize the overall defect density during SDLC. In later stages, the implementation of proposed model can also be planned with more variables with different values that may lead to better feasible solution. In all such cases the trade-offs with the complexity of the system should also be taken care off. The implication for minimization of defect density can be further analyzed to investigate the rate of failure of software.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] M. Staron and W. Meding, “Predicting weekly defect inflow in large software projects based on project planning and test status,” *Information and Software Technology*, vol. 50, no. 7-8, pp. 782–796, 2008.
- [2] T. Fehlmann, “Defect density prediction with six sigma,” in *Proceedings of the 6th Software Measurement European Forum*, Rome, Italy, 2009.

- [3] S. Mark, "Utilizing verification and validation certificates to estimate software defect density," in *Proceedings of the 10th European Software Engineering Conference Held Jointly with 13th ACM SIGSOFT International Symposium on Foundations of Software Engineering*, pp. 381–384, 2005.
- [4] M. Takahashi and Y. Kamayachi, "An empirical study of a model for program error prediction," in *Proceedings of 8th International IEEE Conference on Software Engineering*, pp. 330–333, August 1985.
- [5] J. C. Munson and T. M. Khoshgoftar, "Software metrics in reliability assessment," in *Handbook of Software Reliability Engineering*, M. R. Lyu, Ed., IEEE-CS Press/McGraw-Hill, 1996.
- [6] W. Farr, "Software reliability modeling survey," in *Handbook of Software Reliability Engineering*, M. R. Lyu, Ed., IEEE-CS Press/McGraw-Hill, 1996.
- [7] Y. K. Malaiya and J. Denton, "What do the software reliability growth model parameters represent?" in *Proceedings of the 8th International Symposium on Software Reliability Engineering (ISSRE '97)*, pp. 124–135, Albuquerque, NM, USA, November 1997.
- [8] N. Li and Y. K. Malaiya, "ROBUST: a next generation software reliability engineering tool," in *Proceedings of the 6th International Symposium on Software Reliability Engineering*, pp. 375–380, October 1995.
- [9] Y. K. Malaiya and J. Denton, "Requirements volatility and defect density," in *Proceedings of the 10th International Symposium on Software Reliability Engineering*, pp. 285–294, Boca Raton, Fla, USA, November 1999.
- [10] A. E. Ferdinand, "A theory of system complexity," *International Journal of General System*, vol. 1, no. 1, pp. 19–33, 1974.
- [11] V. R. Basili and B. R. Perricone, "Software errors and complexity," *Communications of the ACM*, vol. 27, no. 1, pp. 42–52, 1984.
- [12] I. Alsmadi and H. Najadat, "Evaluating the change of software fault behavior with dataset attributes based on categorical correlation," *Advances in Engineering Software*, vol. 42, no. 8, pp. 535–546, 2011.
- [13] C. Withrow, "Error density and size in Ada software," *IEEE Software*, vol. 7, no. 1, pp. 26–30, 1990.
- [14] R. D. Banker and C. F. Kemerer, "Scale economies in new software development," *IEEE Transactions on Software Engineering*, vol. 15, no. 10, pp. 1199–1205, 1989.
- [15] L. Hatton, "Reexamining the fault density-component size connection," *IEEE Software*, vol. 14, no. 2, pp. 89–97, 1997.
- [16] Y. K. Malaiya and J. A. Denton, "Module size distribution and defect density," in *Proceedings of the 11th International Symposium on Software Reliability Engineering (ISSRE '00)*, pp. 62–71, October 2000.
- [17] J. Rosenberg, "Some misconceptions about lines of code," in *Proceedings of the 4th International Software Metrics Symposium*, pp. 137–142, November 1997.
- [18] V. Y. Shen, T. Yu, S. M. Thebaut, and L. R. Paulsen, "Identifying error-prone software—an empirical study," *IEEE Transactions on Software Engineering*, vol. 11, no. 4, pp. 317–324, 1985.
- [19] <http://promisedata.org/>.
- [20] SPSS, IBM SPSS Statistics Version 20 64-bit, <http://www.spss.com/statistics>.

