Quantitative OSS Project Assessment Based on Process Capability Index

Shigeru Yamada¹, Yoshinobu Tamura²

¹Tottori University
Minami 4-101, Koyama, Tottori-shi, 680-8552, Japan
¹yamada@tottori-u.ac.jp
²Tokyo City University
Tamazutsumi1-28-1, Setagaya-ku, Tokyo 158-8557, Japan
²tamuray@tcu.ac.jp

Abstract: This paper focuses on the process capability index considering the irregular fluctuation of the operation of OSS. Recently, many open source software (OSS) are developed under various open source projects. Considering the characteristics of OSS, many OSS are used via the internet network, such as the software update service, software version upgrade, cloud service, HTTP server, mobile devices, etc. From the above-mentioned characteristics of OSS, the debugging process and operating phase of OSS development will take an irregular fluctuation in the long term of operation. In particular, we apply Wiener process model to the noisy cases in the operation of OSS. Moreover, we compare the method of process capability assessment based on Wiener process model considering the irregular fluctuation for OSS projects with several nonhomogeneous Poisson process (NHPP) models. Also, we analyze actual data to show numerical examples of the proposed method considering the characteristics of recent OSS projects and the existing NHPP models.

Keywords: Process capability index, open source software (OSS), reliability, software reliability model, Wiener process

I. INTRODUCTION

The research paper in terms of open source software (OSS) have been proposed by several researchers[1-3]. However, the research focused on the usage environment via internet network have been not proposed. In particular, the method of OSS reliability assessment considering the external factors from the triggers by internet network have been not presented.

At present, many OSS have been used in company, government, university, etc., because the OSS are helpful for various users to make cost reduction, standardization, quick delivery. In particular, the OSS are frequently used via the internet network. For example, OpenStack and Hadoop are well known as the software for cloud computing and big data targeted database. Also, many OSS will be upgraded according to the release of new software version. Moreover, Apache HTTP server is used as a server software.

Moreover, a difference in the levels of developer skill will have an affect on the software quality and the software debugging activity. In terms of the software quality, the methods of software reliability assessment based on the software reliability growth models have been proposed several researchers [1-3]. Also, several methods of reliability assessment for the OSS have been also proposed[4]. However, the method of OSS reliability assessment considering the external factors from the triggers by internet network have been not proposed, because it is difficult to comprehend the external factors such as the network environment, the difference in developer skills, and the variety of fault reporters.

In this paper, we discuss on the process capability index considering the difference between the stochastic differential equation model and nonhomogeneous Poisson process (NHPP). Then, we consider the effectiveness of OSS reliability assessment method considering irregular fluctuation resulting from the external factors of usage environment. In particular, a reliability assessment method based on Wiener process model is developed in order to comprehend the external factors in this paper. Also, a set of actual software fault-count data is analyzed in order to show numerical examples of application of software reliability analysis for the OSS. Then, we show the comparison results of process capability index in terms of the stochastic differential equation model and NHPP model.

II. SOFTWARE RELIABILITY MODELS

Considering the characteristic of the operation phase of OSS projects, the software fault-reporting phenomena keep an irregular state in the operation phase, because the following reasons:

- There is variability among the levels of developer skill, because the OSS are developed by several developers and users. Then, the progress of debugging process becomes unstable.
- In recent years, the OSS are used via internet network. Then, the operation phase of OSS depends on the unstable condition of network environment.
The operation phases of many open source projects are influenced from external factors such as above mentioned triggers. Considering the above points, we apply the noisy model to the OSS project.

Fig. 1. The development paradigm of OSS.

We apply the existing Wiener process model to the operational phase of open source project. Then, let \( N \) be the cumulative number of detected faults up to operational time \( t \) \((t \geq 0)\) in the open source development project. Suppose that \( N \) takes on continuous real values. Since latent faults are detected and eliminated during the operational phase of the OSS project, \( N \) gradually increases as the operational procedures go on. Thus, under common assumptions for software reliability growth modeling\[2\], the following linear differential equation can be formulated:

\[
\frac{dN(t)}{dt} = b(t)(a - N(t)), \tag{1}
\]

where \( b \) is the software fault-detection rate at operational time and a non-negative function, \( a \) means the latent faults prior to operation.

Generally, the official version, beta version, and alpha version are developed under the OSS project as shown in Fig.1. We assume that the proposed method is applied to the official version. As an example, the latent faults prior to operation is \( N_{\text{v}} \) of the version \( V \) in Fig.1. Therefore, we can obtain from Eq.(1) by using Ito's formula\[5-8\]:

\[
N(t) = a \left[ 1 - \exp \left( - \int_0^t b(s) ds - \sigma \omega(t) \right) \right], \tag{2}
\]

where \( a \) is a positive constant representing a magnitude of the irregular fluctuation, \( \omega \) one-dimensional Wiener process.

Moreover, we define the software fault-detection rate per fault in case of \( b \) defined as:

\[
\int_0^t b_\omega(s) ds = \lim_{a \to 0} \frac{dN_\omega(t)}{dx} = b, \tag{3}
\]

where \( N_\omega(t) = a(1 - e^{-c(t)}) \) means the mean value functions for the exponential SRGM based on NHPP\[2\], the expected number of latent faults prior to operation for SRGM, and the fault detection rate per fault.

Therefore, the cumulative numbers of detected faults up to time \( t \) are obtained as follows:

\[
N(t) = a[1 - \exp(-bt - \sigma \omega(t))]. \tag{4}
\]

In this model, we assume that the parameter \( \sigma \) depends on several noises by external factors from several triggers in OSS projects.

Moreover, we compare the Wiener process model with the existing NHPP models. We consider the following exponential type and S-shaped type models as the compared models.

\[
E(t) = a[1 - \exp(-bt)], \tag{5}
\]

\[
S(t) = a[1 - (1 + bt)\exp(-bt)], \tag{6}
\]

where \( E(t) \) is the expected number of latent faults prior to operation for SRGM, and \( S(t) \) the fault detection rate per fault.

III. PROCESS CAPABILITY INDEX

The process capability index is well known as the measurement of product standards in the area of statistical quality control\[9\]. In particular, it is important for the project managers to understand the product quality in operation phase of the OSS product developed under the open source project. In this paper, we use the process capability index as the measurement of software product quality in the operation of OSS.

In particular, the process capability index is based on the normal distribution. Then, we can use a software reliability model based on the stochastic differential equation because the solution process of Wiener process model takes on the normal distribution.

Generally, the process capability index is given by

\[
C_p = \frac{U_{sl} - L_{sl}}{3\sigma}, \tag{7}
\]

where \( C_p \) is called as the process capability index. \( U_{sl} \) and \( L_{sl} \) in Eq.(5) mean the upper and lower specification limits, respectively. Moreover, we consider the time-dependent process capability index. The time-dependent process capability index in terms of cumulative numbers of detected faults based on the Wiener process model is given by the following equation.

\[
C_p(t) = \frac{U_{sl} - L_{sl}}{6 \sqrt{\frac{\sigma}{\int_{\omega_{t}}^t N_{\omega}(t) dt}}}, \tag{8}
\]

where the variance of \( N_\omega \) is derived from the exponential SRGM based on the stochastic differential equation as follows:

\[
\text{Var}[N_\omega(t)] = E[N_\omega(t) - \{N_\omega(t)\}^2] = a^2 \exp(-2bt + \sigma^2t)(\exp(\sigma^2t) - 1). \tag{9}
\]

From the range \( U_{sl} - L_{sl} \) of upper and lower specification limits, each specification limit is given by

\[
C_{p,upper}(t) = \frac{U_{sl} - \mu}{3\sigma} \tag{10},
\]

\[
C_{p,lower}(t) = \frac{L_{sl} - \mu}{3\sigma} \tag{11},
\]
where \( \mu \) is the mean. Considering the time-dependent process capability index based on Wiener process model, each specification limit is given by
\[
C_{p,\text{upper}}(t) = \frac{ \mu_{\text{spec}} - \mu_{\text{spec}}(t) }{ 3 \sigma_{\text{spec}}(t) },
\]
\[
C_{p,\text{lower}}(t) = \frac{ \mu_{\text{spec}}(t) - \mu_{\text{spec}} }{ 3 \sigma_{\text{spec}}(t) },
\]
where the expected value of \( N_{\text{spec}}(t) \) is derived from the exponential SRGM based on the stochastic differential equation model in the following:
\[
E[N_{\text{spec}}(t)] = a \left[ 1 - \exp \left( -bt + \frac{\sigma^2}{2} t \right) \right].
\]
Similarly, \( C_p \) and \( C_{p,\text{upper}} \) in case of NHPP models can easily obtained from Eqs.(5) and (6).

\[
C_{p,\text{lower}}(t) = \frac{ \mu_{\text{spec}}(t) - \mu_{\text{spec}} }{ 3 \sigma_{\text{spec}}(t) }.
\]

Fig. 2. The estimated process capability index in case of exponential NHPP model.

Fig. 3. The estimated process capability index.

Fig. 4. The estimated process capability index in case of S-shaped NHPP model.

Fig. 5. The estimated upper and lower specification limits in case of S-shaped NHPP model.

Fig. 6. The estimated upper and lower specification limits in case of exponential NHPP model.
Fig. 7. The estimated upper and lower specification limits.

Fig. 8. The estimated sample path of upper and lower specification limits.

IV. NUMERICAL EXAMPLES

We focus on the Apache HTTP Server[10] known as the OSS developed under Apache Software Foundation. In particular, this section utilizes the fault data of version 2.2 in Apache HTTP server. The development period of version 2.2 in Apache HTTP server is from December 2005 to July 2017. The fault-count data used in this paper are obtained from the bug tracking system on the website of Apache Software Foundation.

Fig.2 shows the estimated process capability index of exponential NHPP model in Eq.(5).Fig.1 shows the estimated process capability index of exponential NHPP model in Eq.(5).In Fig. 2, the vertical axis means $C_p$, the horizontal axis is the range $(U_{sl} -)$ of upper and lower specification limits. As the indication of the quality, the estimated process capability index is greater than equal 1.33, i.e., $C_p(t) \geq 1$. From Fig.2, we found that the range $(U_{sl} -)$ of upper and lower specification limits is better in case of 150 or 200 through the operating phase. Then, the estimated one side specification limit in case of $U_{sl} = L_{sl} = 1$ show the Fig.3. From Fig. 3, we have found that the estimated one side specification limit tends to decrease with time procedures go on in case of the exponential NHPP model.

Similarly, Fig.4 shows the estimated process capability index of S-shaped NHPP model in Eq. (6). Fig.4 shows the estimated process capability index of S-shaped NHPP model in Eq.(6).In Fig. 4, the vertical axis means $C_p$, the horizontal axis is the range $(U_{sl} -)$ of upper and lower specification limits. As the indication of the quality, the estimated process capability index is greater than equal 1.33, i.e., $C_p(t) \geq 1$. From Fig.4, we have found that the range $(U_{sl} -)$ of upper and lower specification limits is better in case of 150 or 200 through the operating phase. Then, the estimated one side specification limit in case of $U_{sl} = L_{sl} = 1$ show the Fig.5. From Fig. 5, we have found that the estimated one side specification limit tends to converge in case of the exponential NHPP model.

Similarly, Fig.6 shows the estimated process capability index of S-shaped NHPP model in Eq.(14). Fig.6 shows the estimated process capability index of S-shaped NHPP model in Eq.(14).In Fig. 6, the vertical axis means $C_p$, the horizontal axis is the range $(U_{sl} -)$ of upper and lower specification limits. As the indication of the quality, the estimated process capability index is greater than equal 1.33, i.e., $C_p(t) \geq 1$. From Fig.6, we have found that the range $(U_{sl} -)$ of upper and lower specification limits is better in case of 150 or 200 through the operating phase. Then, the estimated one side specification limit in case of $U_{sl} = L_{sl} = 1$ show the Fig.7. In particular, we can show the sample path of estimated one side specification limit by using the model based on Wiener process. The sample paths of estimated one side specification limit show in Fig.8. Considering the amount of noise in Fig.8, the case of $U_{sl} = L_{sl} = 1$ is better throughout the whole of operating phase, i.e., $C_p(t) \geq 1$ in case of $U_{sl} = L_{sl} = 1$ considering the noise. Therefore, the software managers will take notice of the gap between 125 faults from the estimated cumulative numbers of detected faults in this case.

We discuss the process capability index in case of the stochastic differential equation model. Fig.6 shows the estimated process capability index of S-shaped NHPP model in Eq.(14). Fig.6 shows the estimated process capability index of S-shaped NHPP model in Eq.(14).In Fig. 6, the vertical axis means $C_p$, the horizontal axis is the range $(U_{sl} -)$ of upper and lower specification limits. As the indication of the quality, the estimated process capability index is greater than equal 1.33, i.e., $C_p(t) \geq 1$. From Fig.6, we have found that the range $(U_{sl} -)$ of upper and lower specification limits is better in case of 150 or 200 through the operating phase. Then, the estimated one side specification limit in case of $U_{sl} = L_{sl} = 1$ show the Fig.7. In particular, we can show the sample path of estimated one side specification limit by using the model based on Wiener process. The sample paths of estimated one side specification limit show in Fig.8. Considering the amount of noise in Fig.8, the case of $U_{sl} = L_{sl} = 1$ is better throughout the whole of operating phase, i.e., $C_p(t) \geq 1$ in case of $U_{sl} = L_{sl} = 1$ considering the noise. Therefore, the software managers will take notice of the gap between 125 faults from the estimated cumulative numbers of detected faults in this case.
From the above results, we have found that the proposed method can assist the OSS project managers to understand the gap from the estimated cumulative numbers of faults. It is important for the software managers to understand the range of the prediction. Thereby, the software managers will keep a schedule according to the future planning of OSS projects by using the proposed method.

In particular, the process capability index proposed in this paper is useful for the OSS managers to assess the stability of OSS quality. From the standpoint of process capability index, we have found that the stochastic differential equation model is better than the NHPP model in case of OSS project. We discuss the stability assessment of quality in OSS project. The estimated process capability index is shown in Fig.6. The time point of minimum process capability index satisfied with $C_p(t) \geq 1.33$ is about 50 months in Fig.6. In the case of $U_{sel} = L_{sel} = 200$, the estimated process capability index gradually increases after 50 months. Therefore, the OSS project managers can judge that the version of Apache HTTP Server focused on the numerical example is satisfied with the high stability of OSS quality. On the other hand, we consider the case of $U_{sel} = L_{sel} = 100$, i.e., $U_{sel} = L_{sel} = 100$ assumes the strict conditions. In the case of strict conditions in $U_{sel} = L_{sel} = 100$, the OSS project managers can judge that the Apache HTTP server keeps the stability of OSS quality after 200 months as shown in Fig.9. Therefore, it is identified as suitable for use with after 200 months from the release of this version, if the software manager use the Apache HTTP server of version focused on this numerical example.

V. CONCLUDING REMARKS

This paper has discussed the comparison of OSS reliability assessment method considering the irregular fluctuation from these external influenced operations and the method based on NHPP models. In particular, it is difficult for the OSS project managers to assess the reliability considering the range of the prediction and the gap from the estimated cumulative numbers of faults. From the comparison results in numerical illustrations, the proposed method based on the stochastic differential equation may be useful as the assessment method of the product quality in operation phase of the OSS product. Also, a set of actual software fault-count data has been analyzed to show numerical examples of software reliability analysis for the OSS. In particular, we have compared the method of process capability assessment based on Wiener process model considering the irregular fluctuation for OSS projects with NHPP models of exponential and S-shaped types. From the results, we have found that the method based on stochastic differential equation fits the actual operating situation better than the method of existing NHPP models. The process capability index of OSS, proposed in this paper, will be helpful for the OSS project managers as the measurement of software product quality in the operation of OSS.

ACKNOWLEDGMENT

This work was supported in part by the JSPS KAKENHI Grant No. 15K00102 and No. 16K01242 in Japan.

REFERENCES