Predicting Release Time for Open Source Software based on the Generalized Software Reliability Model

Hironori Washizaki*, Kiyoshi Honda* and Yoshiaki Fukazawa*

* Global Software Engineering Laboratory, Waseda University
3-4-1 Ohkubo, Shijuku-ku, 169-8555 Tokyo, JAPAN
Email: washizaki@waseda.jp, khonda@ruri.waseda.jp, fukazawa@waseda.jp

Abstract—There is a significant challenge that how to predict the possible release date of the target software having enough reliability in agile development where incremental development and small software releases are key characteristics. Existing approaches targeting agile development usually use release backlogs for predicting and setting delivery windows; however these do not consider the reliability of software for release date prediction so that there is a possibility that software at the predicted release date have poor reliability. Previously we proposed a generalized software reliability model (GSRM) based on a stochastic process and compared it with other models to evaluate recent software developments. However, we did not directly evaluate the accuracy of the predicted release time by model. In this paper, towards prediction of release dates in agile development, we focus on the release dates of open source software (OSS) developments and the number of detected issues (faults) since OSS developments comply well with the definition of the agile development in terms of incremental process and frequent releases. We define the accuracy of the predicted release time using the given development terms and the number of issues. Additionally, we propose a method to evaluate the accuracy of the predicted release time. In the best case, GSRM shows only 0.572% Error Rate, which corresponds to a predicted release date of two days prior to the actual release date. We believe that our method should be applicable to agile developments too.

I. INTRODUCTION

Agile development is aimed at minimizing overall risk and encouraging rapid and flexible response to specification changes by using iterative and incremental process model[1]. Incremental development and small software releases are key characteristics of agile software development processes. In such processes, there is a significant challenge that how to predict the possible release date of the target software having enough reliability. Existing approaches such as Cumulative Flow Diagrams targeting agile development[2] usually use release backlogs for predicting and setting delivery windows; however these do not consider the reliability of software for release date prediction so that there is a possibility that software at the predicted release date have poor reliability.

Software reliability is a critical component of computer system availability. Software reliability growth models can be used to indicate whether enough faults have been removed to release the software. Although the logistic curve and Gompertz curve[3] are well-known software reliability growth curves, they cannot account for the dynamics of software development. Development is affected by various elements of the development environment, including the skills of the development team and changes in requirements. Especially in agile developments, there could be various dynamic situations such as changing and prioritizing requirements frequently.

Examples of software reliability models include the "Times Between Failures Models" and "Failure Count Models"[7]; among them we used the "Failure Count Model," which is based on counting failures (issues) and probability methods. The Goel-Okumoto NHPP Model and the Musa Execution Time Model are examples of this type of model[7]. Some of recent studies by Tamura[8], Yamada[9], Cai[10], Kamei[11], Dohi[7], Schneidewind[12], Nguyen[13], and Okamura[14] have attempted to describe the dynamics of developments using stochastic differential equations. Although many models have been proposed, surveyed, and compared[15], [16], [23], most failure count models cannot account for the dynamics of development, such as drastic changes in the development team composition or significant reductions in the development time. And these conventional models cannot precisely predict when developments will end.

To predict the time range that a development will end, here we propose a method employing our model, a generalized software reliability model (hereafter GSRM[5], [6]), which can describe several development situations involving random factors, such as the skills of teams and the development environment. Previous studies[17] have employed only linear stochastic differential equations, but our research indicates that non-linear stochastic differential equations lead to elaborate equations that can model situations more realistically[6].

Moreover, towards prediction of release dates in agile development, we applied our method to several versions in

In our previous papers[4], [5], [6], we proposed GSRM and applied it to several versions of OSS; however we have not evaluated the results in detail. In this paper we suggest a method to predict the release time using data for six months after the first issue is detected by GSRM and another model in detail.
a certain open source software (OSS) development to reveal
the development time frame of typical OSSs. As a result, we
successfully predicted the release dates of versions of OSS
more accurately than a representative conventional approach.

Agile developments and OSS developments share many
principles and values[18]. We believe that our method should
be applicable to agile developments since OSS developments
comply well with the definition of the agile development
in terms of incremental process and frequent releases[19],
and it is indicated that all the agile methods are in essence
applicable to open source software development because of
their iterative and incremental character[20]. Although in some
agile development methods release dates are fixed to some
degree due to fixed length of iterations, our method should be
still beneficial for various purposes such as predicting which
releases will have enough reliability.

This paper aims to answer the following research questions.
RQ1: Is GSRM better than other models (e.g., NHPP)
from the viewpoint of prediction of number of
issues?
RQ2: Is GSRM better than other models (e.g., NHPP)
from the viewpoint of prediction of release dates?

Our contributions are as follows.

- A two-step method to predict the release time of OSS:
  separation of development time periods into different
  versions, and, application of GSRM for prediction.
- A method to evaluate the prediction accuracy in terms of
  release date by defining a Error Rate as a relative amount
  of prediction error of the release date.
- An evaluation result targeting a OSS front-end framework
  “foundation”[21] showing that our prediction method
  works better than NHPP.

The remainder of the paper is organized as follows. Section
II describes our method, while section III evaluates our
method. Section IV discusses related works. Finally, section
V provides a conclusion and future direction.

II. PROPOSED METHOD

In this section, we propose a two-step method to predict the
release time and a method to evaluate the prediction accuracy
in terms of the release time.

A. Prediction of release time

To determine when OSS can be released with respect to the
number of detected faults, we propose the two-step method
described below: 1) and, 2) Using GSRM to predict the
number of faults and the release date.

1) Separation into time periods: The upper graph in
Fig. 1 indicates the number of detected faults for the
"foundation,"[21] which is a OSS front-end framework, di-
vided by each version. The curve shape is sharper when a
newer version is released. Therefore, the versions are separated
based on the changing points before applying our model
(GSRM) because such separation allows the model to more
precisely approximate the data.

2) GSRM: For our software reliability model, we extend a
nonlinear differential equation that describes the fault content
as a logistic curve to an Ito-type stochastic differential equa-
tion. We start with the logistic differential equation, which is
expressed as

\[ dN(t)/dt = N(t)(a + bN(t)) \]  

(1)

\( N(t) \) is the number of detected faults at time \( t \), \( a \) defines
the growth rate, and \( b \) is the carrying capacity. If \( b = 0 \),
then the solutions of this equation are exponential functions.
We extend equation (1) into a stochastic differential equation
because actual developments do not correctly obey equation
(1) due to numerous uncertainties and dynamic changes.

We consider such dynamic elements to be time-dependent
and to contain uncertainty, which are expressed using \( a \).
The time-dependence of \( a \) can be used to describe situations such
as the improved skills of development members and increased
growth rate. The uncertainty of \( a \) can describe parameters such
as the variability of development members and environment.
We analyze the growth of software with an emphasis on
the test phase by simulating the number of detected faults.
We assume that the software development has the following
properties.

- The total number of faults is constant.
- The number of faults that can be found depends on time.
- The number of faults that can be found contains uncer-
tainty that can be simulated with Gaussian white noise.

Considering these properties, we extend equation (1) to an
Ito-type stochastic differential equation with \( a(t) = \alpha(t) + \sigma dw(t) \) as shown below.

\[ dN(t) = (\alpha(t) + \beta N(t))N(t)dt + N(t)\sigma dw(t) \]  

(2)
We use GSRM to predict the release time. For comparison, we also predict the release date using the Non-Homogeneous Poisson Process (NHPP) model[22], [23] since the NHPP model is the most popular one[6]. Herein we assume that the release time is defined as the time where 95% of the maximum number of predicted issues are detected. It should be noted that this definition depends on the development team’s policy.

Fig. 2 shows that the predicted release times driven by GSRM and NHPP model. Herein the predicted release time is defined as the intersection between the model line and the point where 95% of the maximum number of predicted issues is detected. Additionally, we propose a method to evaluate the accuracy of the predicted date using limited datasets of issues and dates. For example, here we define the terms as the data from the date when the first issue was detected to six months later (180 days).

Figure 3 graphically depicts this method where the data in the blue dotted line is used to predict the release time in the red box.

The table indicates that the reliability growth models can be applied to nine types of development situations. Existing models can describe only one of these situations with additional limitations. In contrast, GSRM can describe several of these situations. This is primarily because existing models cannot handle time-dependent growth rates without limitations, whereas GSRM can handle the time-dependence growth rates.

### B. Evaluation of accuracy

We use GSRM to predict the release time. For comparison, we also predict the release date using the Non-Homogeneous Poisson Process (NHPP) model[22], [23] since the NHPP model is the most popular one[6]. Herein we assume that the release time is defined as the time where 95% of the maximum number of predicted issues are detected. It should be noted that this definition depends on the development team’s policy.

Fig. 2 shows that the predicted release times driven by GSRM and NHPP model. Herein the predicted release time is defined as the intersection between the model line and the point where 95% of the maximum number of predicted issues is detected. Additionally, we propose a method to evaluate the accuracy of the predicted date using limited datasets of issues and dates. For example, here we define the terms as the data from the date when the first issue was detected to six months later (180 days).

| $\alpha_1(t) = a_1$ (const.) | $\gamma_1(t) = N(t)\sigma dw(t)$ | The number of issues per unit time is constant and the uncertainty is greater at the start of project than at the end. (Model 1-1) |
| $\alpha_2(t) = a_2(t < t_1)$ | $\gamma_2(t) = \sigma dw(t)$ | The number of issues per unit time changes at $t_1$, and the uncertainty increases near the end (e.g., new members join the project at time $t_1$). (Model 1-2) |
| $\alpha_3(t) \propto t$ | $\gamma_3(t) = 1/N(t)\sigma dw(t)$ | The number of issues per unit time changes at $t_1$ but the uncertainty is constant at any given time. (Model 1-3) |

Table I summarizes the types of $\alpha(t)$, the coefficient of $dw(t)$, and the corresponding situations.

The table indicates that the reliability growth models can be applied to nine types of development situations. Existing models can describe only one of these situations with additional limitations. In contrast, GSRM can describe several of these situations. This is primarily because existing models cannot handle time-dependent growth rates without limitations, whereas GSRM can handle the time-dependence growth rates.

### B. Evaluation of accuracy

We use GSRM to predict the release time. For comparison, we also predict the release date using the Non-Homogeneous Poisson Process (NHPP) model[22], [23] since the NHPP model is the most popular one[6]. Herein we assume that the release time is defined as the time where 95% of the maximum number of predicted issues are detected. It should be noted that this definition depends on the development team’s policy.

Fig. 2 shows that the predicted release times driven by GSRM and NHPP model. Herein the predicted release time is defined as the intersection between the model line and the point where 95% of the maximum number of predicted issues is detected. Additionally, we propose a method to evaluate the accuracy of the predicted date using limited datasets of issues and dates. For example, here we define the terms as the data from the date when the first issue was detected to six months later (180 days).

| $\alpha_1(t) = a_1$ (const.) | $\gamma_1(t) = N(t)\sigma dw(t)$ | The number of issues per unit time is constant and the uncertainty is greater at the start of project than at the end. (Model 1-1) |
| $\alpha_2(t) = a_2(t < t_1)$ | $\gamma_2(t) = \sigma dw(t)$ | The number of issues per unit time changes at $t_1$, and the uncertainty increases near the end (e.g., new members join the project at time $t_1$). (Model 1-2) |
| $\alpha_3(t) \propto t$ | $\gamma_3(t) = 1/N(t)\sigma dw(t)$ | The number of issues per unit time changes at $t_1$ but the uncertainty is constant at any given time. (Model 1-3) |

Table I summarizes the types of $\alpha(t)$, the coefficient of $dw(t)$, and the corresponding situations.

The table indicates that the reliability growth models can be applied to nine types of development situations. Existing models can describe only one of these situations with additional limitations. In contrast, GSRM can describe several of these situations. This is primarily because existing models cannot handle time-dependent growth rates without limitations, whereas GSRM can handle the time-dependence growth rates.

### B. Evaluation of accuracy

We use GSRM to predict the release time. For comparison, we also predict the release date using the Non-Homogeneous Poisson Process (NHPP) model[22], [23] since the NHPP model is the most popular one[6]. Herein we assume that the release time is defined as the time where 95% of the maximum number of predicted issues are detected. It should be noted that this definition depends on the development team’s policy.

Fig. 2 shows that the predicted release times driven by GSRM and NHPP model. Herein the predicted release time is defined as the intersection between the model line and the point where 95% of the maximum number of predicted issues is detected. Additionally, we propose a method to evaluate the accuracy of the predicted date using limited datasets of issues and dates. For example, here we define the terms as the data from the date when the first issue was detected to six months later (180 days).
We evaluated the prediction accuracy using the Error Rate, which is defined as

$$\text{Error Rate} = \frac{\text{Predicted Release Day} - \text{Release Day}}{\text{Release Day}}$$

In equation (3), the Error Rate means the relative amount of prediction error of the release date. The Predicted Release Day is defined as the value where the model detects 95% of the maximum number of predicted faults. The dataset in the model includes data from the date when the first issue was detected to six months later. If the Error Rate is less (greater) than 0, the model predicts a release day earlier (later) than the actual release day.

If one model precisely approximates the release date for one OSS, it may not be accurate for future values. Thus, estimating the accuracy of model predictions is important, which is why we evaluate the prediction accuracy of a model.

III. EVALUATION

To answer the following research questions, we conducted a case study targeting actual development data of OSS ("foundation") in a given time-independent situation obtained from Github site[24], and evaluated our method employing GSRM and general NHPP models.

RQ1: Is GSRM better than other models (e.g., NHPP) from the viewpoint of prediction of number of issues?

RQ2: Is GSRM better than other models (e.g., NHPP) from the viewpoint of prediction of release dates?

The time limitation is necessary because the NHPP model cannot be applied to time-dependent situations. We also compare the predicted numbers of issues driven by NHPP and GSRM at the end of development using six months (180 days) of data.

As GSRM, we chose Model 1-2 (i.e. both of the number of issues per unit time and the uncertainty are constant) among all of nine model types in Table I because of two reasons. Firstly it was hard to identify specific uncertainty type and specific type of dependencies of $\gamma(t)$ on $N(t)$. Secondly we wanted to make the comparison and discussion simple. Comparison with other model types (such as Model 1-3) is one of our future works.

A. Prediction of Number of Issues (RQ1)

Table II[5] shows the number of predicted issues, days of development for each version, the residual sum of squares (RSS), and the Akaikes Information Criterion (AIC) for each model. The results show GSRM is better than NHPP from the viewpoint of prediction as GSRM more precisely predicts the number of detected issues than NHPP although RSS and AIC of GSRM are slightly larger than those of NHPP. However, Table II cannot describe the accuracy of predictions because it shows the qualities of the models and not the release day. In the next section, we evaluate the prediction accuracy of the models.

B. Prediction of Release Date (RQ2)

Table III shows the actual release date and predicted release date by model. The results show that GSRM is superior to NHPP from the viewpoint of accuracy. Hence, error rates and predicted release of GSRM are better than those of NHPP.

C. Threats to Validity

For the case study, we used the actual development data of OSS as it is so that the data could contain inappropriate issue reports or some other false elements such as duplicate reports[25] and a single issue report actually containing multiple different faults. That might affect the internal validity. As our future work, we intend to confirm the validity of the data in detail.

As a threat to external validity, we only tested our method employing GSRM with single OSS, which is insufficient to make generalizations about our method. As our future work, we intend to apply our method to other OSSs. Moreover, we only compared our method with the NHPP model; although other conventional models are similar to the NHPP model, our method should also be compared to them.

IV. RELATED WORK

Power proposed an approach for predicting and setting delivery windows based on release backlogs targeting agile development[2]; the approach does not consider the reliability of software for release date prediction so that there is a possibility that software at the predicted release date have poor reliability.

Many different types of software reliability growth models exist. Yamada et al. proposed an extended NHPP model, which is related to testing-domain[26]. The test-domain dependent
model includes the notion that the tester’s skills should improve by degrees; thus, skills grow over time. The test-domain dependent model adds additional assumptions to the NHPP model. However they did not confirm the approach is useful for OSS developments and/or agile developments.

Typical software reliability models use waterfall development, but Fuji et al. developed a quantitative software reliability assessment method based on the familiar non-homogeneous Poisson processes for incremental development processes[27]. Fuji et al. employed both the number of faults and software metrics to demonstrate the reliability prediction through a case study. Although there method could be applicable to OSS developments and agile developments, metrics measurement results in addition to data of faults are needed; often it is hard to obtain those additional measurement results.

Aman proposed a multistage model that divides the whole development period of OSS into multiple stages, and applies a different growth curve to a different stage[28]. Although its concept is related to our method involving separation of development time periods, target types of data for growth model applications are quite different; its target is code change events while our target is number of issues (faults). We have a plan to compare our method with the multistage model against same issue data.

There is an ongoing challenge to monitor bug-fixing process after releases in OSSs[29]. Our future work could include an investigation of relationship between the bug-fixing process during development and the process after releases in OSSs.

V. CONCLUSION

Using GSRM, we successfully predict the release dates of OSS. Additionally, we propose a method to evaluate the prediction accuracy, which confirms that GSRM can precisely predict the release date. We believe that our method should be applicable to agile developments because of their iterative and incremental character.

This paper is limited to time-independent development situations in order to compare the two models because NHPP cannot handle time-dependent variables. In the future, we plan to adjust the time-dependence of the models, which may allow GSRM to more accurately predict the number of issues detected. Moreover, we plan to apply our method to agile development projects.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of this paper.

REFERENCES