Software Reliability Growth Model for Imperfect Debugging Process Considering Testing-Effort and Testing Coverage

Zang Sicong，*Pi Dechang* *

College of Computer Science and Technology，Nanjing University of Aeronautics and Astronautics，
Nanjing 211106，P. R. China

Abstract：Because of the inevitable debugging lag，imperfect debugging process is used to replace perfect debugging process in the analysis of software reliability growth model. Considering neither testing-effort nor testing coverage can describe software reliability for imperfect debugging completely，by hybridizing testing-effort with testing coverage under imperfect debugging，this paper proposes a new model named GMW-LO-ID. Under the assumption that the number of faults is proportional to the current number of detected faults，this model combines generalized modified Weibull（GMW）testing-effort function with logistic（LO）testing coverage function，and inherits GMW's amazing flexibility and LO's high fitting precision. Furthermore，the fitting accuracy and predictive power are verified by two series of experiments and we can draw a conclusion that our model fits the actual failure data better and predicts the software future behavior better than other ten traditional models，which only consider one or two points of testing-effort，testing coverage and imperfect debugging.

Key words：software reliability；testing-effort；testing coverage；imperfect debugging（ID）；non-homogeneous Poisson process（NHPP）

0　Introduction

With the rapid development of software engineering，software reliability，one of the most important text indexes used for ensuring the reliability of software products during and after software development，is of great concern. During the last thirty years，large numbers of software reliability growth models（SRGMs）are proposed to track the reliability growth trend of the software testing process[1-3]. SRGMs can be used for defects detection，failure rate calculation and failure prediction. Especially，due to the understandable and simple formula，non-homogeneous Poisson process（NHPP）SRGMs are most widely-used.

During the software reliability analysis，large amounts of manpower and CPU hours，which are called testing-effort（TE），are consumed. Obviously，the consumption rate of TE cannot be a constant and it may change the shape of software reliability growth curve. Testing-effort function （TEF），such as logistic（LO）TEF[4]，was established for quantitatively describing the distribution of testing resources. Zhang et al.[5] added finite queuing model into generalized modified Weibull（GMW）TEF with respect to the testing-effort in failure detection process（FDP）and failure correction process（FCP）. Lin[6] proposed a software reliability modeling framework used to gauge the influence of test phase transitions and got a significant effect on fault detection. Pachauri et al.[7] used genetic algorithm（GA）and multi-attribute utility theory（MAUT）to keep fault detection rate as a constant.

Moreover，just like TE，testing coverage

(TC) can also help programmers detect defects. Cai[8] found that there would be some correlations between TC and software reliability. Thus, testing coverage function (TCF) was established for quantitatively describing TC's transformation, such as LO TCF[9]. Li et al. [10] proposed an IS-LO-SRGM combining inflected S-shaped TEF with LO TCF and yielded a wonderful fitting. Chatterjee and Singh[11] incorporated a logistic-exponential TCF with imperfect debugging and the new model's prediction was very close to actual software faults.

Furthermore, a detected fault cannot be corrected immediately and the time required to correct a detected fault is usually called debugging lag/delay[1]. One or several new faults may appear during the fault correcting. In other words, the number of faults is not a constant and it will grow during the detection, which is called the imperfect debugging (ID). Peng[1] and Zhang[12] both built their SRGMs under ID, and the new models had better descriptive and predictive power than the others. Wang et al. [13] proposed a SRGM under ID considering log-logistic distribution fault content function and achieved good performance.

In a word, the introduction of TE, TC and ID, not only makes the SRGMs reliable, but also improves the defect detection rate. However, little research has been conducted to SRGMs combining ID with both TE and TC. Thus, it is desired to put forward an ID dependent SRGM considering both TE and TC to better fit the actual data.

To address this problem, a new SRGM model named GMW-LO-ID is proposed in this paper. In contrast to most of the existing models, our model not only combines the GMW TEF and LO TCF, but also assumes that the current number of faults is dynamic. The fitting accuracy and predictive power are improved by means of considering TE and TC under ID.

1   The Proposed GMW-LO-ID Model

In this section, a new model named GMW-LO-ID, which considers both TE and TC under

ID, is proposed in this section. GMW-LO-ID considers testing resources' allocation, code coverage analysis and debugging lag. Moreover, it has GMW TEF's great flexibility and LO TCF's highly fitting precision.

The NHPP SRGMs considering ID are formulated based on these following assumptions[10,11,12]:

(1) The faults detected process follows the NHPP process;

(2) The current failure is caused by the rest faults in software;

(3) During time interval $[t, t+\Delta t]$, the mean value of detected number of faults from the current TE is proportional to the rest number of faults. In addition, the ratio between them is called the current fault detection rate $r(t)$ and it can be calculated by

$$r(t) = \frac{C'(t)}{1 - C(t)} \tag{1}$$

where $C'(t)$ denotes coverage rate;

(4) Whenever a fault is detected, it can be corrected immediately;

(5) The micro updating of code coverage function $C(t)$ can be neglected when new faults appear.

Based on these assumptions above, GMW TEF, LO TCF and the form of ID are introduced in the following.

GMW TEF is based on generalized modified Weibull distribution, and the cumulative TE[7] is given by

$$W(t) = a(1 - \exp(-\beta t^m e^{\lambda t}))^\theta \tag{2}$$

where a denotes the total effort expenditure; m and θ are shape parameters; β is a scale parameter and λ is an accelerating factor.

LO TCF is presented in Eq. (3) as

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} \frac{1}{w(t)} = \frac{C'(t)}{1 - C(t)}(N - m(t)) \tag{3}$$

where N denotes the total number of faults; $m(t)$ denotes the mean value function of detected faults during time interval $[t, t+\Delta t]$; $C(t) = \dfrac{1}{(1 + A e^{-\alpha t})}$ reveals the final coverage; α denotes the parameter of coverage rate.

The parameters of TEF and TCF can be esti-

mated by means of least square estimation (LSE)[5,10]. Two reasons why we do not utilize maximum likelihood estimation (MLE) are listed in the following. On the one hand, after we transform Eq. (3) into

$$m(t) = Ne^{-B(W(t))} \int_0^t e^{B(W(\tau))} r(W(\tau)) \mathrm{d}\tau$$

where $r(t) = \dfrac{C'(t)}{(1-C(t))}$ and $B(t) = \int_0^t r(W(\tau)) \mathrm{d}\tau$, the datasets offer us a series of t and $m(t)$, which can be regarded as independent variable and dependent variable, respectively. LSE is more likely to be used in x-/y- coordinate system. On the other hand, the number of data containing in the datasets is quite small. MLE, which is based on statistics, may achieve a larger estimation error. Hence, the parameters are estimated by LSE.

How to describe the dynamic function $N(t)$ is the key point in SRGMs considering ID. In this paper, we assume that the current number of faults is proportional to the number of detected faults[14], i. e.

$$N(t) = N + \delta m(t) \qquad (4)$$

where δ denotes a proportionality coefficient.

By combining these three parts together, the proposed model GMW-LO-ID is presented in the following

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} \frac{1}{w(t)} = \frac{\alpha(N + (\delta-1)m(t))}{1 + Ae^{-\alpha t}} \qquad (5)$$

where $m(t)$ denotes the mean value function of detected faults during time interval $[t, t + \Delta t]$; $w(t)$ reveals the function of TE consumption rate, and it can be calculated as the derivative of cumulative TE, i. e. $w(t) = \dfrac{\mathrm{d}W(t)}{\mathrm{d}t}$.

After solving the differential equation with the boundary condition of $m(0) = 0, W(0) = 0$ and the assumption of $\delta \neq 1$, we get GMW-LO-ID shown as

$$m(t) = \frac{N}{1-\delta} \left(1 - \exp\left(\frac{(\delta-1)\alpha a(1 - e^{-\beta t^m e^{\lambda t}})^\theta}{1 + A\exp(-\alpha a(1 - e^{-\beta t^m e^{\lambda t}})^\theta)}\right)\right) \qquad (6)$$

GMW-LO-ID has great flexibility. By assigning the scale parameters and shape parameters in Eq. (6), it will degenerate to other simple kinds of SRGMs considering TE, TC and ID. For ex-

ample, when $\lambda = 0, \theta = 1, m = 2$ we get the Rayleigh-LO-ID shown in Eq. (7).

$$m(t) = \frac{N}{1-\delta} \left(1 - \exp\left(\frac{(\delta-1)\alpha a(1 - e^{-\beta t^2})}{1 + A\exp(-\alpha a(1 - e^{-\beta t^2}))}\right)\right) \qquad (7)$$

Obviously, GMW-LO-ID has nine parameters. Thus, we can hardly estimate them together. Li[10] mentioned a similar method of parameter estimation: Estimate the parameters related to TEF first and calculate the rest parameters with the estimated ones. Hereon, we get our method of parameter estimation:

(1) Estimate the five parameters relating to TEF, i. e. parameters α, β, m, λ, θ, by TE datasets;

(2) After the substitution of five parameters in step (1) into Eq. (7), there will be only one general NHPP SRGM parameter N, two parameters A and α related to TCF and a proportionality coefficient δ in the formula. Using LSE to estimate the rest four parameters can obtain an accurate estimation set easily.

2 Experiments and Analysis

In this section, the fitting accuracy and predictive power of our proposed model are validated in constrast to ten SRGMs under two classical datasets. LSE is used to estimate their parameters.

2.1 Dataset description

In this paper, two datasets are used for verification that our GMW-LO-ID model has a better descriptive and predictive power than other models. The first dataset is from the System T1 data of the Rome Air Development Center (RADC)[1]. This dataset, which contains running time, testing-effort (computer time) and number of detected faults, is widely used in the latest studies, such as Zhang[5] and Peng[1]. The second dataset is from the PL/I application program test data. Its structure is similar to the first dataset and was used by Pachauri[7], Chatterjee[11]. As both of the datasets are widely used in the latest researches, the comparison between our proposed model and the others can be easily arranged.

2.2　Experimental preparation

In order to find a better descriptive model, five criteria are used to measure the fitting effects, as shown in Table 1.

Table 1　Criteria for model comparisons

Name	Definition
MSE	$\sum_{i=1}^{k}\dfrac{(y_i-m(t_i))^2}{k}$
R-Square	$1-\sum_{i=1}^{k}(y_i-m(t_i))^2\Big/\sum_{i=1}^{k}\Big(y_i-\sum_{j=1}^{k}\dfrac{y_j}{k}\Big)^2$
Bias	$\sum_{i=1}^{k}\dfrac{(w(t_i)-w_i)}{k}$
Variation	$\sqrt{\Big(\sum_{i=1}^{k}(w(t_i)-w_i-\text{Bias})^2\Big)\Big/k}$
RMSPE	$\sqrt{\text{Variance}+\text{Bias}^2}$

To MSE, variation and RMSPE, the smaller the criterion is, the better the model performs. In addition, a better performance achieves when R-Square and bias is more close to constant 1 and 0, respectively.

Moreover, ten SRGMs considering TE, TC or both are introduced for the model comparison. Goel-Okumoto is introduced as a representation of the traditional SRGM without considering TE and TC. Seven SRGMs, such as inflected S-shaped and GMW[5], are introduced as the representations of models considering TE only. Besides, IS-LO[10] is introduced as the SRGM considering both TE and TC. SRGM-GTEFID[12] is the representation of SRGM considering TE under ID. Furthermore, if an SRGM considers more than two parts of TE, TC and ID, it definitely contains more parameters, which may have better fitting and predictive power.

2.3　Comparison of fitting power

Ten SRGMs considering TE, TC or both are introduced to validate the fitting performance of GMW-LO-ID in this section. Table 2 and Table 3 show the estimated parameters and fitting results of all eleven models for DS1 and DS2, respectively.

Table 2　Estimated parameters and fitting results of eleven SRGMs in DS1

| SRGM | $W(t)$ | DS1 | | | | |
		Parameters (LSE)	MSE	R-Square	Bias	Variation	RMSPE
Goel-Okumoto	$m(t)=\alpha(1-e^{-bt})$	$a=18.941\,2,$ $b=0.024\,8$	109.709 2	0.917 2			
Logistic[9]	$\dfrac{N}{(1+Ae^{-\alpha t})}$	$N=29.109\,5,\;A=424.89,$ $\alpha=0.493\,5$	61.720 6	0.974 9	−0.004 9	0.398 0	0.630 9
Exponential Weibull[15]	$N\,(1-e^{-\beta t^m})^{\theta}$	$N=27.186\,0,\beta=2.71\times10^{-6}$ $m=4.587\,8,\theta=1.778\,3$	115.374 6	0.953 1	−0.006 5	0.404 0	0.635 6
Inflected S-shaped[16]	$\dfrac{N(1-e^{-\beta t})}{1+\psi e^{-\beta t}}$	$N=29.114\,2,\beta=0.493\,2,$ $\psi=4\,601.1$	60.196 5	0.975 6	−0.004 8	0.398 0	0.630 9
Weibull	$N(1-e^{-\beta t^m})$	$N=32.581\,8,\beta=1.73\times10^{-7},$ $m=5.281\,0$	50.978 0	0.979 3	−0.052 3	0.533 0	0.731 9
Generalized exponential	$N\,(1-e^{-\beta t})^m$	$N=40.846\,1,\beta=0.215\,3,$ $m=39.846\,4$	139.644 0	0.943 3	−0.287 0	0.486 0	0.753 9
GMW[5]	$\alpha\,(1-e^{-\beta t^m e^{\lambda t}})^{\theta}$	$\alpha=30.264\,8,\beta=0.690\,6,$ $m=0.004\,0,\lambda=0.093\,1,$ $\theta=20.731\,0$	100.398 6	0.959 2	−0.013 6	0.426 2	0.653 0
Burr type XII	$\alpha(1-(1+(\beta t)^{\delta})^{-m})$	$\alpha=28.002\,5,\beta=0.043\,9$ $\delta=6.561\,7,m=5.149\,5$	86.857 0	0.964 7	−0.004 9	0.408 3	0.639 0
IS-LO[10]	$\dfrac{N(1-e^{-\beta t})}{1+\psi e^{-\beta t}}$	$N=29.114\,2,\beta=0.493\,2,$ $\psi=4\,601.1$	27.399 9	0.988 9	**−0.004 8**	**0.398 0**	**0.630 9**
SRGM-GTEFID[12]	$W\alpha b(k+l)F$	$W=28.846\,2,\alpha=0.503\,3,$ $\beta=0.568\,1,b=0.000\,3,$ $n_1=2,k=3.382\,4,$ $l=0.671\,2,n_2=2,$ $\bar{\omega}=-0.001\,1,\delta=-0.474\,4$	30.539 8	0.978 1	−0.025 8	0.407 0	0.638 5
GMW-LO-ID	$\alpha\,(1-e^{-\beta t^m e^{\lambda t}})^{\theta}$	$\alpha=30.264\,8,\beta=0.690\,6,$ $m=0.004\,0,\lambda=0.093\,1,$ $\theta=20.731\,0$	**26.649 8**	**0.989 2**	−0.036 8	0.432 4	0.658 6

Table 3   Estimated parameters and fitting results of eleven SRGMs in DS2

SRGM	Parameters (LSE)	DS2				
		MSE	R-Square	Bias	Variation	RMSPE
Goel-Okumoto	$a = 1\ 482, b = 3.44 \times 10^{-5}$	160. 149 2	0. 991 3			
Logistic[9]	$N = 54.837\ 2, A = 13.032, \alpha = 0.226\ 3$	1 130. 600 0	0. 890 5	0. 055 5	1. 185 3	1. 090 1
Exponential Weibul[15]	$N = 112.911\ 7, \beta = 3.1 \times 10^{-5}$ $m = 2.8114, \theta = 0.396\ 8$	113. 113 3	0. 989 0	−0. 010 3	0. 960 5	0. 980 1
Inflected S-shaped[16]	$N = 91.583\ 8, \beta = 0.090\ 1, \psi = 3.164\ 2$	1 006. 900 0	0. 902 4	−0. 006 1	0. 949 3	0. 974 3
Weibull	$N = 269.895\ 6, \beta = 0.006\ 5, m = 1.156\ 0$	113. 222 4	0. 989 0	−0. 014 6	0. 965 2	0. 982 6
Generalized exponential	$N = 300.649\ 4, \beta = 0.012\ 3, m = 1.175\ 0$	166. 068 4	0. 983 9	0. 001 7	0. 966 9	0. 983 3
GMW[5]	$\alpha = 64.971\ 7, \beta = 0.043\ 4, m = 0.821\ 9$ $\lambda = 0.055\ 0, \theta = 1.077\ 3$	380. 998 7	0. 963 1	−0. 001 8	0. 941 9	0. 970 5
Burr type XII	$\alpha = 437.963\ 7, \beta = 0.012\ 3$ $\delta = 1.182\ 4, m = 0.702\ 1$	652. 003 8	0. 936 8	−0. 012 0	0. 968 3	0. 984 1
IS-LO[10]	$N = 91.583\ 8, \beta = 0.090\ 1, \psi = 3.164\ 2$	**94. 801 3**	0. 990 8	−0. 006 1	0. 949 3	0. 974 3
SRGM-GTEFID[12]	$W = 67.251\ 3, \alpha = 0.142\ 5,$ $\beta = 0.536\ 4, b = 2.7 \times 10^{-5},$ $n_1 = 5, k = 5.081\ 4,$ $l = 0.896\ 9, \bar{\omega} = -0.9 \times 10^{-6},$ $\delta = -0.574\ 5, n_2 = 2$	559. 428 5	0. 955 8	−0. 011 7	0. 960 3	0. 980 0
GMW-LO-ID	$\alpha = 64.971\ 7, \beta = 0.043\ 4, m = 0.821\ 9$ $\lambda = 0.055\ 0, \theta = 1.077\ 3$	108. 402 2	0. 989 5	**−0. 001 8**	**0. 941 9**	**0. 970 5**

The numbers highlighted in bold are the best results of each column. And in the tables, Goel-Okumoto is a traditional SRGM without considering TE, which means there will be no $W(t)$ or $w(t)$. That is why the last three criteria (bias, variation and RMSPE) in Table 1 are missing and $m(t)$ is used to fill the blank of $W(t)$ instead. The F in $W(t)$ of SRGM-GTEFID is short for $F = \sum\limits_{n_1=0}^{\infty} \sum\limits_{n_2=0}^{\infty} \left\{ \dfrac{-(-\beta)^{n_1}(-k)^{n_2}(n_2+1)\left[e^{-(n_1\bar{\omega}+(n_2+1)\alpha+\delta)t}-1\right]}{n_1\bar{\omega}+(n_2+1)\alpha+\delta} \right\}$. $W(t)$ is utilized here as the substitution of full expression of SRGMs. Two reasons are shown in the following: On the one hand, all 11 SRGMs in Table 2 follow the NHPP process, which means all models are based on the structure of $\dfrac{dm(t)}{dt} \dfrac{1}{w(t)} = r(t)(N-m(t))$. The main differences between 11 SRGMs are the formula of

$W(t)$. On the other hand, the full expressions of SRGMs are not intuitive. The 11 long formulas may make Table 2 chaotic for understanding. Obviously, in both experiments, IS-LO and GMW-LO-ID almost achieve all the best performance, which shows that SRGMs considering both TE and TC have better curve fitting effects than other SRGMs.

The best three, GMW-LO-ID, IS-LO and Weibull, are plotted in Fig. 1 and Fig. 2 to show the comparison between real and estimated results for each dataset, respectively. The subfigures in left plot the observed/estimated TE curves and the subfigures in right plot the cumulative failure curves. Obviously, all three SRGMs yield good fittings, but we can hardly tell which one has the best fitting power. Thus, we use relative error (RE) to analyze the transformation of fitting accuracy[12], and the results of fitting accuracy vs

time are shown in Figs. 3, 4.

$$RE_i = \frac{m_{estimated}(t_i) - m_{real}(t_i)}{m_{real}(t_i)} \qquad (8)$$

As shown in Fig. 3 and Fig. 4, GMW-LO-ID has the fastest convergence speed and gets close to the x-axis in the shortest time. Hereon, it has the best fitting power of all three SRGMs and can be used to describe the real testing-effort expenditure.
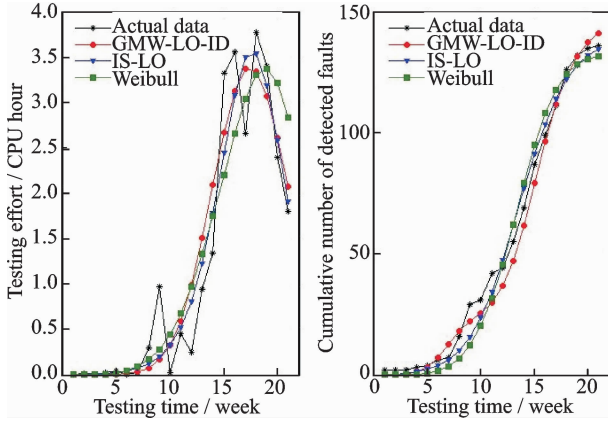

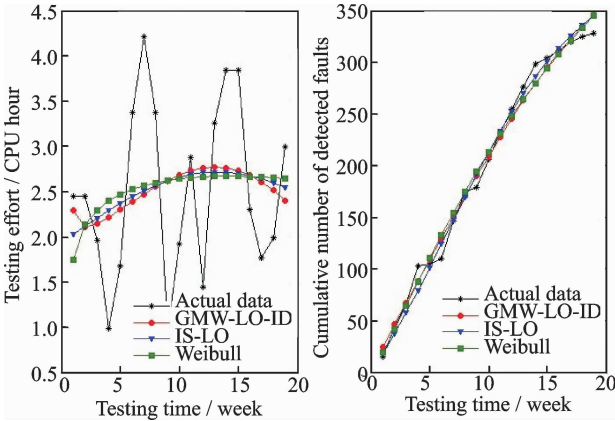Fig. 1   Observed/estimated TE and cumulative failure curves of three SRGMs vs time in DS1

Fig. 2   Observed/estimated TE and cumulative failure curves of three SRGMs vs time in DS2
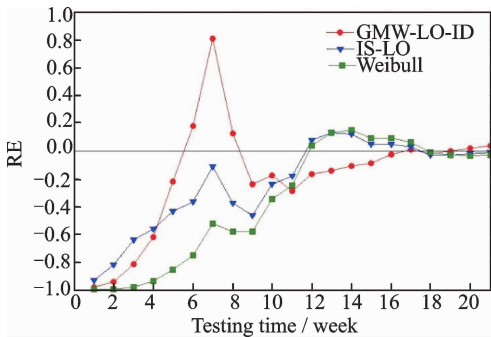
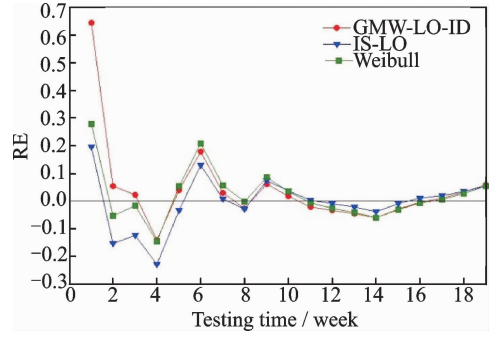Fig. 3   RE curves of three SRGMs in DS1

Fig. 4   RE curves of three SRGMs in DS2

2.4   Comparison of predictive power

The predictive performance of the proposed model is validated in this section. In order to measure the predictive power of SRGMs, we divide the dataset into training set and testing set. Training set is used to estimate models' parameters and testing set is for validating the predictive power of models. In terms of the MSE in testing set, we can tell the differences between the predictive power of SRGMs.

To each group of the experiments, the training set contains 70%, 80% or 90% of data, respectively. Table 4 and Table 5 respectively show the predictive results of all 11 models for DS1 and DS2. The numbers in bold are the best results of each column.

Table 4 shows that the best model for 80% and 90% of DS1 is the proposed model because it has the smallest MSE (46.855 2 and 6.712 1).

Table 4   Comparison results of the predictive power in DS1

Model name	MSE (Predicted part)		
	70% of data	80% of data	90% of data
Goel-Okumoto	1 975.667 2	1 702.751 8	357.555 2
Logistic	**8.268 6**	803.828 0	267.291 1
Exponential Weibull	1 721.952 7	1 384.414 4	246.281 2
Inflected S-shaped	1 512.624 6	1 076.183 7	273.754 0
Weibull	1 681.710 0	1 080.109 2	160.520 8
Generalized exponential	1 755.963 2	1 339.472 0	189.167 4
GMW	1 570.191 0	1 254.475 5	226.019 7
Burr type XII	1 696.989 6	1 387.383 1	196.121 1
IS-LO	383.800 3	134.935 3	52.455 8
SRGM-GTEFID	1 143.008 5	873.967 1	198.342 0
GMW-LO-ID	356.329 5	**46.855 2**	**6.712 1**

Table 5   Comparison results of the predictive power in DS2

Model name	MSE（Predicted part）		
	70％ of data	80％ of data	90％ of data
Goel-Okumoto	2 155. 220 6	1 344. 917 2	601. 079 5
Logistic	179. 549 2	**62. 322 0**	132. 747 5
Exponential Weibull	1 752. 292 1	873. 060 8	448. 704 9
Inflected S-shaped	2 286. 787 4	1 029. 705 2	488. 341 7
Weibull	940. 586 2	694. 577 9	413. 726 7
Generalized exponential	1 117. 800 1	803. 638 3	393. 806 8
GMW	1 679. 898 4	880. 751 3	463. 751 8
Burr type XII	1 445. 675 4	743. 424 2	420. 069 5
IS-LO	1 879. 181 7	559. 546 6	143. 442 3
SRGM-GTEFID	494. 184 6	349. 316 6	165. 748 3
GMW-LO-ID	**165. 937 7**	234. 587 6	**60. 566 1**

The other MSE can be 2. 88 times （IS-LO's 134. 935 3 for 80％）, even 39. 82 times （Logistic's 267. 291 1 for 90％） larger than GMW-LO-ID's. However，Logistic gets an amazing performance （8. 268 6） for 70％ of DS1 while the MSE of GMW-LO-ID （356. 329 5） is 43 times larger. Figs. 5—7 are plotted to reveal the reason why Logistic performs so terrific in this situation.

Figs. 5—7 show the predictive performance of Logistic and GMW-LO-ID for 70％，80％ and 90％ of DS1，respectively. The x- and y-axes denote the testing time and the cumulative number of detected faults，respectively. The circle and square dots are the training and testing sets. In addition，the blue solid curves represent the predictive performances of Logistic and the red dash ones represent the performances of our proposed model. We can summarize that if the training set is 80％ or 90％ of DS1，our model has a better performance. Furthermore，the result goes opposite if the training set is 70％ of DS1.

Here follows the reason why Logistic performs so terrific in Fig. 5. As can be seen in Fig. 5，we cannot tell the tendency of DS1 just focusing on the training set because the increasing rate of DS1 only decreases in testing set. It is the reason that GMW-LO-ID makes a mistake in judging the tendency of data，which leads to the huge MSE unfortunately.

Besides，Logistic's S-shape helps to decide the tendency of curve and gets a terrific result. Though the MSE of GMW-LO-ID is far larger than Logistic's，it still performs better than other nine SRGMs. Thus，the proposed model has a better predictive power. However，when the training set is 80％ or 90％ of DS1，these time training sets contain the data part in which the increasing rate decreases. Our model recognizes it and achieves better predictive performance in Figs. 6，7.
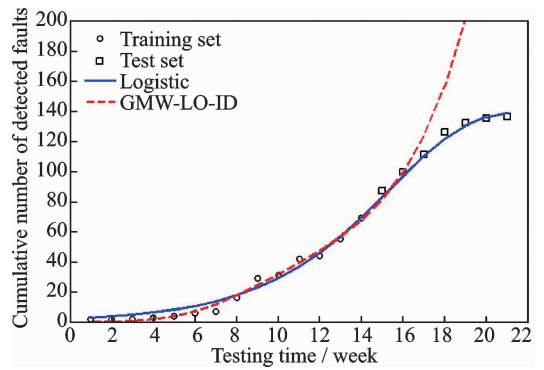

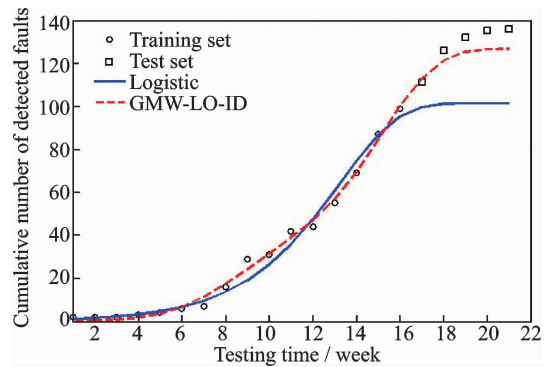Fig. 5   Logistic and GMW-LO-ID's performances for 70％ of DS1

Fig. 6   Logistic and GMW-LO-ID's performances for 80％ of DS1

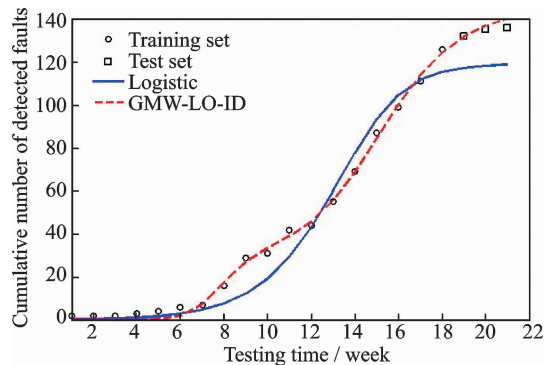Fig. 7   Logistic and GMW-LO-ID's performances for 90％ of DS1

As for DS2，Table 5 shows that the best SRGM for 70% and 90% of DS2 is the proposed model because it has the smallest MSE (165.937 7 and 60.566 1). The other MSE can be 5.67 times (Weibull's 940.586 2 for 70%), even 9.92 times (Goel-Okumoto's 601.079 5 for 90%) larger than GMW-LO-ID's. The reason why Logistic (62.322 0) performs better than the proposed model (234.587 6) for 80% of DS2 is the same just mentioned above and it still performs better than other nine SRGMs as well.

In a word，the proposed model has a better predictive power and can effectively predict software failure behavior.

3 Conclusions

As the dynamic allocation of testing resources，the raise of code coverage rate and the existence of debugging lag will affect software reliability growth curve. By incorporating TE and TC into traditional SRGMs，we use debugging lag to update the constant number of faults with a dynamic variety. Under the assumption that the number of faults is proportional to the current number of detected faults，we propose a new model named GMW-LO-ID combining GMW TEF with LO TCF. LSE is utilized to figure out the estimated values of parameters involved.

After analyzing the experimental results based on MSE，RE and other four kinds of criteria in two testing datasets，we can draw a conclusion that GMW-LO-ID fits the real failure data better and has a better predictive power than other ten SRGMs only considering one or two points of TE，TC and ID. Meanwhile，due to the assignment of some scale parameters and shape parameters，it is more flexible than most of the SRGMs and it can degenerate to other simple kinds of SRGMs，such as Rayleigh-LO-ID. Consequently，this hybrid model brings us a good performance.

It will be worthwhile to do further research on extending or finding new TEF，TCF and other forms of ID. Moreover，trying to find a more accuracy method of parameter estimation may be another study direction.

References：

[1] PENG R，LI Y F，ZHANG W J，et al. Testing effort dependent software reliability model for imperfect debugging process considering both detection and correction [J]. Reliability Engineering and System Safety，2014，126：37-43.

[2] YANG X，FANG Z，TAO L. Life evaluation GERT network model of equipment reliability test [J]. Journal of Nanjing University of Aeronautics and Astronautics，2016，48(5)：689-695. (in Chinese)

[3] KAPUR P K，GOSWAMI D N，BARDHAN A，et al. Flexible software reliability growth model with testing effort dependent learning process [J]. Applied Mathematical Modelling，2015，32(7):1298-1307.

[4] HUANG C Y，KUO S Y. Analysis of incorporating logistic testing-effort function into software reliability modeling [J]. IEEE Transactions on Reliability，2002，51(3)：261-270.

[5] ZHANG N，CUI G，LIU H W. A finite queuing model with generalized modified Weibull testing effort for software reliability [C]//Proceedings of International Conference on Computer Science and Network Technology (ICCSNT 2011). Harbin，China：IEEE，2011;401-406.

[6] LIN C T. Enhancing the accuracy of software reliability prediction through quantifying the effect of test phase transition [J]. Applied Mathematics and Computation，2012，219(5)：2478-2492.

[7] PACHAURI B，KUMAR A，DHAR J. Software reliability growth modeling with dynamic faults and release time optimization using GA and MAUT [J]. Applied Mathematics and Computation，2014，242：500-509.

[8] CAI X，LYU M R. Software reliability modeling with test coverage：Experimentation and measurement with a fault-tolerant software project [C]//Proceedings of the 18th IEEE International Symposium on Software Reliability. Trollhattan，Sweden：IEEE，2007;17-26.

[9] LI H F，LI Q Y，LU M Y. Software reliability modeling with logistic test coverage function [J]. Journal of Computer Research and Development，2011，48(2)：232-240. (in Chinese)

[10] LI H F，WANG S Q，LIU C，et al. Software reliability model considering both testing effort and testing coverage [J]. Journal of Software，2013，24(4)：749-760. (in Chinese)

[11] CHATTERJEE S，SINGH J B. A NHPP based soft-

ware reliability model and optimal release policy with lo-gistic-exponential test coverage under imperfect debug-ging [J]. International Journal of System Assurance En-gineering and Management，2014，5(3)：399-406.

[12] ZHANG C，CUI G C，LIU H W，et al. A unified and flexible framework of imperfect debugging dependent SRGMs with testing-effort [J]. Journal of Multimedia，2014，9(2)：310-317.

[13] WANG J Y，WU Z B，SHU Y J，et al. An imperfect software debugging model considering log-logistic distri-bution fault content function [J]. Journal of Systems and Software，2015，100：167-181.

[14] HUANG C Y，KU S Y，LYU M R. An assessment of testing-effort dependent software reliability growth mod-els [J]. IEEE Transactions on Reliability，2007，56(2)：198-211.

[15] SAGAR B B，SAKET R K，SINGH C G. Exponentiat-ed Weibull distribution approach based inflection S-shaped software reliability growth model [J]. AIN Shams Engineering Journal，2016，7(3)：973-991.

[16] JIN C，JIN S W. Parameter optimization of software re-liability growth model with S-shaped testing-effort func-tion using improved swarm intelligent optimization [J]. Applied Soft Computing，2016，40(C)：283-291.

Mr. **Zang Sicong** received the B. S. and M. S. degrees in com-puter science and technology from Nanjing University of Aero-nautics and Astronautics (NUAA)，Nanjing，China in 2014 and 2017，respectively. Now he is working for his Ph. D. de-gree in Shanghai Jiao Tong University. His research is focused on deep learing and generative model.

Prof. **Pi Dechang** received the Ph. D. degree in mechatronic en-gineering from NUAA，Nanjing，China in 2002. After that，he joined the College of Computer Science and Technology，NUAA，Nanjing，China. His research is focused on data min-ing and big data analytics.

（Production Editor：Zhang Huangqun）