Software Reliability Growth Modelling using a Weighted Laplace Test Statistic

Yan Luo[⋆] Torsten Bergander[⋄] A. Ben Hamza[⋆]

[⋆]Concordia Institute for Information Systems Engineering
Concordia University, Montréal, QC, Canada

[⋄]SAP Labs Canada Inc., Montréal, QC, Canada

Abstract

We introduce a new weighted Laplace test statistic for software reliability growth modelling. The proposed model not only takes into account the activity in the system but also the proportion of reliability growth within the model. This generalized approach is defined as a weighted combination of a growth reliability model and a non-growth reliability model. Experimental results illustrate the effectiveness and the much improved performance of the proposed method in software reliability modelling.

1 Introduction

During the development process of computer software systems, many software defects may be introduced and often lead to critical problems and complicated breakdowns of computer systems [1, 2]. Hence, there is an increasing demand for controlling the software development process in terms of quality and reliability. Software reliability can be evaluated by the number of detected faults or the software failure-occurrence time in the testing phase which is the last phase of the development process, and it can be also estimated for the operational phase. A software failure is defined as an unacceptable departure of program operation caused by a software fault remaining in the software system [1–3].

It is, however, very difficult for developers to produce highly reliable software systems efficiently because of the diversified and complicated software requirements. Software reliability models can provide quantitative measures of the reliability of software systems during software development processes [4, 5]. In recent years, several software reliability models have been proposed [6, 7]. In particular, software reliability models that describe software fault-detection or software failure-occurrence phenomena in the testing phase are referred to as software reliability growth models (SRGMs). The SRGMs have been proven to be successful in estimating the software reliability and the number of errors remaining in the software, and are very useful to assess the reliability for quality control and testing-process control of software development [4–9].

The rest of this paper is organized as follows. In the next section, we formulate the problem and we briefly review the mathematical aspects of non-homogeneous Poisson processes. In Section 3, the likelihood function of the cumulative number of failures is derived. In Section 4, we propose a weighted Laplace test statistic which is defined in terms of a weighted combination of a growth reliability model and a non-growth reliability model. Section 5 presents experimental results to demonstrate the much improved performance of the proposed approach in software reliability growth modelling. Finally, we conclude in Section 6.

2 Problem Formulation

Software failure data are usually available to the user in three basic forms:

1. in the form of a sequence of ordered failure times $0 < t_1 < t_2 < \ldots < t_n$
2. in the form of a sequence of interfailure times τ_i where $\tau_i = t_i - t_{i-1}$ for $i = 1, \ldots, n$
3. in the form of cumulative number of failures.

It is easy to verify that the failure and interfailure times are related by $t_i = \sum_{j=1}^{i} \tau_j$.

The cumulative number of failures $N(t_i)$ detected by time t_i (i.e. the cumulative number of failures over the period $[0, t_i)$) defines a non-homogeneous Poisson process (NHPP) with failure intensity or rate function $\lambda(t_i)$ such that the rate function of the process is time-dependent. The mean value function $m(t_i) = E(N(t_i))$ of the process is given by $m(t_i) = \int_0^{t_i} \lambda(u)du$.

31st Annual International Computer Software and Applications Conference(COMPSAC 2007)
0-7695-2870-8/07 $25.00 © 2007 IEEE

3 The Likelihood Function

Assume we model the failure times using an NHPP with failure intensity function $\lambda(t;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown parameter vector. Table 1 shows examples of NHPP models with different failure intensity functions $\lambda(t;\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \beta)$.

Model name	$m(t)$	$\lambda(t)$
Log-linear	$\frac{\exp(\alpha+\beta t)}{\beta}$	$\exp(\alpha+\beta t)$
Exponential	$\alpha(1-\exp(-\beta t))$	$\alpha\beta\exp(-\beta t)$
Power law	$\left(\frac{t}{\alpha}\right)^{\beta}$	$\frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$

Table 1. NHPP models.

Under the NHPP assumption, the failure times t_i define intervals for which only failure counts $n_i = N(i)-N(i-1)$ in the interval (t_{i-1}, t_i) are recorded, that is n_i is the number of failures during the i^{th} unit of time. The probability of seeing n_i events in the interval (t_{i-1}, t_i) is then given by

$$P(N(t_{i-1}) - N(t_i) = n_i) = \frac{(m(t_i) - m(t_{i-1}))^{n_i}}{n_i!} \exp(-(m(t_i) - m(t_{i-1}))).$$

If we consider k time intervals, then the likelihood function is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{k} \frac{(\int_{t_{i-1}}^{t_i} \lambda(u;\boldsymbol{\theta})du)^{n_i}}{n_i!} \exp\left(-\int_{t_{i-1}}^{t_i} \lambda(u;\boldsymbol{\theta})du\right)$$

and the marginal probability that there are exactly $N = \sum_{i=1}^{k} n_i$ events is given by

$$G(\boldsymbol{\theta}) = \frac{(m(t_n))^N}{N!} \exp(-m(t_n))$$

Hence the conditional log-likelihood function is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \log(L(\boldsymbol{\theta})) - \log(G(\boldsymbol{\theta})) \\ &= \sum_{i=1}^{k} n_i \log\Big(\int_{t_{i-1}}^{t_i} \lambda(u;\boldsymbol{\theta})du\Big) \\ &\quad - N\log\Big(\int_0^{t_k} \lambda(u;\boldsymbol{\theta})du\Big) + C,\end{aligned}$$

where $C = \log(N!) - \log(\prod_{i=1}^{k} n_i!)$ is a constant.

4 Proposed Method

During testing and development of new systems, reliability trend analysis is needed to evaluate the progress of the development process [4, 5, 10, 11]. The hypotheses we wish to test are:

$$\begin{aligned} H_0 &: \text{HPP} \\ H_1 &: \text{NHPP} \end{aligned}$$

where H_0 and H_1 are the null and the alternative hypotheses respectively.
Under the null hypothesis, we define the Laplace trend as

$$U = \frac{\mathcal{L}(\theta_0)'}{E(-\mathcal{L}(\theta_0)'')},$$

where θ_0 is a component of the vector $\boldsymbol{\theta}$ such that its value makes the intensity function $\lambda(t;\boldsymbol{\theta})$ time independent.

Assuming a type I error probability $\alpha = P\{\text{reject } H_0 | H_0 \text{ is true}\}$, the Laplace trend values may be interpreted as follows:

- $U < -z_\alpha$: reliability growth
- $U > z_\alpha$: reliability deterioration
- $-z_\alpha < U < z_\alpha$: stable reliability,

where z_α is the is the upper α percentage of the standard normal distribution Z such that $P\{Z \geq z_\alpha\} = \alpha$ (i.e. z_α is the $100(1-\alpha)$ percentage point of the standard normal distribution). If "H_0 : HPP" is true, the distribution of the Laplace test statistic U is approximately normal $N(0,1)$. Hence, if "H_0 : HPP" is true, the probability is $1-2\alpha$ that a value of the test statistic U falls between $-z_\alpha$ and z_α.

The objective of system reliability trend tests is to determine whether the pattern of failures is significantly changing with time. For example, when the occurrence of the events is an NHPP with a log-linear failure intensity function $\lambda(t) = \exp(\alpha + \beta t)$, then the null hypothesis may be expressed as $H_0 : \beta = 0$. Moreover, it can be shown that in the case of a log-linear failure intensity function [4, 5, 11], the Laplace test statistic is given by

$$U(k) = \frac{\sum_{i=1}^{k}(i-1)n_i - \frac{k-1}{2}\sum_{i=1}^{k} n_i}{\sqrt{\frac{k^2-1}{12}\sum_{i=1}^{k} n_i}}$$

4.1 Anisotropic Laplace trend

The main limitation of the Laplace trend is that it does not take into account the presence or the absence of activity in the system. To circumvent this problem, we replace the Laplace trend factor $U(k)$ with an anisotropic Laplace trend factor $A(k)$ that is defined as follows

$$A(k) = \begin{cases} g(U(k)) & \text{if no activity} \\ U(k) & \text{otherwise,} \end{cases}$$

where g is a "reliability growth-stopping" function as shown in Table 2. The g-function is chosen to satisfy $g(x) \to 0$ when $x \to \infty$ so that the reliability growth is stopped when there is no activity in the system [12].

Function	$g(x)$
Green [13]	$\dfrac{\tanh(x)}{2x}$ (if $x \neq 0$)
Gaussian [14]	$\exp\left(-\dfrac{x^2}{2\sigma^2}\right)$
Lorentzian [14]	$\dfrac{1}{1 + x^2/\sigma^2}$

Table 2. reliability growth-stopping functions.

The parameter σ of the Gaussian and Lorentzian g-functions may be estimated using tools from robust statistics as follows

$$\hat{\sigma} = 1.4826 \text{ MAD}\{(U(k) - U(k-1))_k\},$$

where MAD denotes the median absolute deviation [15].

4.2 Weighted Laplace trend

Let $w \in [0, 1]$ be a weight parameter denoting the proportion of software reliability growth during the period $[t_\ell, t_\ell + t_w]$. Then, we define a weighted failure intensity function as follows

$$\begin{aligned} \lambda_w(t) &= \lambda(t)\, \mathbf{1}_{(0 \le t \le t_\ell)} + w\lambda(t)\, \mathbf{1}_{(t_\ell \le t \le t_\ell + t_w)} \\ &\quad + (1-w)\lambda(t_\ell)\, \mathbf{1}_{(t_\ell \le t \le t_\ell + t_w)}, \end{aligned}$$

where $\lambda(t)$ is the failure intensity function, and $\mathbf{1}_S$ denotes the indicator function of a subset S.

When $w = 1$, the weighted failure intensity function reduces to the original intensity function, and when $w = 0$, the function λ_w becomes a constant (straight line). Moreover, for $w \in (0, 1)$, the weighted failure intensity function λ_w has a less heavier tail that $\lambda(t)$ in the interval $[t_\ell, t_\ell + t_w]$ indicating a slow reliability growth of the Laplace trend as illustrated in Figure 1.

Therefore, we may define a weighted anisotropic Laplace test statistic as follows

$$\begin{aligned} A_w(k) &= A(k)\, \mathbf{1}_{(0 \le k \le t_\ell)} + wA(k)\, \mathbf{1}_{(t_\ell \le k \le t_\ell + t_w)} \\ &\quad + (1-w)A(t_\ell)\, \mathbf{1}_{(t_\ell \le k \le t_\ell + t_w)}, \end{aligned}$$

where $A(k)$ is the anisotropic Laplace test statistic.

5 Experimental Results

We tested our proposed anisotropic Laplace test statistic on a real software failure data which was taken from an SAP development system. The data contains daily software failures that was recorded for a period of 175 days. Moreover, there are no activities in the system during the test phase process on the days 121, 122, 128, 142, 143, 144, 145, 146, 147, 148, 149, and 150.

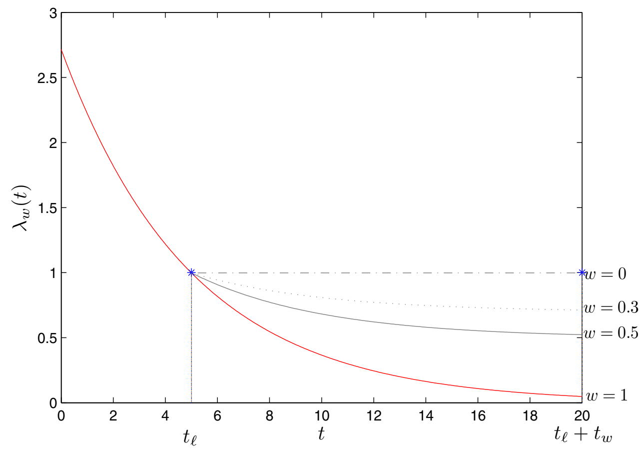

Figure 1. Weighted failure intensity function.

Figure 2 displays the scatter plot of cumulative failure number versus failure time, and it clearly illustrates an improving system since the probability of failures stabilizes substantially after a period of 150 days.

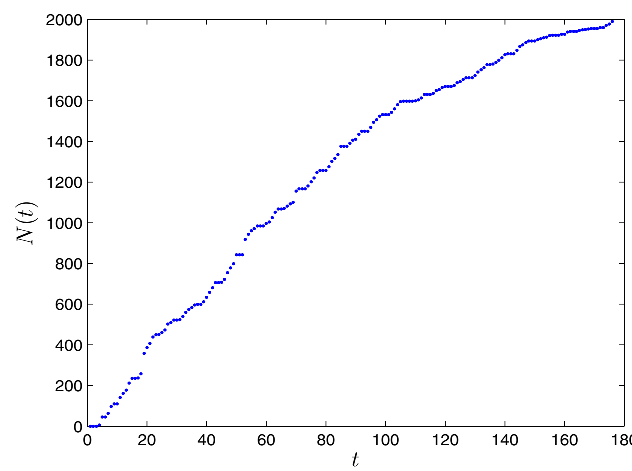

Figure 2. Cumulative Number of Failures vs. Failure Time.

5.1 Weighted Laplace trend

Figure 3 through Figure 5 depict the much improved performance of the weighted anisotropic Laplace trends $A_w(k)$

with a weight parameter $w = 0.5$ in comparison with the anisotropic Laplace trends A_k. The no-activity periods of the weighted anisotropic Laplace trend are displayed with black-star points, whereas the no-activity periods of the anisotropic Laplace trend are displayed with red-star points.

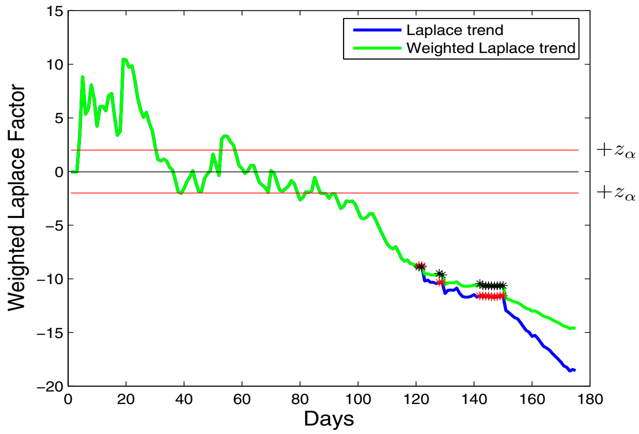

Figure 3. Weighted anisotropic Laplace trend using Green's function, with $w = 0.5$.

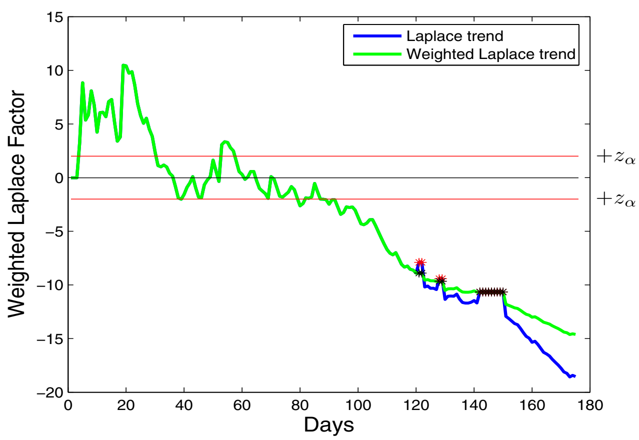

Figure 4. Weighted anisotropic Laplace trend using Gaussian function, with $w = 0.5$.

5.2 Weighted adjusted anisotropic Laplace trend

The Laplace test statistic is a test for the null hypothesis H_0 that the data come from an HPP. Thus rejection of H_0 means that the process is not an HPP, but it could still in principle be a renewal process and hence still has no trend. In order to improve the test performance when

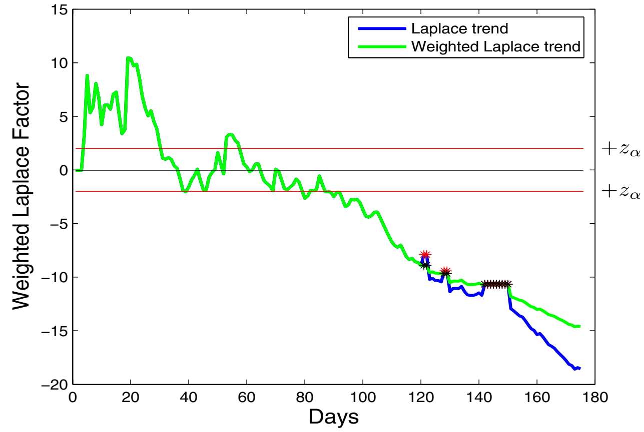

Figure 5. Weighted anisotropic Laplace trend using Lorentzian function, with $w = 0.5$.

the null hypothesis is a more general renewal process, the Lewis-Robinson (LR) test should be used [16]. The LR test is basically a scaled version of the Laplace test and it is defined as

$$U_{LR} = \frac{U(k)}{\widehat{CV}(\tau)},$$

where $\widehat{CV}(\tau)$ is an estimate of the coefficient of variation of the interfailure times τ_i, and it is calculated in terms of the mean and the standard deviation of interfailure times as follows

$$\widehat{CV}(\tau) = \frac{\sigma_\tau}{\bar{\tau}},$$

with τ representing the variable of interfailure times.

The reason for dividing the Laplace trend by the coefficient of variation is to account for non-exponential distributions of the interfailure times and also in order to insure that U_{LR} follows a standard normal distribution whenever the data come from a renewal process. Moreover, when the null-hypothesis is a renewal process model with non-exponential interarrival times, this adjustment maintains the type-I error probability better than the Laplace test [16].

Figure 6 through Figure 8 show the weighted adjusted anisotropic Laplace trends with a weight parameter $w = 0.5$. The weighted adjusted anisotropic Laplace trends with a weight parameter $w = 0.1$ are depicted in Figure 9 through Figure 11. Note that the no-activity periods of the weighted adjusted anisotropic Laplace trends are displayed with black-star points, whereas the no-activity periods of the adjusted anisotropic Laplace trends are displayed with red-star points.

31st Annual International Computer Software and Applications Conference(COMPSAC 2007)
0-7695-2870-8/07 $25.00 © 2007 IEEE

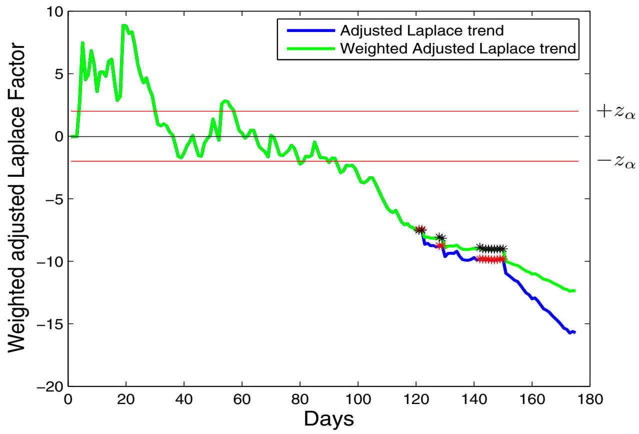

Figure 6. Weighted adjusted anisotropic Laplace trend using Green's function ($w = 0.5$).

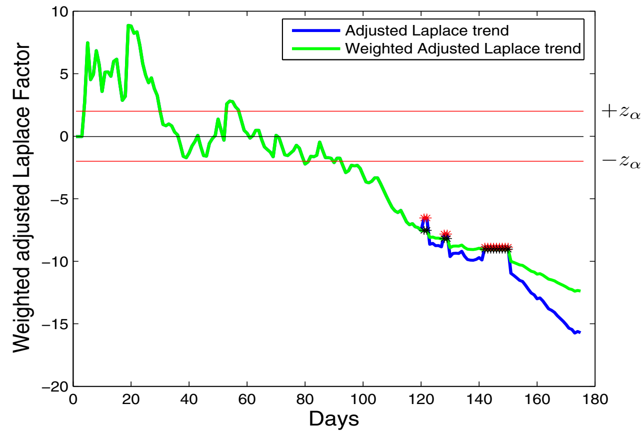

Figure 8. Weighted adjusted anisotropic Laplace trend using Lorentzian function ($w = 0.5$).

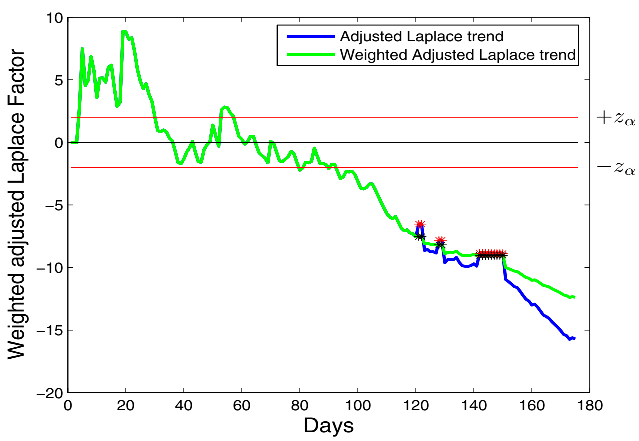

Figure 7. Weighted adjusted anisotropic Laplace trend using Gaussian function ($w = 0.5$).

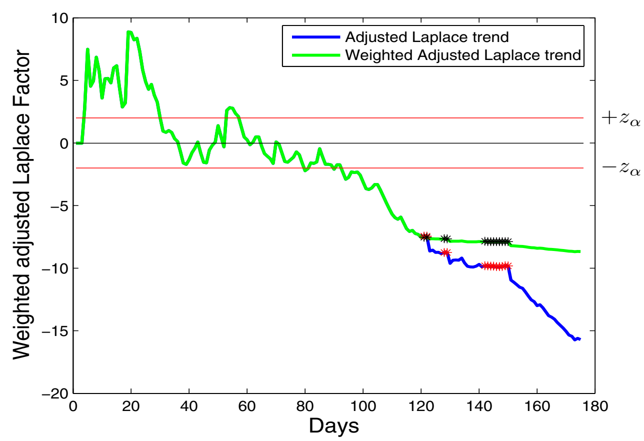

Figure 9. Weighted adjusted anisotropic Laplace trend using Green's function ($w = 0.1$).

6 Conclusions

In this paper, we proposed a new weighted Laplace test statistic for software reliability growth modelling. The proposed model not only takes into account the activity in the system but also the proportion of reliability growth within the model. This generalized approach is defined as a weighted combination of a growth reliability model and a non-growth reliability model. The experimental results clearly indicate a much improved performance of the proposed anisotropic Laplace test statistic in software reliability growth modelling.

Acknowledgments

This work was supported by SAP Labs Canada Inc.

References

[1] J.D. Musa, "A theory of software reliability and its application," *IEEE Trans. Software Engineering*, vol. 1, no. 1, pp. 312-327, 1975.

[2] J.D. Musa, A. Iannino, and K. Okumoto, *Software Reliability: Measurement, Prediction, Application*,

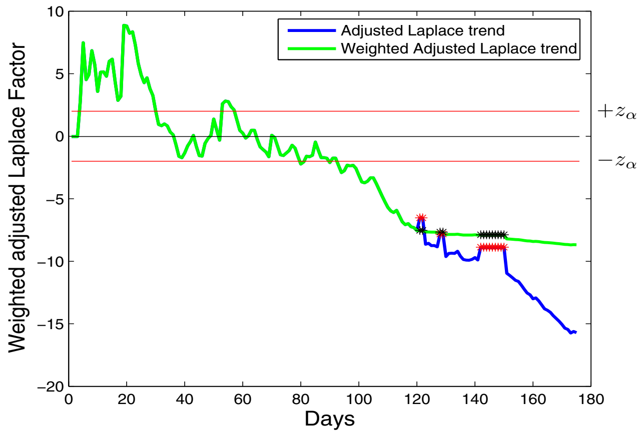

Figure 10. Weighted adjusted anisotropic Laplace trend using Gaussian function ($w = 0.1$).

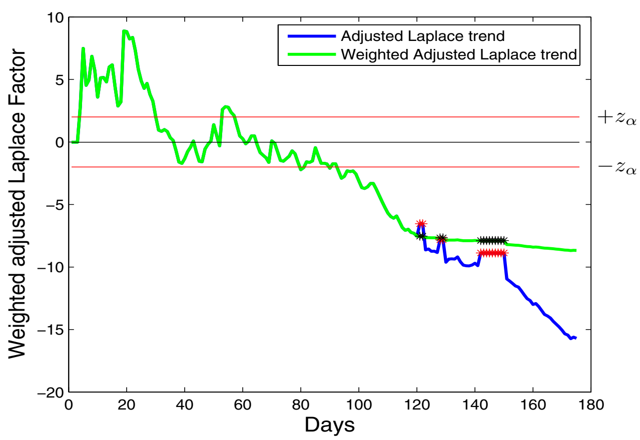

Figure 11. Weighted adjusted anisotropic Laplace trend using Lorentzian function ($w = 0.1$).

McGraw-Hill Book Company, 1987.

[3] A.L. Goel and K. Okumoto, "Time-dependent error detection rate models for software reliability and other performance measures," *IEEE Trans. Reliability*, vol. 28, no. 3, pp. 206-211, 1979.

[4] M.R. Bastos Martini, K. Kanoun, and J. Moreira de Souza, "Software-reliability evaluation of the TROPICO-R switching system," *IEEE Trans. Reliability*, vol. 39, no. 3, pp. 369-379, 1990.

[5] K. Kanoun and J.C. Laprie, "Software reliability trend analysis from theoretical to practical considerations," *IEEE Trans. Software Engineering*, vol. 41, no. 4, pp. 525-532, 1992.

[6] A.L. Goel, "Software reliability models: assumptions, limitations and applicability," *IEEE Trans. Software Engineering*, vol. 11, no. 12, pp. 1411-1423, 1985.

[7] S. Yamada, M. Ohba, and S. Osaki, "S-shaped reliability growth modeling for software error detection," *IEEE Trans. Reliability*, vol. 32, no. 5, pp. 475-485, 1983.

[8] J.H. Lo and C.Y. Huang, "An integration of fault detection and correction processes in software reliability analysis," *Journal of Systems and Software*, vol. 79, pp. 1312-1323, 2006.

[9] X. Zhang and H. Pham, "Software field failure rate prediction before software deployment," *Journal of Systems and Software*, vol. 79, pp. 291-300, 2006.

[10] D.R. Cox and P.A.W. Lewis, *The Statistical Analysis of a Series of Events*, Chapman and Hall, London, 1978.

[11] O. Gauodin, "Optimal properties of the Laplace trend test for software-reliability models," *IEEE Trans. Reliability*, vol. 20, no. 9, pp. 740-747, 1992.

[12] A. Ben Hamza, H. Krim, and G. Unal, "Unifying probabilistic and variational estimation," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 37-47, 2002.

[13] P.J. Green, "Bayesian Reconstruction from Emission Tomography Data Using Modified EM Algorithm," *IEEE Trans. Medical Imaging*, vol. 9, no. 1, pp. 84-93, 1990.

[14] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Analysis and Machine intelligence*, vol. 12, no. 7, pp. 629-639, 1990.

[15] P.J. Rousseeuw and A.M. Leroy, *Robust Regression and Outlier Detection*, Wiley, NY, 1987.

[16] H.E. Ascher, C.K.Hansen, "Spurious exponentiality observed when incorrectly fitting a distribution to nonstationary data," *IEEE Trans. Reliability*, vol. 47, no. 4, pp. 451-45, 1998.