

融合小波变换和非齐次泊松过程的开源软件故障预测模型

顾子欣¹, 杨剑锋^{1,2*}, 李柳英¹, 肖雪珊¹

¹南宁师范大学数学与统计学院, 广西 南宁

²南宁师范大学广西应用数学中心, 广西 南宁

收稿日期: 2025年7月19日; 录用日期: 2025年8月12日; 发布日期: 2025年8月21日

摘要

开源软件凭借其开放源代码、社区协作开发等特性, 拥有极为广泛的应用。但开源软件故障频发, 传统基于非齐次泊松过程(NHPP)的预测方法在处理开源软件故障数据时, 由于故障数据具有非平稳性, 难以准确捕捉数据中的复杂特征和规律。为了提高软件故障预测效果, 本文提出了一种融合小波变换和非齐次泊松过程的开源软件故障预测模型。利用小波变换提取故障数据多尺度特征, 提取故障数据的低频趋势与高频波动分量, 来克服传统模型缺乏多尺度分析的缺陷。结合Delayed S-Shaped (DSS)可叠加可靠性模型进行故障预测。以Tomcat 9开源软件故障数据为例, 通过极大似然估计法对模型进行参数估计, 用提出的模型与DSS模型进行性能对比。结果显示, 本文提出的模型故障拟合效果优于DSS模型。

关键词

开源软件, 小波变换, 非齐次泊松过程, 故障预测

Open Source Software Fault Prediction Model Integrating Wavelet Transform and Non-Homogeneous Poisson Process

Zixin Gu¹, Jianfeng Yang^{1,2*}, Liuying Li¹, Xueshan Xiao¹

¹School of Mathematics and Statistics, Nanning Normal University, Nanning Guangxi

²Guangxi Applied Mathematics Center of Nanning Normal University, Nanning Guangxi

Received: Jul. 19th, 2025; accepted: Aug. 12th, 2025; published: Aug. 21st, 2025

*通讯作者。

文章引用: 顾子欣, 杨剑锋, 李柳英, 肖雪珊. 融合小波变换和非齐次泊松过程的开源软件故障预测模型[J]. 应用数学进展, 2025, 14(8): 227-238. DOI: 10.12677/aam.2025.148385

Abstract

Open source software has gained extensive applications due to its open-source code and community-driven development. However, frequent failures in such software pose challenges for traditional Non-Homogeneous Poisson Process (NHPP)-based prediction methods, which struggle to capture complex patterns in non-stationary failure data. To address this, this paper proposes a hybrid model integrating Wavelet Transform (WT) and NHPP for open source software failure prediction. The WT decomposes failure data into multi-scale components (low-frequency trends and high-frequency fluctuations) to overcome the limitations of single-scale NHPP models. By combining this with the Delayed S-Shaped (DSS) reliability model, the hybrid approach enables more accurate forecasting. Using Tomcat 9 failure data as a case study, parameters are estimated via maximum likelihood estimation, and performance is compared against the standalone DSS model. The results show that the failure fitting effect of the model proposed in this paper is better than that of the DSS model.

Keywords

Open Source Software, Wavelet Transform, Non-Homogeneous Poisson Process, Failure Prediction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

开源软件系统凭借其开放协作、快速迭代的优势，已成为云计算、分布式计算等领域的核心支撑。然而，其开放性也带来显著挑战：代码复杂度高、社区贡献代码质量参差不齐，导致系统缺陷隐蔽性强、维护难度大。现有的开源软件系统故障预测方法多基于非齐次泊松过程类模型，但故障数据具有非平稳性(如突发故障簇、周期性负载波动等)，而传统方法缺乏多尺度分析能力，无法有效解析短期异常与长期趋势的关联，使模型的拟合效果下降。通过对开源软件系统可能出现的故障进行预测，提前采取相应的措施，如修复潜在的漏洞、优化系统配置等，能够显著降低系统故障出现的频率，进而提升系统的可靠性与稳定程度。

在软件可靠性建模与故障预测研究领域，NHPP 类可靠性模型以及基于小波变换预测模型的研究长期受到关注，众多学者围绕这些方向展开多维度的深入研究，收获了丰硕成果，但仍存在部分尚未攻克的难题。

在 NHPP 类软件可靠性模型领域，陈静等人[1]对 NHPP 类开源软件可靠性增长模型的极大似然估计进行了深入探讨。Das. S 等人[2]基于 NHPP 构建连续时间软件可靠性模型，解决了传统可靠性模型与工程实践的脱节问题。文献[3][4]围绕软件可靠性模型展开优化工作，均通过实验或方法验证模型的有效性。文献[5][6]用不同方法应用于软件可靠性建模，突破了传统 NHPP 框架。在小波分析预测模型领域，文献[7][8]运用小波变换提取数据特征进行模型优化，解决传统模型在处理数据上的难题。文献[9]-[13]使用小波变换进行模型分解重构，提取数据特征，建立优化模型解决实际预测问题。小波变换除了在数据提取和模型重构上有重要作用之外，还在与模型结合上有重要用途。Nasser 等人[14]引入小波注意力计算方法(WABC)分析各外部因素重要性并加权，提出基于天气的交通分析方法(WBTA)探究交通流与天气因素在

不同频率和时间间隔的时间相关性。Samani 等人[15]将小波变换与 5 种机器学习模型(ANN、ANFIS、GMDH 等)结合,验证了小波变换对中长期预测的优化作用。Dincer 等人[16]通过最大重叠离散小波变换(MODWT)将时间序列分解为子序列并分别构建模糊时间序列(FTS)模型,利用 MODWT 降噪能力和 FTS 优势提升预测性能。

综合来看,传统 NHPP 类故障预测模型在处理开源软件复杂、非平稳的故障数据时存在局限性,难以精准捕捉故障演化规律,而小波变换在提取数据尺度特征上有重要作用。因此,本文提出了一种融合小波变换和 NHPP 的开源软件故障预测模型,旨在通过小波变换来提升模型预测效果。

2. 基于小波变换的非齐次泊松过程故障预测模型

2.1. 非齐次泊松过程的故障预测模型

2.1.1. 模型构建

Delayed S-Shaped (DSS)模型基于软件缺陷的发现和修复过程,认为软件中的缺陷随时间逐渐被发现和修复,即软件的可靠性得以提高。

DSS 模型能够较好地拟合软件测试过程中故障发现的情况。在测试初期,由于测试范围有限、测试用例尚未全面覆盖等原因,缺陷发现率较低;随着测试的深入,测试用例的增加和测试方法的改进,缺陷发现率逐渐提高;当测试进行到一定阶段后,大部分容易发现的缺陷已被找出,缺陷发现率开始下降,呈现出 S 形的变化趋势。

可靠性增长模型的均值函数为

$$m(t) = ae^{-B(t)} \int_0^t b(u) e^{B(u)} du \quad (1)$$

根据上述公式,当故障发现率函数 $b(t)$ 取不同值时,可以得到不同类型的 NHPP 可靠性模型。

当 $b(t) = \frac{b^2 t}{1+bt}$ 时, DSS 模型的均值函数为

$$m(t) = a[1 - (1+bt)e^{-bt}] \quad (2)$$

在传统模型的基础上,可以进一步变换为可叠加可靠性增长模型的均值函数:

$$m(t) = \sum_{i=1}^n a_i [1 - (1+b_i t)e^{-b_i t}] \quad (3)$$

式中: n 为组件的总个数; i 表示第 i 个组件,每个组件的失效过程服从 NHPP。

2.1.2. 模型参数的极大似然估计

本次用于估计模型未知参数的数据为分组数据, m_k 为在 $[0, t_k]$ 时间段上被检测出来的累计失效数,记 (t_k, m_k) 为 n 个检测到的失效数据对, $k=1, 2, \dots, n$ 。当 $t=0$ 时, $m_0=0$ 。在 NHPP 中, $N(t_k) - N(t_{k-1})$ 服从参数为 $m(t_k) - m(t_{k-1})$ 的泊松分布,可得

$$p\{N(t_k) - N(t_{k-1}) = m_k - m_{k-1}\} = \frac{[m(t_k) - m(t_{k-1})]^{m_k - m_{k-1}} \exp\{-[m(t_k) - m(t_{k-1})]\}}{(m_k - m_{k-1})!} \quad (4)$$

得到分组数据下 DSS 模型的似然函数为

$$\begin{aligned} L(\theta) &= \exp\left\{-a[1 - (1+bt_n)e^{-bt_n}]\right\} \prod_{k=1}^n \frac{\left\{a[(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}]\right\}^{m_k - m_{k-1}}}{(m_k - m_{k-1})!} \\ &= e^{-a[1 - \exp(-bt_n)]} a^{m_n} \prod_{k=1}^n \frac{[(1+bt_{k-1})e^{-bt_{k-1}} - (1+bt_k)e^{-bt_k}]^{m_k - m_{k-1}}}{(m_k - m_{k-1})!} \end{aligned} \quad (5)$$

得到 DSS 模型的似然方程组为

$$\begin{cases} a = \frac{m_n}{1 - (1 + bt_n)e^{-bt_n}} \\ \sum_{k=1}^n \left[\frac{(m_k - m_{k-1})(bt_k^2 e^{-bt_k} - bt_{k-1}^2 e^{-bt_{k-1}})}{(1 + bt_{k-1})e^{-bt_{k-1}} - (1 + bt_k)e^{-bt_k}} \right] - abt_n^2 e^{-bt_n} = 0 \end{cases} \quad (6)$$

2.2. 小波分解重构

小波分解采用快速变换算法实现, 其中 Mallat 算法(塔式算法)和 ATrous 算法是两种主流方法。Mallat 算法通过逐层下采样实现多尺度分解, 计算效率高但可能引入频率混叠; 而 ATrous 算法通过保持数据长度不变, 避免下采样带来的信息损失, 适合需要精确重构的应用。本文选用 Mallat 算法, 因其在计算效率与分解精度间的综合表现更优。其中 Mallat 分解算法的数学表达如下

$$\begin{cases} W_{j+1} = G^* W_j \\ V_{j+1} = H^* V_j \end{cases} (j = 0, 1, 2, \dots, J-1) \quad (7)$$

式中: H^* 为低通滤波器;

G^* 为高通滤波器。

其设计直接影响分解效果。对于 Daubechies 小波, 滤波器系数可通过递归方程求解, 确保满足正交性和正则性条件。重构过程通过逆滤波器组实现

$$V_j = HV_{j+1} + GW_{j+1}, j = J-2, J-3, \dots, 0 \quad (8)$$

重构到原尺度上, 即得单支重构后的信号, 记为 $W_1, W_2, \dots, W_{J-1}$ 和 $V_1, V_2, \dots, V_{J-1}$ 。其中 H, G 分别是 H^*, G^* 的对偶算子。

2.3. 软件故障预测模型

记 $Y = \{X_1, X_2, \dots, X_N\}$ 是原始信号, 对其进行小波分解以及对各层时间序列分别进行单支重构

$$Y = W_1 + W_2 + \dots + W_{J-1} + V_{J-1} \quad (9)$$

式中: $W_j = \{W_{j,1}, W_{j,2}, \dots, W_{j,N}\}$ 为 j 层 ($j = 0, 1, 2, \dots, J-1$) 高频细节分量的单支重构结果, 表征信号的短期波动与噪声; $V_{J-1} = \{V_{J-1,1}, V_{J-1,2}, \dots, V_{J-1,n}\}$ 是第 $J-1$ 层的低频近似分量的单支重构结果, 反映系统失效的长期趋势。因此, 任一时刻的失效数据可分解为

$$X_i = W_{1,i} + W_{2,i} + \dots + W_{J-1,i} + V_{J-1,i} \quad (10)$$

2.3.1. 分量建模

低频趋势建模: 对 V_{J-1} 的累积序列建立 DSS 模型, 其均值函数表达式为

$$m_{low}(t) = a_{low} [1 - (1 + b_{low}t)e^{-b_{low}t}] \quad (11)$$

高频波动建模: 对每一高频分量建立适配的时间序列模型 $ARIMA(p, d, q)$, 具体步骤如下:

平稳化处理: 通过 ADF 检验确定差分阶数 d 。若 $p\text{-value} < 0.05$, 接受平稳性假设。

自相关分析: 根据 ACF 图和 PACF 图确定 p 和 q 。

模型诊断: 通过 Ljung-Box 检验验证残差是否为白噪声

$$Q = N(N+2) \sum_{k=1}^K \frac{\hat{p}_k^2}{N-k} \sim \chi^2(K-p-q) \quad (12)$$

若 $Q < \chi_{0.95}^2(K - p - q)$ ，接受原假设。

2.3.2. 融合预测

通过独立预测各分量的未来值 $\hat{W}_{1,t}$ 和 $\hat{V}_{J-1,t}$ ，叠加得到最终预测结果

$$\hat{X}_t = \hat{W}_{1,t} + \hat{W}_{2,t} + \cdots + \hat{W}_{J-1,t} + \hat{V}_{J-1,t} \tag{13}$$

该方法通过分离信号的时频特性，分别捕捉长期趋势与短期波动，从而提升非平稳失效数据预测精度。

3. 案例分析

3.1. 故障数据

本次开源软件系统故障采用的数据来自 Tomcat 服务器软件缺陷追踪系统 <https://bz.apache.org/bugzilla/>。Tomcat 是由 Apache 软件基金会下属的 Jakarta 项目开发的。具有成本低，且社区活跃，资源丰富，占用系统资源少，启动速度快，适合各种规模的项目，从小型的个人应用到大型的企业级系统都能胜任等优点。但是 Tomcat 服务器存在部分缺陷，导致存在软件故障。由于 Tomcat AJP 协议设计上的缺陷，攻击者通过 Tomcat AJP Connector 可以读取或包含 Tomcat 上所有 Webapp 目录下的任意文件。因此，提高开源软件系统可靠性具有重要意义。

本文采用的数据是来自 Tomcat 9 的用户缺陷跟踪系统的故障数据，故障数据中包含有：Bug ID、Product、Component、Assignee、Status、Resolution、Summary、Changed 等主要内容。分别表示意思如下：

- Bug ID: 故障的唯一标识符，用于区分不同的故障。
- Product: 出现故障的产品名称，这里均为 Tomcat 9。
- Component: 故障所在的组件。
- Assignee: 负责处理故障的人员或团队。
- Status: 故障的当前状态，如 NEW (新的)、REOPENED (重新打开的)等。
- Resolution: 故障的解决情况。
- Summary: 对故障的简要描述。
- Changed: 故障信息的最后更改时间。

对整体故障数据通过分组数据的方式进行筛选，按软件故障发生时间以月为单位提取了 Tomcat 9 2015 年 1 月到 2025 年 3 月一共 123 个月的失效数据，失效数代表具体月份中发生的故障总个数。开源软件 Tomcat 9 的失效数据见表 1。

Table 1. Failure data of the open source software Tomcat 9
表 1. 开源软件 Tomcat 9 失效数据

时间/月	失效数	时间/月	失效数	时间/月	失效数	时间/月	失效数	时间/月	失效数	时间/月	失效数
1	2	25	5	49	13	73	7	97	6	121	5
2	2	26	2	50	2	74	8	98	4	122	11
3	2	27	3	51	14	75	12	99	7	123	5
4	2	28	1	52	11	76	5	100	6		
5	0	29	4	53	8	77	7	101	6		
6	5	30	9	54	6	78	14	102	6		
7	0	31	5	55	13	79	10	103	3		

续表

8	1	32	3	56	7	80	7	104	5
9	1	33	10	57	16	81	10	105	8
10	1	34	6	58	15	82	3	106	9
11	8	35	9	59	17	83	5	107	7
12	1	36	7	60	13	84	3	108	1
13	1	37	7	61	10	85	11	109	4
14	4	38	6	62	10	86	4	110	10
15	6	39	6	63	18	87	7	111	11
16	6	40	6	64	10	88	11	112	8
17	8	41	7	65	8	89	13	113	6
18	6	42	11	66	15	90	7	114	13
19	1	43	7	67	5	91	1	115	25
20	7	44	12	68	7	92	10	116	7
21	1	45	10	69	13	93	9	117	6
22	6	46	9	70	9	94	9	118	21
23	7	47	11	71	14	95	8	119	10
24	2	48	7	72	8	96	5	120	1

将前 100 个月的失效数据作为训练集，用来估计模型的未知参数，后 23 个失效数据作为预测集，用于检测模型的预测效果，并对 2025 年 3 月到 12 月的 Tomcat 9 故障数进行预测。

由原故障数据可以得知，Tomcat 9 的失效数据由多个组件的失效数据构成。本文对前 100 个月的失效数据用分组件的方式来分别进行未知参数的计算。通过对原数据的筛选，选出前两个失效数最多的组件，分别为组件 Catalina 和组件 Connectors，剩余的组件用 Others 来代替。

3.2. 小波分解处理

小波变换作为一种强大的时频分析工具，在处理非平稳时间序列数据方面具有独特优势。由图 1 可知 Tomcat 9 的失效数据序列是一个非平稳过程，传统分析方法难以有效捕捉其中的规律。因此，采用小波分解与重构技术对数据进行处理。对 Tomcat 9 的月失效数据利用 Daubechies 小波进行小波分解，本文选择 Daubechies 4 小波(db4)进行分解，其支撑长度为 4，具有 4 阶消失矩，既能捕捉信号的局部突变特征，又能有效抑制边界效应。

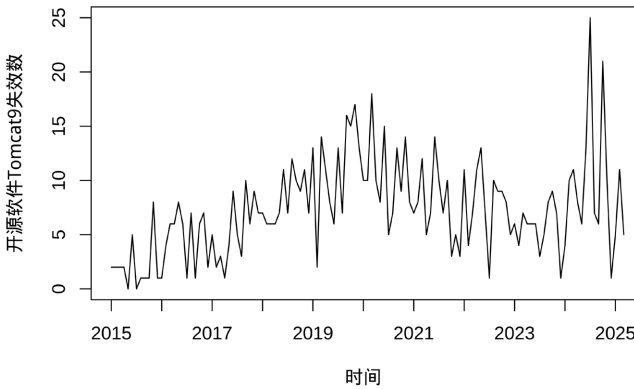


Figure 1. The number of failures of the open source software Tomcat 9 from January 2015 to March 2025
图 1. 2015 年 1 月至 2025 年 3 月的开源软件 Tomcat 9 的失效数

以组件 Catalina 的数据处理为例，由于小波分解算法对数据长度有特定要求，通常需要数据长度为 2 的幂次方。因此，要确定最接近且大于该长度的 2 的幂次方，本文通过在原始数据末尾补零的方式，将数据长度扩展，以满足小波分解的输入要求。然后，利用对扩展后的序列进行小波分解，得到小波系数。通过计算获取分解的层数(本文中 Catalina 月度失效数据(长度 $N = 100$)，分解层数 $J = 4$)。

在小波系数中，低频系数反映了数据的总体趋势，高频系数则包含了数据的细节和噪声信息。为提取低频趋势分量，采用软阈值方法对小波系数进行处理。高频波动分量则通过原始序列减去低频趋势分量获得。处理完成后，将低频和高频分量截断回原始数据长度，以去除因补零操作引入的额外数据点。

识别出高频波动分量的时间序列模型为 ARIMA(3,0,0)，其函数表达式为

$$m_{high}(t) = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} + c \tag{14}$$

3.3. 模型评估

3.3.1. 参数估计

Tomcat 9 组件的累计失效数分别通过传统 DSS 和小波分析后计算未知参数，组件 Catalina、组件 Connectors 和组件 Others 的未知参数估计值由表 2 和表 3 所示。

Table 2. Parameter estimates of Tomcat 9 components

表 2. Tomcat 9 组件的参数估计值

组件	传统参数 a	传统参数 b	小波低频参数 a'	小波低频参数 b'
Catalina	1069.618	0.01184861	998.6647	0.01186058
Connectors	373.711	0.0105276	306.8063	0.0106954
Others	742.9984	0.01208692	730.2753	0.01208636

Table 3. AR coefficients of the fusion model for Tomcat 9 components

表 3. Tomcat 9 组件的融合模型 AR 系数

组件	AR 系数 ϕ_1	AR 系数 ϕ_2	AR 系数 ϕ_3
Catalina	1.994044	-1.398423	0.3084136
Connectors	1.957477	-1.42516	0.3695355
Others	1.864988	-1.309163	0.3101134

则组件基于融合小波模型函数为

$$m_{wavelet}(t) = a' \left[1 - (1 + b't) e^{-b't} \right] + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \phi_3 W_{t-3} \tag{15}$$

3.3.2. 模型性能对比

均方误差(Mean Squared Error, MSE)是一种衡量预测值与真实值之间差异的指标，其值用 M 表示， M 越小表示模型拟合度越好。对于一组预测值 $\hat{y}_i$ 和真实值 y_i ， $i = 1, 2, \dots, n$ ，则均方误差计算公式为

$$M = \frac{1}{n} \sum_{k=1}^n \left[m(t_k) - m_k \right]^2 \tag{16}$$

赤池信息准则(Akaike Information Criterion, AIC)是用来衡量统计模型拟合优良性的一种标准。其值用 A 表示， A 越小表示模型拟合度越好。计算公式为

$$A = 2K - 2\log L \quad (17)$$

式中: K 为模型参数数量; $\log L$ 为模型的对数似然值。

相对误差(Relative Error, RE)是一种衡量测量值与真实值之间差异的指标,是绝对误差与真实值的比值,其值用 E 表示。相对误差越小,说明预测值越接近真实值,预测的准确度越高。RE 的计算公式为

$$E = \frac{m(t_q) - m_q}{m_q} \quad (18)$$

式中: $m(t_q)$ 是模型的估计值; m_q 是真实值。

MAE (Mean Absolute Error, 平均绝对误差)和 MAPE (Mean Absolute Percentage Error, 平均绝对百分比误差)用于衡量模型预测值与实际值之间的偏差程度,计算公式分别为

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (19)$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (20)$$

式中: n 为样本数量;

y_i 为第 i 个样本的实际值;

$\hat{y}_i$ 为第 i 个样本的预测值。

本次研究的模型评估分为两部分,分别是基于 NHPP 参数估计的模型 MSE、AIC、MAE 和 MAPE,以及融合小波分析参数估计的模型 MSE、AIC、MAE 和 MAPE。通过对比两种方法的 M 值和 A 值,可以得出拟合度较好的模型。

通过上节计算出的参数估计值,可以分别求出每组的评估值,由表 4 所示。

Table 4. Model comparison results

表 4. 模型比较结果

组件	DSS MSE	WTDSS MSE	DSS AIC	WTDSS AIC	DSS MAE	DSS MAEP	WTDSS MAE	WTDSS MAEP
Catalina	452.2688	404.1873	615.4277	608.1878	18.80676	21.76459%	18.44811	24.0507%
Connectors	12.70266	10.16581	258.1812	239.903	2.780294	\	2.678012	\
Others	203.0481	186.3737	535.3443	530.7754	12.03573	53.78229%	11.71335	52.3908%

可以看出,相较于 DSS 模型,本文所提出的融合小波变换和非齐次泊松过程的开源软件故障预测模型(Wavelet Transform DSS, WTDSS)的 M 值、 A 值和 MAE 会更小,这说明融合模型的故障拟合效果更好。

进一步对 DSS 与 WTDSS 进行故障拟合效果对比以及故障预测效果对比,由图 2 和图 3 所示。

由图 3 可以看出, WTDSS 相较于 DSS, 有更优秀的未来故障预测能力,使得故障预测更加准确可靠。

3.3.3. 故障预测

根据式(3),可以得到基于 Tomcat9 数据下的融合小波变换和 NHPP 的模型,即本文所提出的融合小波变换和非齐次泊松过程的开源软件故障预测模型(WTDSS),其表达式为

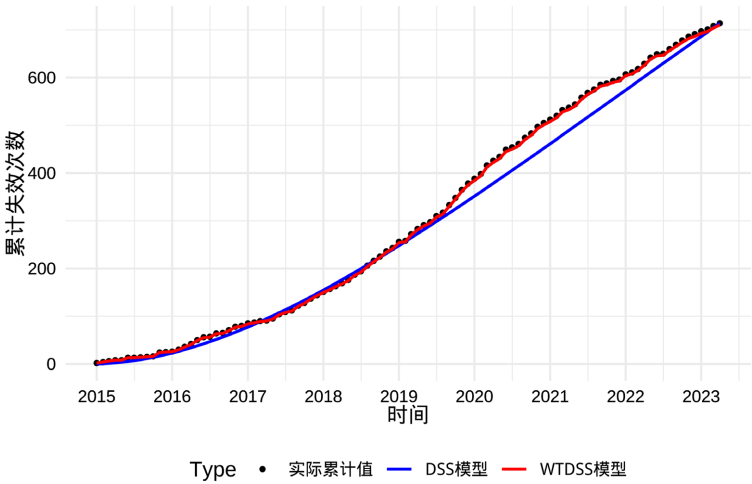


Figure 2. Fault fitting effect diagram
图 2. 故障拟合效果图

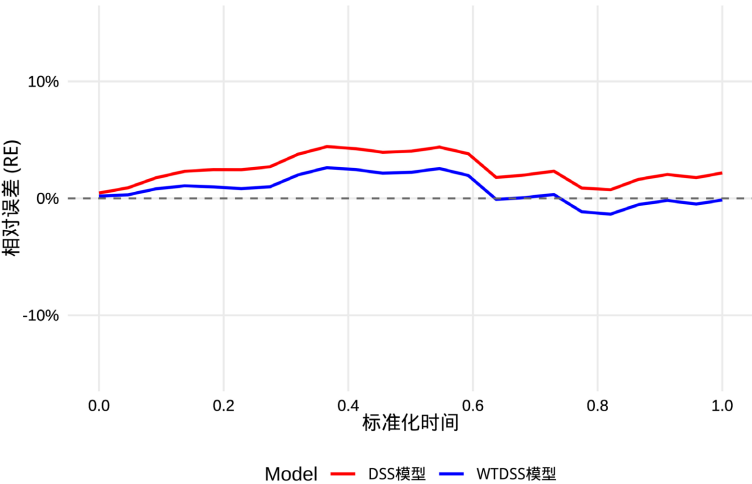


Figure 3. Fault prediction effect diagram
图 3. 故障预测效果图

$$M(t)=\sum_{i=1}^nm_{wavelet_i}(t)$$

(21)

分别用 DSS 和 WTDSS 对 Tomcat 9 2025 年 4 月到 12 月的故障数进行分组件故障预测，得到的结果由表 5 所示。

Table 5. AR coefficients of the fusion model for Tomcat 9 components
表 5. Tomcat 9 组件的融合模型 AR 系数

月份	Catalina		Connectors		Others		总故障数	
	故障数		故障数		故障数		故障数	
	DSS	WTDSS	DSS	WTDSS	DSS	WTDSS	DSS	WTDSS
4	4.29	3.22	1.39	0.16	3.01	2.7	8.69	6.08
5	4.28	3.05	1.39	0.25	3	2.95	8.67	6.26

续表

6	4.26	2.84	1.39	0.36	2.99	2.95	8.64	6.16
7	4.24	2.85	1.38	0.58	2.98	2.95	8.6	6.39
8	4.23	3.05	1.38	0.84	2.96	2.93	8.57	6.82
9	4.21	3.33	1.38	1.05	2.95	2.91	8.54	7.3
10	4.19	3.63	1.37	1.19	2.94	2.89	8.5	7.72
11	4.17	3.9	1.37	1.25	2.93	2.88	8.47	8.03
12	4.16	4.09	1.36	1.25	2.91	2.86	8.43	8.21

由表 5 可得, Catalina 作为系统的核心组件, 发生的故障数占绝大部分。通过对比 DSS 和 WTDSS 的预测结果可以看出, WTDSS 的预测存在小幅波动, 更加精准。各组件的累计预测数对比图由图 4~6 所示。

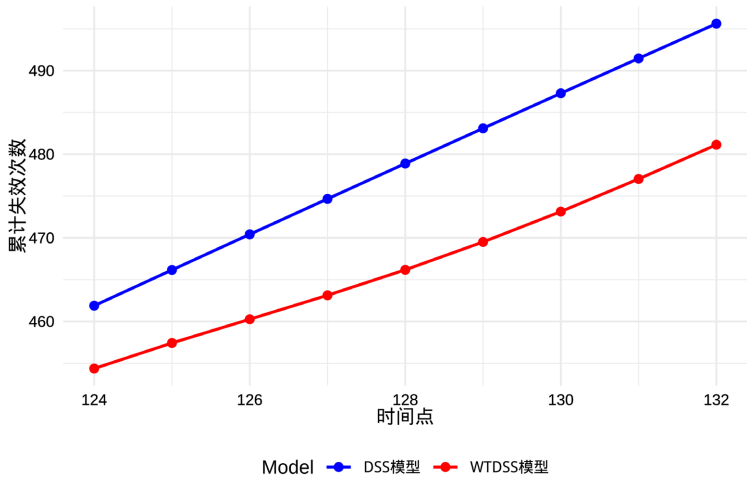


Figure 4. Prediction comparison chart of the Catalina component
图 4. Catalina 组件预测对比图

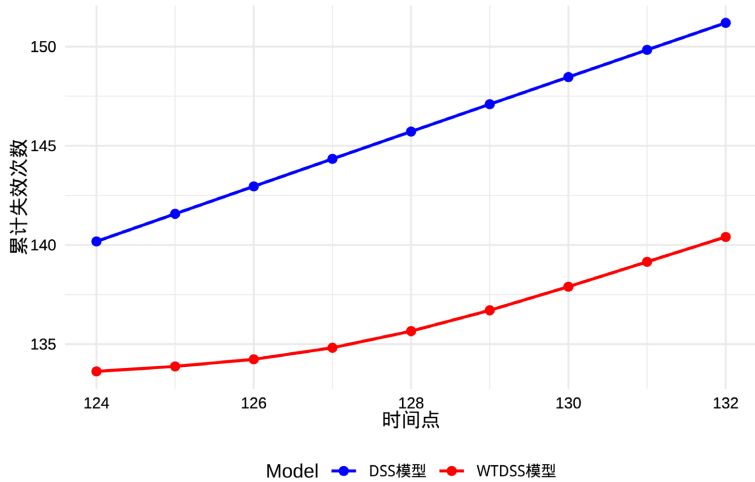


Figure 5. Prediction comparison chart of the Connectors component
图 5. Connectors 组件预测对比图

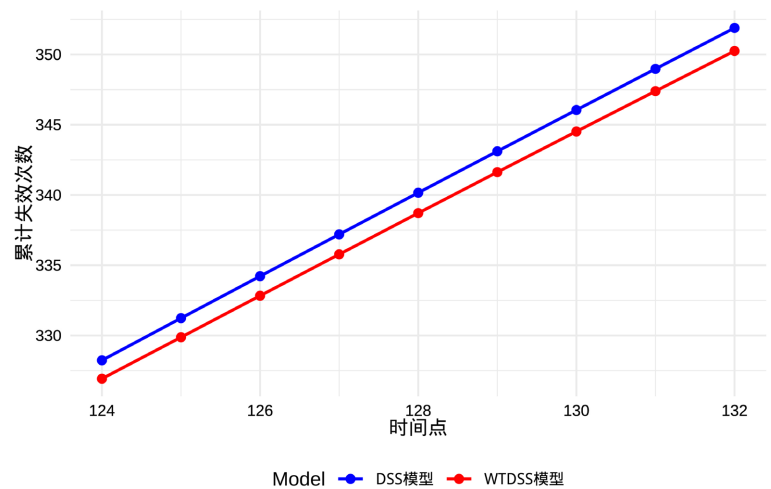


Figure 6. Prediction comparison chart of the Connectors component
图 6. Connectors 组件预测对比图

另外，用 WTDSS 对系统整体的故障数进行预测，得到的理论故障数由表 6 所示。

Table 6. AR coefficients of the fusion model for Tomcat 9 components
表 6. Tomcat 9 组件的融合模型 AR 系数

月份	4	5	6	7	8	9	10	11	12
故障数	4	6	11	13	13	12	8	6	6

预测结果如图 7 表示。

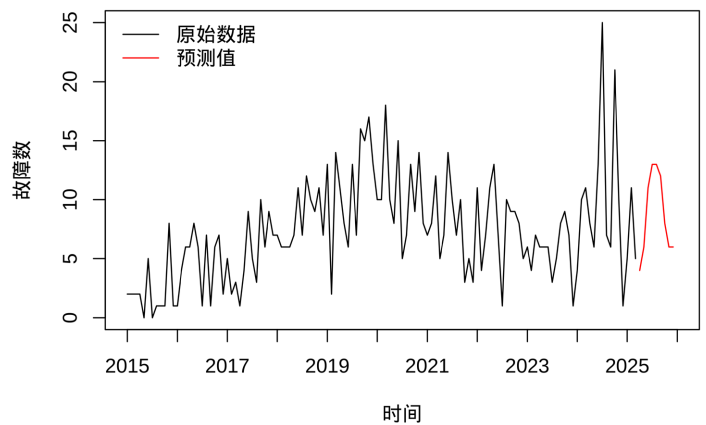


Figure 7. Prediction chart of Tomcat 9 from April to December 2025
图 7. Tomcat 9 2025 年 4 月至 12 月预测图

4. 结论与展望

本论文成功构建了融合小波变换和非齐次泊松过程的开源软件故障预测模型(WTDSS)，有效解决了传统模型在处理开源软件故障数据时面临的难题，显著提升了预测效果。首先利用极大似然估计法对 DSS 进行参数估计，并通过小波分析对数据处理后再次进行参数估计，所得参数估计值合理。接着通过对 Tomcat 9 故障数据的分析，WTDSS 在 MSE、AIC 和 MAE 指标上均优于 DSS，表明 WTDSS 能较好地适

应故障数据的非平稳性,有效预测未来故障趋势。

鉴于本文所选的数据集限制,有可能存在小样本量对模型稳定性的影响,在未来相关研究中可以采用更加复杂的数据集进行模型训练提高模型稳定性。

基金项目

国家自然科学基金项目(72361008),广西自然科学基金面上项目(2025GXNSFAA069686),贵州省科技计划项目(编号:黔科合支撑[2023]一般 268)。

参考文献

- [1] 陈静,杨剑锋,王喜宾,等. NHPP 类开源软件可靠性增长模型的极大似然估计[J]. 广西大学学报(自然科学版), 2022, 47(1): 174-184.
- [2] Das, S., Kundu, D. and Dewanji, A. (2022) Software Reliability Modeling Based on NHPP for Error Occurrence in Each Fault with Periodic Debugging Schedule. *Communications in Statistics—Theory and Methods*, **51**, 4890-4902. <https://doi.org/10.1080/03610926.2020.1828462>
- [3] 吕勋. 基于 NHPP 的软件可靠性模型研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2017.
- [4] 周波. 基于 NHPP 软件可靠性模型的预测研究及实现[D]: [硕士学位论文]. 成都: 电子科技大学, 2016.
- [5] Kim, Y.S., Song, K.Y. and Chang, I.H. (2023) Prediction and Comparative Analysis of Software Reliability Model Based on NHPP and Deep Learning. *Applied Sciences*, **13**, Article 6730. <https://doi.org/10.3390/app13116730>
- [6] Aggarwal, A., Kumar, S. and Gupta, R. (2024) Testing Coverage Based NHPP Software Reliability Growth Modeling with Testing Effort and Change-Point. *International Journal of System Assurance Engineering and Management*, **15**, 5157-5166. <https://doi.org/10.1007/s13198-024-02504-7>
- [7] Lin, J.C. and Zhu, A.M. (2022) CNN-Wavelet-Transform-Based Model for Solar Photovoltaic Power Prediction. *19th IEEE International Computer Conference on Wavelet Active Media Technology and Information Processing*, Chengdu, 16-18 December 2022, 1-5.
- [8] Serhal, H., Abdallah, N., Marion, J., Chauvet, P., Oueidat, M. and Humeau-Heurtier, A. (2022) Wavelet Transformation Approaches for Prediction of Atrial Fibrillation. *2022 30th European Signal Processing Conference (EUSIPCO)*, Belgrade, 29 August-2 September 2022, 1188-1192. <https://doi.org/10.23919/eusipco55093.2022.9909695>
- [9] Cheng, R., Liu, Z. and Zhang, M. (2022) An Improved Traffic Flow Prediction Model: Spatial-Temporal Network Based on Wavelet and LSTM. *2022 16th IEEE International Conference on Signal Processing (ICSP)*, Beijing, 21-24 October 2022, 276-280. <https://doi.org/10.1109/icsp56322.2022.9965226>
- [10] Wang, Y.T., Liu, J., Li, R., Suo, X.Y. and Lu, E.H. (2022) Medium and Long-Term Precipitation Prediction Using Wavelet Decomposition-Prediction-Reconstruction Model. *Water Resources Management*, **36**, 971-987. <https://doi.org/10.1007/s11269-022-03063-x>
- [11] Yuan, M.X., Wei, S.K., Sun, M. and Zhao, J.D. (2022) Wavelet Decomposition and Seq2Seq Hybrid Models for Water Quality Prediction. *Water Resources*, **49**, 743-752. <https://doi.org/10.1134/s0097807822040212>
- [12] Feng, Q.M. and Qian, S.P. (2022) Research on the Prediction of Short-Term Wind Power Based on Wavelet Neural Networks. *Energy Reports*, **8**, 553-559. <https://doi.org/10.1016/j.egyr.2022.05.081>
- [13] Yu, C.J., Li, Y.L., Chen, Q., Lai, X.P. and Zhao, L.Y. (2022) Matrix-Based Wavelet Transformation Embedded in Recurrent Neural Networks for Wind Speed Prediction. *Applied Energy*, **324**, Article 119692. <https://doi.org/10.1016/j.apenergy.2022.119692>
- [14] Nasser, A. and Simon, V. (2022) Wavelet-Attention-Based Traffic Prediction for Smart Cities. *IET Smart Cities*, **4**, 3-16. <https://doi.org/10.1049/smc2.12018>
- [15] Samani, S., Vadiati, M., Delkash, M. and Bonakdari, H. (2023) A Hybrid Wavelet-Machine Learning Model for Qanat Water Flow Prediction. *Acta Geophysica*, **71**, 1895-1913. <https://doi.org/10.1007/s11600-022-00964-8>
- [16] Dincer, N.G., Yalcin, M.O. and Guneri, O.I. (2022) A Hybrid Time Series Prediction Model Based on Fuzzy Time Series and Maximal Overlap Discrete Wavelet Transform. *Gazi University Journal of Science*, **35**, 1152-1169. <https://doi.org/10.35378/gujs.798423>