



码头结构健康监测传感器故障数据的识别与修复研究

林 猛¹, 高刚刚²

(1. 中交三航局第六工程(厦门)有限公司, 福建 厦门 361006;
2. 中交上海三航科学研究院有限公司, 上海 200032)

摘要: 针对码头结构健康监测系统中因传感器故障导致数据异常、进而影响结构安全评估准确性的问题, 提出一种基于改进箱型图法的故障数据识别与修复方法。通过融合改进箱型图法与散点图分析技术, 构建了包含无故障、尖峰故障、漂移故障和偏置故障 4 类数据的统计特征识别模型, 实现了故障类型的精确分类。在试验验证阶段, 该方法在 56 组实测数据段中表现出 91.07% 的故障识别准确率。针对不同故障类型特征, 研究分别设计了差异化修复策略: 采用拉格朗日插值法处理尖峰故障, 运用移动窗口均值拟合校正漂移和偏置故障。最后, 通过修复前后数据的自相关性对比, 验证了数据修复的有效性。研究成果为码头结构健康监测系统提供了可靠的数据质量保障方法, 对提升重大基础设施安全监测水平具有重要的实践意义。

关键词: 结构监测; 传感器故障数据; 数据异常修复; 箱型图法

中图分类号: U653.5

文献标志码: A

文章编号: 1002-4972(2025)12-0054-08

Identification and correction of sensor fault data for dock structural health monitoring

LIN Meng¹, GAO Ganggang²

(1. China Communications Third Navigation Bureau Sixth Engineering (Xiamen) Co., Ltd., Xiamen 361006, China;
2. CCCC Shanghai Third Harbor Engineering Research Institute Co., Ltd., Shanghai 200032, China)

Abstract: To address the issue of data anomalies in the structure health monitoring system of a dock due to sensor failures, which subsequently affects the accuracy of structural safety assessment, a fault data identification and correction method is proposed based on an improved box plot method. By integrating the improved box plot method with scatter plot analysis technique, a statistical feature recognition model is constructed that includes four types of data: no fault, spike fault, drift fault, and bias fault, achieving precise classification of fault types. In the experimental verification phase, the method exhibits a 91.07% fault recognition accuracy rate among 56 sets of actual measurement data segments. For different fault types, the research designs differentiated correction strategies: using Lagrange interpolation for spike faults, applying moving window mean fitting to correct drift and bias faults. Finally, by comparing the autocorrelation of data before and after correction, the effectiveness of data correction is verified. The research results provide reliable data quality assurance method for the dock structure health monitoring system and have significant practical significance for improving the level of safety monitoring of major infrastructure.

Keywords: structural monitoring; sensor fault data; data anomaly correction; boxplot method

当前, 码头结构的全寿命周期健康监测已成为国内工程界关注的焦点, 其核心目标在于实现评估结构的损伤程度及其剩余寿命^[1-2]。在港口工

程结构健康监测系统中, 传感设备的长期服役性能是决定评估可信度的关键要素^[3]。然而, 由于码头结构所处的特殊环境, 传感器需长期暴露于

恶劣条件下, 相较于其他应用场景, 其发生故障的概率显著提高, 这对码头结构健康监测工作造成了严峻挑战^[4]。

一旦原始数据中出现严重的传感器故障, 将对结构健康监测分析工作以及基于此的后续决策制定与状态评估产生不容轻视的负面影响。当前, 传感器主要存在传感器失效与传感器故障两类问题^[5]。传感器失效表现为传感器节点无法回传采集的数据或对用户指令毫无响应^[6], 如数据丢失等常见情形, 这种情况下在埋入式传感器布置场景中通常只能弃用。而传感器故障则是节点能正常报告数据但传输数据在某些方面受损导致数据质量下降, 可通过数据修复技术恢复其功能。

尽管现有研究已对传感器故障的分类和处理有所涉及, 但在如何从有效角度准确识别和处理传感器故障, 以提升码头结构健康监测结果的说服力方面, 仍有进一步研究的空间。基于此, 本文研究主要从基于数据的角度出发, 深入探讨传感器故障引起数据异常的类型, 并针对性地提出解决方案, 旨在为提高码头结构健康监测的准确性和可靠性提供理论与实践支持。

1 监测数据异常分类

传感器故障可以通过两种方式进行检查: 以系统为中心和以数据为中心。从以系统为中心的角度来看, 传感器故障是根据故障数据的根本原因进行分析的, 可以分为校准故障、连接故障和电源电压低等。从以数据为中心的角度出发, 基于传感器测量数据的特征来研究传感器故障, 通常分为尖峰、漂移、偏置。本文研究主要将从以数据为中心的角度解决传感器故障。

已有研究表明, 振动测量中报告了一些因传感器故障而损坏的数据集, 损坏的传感器数据具有三种类型故障之一: 漂移、尖峰和偏置^[7]。高桩码头结构监测数据中传感器的典型故障见图 1。漂移故障是最常见的传感器故障, 其中大部分是由温度影响引起的。尖峰故障是第 2 常见的故障,

可能由多个问题引起, 如电池电源不足、电气接触松动^[8]和传感器饱和^[9]等。具体来说, 当超过测量范围时, 信号将在原始时间历程数据中被削波, 如果应用低通滤波器对信号进行采样, 则在信号被削波的地方会出现尖峰。偏置故障相对较少, 可能是由安装问题引起的。

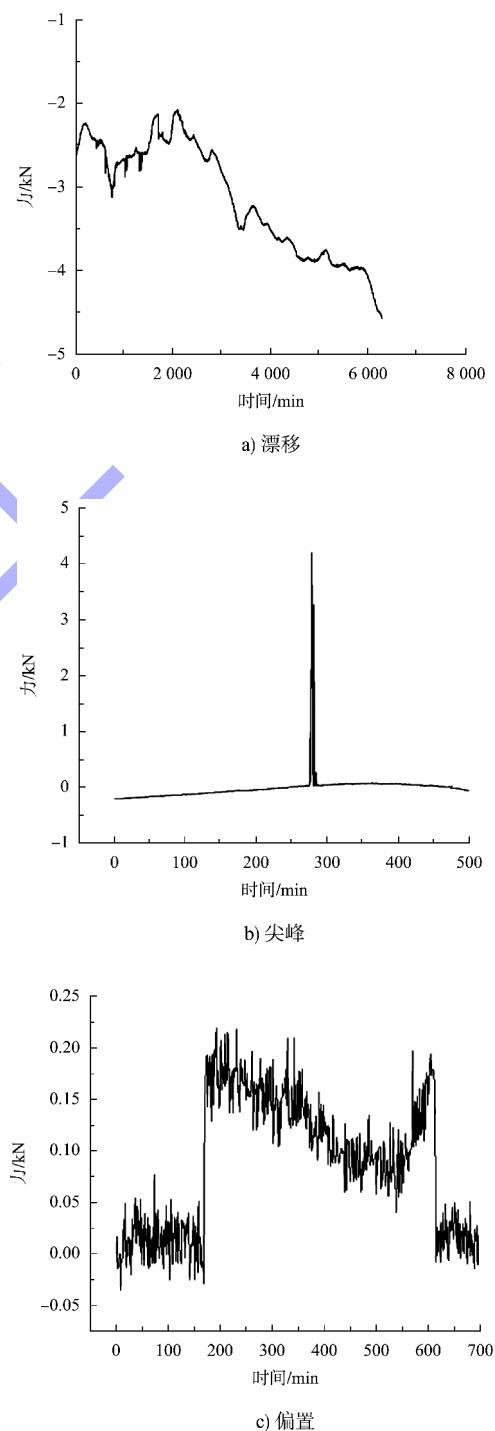


图 1 健康监测中的故障数据分类
Fig. 1 Fault data classification in health monitoring

2 基于改进箱型图法的异常数据段的识别

对于高桩码头结构监测数据，通过提取数据的统计特征为故障数据建立标签。总共考虑 50 组固定长度 180 min 的数据段，手动将其分配为 4 种标签之一：无故障数据、带有尖峰故障的数据、带有漂移故障的数据和带有偏置故障的数据，其中无故障数据 20 组，其余故障数据各 10 组。本文依据改进箱型图法对故障数据和无故障数据段进行分析，确定不同标签的数据段所具有的特征，箱型图见图 2。

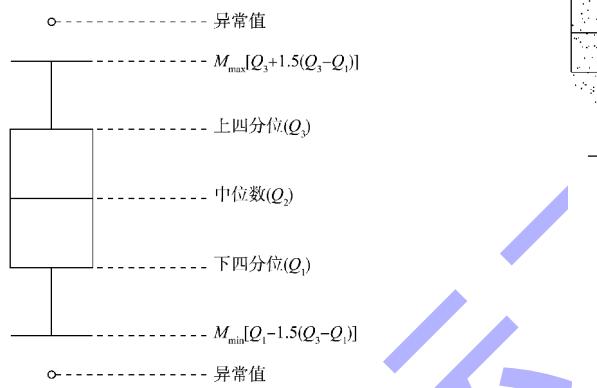
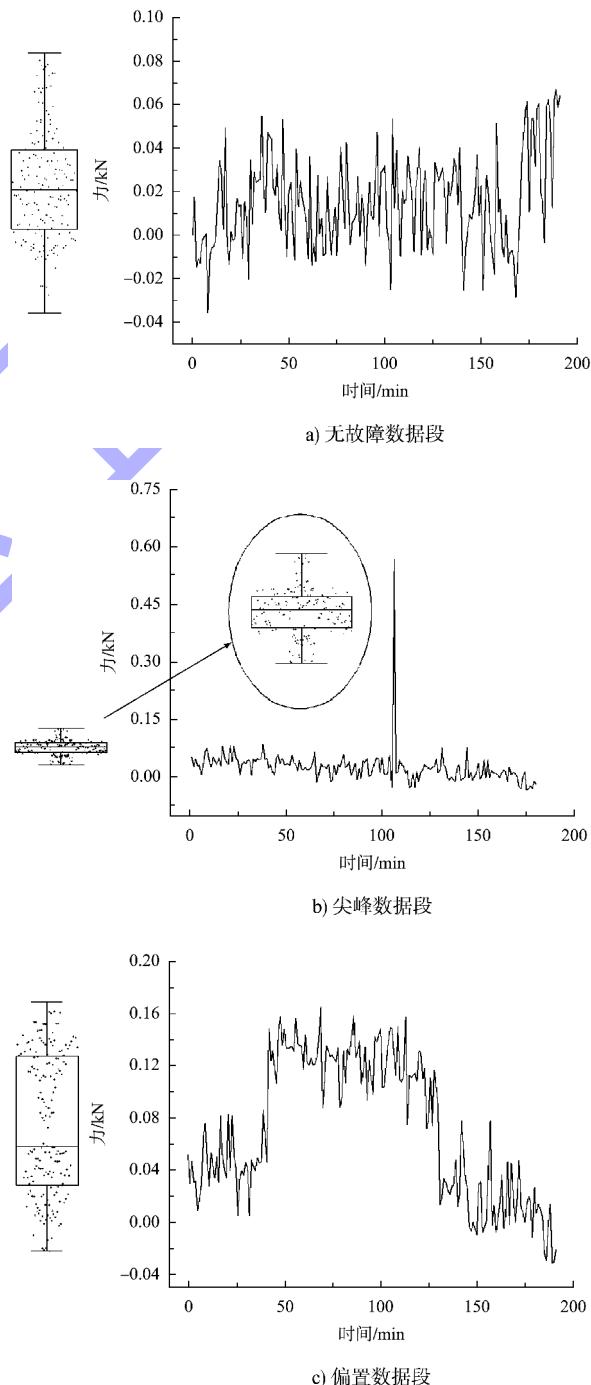


图 2 箱型图

Fig. 2 Box plot diagram

箱型图又称为箱线图，是一种用于显示一组数据分布情况的统计图，它的优点在于能够以一种非常紧凑和直观的方式概括数据的中心位置、离散程度、偏态和异常值，同时方便比较不同数据集的分布特性^[10]。它可以展示数据的最大值 M_{\max} 、最小值 M_{\min} 、中位数 Q_2 、下四分位数 Q_1 和上四分位数 Q_3 。图 2 中，中位数表示数据集的中位数，在箱型图中通常用一条线表示，它将箱子分为两部分；下四分位数 Q_1 表示 25% 的数据小于此值；上四分位数 Q_3 即表示 75% 的数据小于此值。最大值与最小值用以识别数据集中的异常值，最大值 $M_{\max} = Q_3 + 1.5(Q_3 - Q_1)$ ，最小值 $M_{\min} = Q_1 - 1.5(Q_3 - Q_1)$ ，小于最小值和大于最大值的数据会被认定为异常值^[11]。

本文通过箱型图结合散点图的方式对传统的箱型图进行改进，一方面可以通过箱型图识别出数据异常值，另一方面可通过数据散点在箱型图中的分布情况对 4 种数据段进行识别。通过 50 组固定长度为 180 min 的数据段，采用改进的箱型图法对数据段进行分析，这里仅展示不同标签数据段的一组数据，见图 3。



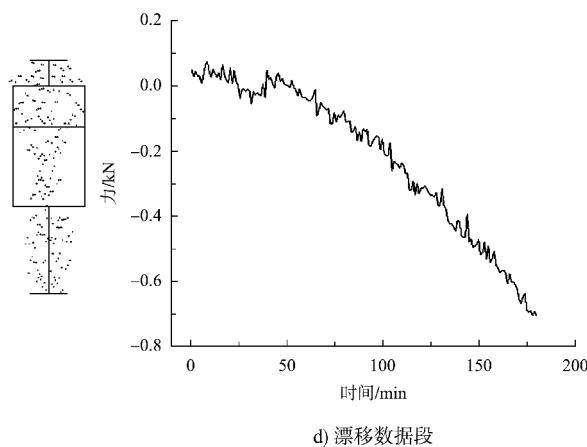


图 3 4 种数据段识别

Fig. 3 Identification of four types of data segment

通过对比不同数据段下改进箱型图和折线图的组合图示, 分析得出:

1) 无故障数据段, 箱型图分布较为平稳, 数据散点集中分布在箱体内。此数据分布为无偏态的数据分布, 近似于正态分布。在无荷载输入的情况下, 数据波动为随机波动, 数据分布为正态分布, 符合实际情况。

2) 对于带有尖峰故障的数据段, 峰值点在箱型图中会被准确识别为异常值, 且排除异常值后的箱型图中的数据分布特征与无故障数据分布相似。在个别情况下, 180 min 数据段具有尖峰的数据可能不止一个, 筛选出具有多个尖峰的非典型数据段用箱型图法进行识别, 见图 4。对于多峰故障数据段, 箱型图法也能准确识别出异常值。在选用合适的样本数据情况下, 箱型图法对于单个异常值的识别效果受数据值的大小变化而受到的影响可忽略不计, 对于异常值的识别界限大概为 2.698σ (σ 为数据样本总量的标准差), 近似于拉伊达准则(3σ 准则), 但相对于拉伊达准则, 箱型图法具有计算量少的优势。

3) 对于偏置数据段, 散点图显示数据集中在箱体上下四分位线附近, 箱型图数据具有明显的上偏且无异常值。在统计特征上, 这类数据段的中位数 Q_2 会显著偏移。

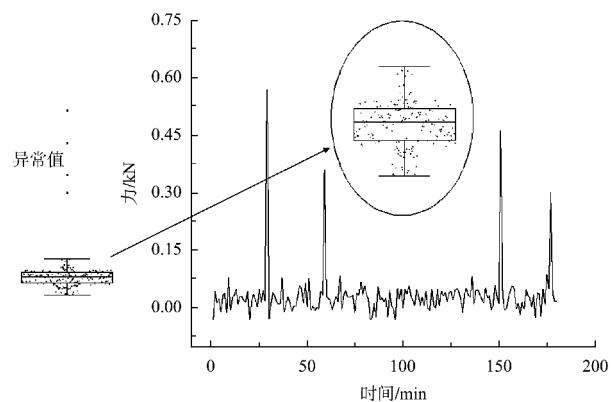


图 4 多峰故障数据段识别

Fig. 4 Identification of multi-peak fault data segment

4) 漂移数据段散点均匀分布在箱型图中而不是集中分布在箱体内, 且会因数据漂移节点和上下漂移的方向不同而产生上偏或下偏, 甚至无偏。在统计特征上, 这类数据段的离散性会偏大, 四分位数距($Q_3 - Q_1$)的差值会偏大。

对于以上 4 类数据段, 依据改进箱型图法可以实现准确的识别。为进一步实现 4 类数据段大批量的识别, 引入箱型图法的四个统计特征参数: 异常值、值域范围($M_{\max} - M_{\min}$)、中位数 Q_2 和四分位数距($Q_3 - Q_1$)。提取 50 组已建立标签的数据统计特征, 通过对故障数据与无故障数据统计特征参数的差异性进行区分, 见表 1, 统计参数取组中数据的众数值。表中以无故障数据的统计指标为标准, 异常值的存在与否决定是否为显著差异, 其余 3 个指标采用式(1)计算出相对差异率 Y , 以相对差异率 $0 \sim 30\%$ 、 $30 \sim 50\%$ 、 $50 \sim 100\%$ 为界限分为无差异、有差异、显著差异。

$$Y = \frac{2 |P_f - P_n|}{|P_f + P_n|} \times 100\% \quad (1)$$

式中: P_f 为故障数据统计特征参数; P_n 为无故障数据统计特征参数。

表 1 故障数据段差异性对比

Tab. 1 Comparison of differences in fault data segments

数据段	异常值	值域范围 ($M_{\max} - M_{\min}$)	中位数 Q_2	四分位数距 ($Q_3 - Q_1$)
尖峰数据段	×	√	√	√
偏置数据段	√	○	○和×	○
漂移数据段	√	×	○	×

注: “√”表示无差异; “○”表示有差异; “×”表示存在显著差异。

由表 1 可以得出不同数据段的故障数据识别方法。对于尖峰数据段可以通过箱型图法识别异常值的方式确定；对于漂移数据段，可由值域范围和四分位数距的显著差异性确定；对于偏置数据段，一方面可以通过分析中位数是否具有显著差异确定，另一方面通过确定值域范围和四分位数距是否具有差异性与漂移数据进行区分。

确定 3 类故障数据段的统计特征参数界限值后，建立故障数据段识别模式，见图 5。通过该模型对时间跨度为 1 d 的数据段进行故障识别，共 56 组数据段，其中 49 组数据段被准确识别出，故障识别准确率为 91.07%。由此可见，基于箱型图原理对故障数据段建立的统计特征参数界限值是有效的。但因特征提取所用样本量过少以及相对差异性的界定范围不够准确，导致故障数据识别存在一定的误差。

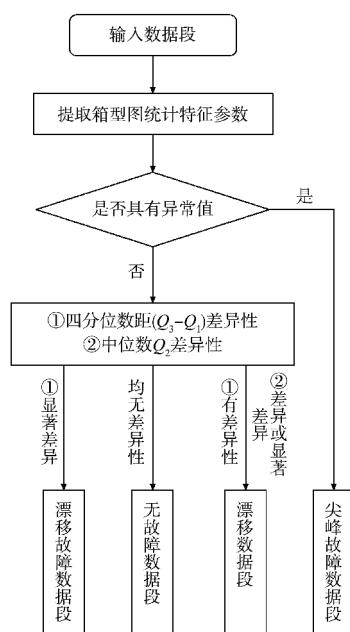


图 5 故障数据段识别流程

Fig. 5 Identification process for fault data segment

3 异常数据段的修复

在成功识别传感器故障类型后，可以采取针对性的手段对故障数据段进行修复。使用不同的策略修复不同的故障。对于具有尖峰数据的故障

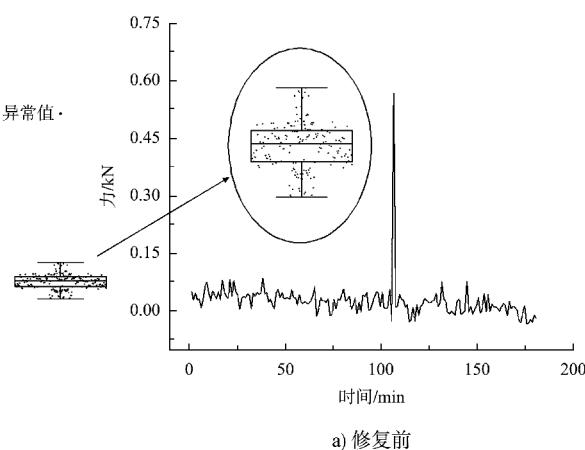
恢复，通过改进箱型图法识别出时域中尖峰故障的位置后，可以剔除该异常值再采用拉格朗日多项式插值法进行修复。拉格朗日多项式插值法是数值分析中的一种插值方法，即构建通过一组离散点的多项式函数进行插值。拉格朗日插值多项式的基本思想是通过已知的 $n+1$ 个数据点构造一个最高次数为 n 的多项式，该多项式恰好在这些数据点上取得相应的函数值。对于 $n+1$ 个样本点 $(x_0, y_0), (x_1, y_1), \dots, (x_n, y_n)$ 。可以通过给定时间节点计算出该节点的估计值，见式(2)、(3)。

$$L_n(x) = \sum_{i=0}^n y_i l_i(x) \quad (2)$$

$$l_i(x) = \prod_{j=0, j \neq i}^n (x - x_j) / (x_i - x_j) \quad (3)$$

式中： $L_n(x)$ 为拉格朗日多项式， $n+1$ 为数据点总数； x_i, x_j 为不同的时间节点； y_i, y_j 为不同测点数据； $l_i(x)$ 为基函数。

通过拉格朗日多项式进行插值，可以实现对剔除异常值后的数据点修复。对于原始数据段，测值数据含有一定的误差，在使用拉格朗日多项式进行插值时，应注意函数次项 n 的选取。使用较高次数的多项式可能会放大误差，造成插值修复的不理想。在取 $n=2$ 时，选取异常值附近 5 个单位的数据点构建拉格朗日多项式进行插值，插值效果见图 6。修复后的数据可再次使用箱型图法进行数据故障识别或检验数据修复质量。



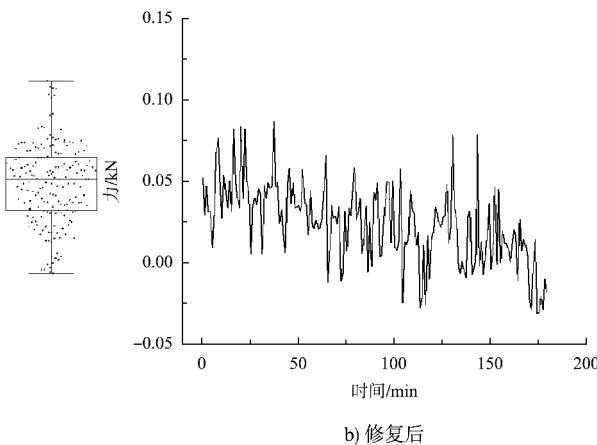
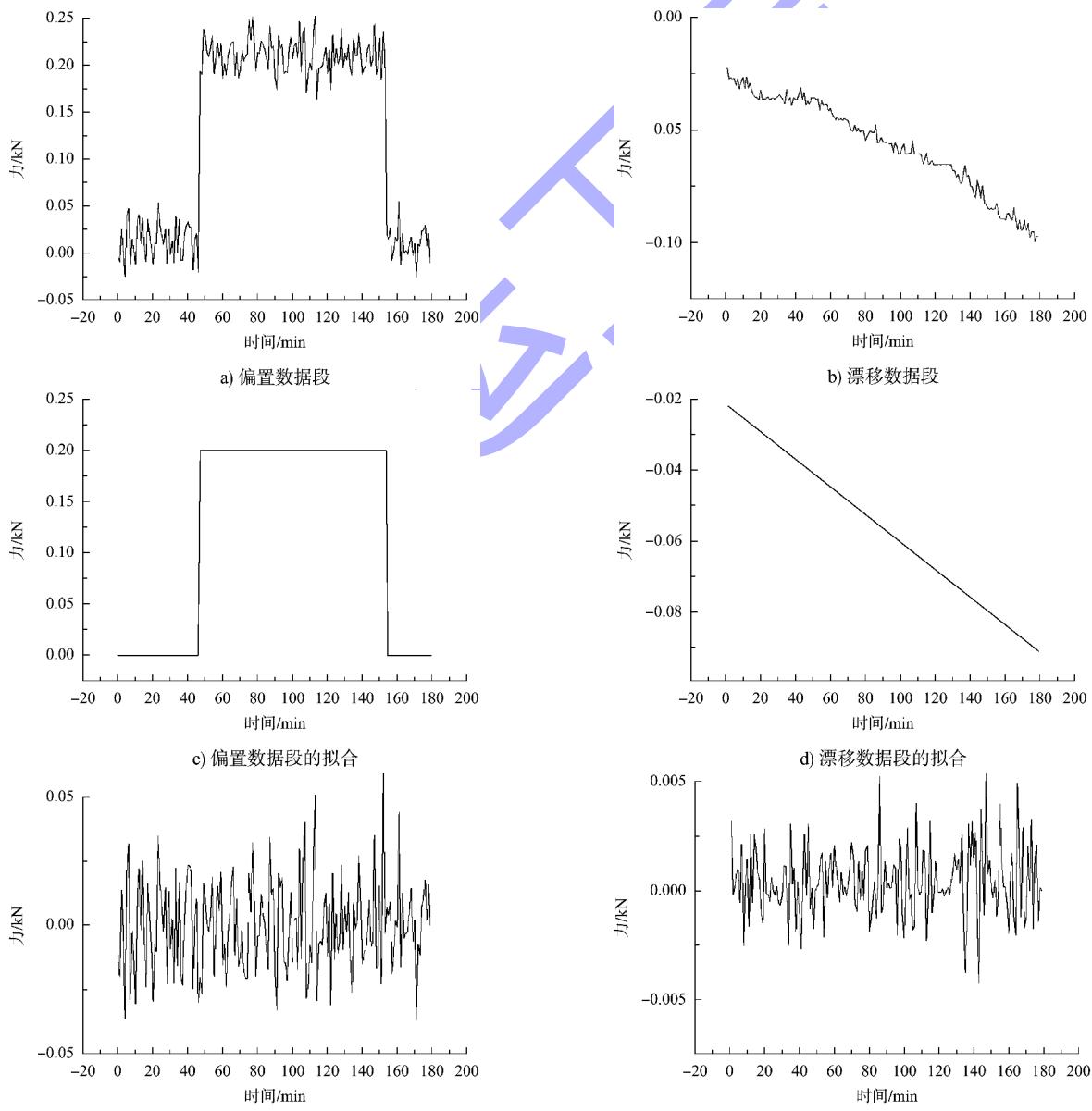


图 6 尖峰数据段修复前后

Fig. 6 Spike data segment before and after correction

图 7 偏置与漂移数据段修复
Fig. 7 Correction of bias and drift data segments

对于偏置和漂移数据段, 通过箱型图法识别后, 可以通过应用校正函数完成故障恢复。基于 5 min 固定长度的移动窗口的数据均值为数据点, 得到光滑多项式拟合直线。再通过将原始数据减去该校正函数数据来恢复数据。无论原始数据是否有漂移, 此简单的恢复策略都会起作用。此外, 如果无故障数据被错误地识别为有漂移的数据, 这种修正不会对结果产生负面影响。对于偏置故障数据段, 可以通过 5 min 移动窗口识别出均值变化点的方式确定偏置部分的数据起点与终点, 再以此构建多段拟合函数进行修复, 数据处理前后效果见图 7。

综上所述, 针对传感器故障引起的监测数据异常, 其分类、识别与修复流程模型见图 8。从人工识别建立故障数据段标签对异常数据进行分类到通过改进箱型图法实现对故障数据段的准确识

别。结合拉格朗日多项式插值实现对尖峰故障数据段的修复, 基于移动窗口均值建立拟合函数实现对偏置和漂移故障数据段的修复。

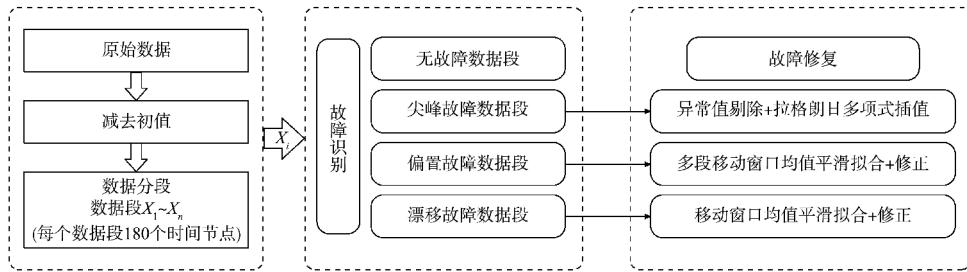


图 8 故障数据修复模型
Fig. 8 Fault data correction model

4 数据段修复效果评估

码头结构监测数据属于时序数据, 可划分为稳定和非稳定两类, 其判别依据在于统计特征是否随时间变化, 尤其是自相关系数的稳定性。若监测数据的自相关模式保持稳定, 则表明其变化规律具有一致性^[12]。因此, 在评估修复后数据段的修复效果时, 可基于自相关性指标进行分析, 以验证数据是否恢复稳定趋势。数据的自相关性可依据以下步骤进行计算。

对于监测时间序列 $\{X_t\}$, 将其分成 $(t-k)$ 对数据 (X_n, X_{n+k}) , 其中, $n=1, 2, \dots, t-k; k=1, 2, \dots, t/m$ 。分组后的据可通过式(4)、(5)计算每组数据的自相关系数。

$$\bar{X} = \frac{1}{n} \sum_{n=1}^t X_n \quad (4)$$

$$r_k = \frac{\sum_{n=1}^{t-k} (X_n - \bar{X})(X_{n+k} - \bar{X})}{\sum_{n=1}^t (X_n - \bar{X})^2} \quad (5)$$

式中: r_1 为两个相邻时刻数据之间相关关系的定量化表示; r_k 为某起始时刻与 k 时刻之间数据相关关系的定量化表示。 m 值一般取 4。

依据式(5)可算出 k 个自相关系数, 对于得到的自相关系数集, 可以依据以下条件进行判断确定: 1) 当所得的自相关系数均接近 0 时, 可以认定该时间序列是完全随机的; 2) 如果 r_1 与 0 相差较大, 且 r_2, r_3, \dots, r_k 如果逐渐减小并接近 0, 表

明时间序列是平稳的; 3) 如果自相关系数定期达到最大或最小值, 而其余部分接近 0, 则可以推断该序列是周期性变化的时序数据。4) 如果计算得到的自相关系数没有趋于 0, 而是逐渐减小, 那么可以推断该时间序列是不稳定的, 并且存在趋势变化。

在结构监测时序数据的修复评估中, 若修复后的数据满足以上条件 1) ~ 3), 则可认定其已恢复为稳定的时序数据, 符合监测需求。研究发现, 不同类型的故障数据表现出不同的相关性特征: 尖峰型故障数据相关性符合条件 1) 或 2), 且修复后的数据段仍保持原有的相关性特征; 偏置及漂移型故障数据相关性仅满足条件 4), 但经修复后, 数据段的相关性可恢复至条件 1) 或 2) 的稳定状态。经过修复处理的监测数据能够满足时序稳定性的各项要求, 其统计特性均呈现稳定的时不变特征。这一修复效果不仅验证了上述修复方法的有效性, 更为后续的结构健康监测、性能评估及预警分析等应用提供了可靠的数据基础。

5 结论

1) 针对码头结构健康监测系统中传感器故障导致的数据异常问题, 本文提出一种基于改进箱型图法的多模态故障识别与修复方法, 并通过实验验证了其可行性和有效性。

2) 通过融合箱型图的统计特征分析与散点图的空间分布特性, 构建了针对监测数据故障类型

的统计特征识别模型。试验数据表明, 该方法对尖峰故障、漂移故障和偏置故障数据均具有显著区分能力, 整体识别准确率达 91.07%。

3) 提出采用二次拉格朗日多项式插值对尖峰异常进行局部重建, 通过设置 5 个邻域数据点的滑动窗口, 实现了峰值区域的平滑过渡。试验表明, 修复后的数据自相关性符合稳定时序特征, 满足监测需求。

4) 基于 5 min 移动窗口均值构建拟合函数, 成功校正了漂移和偏置故障数据。该方法通过消除趋势性偏差, 使数据恢复至稳定状态, 且对误判的无故障数据无负面影响。通过计算修复前后数据的自相关系数, 证明了该修复方法的有效性。

5) 本文提出的方法为码头结构健康监测提供了实用的数据异常处理方案, 显著提升了监测数据的准确性和可靠性, 为后续结构状态评估与决策制定奠定了坚实基础。

参考文献:

- [1] 吴锋, 卓杨, 刘旭, 等. 高桩码头全寿命周期的健康监测技术[J]. 水运工程, 2023(2): 44-48, 53.
WU F, ZHUO Y, LIU X, et al. Health monitoring technique of high piled wharf based on life cycle theory [J]. Port & waterway engineering, 2023(2): 44-48, 53.
- [2] 王承强, 王臣, 贾宇, 等. 高桩码头结构安全监测预警模型及工程应用[J]. 水运工程, 2024(4): 15-21.
WANG C Q, WANG C, JIA Y, et al. High-pile wharf structure safety monitoring and forewarning model and its engineering application [J]. Port & waterway engineering, 2024(4): 15-21.
- [3] 陈际丰, 解东升, 章少兰. 光纤光栅监测系统在码头混凝土主体结构中的应用 [J]. 水运工程, 2011(8): 70-72.
CHEN J F, XIE D S, ZHANG S L. Application of FBG monitoring system in the main structure of concrete [J]. Port & waterway engineering, 2011(8): 70-72.
- [4] 王政. 分布式光纤布拉格光栅传感器的故障检测[D]. 西安: 西安电子科技大学, 2023.
WANG Z. Fault detection of distributed fiber Bragg grating sensors [D]. Xi'an: Xidian University, 2023.
- [5] LO C, LYNCH J P, LIU M Y. Distributed model-based nonlinear sensor fault diagnosis in wireless sensor networks [J]. Mechanical systems and signal processing, 2016, 66: 470-484.
- [6] 马现伟, 范朝辉, 聂为之, 等. 对失效传感器具备鲁棒性的故障诊断方法 [J]. 浙江大学学报(工学版), 2024, 58(7): 1488-1497.
MA X W, FAN C H, NIE W Z, et al. Robust fault diagnosis method for failure sensors [J]. Journal of Zhejiang University (engineering science), 2024, 58(7): 1488-1497.
- [7] PARADIS L, HAN Q. A survey of fault management in wireless sensor networks [J]. Journal of network and systems management, 2007, 15(2): 171-190.
- [8] NIL K, RAMANATHAN N, CHEHADE M N H, et al. Sensor network data fault types [J]. ACM transactions on sensor networks, 2009, 5(3): 1-29.
- [9] ASADOLLAHI P, LI J. Statistical analysis of modal properties of a cable-stayed bridge through long-term wireless structural health monitoring [J]. Journal of bridge engineering, 2017, 22(9): 04017051.
- [10] 孙向东, 刘拥军, 陈雯雯, 等. 箱线图法在动物卫生数据异常值检验中的运用 [J]. 中国动物检疫, 2010, 27(7): 66-68.
SUN X D, LIU Y J, CHEN W W, et al. Application of boxplot method in outlier detection of animal health data [J]. China animal health inspection, 2010, 27(7): 66-68.
- [11] 李梦辉, 桑小田, 田振川, 等. 箱形图在长江流域棉区棉花含杂率异常值检验中的应用 [J]. 湖北农业科学, 2016, 55(11): 2895-2898, 2954.
LI M H, SANG X T, TIAN Z C, et al. Application of Box plot in the outlier test of the cotton trash content in the Yangtze River cotton areas [J]. Hubei agricultural sciences, 2016, 55(11): 2895-2898, 2954.
- [12] 葛金龙, 方舒宁, 张登奎. 基于时间序列的码头变形监测数据的分析与预测 [J]. 经纬天地, 2016(6): 32-36, 41.
GE J L, FANG S N, ZHANG D K. Analysis and prediction of wharf deformation monitoring data based on time series [J]. Survey world, 2016(6): 32-36, 41.