

# Discussion on the Identification and Handling Methods of Abnormal Values in Pile Foundation Testing Data

Liqiang Lin

Fujian Provincial Construction Engineering Geophysical Exploration Test and Inspection Center, Fuzhou, Fujian, 362000, China

## Abstract

This paper takes pile foundation inspection as a key link to ensure the quality and safety of pile foundation engineering, and its data accuracy directly affects the judgment of bearing capacity and settlement characteristics. It also considers the situation where abnormal values may result from testing errors, environmental interference or pile body defects, leading to distorted assessment results or even engineering hazards. This paper aims to systematically explore the effective identification and handling methods of abnormal values in pile foundation inspection data. By studying and analyzing the characteristics of various detection data such as static load tests and high strain methods, the specific types and causes of outliers are clarified. In terms of identification, the applicability and advantages and disadvantages of statistical methods such as Grubbs test and machine learning methods such as isolated forest are compared. In terms of processing, the feasibility and application conditions of strategies such as data deletion, correction, and model interpolation are evaluated. The results show that choosing targeted identification and processing procedures by combining data features and abnormal causes can significantly improve data reliability and provide a more accurate basis for the safety evaluation of pile foundations.

## Keywords

Pile foundation testing; Outlier identification; Data processing; Grubbs test; Isolation Forest

## 桩基检测数据异常值识别与处理方法探讨

林立强

福建省建设工程物探试验检测中心, 中国·福建 福州 362000

## 摘要

本文结合桩基检测作为保障桩基工程质量与安全关键环节且其数据准确性直接影响承载力与沉降特性判断, 以及异常值可能源于测试误差、环境干扰或桩身缺陷而导致评估结果失真甚至工程隐患的情况, 本文旨在系统探讨桩基检测数据异常值的有效识别与处理方法, 通过研究分析静载试验、高应变法等多种检测数据特征来明确异常值的具体类型与成因, 在识别方面对比统计方法如Grubbs检验与机器学习方法如孤立森林的适用性与优劣, 在处理方面评估数据删除、修正以及模型插值等策略的可行性与应用条件。结果表明, 结合数据特征与异常成因选择针对性识别与处理流程能显著提升数据可靠性, 为桩基安全性评价提供更准确依据。

## 关键词

桩基检测; 异常值识别; 数据处理; Grubbs检验; 孤立森林

## 1 引言

作为建筑结构与地基之间关键传力构件的桩基, 其工程质量直接关乎整体建筑物安全性与耐久性, 而桩基检测作为控制和评价桩基施工质量的核心手段, 通过静载试验、高应变法、低应变法等获取荷载-沉降曲线、波速等关键数据, 为判定桩的承载力与完整性提供依据, 但在实际检测中, 因仪器设备漂移、现场测试环境干扰、人为操作失误或桩身局部缺陷等因素, 采集数据不可避免掺杂异常值, 这些偏离正

常规律的数据点严重扭曲分析结果, 可能导致桩基承载力误判或掩盖真实桩身缺陷, 给工程安全留隐患, 所以准确识别并有效处理这些异常值, 作为确保桩基检测数据可靠性、得出科学公正结论的首要前提, 对预防工程事故、保障结构安全具有至关重要的现实意义。

## 2 桩基检测数据异常值基础

### 2.1 桩基检测数据来源与类型

桩基检测数据的获取是桩基工程质量评定的直接依据, 其来源严格依赖于现场的原位测试技术。静载试验作为获得桩基竖向抗压、抗拔及水平承载力最权威的方法, 通过在工程桩上施加接近于实际工作状态的荷载, 并精确测量桩顶随

【作者简介】林立强(1981-), 男, 中国福建莆田人, 硕士, 工程师, 从事地基基础检测研究。

之产生的沉降、上拔量或水平位移，从而绘制出荷载 - 沉降 (Q-s) 曲线等核心关系曲线。该曲线是判断桩的极限承载力、变形特性及工作状态的关键。高应变动力检测则通过重锤冲击桩顶，使桩 - 土体系产生足够的塑性位移，安装在桩顶侧的力传感器和加速度计同步采集打击力与质点速度的时程曲线，基于一维波动理论对曲线进行分析，从而推算单桩承载力并评价桩身完整性。

### 2.2 异常值的定义与分类

在桩基检测的数据分析领域，异常值特指那些显著偏离数据主体分布规律或与预期物理模型行为严重不符的观测值或数据片段<sup>[1]</sup>。从误差理论的根源进行剖析，这些异常值主要源于三类误差。粗大误差通常由测试过程中突发、偶发的失误所引起，例如记录员读错刻度、数据录入错误、仪器受到意外撞击、数据采集系统出现瞬时断电或信号跳变。其数学表达式如下，

$$|x_i - \bar{x}| > k \cdot s \quad (1)$$

式 (1) 中， $x_i$  代表待判定的单个观测值， $\bar{x}$  代表数据样本的算术平均值， $s$  代表数据样本的标准差， $k$  是一个根据所需置信水平选定的阈值常数，通常取值在 2 至 3 之间。该公式的含义是，如果某个数据点偏离样本均值的幅度超过了  $k$  倍的标准差，则有理由怀疑该点为异常值。

## 3 异常值识别方法

### 3.1 基于统计的识别方法

基于统计的异常值识别方法建立在坚实的数理统计理论基础之上，其核心在于假设正常数据点来源于一个特定的概率分布模型，而异常值则是那些以小概率出现、显著偏离该模型预期范围的观测点。这类方法的有效性强烈依赖于数据是否符合预设的分布假设，例如正态分布。在桩基检测数据分析中，Grubbs 检验是识别单变量数据集中是否存在一个异常值的经典参数检验方法。它特别适用于样本量不大且数据大致服从正态分布的情形，例如分析一批由低应变法测得的桩身混凝土波速值是否混入异常值。该检验的统计量计算基于样本均值和标准差，量化了可疑观测点与数据中心点的偏离程度。其检验统计量  $G$  的计算公式如下，

$$G = \frac{\max |x_i - \bar{x}|}{s} \quad (2)$$

式 (2) 中， $x_i$  代表待检验的观测值， $\bar{x}$  代表数据样本的算术平均值，它反映了数据集的集中位置， $s$  代表数据样本的标准差，它衡量了数据的离散程度。计算过程是找出所有观测值与均值偏差的绝对值中的最大值，然后将其除以标准差。将计算得到的  $G$  统计量与 Grubbs 检验临界值表中对应样本量  $n$  和选定显著性水平  $\alpha$  (工程上常取  $\alpha=0.05$ ) 的临界值  $G_{critical}$  进行比较。如果  $G > G_{critical}$ ，则在  $\alpha$  显著性水平下，认为该可疑值  $x_i$  是异常值，应予以考虑剔除<sup>[2]</sup>。剔除后，需对剩余数据重新计算均值和标准差，重复上述过程

进行检验，直至不再检出异常值为止。某住宅楼项目采用低应变法检测 30 根灌注桩的桩身完整性。测得混凝土波速值 (m/s) 分别为：3850, 3920, 3880, 4100, 3950, 3780, 3900, 3850, 3950, 4120 (其余 20 根数据在 3800-4000 之间)。采用 Grubbs 检验 ( $\alpha=0.05$ ) 对波速最高值 4120 m/s 进行检验，计算得  $G=3.12$ ，大于临界值 2.89，判定为异常值。经查证，该测点为提供的施工桩长参数有误，重新测试后波速为 3910 m/s，与群体数据一致。

### 3.2 基于机器学习的识别方法

基于机器学习的异常值识别方法为处理复杂、高维、非线性的桩基检测数据提供了强大且灵活的工具集，这类方法通常不依赖于严格的数据分布假设，而是通过算法自动学习数据的内在结构或正常模式，并将不符合该模式的数据实例识别为异常。它们在处理诸如高应变动力检测的完整时程曲线或包含多个特征参数的数据集时，展现出超越传统统计方法的优势。孤立森林是一种高效的无监督异常检测算法，其核心机理非常巧妙，它利用异常值“少而不同”的特性，通过随机选择特征和分割值递归地划分数据空间，孤立森林算法原理示意如图 1 所示。某预制桩工程高应变检测数据包含 200 组打击信号，每条信号曲线包含 1000 个时间点数据。采用孤立森林算法对力 - 速度曲线进行异常检测，成功识别出 3 组异常信号：一组表现为力信号初始段出现高频振荡 (锤击偏心导致)，另两组在曲线后期出现非规律波动 (传感器松动所致)。传统统计方法难以有效识别此类局部异常，而孤立森林通过整体曲线模式对比实现了准确检测。

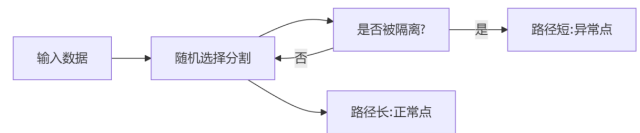


图 1 孤立森林算法原理示意图

### 3.3 识别方法比较与选择

面对多种异常值识别方法，工程技术人员需要根据具体的检测数据类型、数据规模、异常特性以及应用场景的约束条件，审慎地选择最合适的方法或方法组合。统计方法与机器学习方法在原理、适用性、复杂度和可解释性方面存在显著差异。统计方法如 Grubbs 检验和 Dixon 检验，其主要优势在于原理简单明了，计算过程快速，结果具有统计意义上的可解释性，例如可以明确给出在 5% 的显著性水平下某数据点为异常。这对于需要明确统计证据的场合非常重要。它们的局限性同样突出，严重依赖数据分布假设，对多异常点检测效果不佳，且主要适用于单变量数据<sup>[3]</sup>。

## 4 异常值处理方法

### 4.1 数据删除与修正法

数据删除与修正法是应对已识别异常值的基础性技术

策略，其核心决策在于判断是否保留异常点所承载的信息。直接删除法应用于最为极端的情形，即当异常值被确认为由纯粹的偶然性、非再现性的外部干扰引起，且该数据点完全扭曲了物理过程的真实反映时，将其从数据集中移除是最为彻底和合理的处理<sup>[4]</sup>。某商业大厦静载试验中，第7级荷载（5600kN）持荷期间，沉降数据序列为：12.5, 12.8, 13.0, 18.5, 13.2, 13.1mm。第4个数据点18.5mm明显异常，经查为降雨导致基准梁轻微沉降所致。由于该点明显偏离持荷稳定规律，采用中位数替换法，用序列中位数13.0mm替换异常值，修正后曲线平滑合理，真实反映了桩土变形特性。

#### 4.2 基于模型的处理方法

基于模型的处理方法代表了更为精细和科学的数据修复策略，其核心思想是利用数据中内在的规律性、变量间的相关性或序列前后关联性，通过构建数学模型来估计异常点本应具有合理数值。这种方法旨在最大限度地保留数据集的结构完整性和信息量，适用于异常值并非完全随机谬误，而其周围数据呈现出可识别模式的情况<sup>[5]</sup>。插值法则是处理序列数据中异常值的经典方法，特别适用于时间序列如静载试验的沉降-时间曲线，或空间序列如桩身完整性低应变检测的深度-波阻抗曲线，异常值处理流程见图2。某码头工程采用高应变法检测钢管桩，发现某测点力信号在8.5ms处出现异常跌落（传感器受潮导致）。利用正常测点的力-速度相关性建立回归模型，预测该时刻正常力值应为1250kN，替代原异常值1050kN。修正后曲线连续性明显改善，CAPWAP分析得出的承载力与相邻桩结果一致，验证了修正合理性。

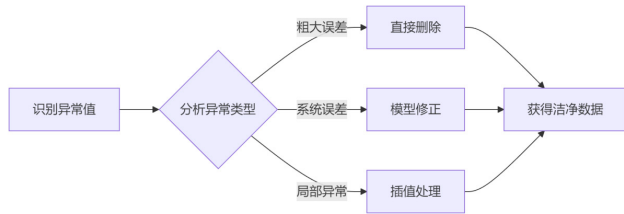


图2 异常值处理流程图

#### 4.3 处理方法的工程应用建议

在桩基检测工程实践中，异常值处理绝非简单的数学运算，而是一个需要融合专业知识、现场经验与严谨判断的决策过程。选择最适宜处理方法的黄金法则是追根溯源，即紧密联系异常值产生的具体原因和工程背景。处理策略的制定应始于对异常现象的合理解释。对于有明确记录证实为偶

然外部因素导致的数据点，例如读数记录笔误、采集系统瞬时断电或雷击引起的信号毛刺，且该点明显脱离物理规律，直接删除是合理且高效的选择。当异常表现出系统性偏差的特征，如所有传感器读数整体漂移，则应考虑系统误差的修正。例如，如果发现低应变检测中所有测点的波速值因计时系统校准误差而整体偏大或偏小，则需要根据校准报告对整个数据集进行统一的系数修正。对于在时间序列或空间序列中出现的孤立异常点，且其前后/相邻数据呈现良好规律性时，插值法是优选方案，它能较好地维持序列的连续性和整体趋势。至关重要的一点是，必须高度警惕那些可能蕴含真实工程意义的“异常”。例如，低应变反射波法波形中出现明显的同相反射波，或高应变信号中表现出桩身某截面阻抗显著降低的特征，这些异常极有可能是桩身存在缩径、混凝土离析甚至断裂等真实缺陷的反映<sup>[6]</sup>。

### 5 结语

本文系统阐述桩基检测数据异常值的成因、识别与处理方法并建立从数据特征分析到统计与智能识别算法选择再到针对性处理策略制定的完整技术框架，其实际意义在于为工程技术人员提供一套清晰实用且兼具理论依据与操作性的异常值管理方案，助其有效甄别因测试误差产生的干扰信号与反映真实桩身缺陷的有价值信息，通过科学处理异常数据显著提升桩基检测数据的可靠性与分析结果的准确性，为桩基承载力评定与完整性判断奠定坚实数据基础，直接服务于工程质量控制与安全评估，对预防工程隐患、优化设计施工具有重要实践价值。

#### 参考文献

- [1] 刘宁,张杰武,严靖,等.基于声呐点云的桥梁水下桩基轴线提取与倾斜检测[J/OL].重庆交通大学学报(自然科学版),1-7[2025-11-03].
- [2] 王曦强,郑剑波,石松文,等.桥梁水下结构物常见病害及多维度检测方法研究——以某跨乌龙江斜拉桥桥梁桩基冲刷检测为例[J].福建建材,2025,(09):38-42.
- [3] 杨钦福.无损检测技术在桩基础检测工作中的应用研究[J].中国新技术新产品,2025,(16):102-104.
- [4] 钱逢龙.桩基施工过程中孔底沉渣特性的光纤检测方法分析[J].安徽建筑,2025,32(09):158-160+192.
- [5] 李守国.路桥工程桩基后注浆施工关键技术及质量检测[J].四川水泥,2025,(08):172-174.
- [6] 王心雨.坦桑尼亚中央线米轨戈代戈代桥梁工程项目低应变桩基完整性检测分析[J].工程建设与设计,2025,(14):174-176.