

基于曲线拟合与MAD法的测斜数据粗差处理方法

姜思源

上海京海工程技术有限公司 上海 200137

摘要：针对测斜数据中粗差干扰变形分析的问题，提出一种基于曲线拟合与MAD法的粗差处理方法。通过对傅立叶函数、高斯函数和多项式函数对测斜数据的拟合效果，结合MAD法进行粗差探测与替换。实验采用某地铁深基坑工程的实测数据，并人工添加6个粗差点以验证方法有效性。结果表明，傅立叶函数的拟合效果最佳且粗差识别准确率达100%，优于其他两种拟合函数，适用于测斜数据的粗差处理。

关键词：测斜数据；粗差处理；曲线拟合；MAD

引言

深基坑工程的安全监测是顺利推进地下空间开发的基本技术保障，测斜技术作为一种高精度的变形监测手段，通过测量支护结构及土体不同深度处的水平位移数据，为动态了解基坑稳定性和变形情况提供了数据支撑^[1]。然而，受仪器误差、现场复杂的施工环境及人员操作等因素影响，原始测斜数据中通常包含粗差，这些异常值会严重干扰变形趋势分析，甚至导致误判。因此，有效地识别和处理粗差是确保监测数据可靠性的关键环节。

在粗差处理研究领域，主要方法是基于统计假设检验的传统准则法和机遇曲线拟合的探测法。楼楠采用多项式曲线拟合建立监测数据的变形趋势模型，通过测量值与拟合值的偏离程度探测粗差^[2]；郭超比较拉依达准则、格拉布斯准则、狄克逊准则和肖维勒准则的适用情况，提出了改进的格拉布斯准则^[3]；鲁铁定等在中位数绝

对偏差法（median absolute deviation, MAD）的基础上结合小波分析，以探测GNSS高程时间序列中偏离度较小的粗差^[4]。上述研究均对测斜数据的粗差处理具有指导作用，多项式的阶次会影响拟合效果，传统粗差探测方法建立在观测值误差服从一定概率统计规律的基础上。基于此，利用傅立叶函数、高斯函数和多项式函数对测斜数据进行拟合，再用MAD法探测粗差，将其剔除后用拟合值代替，对比粗差识别准确率、残差值和评价参数等指标分析三种拟合函数在粗差处理中的适用性，得出最适合测斜数据的拟合模型。

1 曲线拟合模型

1.1 傅立叶函数拟合

傅立叶函数是将周期性数据分解为不同频率的正弦和余弦函数线性组合，具有广泛的适应性，可用下列模型表示：

$$F(x) = a_0 + a_1 \times \cos(\omega x) + b_1 \sin(\omega x) + \dots + a_k \times \cos(k\omega x) + b_k \sin(k\omega x) \quad (1)$$

其中， a_i 、 b_i ， $i=0,1,2,\dots,k$ ，为傅里叶系数； ω 为频率； k 为谐波数阶数。在实际工程应用中，傅立叶函数模型不仅广泛应用于周期信号的分解，还可有效用于变形特征提取和测量误差改正^[5]。对于测斜数据，通过适当选择傅立叶函数的谐波阶数 k ，可以在保留主要特征的同时

有效平滑随机波动，以识别异常值。

1.2 高斯函数拟合

高斯函数拟合利用若干个高斯概率密度函数（即正态分布曲线）逼近数据，可用下列模型表示：

$$F(x) = a_0 + a_1 \times \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \dots + a_m \times \exp\left(-\frac{(x-\mu_m)^2}{2\sigma_m^2}\right) \quad (2)$$

其中， a_i 、 μ_i 、 σ_i ， $i=0,1,2,\dots,m$ ，分别为幅值、均值和标准差， m 为指数项的个数。高斯函数曲线因其钟型对称分布的特点，在拟合具有局部极值特征的数据中表现

出优势。在测斜数据中，支护结构的最大位移点通常呈现类似特征，与高斯函数曲线较吻合。然而，高斯函数拟合存在局限性，当数据在首尾处变化剧烈时易出现龙格现象，在工程数据分析中可能会误判特征点。

1.3 多项式拟合

多项式拟合通过构造多项式函数逼近离散数据点，利用最小二乘法求解最优系数。其核心思想是将非线性

作者简介：姜思源（1998年3月—），女，汉族，河南省固始县人，硕士研究生，助理工程师，主要研究方向为变形监测。

关系转化为线性参数估计问题，可用下列模型表示：

$$P(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \quad (3)$$

其中， $a_i, i=0,1,2,\cdots,n$ ，为待求多项式系数； n 为多项式阶数，可根据实际情况适当选择。多项式拟合的优势在于形式简单，计算效率高，且可以通过调整阶数 n 来适应不同的数据特征。在实际应用中，多项式阶数的选择需要权衡拟合精度和模型复杂度。

2 粗差处理拟合模型及其比较分析

2.1 数据概况

以某地铁站临近深基坑工程围护结构的观测数据为实验对象，选取其中一测斜孔的某次实测值进行分析。该测斜管深度为40m，以0.5m为间隔进行数据采集，共获得80个测量点，测斜曲线如图1所示。观察图像可知，原始测斜曲线呈抛物线型，主要特征是位移量从基坑顶部向下逐渐增大，在深度约11米处达到峰值，最大位移18.1272mm，随后逐渐减小，形成顶部和底部位移小，中部位移大的情况，是测斜曲线中较为常见的线型^[6]。为模拟测量仪器和环境扰动的影响，在原始数据中随机选取6个测量点添加均匀分布随机误差，见表1。

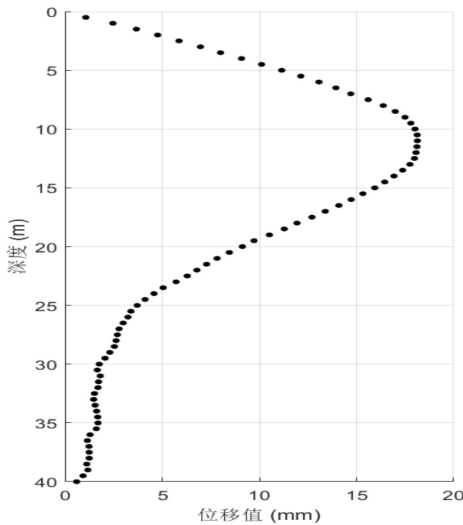


图1 原始测斜数据

经试验和调整，傅立叶函数采用3个谐波项（即 $k=3$ ），高斯函数采用4个指数项（即 $m=4$ ），多项式函数采用6次多项式（即 $n=6$ ），三种函数的拟合情况和残差如图2、3所示，拟合残差见表2。

对比分析傅立叶函数、高斯函数和多项式函数拟合方法，三种模型均能有效反映测斜数据的基本变化趋势，但在具体表现上存在差异。相较于其他两项模型，傅立叶函数生成的拟合曲线更平滑，数据两端没有出现明显的波动。高斯函数和多项式函数在数据首尾表现出

表1 粗差位置

序号	深度 (m)	原始值 (mm)	粗差值 (mm)
1	3	6.97	4.73
2	7	14.70	13.45
3	12.5	17.98	20.58
4	21	7.82	4.21
5	32.5	1.52	2.84
6	37.5	1.24	4.53

2.2 粗差识别与替换

通过Matlab编写数据处理程序，分别利用傅立叶函数、高斯函数和多项式函数三种模型对测斜数据进行拟合。粗差识别采用MAD法，设残差序列为 $e_n = \{e_1, e_2, \cdots, e_i, \cdots, e_n\}$ ，残差序列的中位数为 $\text{median}(e)$ ，则有：

$$MAD = 1.4826 \times \text{median}(|e_i - \text{median}(e)|) \quad (4)$$

其中，1.4826是校正系数。若 e_i 满足：

$$|e_i - \text{median}(e)| > 3 \times MAD \quad (5)$$

则，认为 e_i 为粗差点。识别出的粗差点以对应的函数拟合值代替。

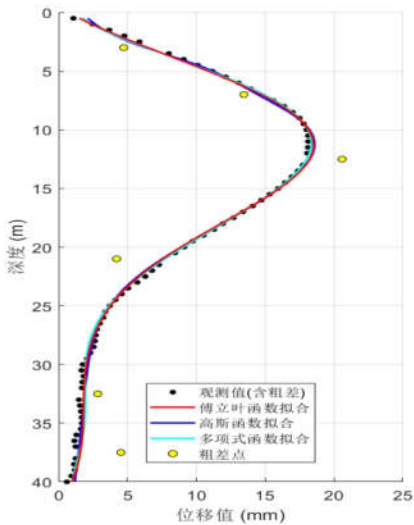


图2 测斜数据（含粗差）拟合

不同程度的龙格现象，导致拟合曲线在首尾区域出现偏离数据的波动，残差明显增大。

粗差识别能力方面，三种函数的表现差异明显。傅立叶函数和多项式函数准确识别出预设的6个粗差点，未出现误判情况。高斯函数由于龙格现象影响，除6个真实粗差之外，另将0.5深处的正常测量值误判为粗差，识别出粗差个数为7个。

结合图表分析表明，三种函数均能有效地拟合测斜数据，傅立叶函数拟合曲线更平滑，高斯函数和多项式

函数首尾两端存在更明显的龙格现象，拟合两端部分残差值更大，存在一定的震荡。从粗差识别分析，傅立叶函数和多项式函数准确识别出预设的6个粗差，高斯函数的拟合曲线因龙格现象发生曲线两端震荡，识别粗差个数为7个。

表2 函数拟合残差表

序号	深度	粗差值	傅立叶函数拟合		高斯函数拟合		多项式函数拟合	
			替换值	残差	替换值	残差	替换值	残差
1	0.5	NaN	NaN	NaN	2.1075	1.0346	NaN	NaN
2	3	4.73	6.4886	-0.4801	6.2659	-0.7028	6.3441	-0.6246
3	7	13.45	14.722	0.0186	14.44	-0.2634	14.918	0.2146
4	12.5	20.58	18.228	0.2461	18.257	0.2751	18.02	0.0381
5	21	4.21	7.5557	-0.2665	7.5093	-0.3129	7.7307	-0.0915
6	32.5	2.84	1.8331	0.3153	1.8966	0.3788	1.9501	0.4323
7	37.5	4.53	1.5731	0.3297	1.4416	0.1982	1.5032	0.2598

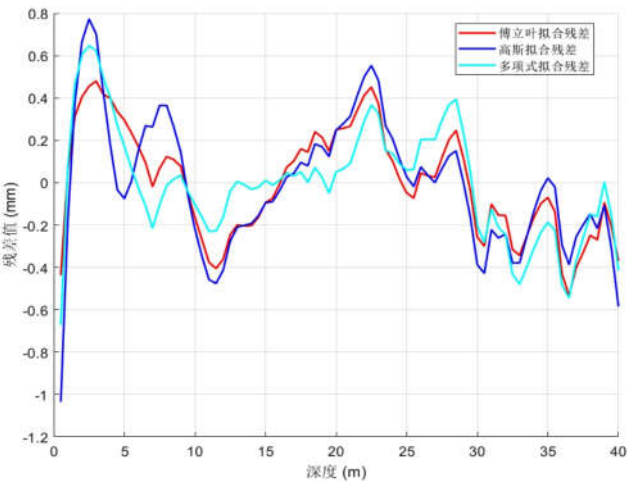


图3 各拟合方法残差比较

为进一步评价不同拟合函数的粗差处理效果，选取SSE（拟合误差平方和）、R-square（相关系数平方值）、RMSE（均方差）作为量化指标，如表3所示。其中，SSE是指拟合误差平方和，反映拟合值与实测值的偏差，数值越接近0拟合越好；R-square是指数据点与拟合值之间的相关系数平方值，反映函数拟合程度，数值越接近1拟合越好；RMES是指均方差，衡量拟合误差的标准差，数值越小拟合越好。比较3个评价参数，傅立叶函数在三项指标中表现最优，其SSE值为35.7599，均小于另外两个拟合函数，更接近于0；R-square值为0.9885，均大于另外两个拟合函数，更接近于1；RMES值为0.7047，更小于另外两个函数。由此表明，傅立叶函数的拟合效果更好。

基于实验结果，在实际工程应用中可优先选用傅立叶函数进行测斜数据拟合，谐波数取3~5之间，可根据实际数据特征调整。对于重要工程，可采用两种以上的拟

合方法进行交叉验证。

表3 评价参数

拟合函数	SSE	R-square	RMSE
傅立叶函数	35.7599	0.9885	0.7047
高斯函数	36.9014	0.9863	0.7376
多项式函数	38.6093	0.9876	0.7273

3 结论

利用傅立叶函数、高斯函数和多项式函数拟合加入粗差的测斜数据，对比分析三种函数在测斜数据粗差处理中效果。结果表明：（1）三种函数均能较好地拟合测斜数据，拟合曲线较为光滑，高斯函数和多项式函数两端存在更明显的龙格现象，残差值更大；（2）高斯函数拟合由于龙格现象发生粗差识别错误；（3）傅立叶函数拟合的贴合度、稳定性和精更高，粗差处理效果最优。

参考文献

[1]李润辰,江玮.地铁深基坑施工变形监测分析[J].住宅与房产,2021,(18):231-232.

[2]楼楠.深基坑工程的综合监测与安全预报[D].解放军信息工程大学,2007.

[3]郭超.地铁地表沉降监测数据的预处理及预测方法的探讨[D].东华理工大学,2015.

[4]鲁铁定,何锦亮,徐华卿,等.一种基于MAD改进的GNSS高程时间序列粗差探测方法[J].大地测量与地球动力学,2022,42(04):349-354.

[5]曾继礼.地面三维激光扫描仪角度测量检校方法研究[J].黑龙江科学,2024,15 (12):62-65.

[6]何钦,陈大江,张记峰,等.基坑支护结构深层水平位移线型及危险点的判定[J].工业建筑,2022,52(02):96-100+150.