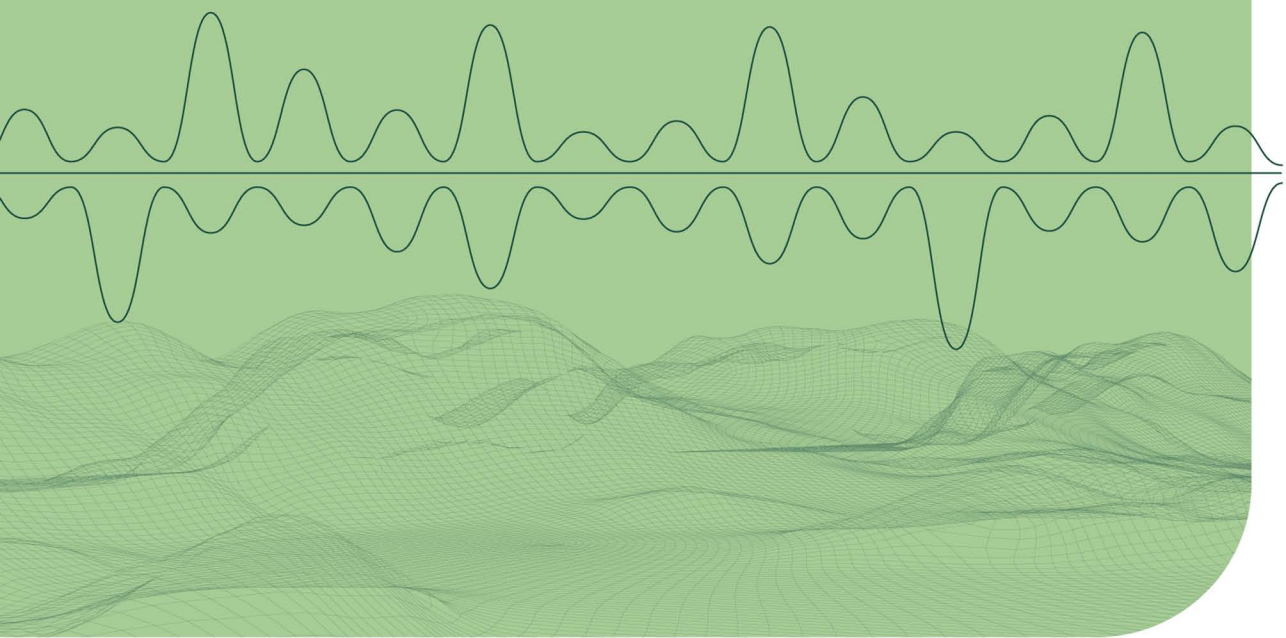


# 管道地质灾害监测 数据挖掘及预警模型

## 研究与应用



刘奎荣 余东亮 周广 朱建平 方迎潮 等 © 著

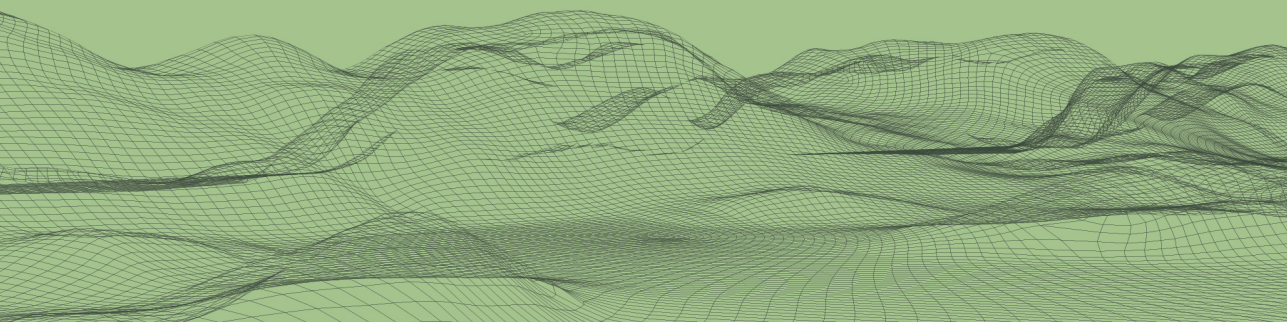


西南交通大学出版社



# 管道地质灾害监测 数据挖掘及预警模型 研究与应用

刘奎荣 余东亮 周广 朱建平 方迎潮 等◎著



西南交通大学出版社

· 成 都 ·

-----  
图书在版编目 ( C I P ) 数据

管道地质灾害监测数据挖掘及预警模型研究与应用 /  
刘奎荣等著. —成都: 西南交通大学出版社, 2022.4  
ISBN 978-7-5643-8383-1

I. ①管… II. ①刘… III. ①管道工程 - 地质灾害 -  
风险管理 - 研究 IV. ①U172

中国版本图书馆 CIP 数据核字 (2021) 第 244125 号  
-----

Guandao Dizhi Zaihai Jiance Shuju Wajue ji Yujing Moxing Yanjiu yu Yingyong  
管道地质灾害监测数据挖掘及预警模型研究与应用

刘奎荣 余东亮 周 广 朱建平 方迎潮 等 著

责任编辑	孟秀芝
封面设计	GT 工作室
出版发行	西南交通大学出版社 (四川省成都市金牛区二环路北一段 111 号 西南交通大学创新大厦 21 楼)
发行部电话	028-87600564 028-87600533
邮政编码	610031
网 址	<a href="http://www.xnjdcbs.com">http://www.xnjdcbs.com</a>
印 刷	四川玖艺呈现印刷有限公司
成品尺寸	170 mm × 230 mm
印 张	11
字 数	152 千
版 次	2022 年 4 月第 1 版
印 次	2022 年 4 月第 1 次
书 号	ISBN 978-7-5643-8383-1
定 价	120.00 元

图书如有印装质量问题 本社负责退换  
版权所有 盗版必究 举报电话: 028-87600562

## 本书作者

刘奎荣 余东亮 周 广 朱建平 方迎潮  
黄 鹏 王 庆 唐 侨 赵 雄 吴东容  
王爱玲 蒋 毅 王彬彬 谢 锐 吴 瑶  
王垒超 兰才富 轩 恒 杨 川 刘宇婷  
梁 栋

# 前 言

随着我国经济的蓬勃发展，对能源的需求越来越大，能源供给已成为影响和制约我国经济发展的重要因素。管道是油气运输的主要方式，承担着我国 70%的原油、99%的天然气运输及 80%的成品油。油气管道作为线性工程，不可避免要穿越地质、地貌和水文条件复杂区域，面临崩塌、滑坡、泥石流、河沟道水毁等地质灾害风险。油气管道沿线人类工程活动频繁，公路、铁路、城镇等基础设施建设与管道频繁交叉重叠，复杂的地质环境变化为管道的管理带来了严峻挑战。长输油气管道安全运营受到自然和社会环境双重制约。

随着管道地质灾害防治关口前移，防治技术向着智能化、智慧化方向发展，防治策略由被动治理向主动防控转变。监测预警作为管道地质灾害防治的有效方式之一，能及时获取油气管道沿线地质灾害变化信息，提前发现并预测风险。现有的管道地质灾害监测预警技术尚存在判据单一，预警阈值设置受人为因素影响大，缺乏自学习、自修正过程，预警模型相对简单，缺乏对多因子的综合考虑，缺乏对灾害发展趋势的分析等不足。因此，有必要对地质灾害监测数据进行深度挖掘，优化预警模型，为山区油气管道安全运营提供技术支撑。

本书以管道沿线最为常见、危害较大的滑坡、水毁及采空区地质灾害为例，系统介绍了管道地质灾害监测数据挖掘及预警模型的研究与应用。第 1 章概述了油气管道地质灾害现状及当前地质灾害监测预警工作中存在的问题。第 2 章主要介绍了当前国内外研究现状及发展趋势。第 3

章详细阐述了异常值过滤、缺失值填补、数据预测、单指标预警算法和报警阈值动态调整的相关算法原理。第 4 章分析了基于实时监测数据的自学习自修正预警预报阈值分析方法并进行应用。第 5 章以地灾形变、外界诱发因素、管道力学三个指标为基础，构建了滑坡、水毁、采空区地质灾害监测指标体系。第 6 章详细阐述了监测预警技术在实际工程中的应用，并以实际的案例验证了预警模型的精度。

著 者

2022 年 4 月

# 目 录

第 1 章 绪 论 .....	001
第 2 章 国内外研究现状 .....	004
2.1 异常值消除和过滤方法研究现状 .....	004
2.2 缺失值填补法研究现状 .....	006
2.3 预测方法研究现状 .....	008
2.4 动态阈值调整研究现状 .....	010
2.5 管道地质灾害监测预警技术研究现状 .....	011
2.6 滑坡灾害预警预报模型研究现状 .....	014
2.7 管道水毁灾害预警预报模型研究现状 .....	019
2.8 采空区灾害预警预报模型研究现状 .....	022
第 3 章 基于实时监测数据的自学习自修正预警预报阈值 分析方法 .....	025
3.1 异常值消除和过滤算法 .....	025
3.2 缺失值填补算法 .....	030
3.3 数据预测算法 .....	045
3.4 监测数据单指标算法 .....	060
3.5 报警阈值动态调整 .....	061
3.6 本章小结 .....	067

第 4 章	基于实时监测数据的自学习自修正预警预报阈值 分析方法应用	068
4.1	$3\sigma$ 异常值过滤	068
4.2	LSSVM 数据补全	072
4.3	LSTM 数据预测	078
4.4	单指标预警算法	091
4.5	阈值动态调整	094
4.6	本章小结	103
第 5 章	地质灾害监测关键指标体系	105
5.1	滑坡灾害地质特征及变形机理	105
5.2	水毁灾害地质特征及变形机理	108
5.3	采空区灾害地质特征及变形机理	110
5.4	滑坡灾害监测预警关键指标体系构建	111
5.5	水毁灾害监测预警关键指标体系构建	112
5.6	采空区灾害监测预警关键指标体系构建	114
第 6 章	管道地质灾害预警预报模型	116
6.1	管道地质灾害监测预警系统总体架构和 workflow	116
6.2	滑坡灾害预警预报模型	119
6.3	水毁灾害预警预报模型	133
6.4	采空区灾害预警预报模型	139
6.5	预警预报成果	150
	参考文献	162

# 第1章 绪论

---

作为生命线工程的高压油气长输管道，担负着油气资源的主要输送任务，由于分布范围非常广阔，沿途区域自然地理和地质环境复杂多样，不可避免地会受到各种地质灾害的威胁和侵害。如西南山区油气管道穿越地质地貌条件复杂区，山高谷深、地形陡峻，地震及活动断裂发育，滑坡、崩塌、泥石流、水毁等地质灾害活跃，管道面临的地质灾害风险大。

油气管道事故统计分析表明，受地理环境、气候及人类活动等因素影响，管道工程中的地质灾害种类繁多、危害程度也各不相同。地质灾害引发土体运动和地表变形，从而导致埋地管道产生弯曲、压缩、扭曲、拉裂、局部屈曲等破坏形式。地质灾害一旦发生，不仅会造成管道变形、断裂和大范围破坏，从而导致油气泄漏、管线停输，带来巨大经济损失，甚至引发火灾、爆炸等事故，对自然环境、生命财产和社会安定造成严重后果。特别是在灾害多发地区，突发性的地质灾害经常造成难以估量的损失。如兰成渝成品油管道投产以来，先后经历了2003年、2006年、2008年、2009年暴雨袭击，东裕沟、响河沟、火烧沟、石亭江等区段发生了多处水毁、泥石流，造成管道露管、漂管、悬管，抢险治理费用高达数千万元。2008年“5·12”汶川地震发生后，兰成渝成品油管道穿越秦岭山区段发生了K0572、K0569、K0528+850等滑坡，给管道安全运营造成威胁。

传统地质灾害治理工作被动，治理金额居高不下。相比工程治理，地质灾害监测投入成本少，可以及时获取地质灾害发展变化信息，提前发现并预测风险，实现地质灾害风险有效管控。但现有的管道地质灾害

预警技术尚存在以下不足：

- (1) 预警判据的选择过于单一；
- (2) 预警阈值设置受人为因素影响大；
- (3) 缺乏自学习、自修正过程；
- (4) 预警模型相对简单，缺乏对多因子的综合考虑；
- (5) 缺乏对灾害的预测分析。

本书在分析管道地质灾害形成机制、危害特征与破坏模式的基础上，开展管道地质灾害预警预报模型研究，构建基于地灾形变指标  $X$ 、外界诱发指标  $Y$ 、管道力学指标  $Z$  的滑坡、水毁、采空区灾害监测关键指标体系。结合工程实际形成滑坡、水毁、采空区灾害一维预警预报模型、二维预警预报模型、三维预警预报模型，使得预警结果更符合实际，进一步提高管道地质灾害的预警精度。本书总体技术路线参见图 1.1。

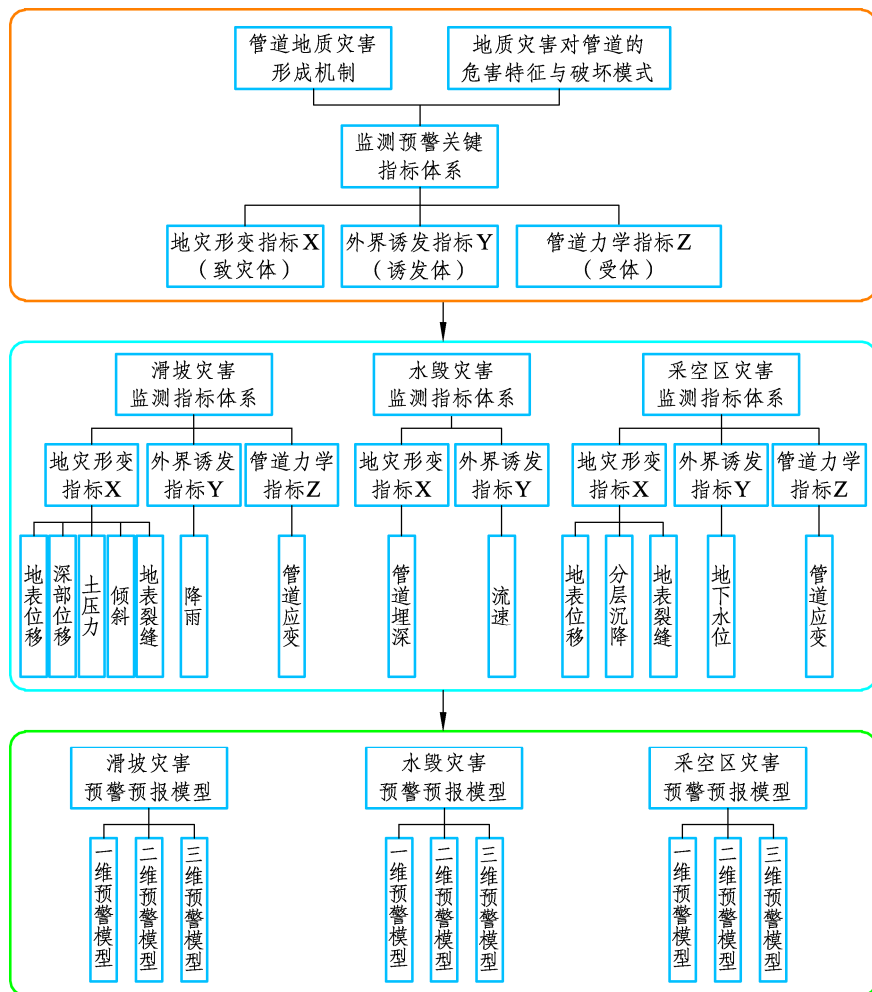


图 1.1 本书技术路线图

## 第 2 章 国内外研究现状

本章对数据异常及缺失的处理、数据预测、单指标预警模型、预警阈值动态调整这几个方面，调研分析了国内外研究现状，掌握了现阶段主流算法的适用性；总结前人的经验，为后期算法选型提供理论基础。

### 2.1 异常值消除和过滤方法研究现状

时间序列中的点异常检测是在一段时间序列数据中找出异常点。一般的做法是挖掘出时间序列中的正常模式或特征，然后与当前数据点和正常模式进行比较来判断是否异常。从目前的研究情况来看，根据数据集是否有标签（正常/异常），异常检测算法可以分为以下三类。

#### 1) 基于有监督学习的异常检测算法

这种类型的算法需要有标签的训练集数据和有标签的测试集数据。基于监督学习进行异常检测的基本思想是：先在有标签的训练集数据上训练一个分类器，通过测试集数据优化分类器，最后运用训练好的分类器去判断异常。这类算法比较典型的有分类回归决策树算法（CART）、支持向量机（SVM）和神经网络模型（如 RNN、LSTM 等）。基于有监督学习的异常检测算法针对的是有标签的数据集，所以异常检测的精确度一般都很高。

## 2) 基于半监督学习的异常检测算法

这种类型的算法同样使用训练集数据和测试集数据，不过训练集数据仅仅含有正常数据，不包含异常数据。基本思想是：根据训练集数据学习正常类的模式，训练得到一个正常类模型，然后运用模型检测到偏离正常模式的数据即为异常数据。这种类型的算法比较典型的有一类支持向量机（One-Class SVM）算法、自动编码器（Auto-Encoder）算法等。也可以使用密度估计算法对正常类的概率密度函数进行建模，得到正常类的概率模型，然后运用模型检测低于正常概率的数据即为异常数据，比如核密度估计（Kernel Density Estimation）算法。

## 3) 基于无监督学习的异常检测算法

这种类型的算法是最灵活的，因为不需要有标签的数据，而且使用无监督学习的训练集数据和测试集数据之间也没有很大差异。无监督异常检测的基本思想是：基于数据集本身的固有属性来进行异常评分。例如，基于距离的 K 均值（K-means）算法等、基于密度的局部异常因子（LOF）算法等都可以用来对数据集进行异常评分，最后根据异常分值判断是否为异常。另外，在系统框架层面，目前已有一些针对时间序列的异常检测算法，如 Yahoo Research 的 N. Laptev 等人于 2015 年提出的 EGADS 框架，通过对时间序列建模及预测，然后判断偏差值进行异常检测，目前已被运用于 Yahoo 内部的监控系统中。其不足在于，只能检测单维度时间序列的异常。Twitter 提出了用于异常监控的 R 数据包 Anomaly Detection，用于分析时间序列数据的异常，使用统计算法 Seasonal Hybrid ESD（SHESD）Algorithm 分析监控曲线的尖峰和低谷，可以分析得出是局部异常（local anomalies）还是全局异常（global anomalies），但只能对单维度时间序列进行异常检测。Solaimnsni 等（2015）提出的基于 Chi-square 的方法，对基于 VMWare 的云数据中心的性能数据时间序列异常检测，同样，只关注于单维度的时间序列异常的检测。L. Retting 等运用 KL 距离和 Pearson 相关度来比较两个时间序列的分布，以检测大数据流中的异

常，也只关注于单维度的时间序列异常的检测。Pavel Filonov (2016) 提出了单独使用 LSTM 进行时间序列数据预测，然后通过偏差值来进行异常检测，能够处理多维度时间序列，但是并没有考虑序列的上下文信息。

从以上研究可知，序列异常检测是在一个或多个时间序列集合中，使用已知的正常时间序列模式进行匹配，检测异常的时间序列模式。目前面向时间序列的异常检测算法，一方面主要是针对单维度时间序列，不能有效地对多维度时间序列进行处理；另一方面部分考虑多维度时间序列的算法仅仅是对序列进行了比较单一的处理，没有对序列进行更深层次的挖掘。

## 2.2 缺失值填补法研究现状

数据是科学研究中最重要的资源，在实际工程领域的监测数据中，经常会出现数据缺失现象。对缺失数据直接删除或者使用均值进行替换是一种简单易行的方法。直接删除方法虽然简单易行，但如果在缺失率较高或者数据变化幅度较大的情况下，不但会丢失一些重要的信息，而且会导致数据集规模减小；使用均值填补有一定的限制条件，且可能会改变数据变化的趋势。目前，常用的缺失数据处理方法可分为两类：一类是基于统计学的方法，另一类是基于机器学习的方法。

在基于统计学的方法中，均值填补在缺失数据小于 5% 时是一种有效的填补方法。在统计理论上，国外学者 Laird 等提出将最大期望算法 (Expectation Maximization, EM) 用于缺失值的填补，EM 算法是通过迭代进行极大似然估计 (Maximum Likelihood Estimate, MLE) 的优化算法，由于其简单性和广泛性特点，它在处理缺失数据方面得到了广泛应用，但是该算法存在计算复杂、收敛速度慢等缺点。

为了克服 EM 算法的弊端，Rubin 等在 EM 算法的基础上进行一系列的优化并提出了多重插补法，在一定程度上改善了单一插补的不足。为了对现有的 EM 方法进行改进，Ghapor 等 (2017) 将 Bootstrap 有放回抽

样思想和 EM 算法结合起来对缺失值进行填补，通过为缺失数据创建多个估计值，从而提高缺失值填补的精确度，此方法更适用于可以用线性函数关系表示的数据集中。

Damare 等（2017）提出了一种基于遗传规划和拉格朗日的时间序列插值方法，利用多元时间序列属性之间的关系来估计缺失值，然而，由于遗传规划过程的随机性，算法必须多次运行才能保证良好的结果。王一蓉等提出了一种基于遗传优化的调度控制系统缺失填补算法，该算法首先利用遗传优化算法估计不完整数据的最优参数，然后在此基础上利用马尔科夫链蒙特卡罗算法对缺失数据进行填补。Little 和 Rubin 等（2002）针对传统的缺失值填补方法做了详细论述，传统方法仅仅适用于缺失率较低的情况，随着缺失率不断提高，插补算法精度下降剧烈。由上述可以发现，随着对缺失值数据处理的要求不断提高，传统的缺失值处理方法已经逐渐失去了其应用价值。

在基于机器学习的方法中，唐慧强等提出了使用粗糙集理论结合径向基（Radial Basis Function, RBF）神经网络对气象缺失数据进行填补的方法，采用粗糙集提取关键属性作为网络的输入，采用径向基神经网络进行缺失值的估计。目前，常用的机器学习方法大多是 EM 结合聚类方法，或者是朴素贝叶斯、贝叶斯网络结合支持向量机或者决策树方法进行缺失数据的填补。Liu S.等（2014）提出了一种基于向量自回归（Vector Autoregression, VAR）模型的多变量缺失数据处理方法，并进行了仿真研究。卜范玉等提出了一种深度填充网络模型，该模型以自动编码器为基础模块构建三层网络模型来提取大数据的深度特征，进而对缺失数据进行填补。梁秉毅等提出一种基于优化决策树和 EM 的缺失数据填补算法，该算法首先利用优化后的决策树对数据缺失进行分类，然后使用 EM 算法对缺失数据进行填补。王妍等提出了一种基于泛化中心聚类的填补方法，采用聚类方法对数据进行分簇，并对缺失数据结合聚类结果一起进行缺失数据的填补。李国和等提出了一种基于聚类的递归填充算法，使用同类簇的均值对缺失值进行首次填充，得到完备数据集进行多次递

归聚类修正上一次填充的缺失值，直至相邻两次填充结果较稳定，实验结果表明，该方法与其他聚类方法相比有着较好的填充精度。

综合以上研究可以看出，缺失数据处理方法分为基于统计学的方法和基于机器学习的方法两类。随着对缺失数据处理的要求不断提高，传统的统计学方法难以满足当下的需求，基于机器学习的方法在不断的发展与创新，已有一些方法能达到较好的缺失值填补精度。

## 2.3 预测方法研究现状

监测及预测一般应用于桥梁、大坝、滑坡等易发生重大灾害的工程领域，通过分析监测数据进行变形预测，这种监测数据具有时间序列特性。根据实际观测到的时间序列数据，通过数学方法来建立预测模型进行曲线拟合，这种预测方法也被称为时间序列模型。常用的时间序列预测模型有自回归移动平均（Auto Regression Moving Average, ARMA）模型和差分整合移动平均自回归（Auto Regression Intergrated Moving Average, ARIMA）模型，其中 ARMA 模型主要针对平稳时间序列，当时间序列不平稳时可对时间序列进行一次或多次差分后再使用 ARIMA 模型进行预测。

在预测方面，唐浩等（2015）首先使用 ARMA 技术对桥梁历史静态监测量进行分析，并使用该技术对西安白蛇峪大桥应变、挠度、裂缝、倾斜监测量进行预测。结果表明，ARMA 单步预测误差小于 10%，拥有较高的预测精度。陆付民等为了实现对滑坡变形的预测，将多因素的预测模型中的模型参数看作带有动态噪声的状态向量，在此基础上建立基于多因素的 Kalman（卡尔曼）滤波模型，对模型参数进行持续更新，实验结果表明，该模型拥有较高的拟合精度及预测精度。王万祥（2018）利用 Kalman 滤波方法对边坡沉降监测数据进行去噪，并使用灰色预测模型进行预测，最后使用遗传算法（Genetic Algorithm, GA）优化的 BP 神经网络对残差进行修正，得到最终的边坡沉降预测值，结果显示拥有较