

DOI:10.19651/j.cnki.emt.2210302

## 大气电场测量数据的异常检测及校正方法研究\*

夏志祥<sup>1,2,3</sup> 李准<sup>4</sup> 徐伟<sup>1,2</sup>

(1.南京信息工程大学气象灾害预警预防与评估协同创新中心 南京 210044;  
2.南京信息工程大学江苏省气象探测与信息处理重点实验室 南京 210044; 3.盐城市第三人民医院 盐城 224000;  
4.江西省气象服务中心 南昌 330096)

**摘要:** 大气电场序列的清洗预处理对后续的挖掘研究具有重要意义。提出一种基于孤立森林结合 Chen-Liu 迭代算法的大气电场异常点检测与校正方法。该方法利用求和自回归移动平均(ARIMA)模型对大气电场时间序列进行拟合并得到拟合残差,基于残差序列构建孤立森林模型以确定异常点位置,最后通过 Chen-Liu 算法进行校正。通过模拟和实测大气电场数据验证所提方法的可靠性,相对于原序列,清洗后大气电场序列预测结果在均方根误差和平均百分比误差分别改善 27.8%和 34.98%。

**关键词:** 大气电场;异常检测与校正;孤立森林;Chen-Liu 迭代算法

**中图分类号:** TM863 **文献标识码:** A **国家标准学科分类代码:** 170.15

## Research on anomaly detection and correction method of atmospheric electric field measurement data

Xia Zhixiang<sup>1,2,3</sup> Li Zhun<sup>4</sup> Xu Wei<sup>1,2</sup>

(1. Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Key Laboratory for Aerosol-Cloud-Precipitation of China Meteorological Administration, Nanjing University of Information Science & Technology, Nanjing 210044, China; 3. Yancheng Third People's Hospital, Yancheng 224000, China; 4. Jiangxi Meteorological Service Center, Nanchang 330096, China)

**Abstract:** The cleaning of the atmospheric electric field is the key step of pretreatment, which is of great significance to the subsequent excavation research. In view of the shortcomings of traditional anomaly detection algorithm, which needs to specify the corresponding parameters and fail to use the relevant information between time series, a new outlier detection and correction method based on the combination of isolation forest and Chen-Liu algorithm is proposed. The method uses ARIMA model to combine the atmospheric electric field to get the fitting residual. The isolation forest model is constructed based on residual sequence to determine the location of the outliers. Finally, the Chen-Liu algorithm is used to correct the outliers. The reliability of the proposed method is verified by simulation series and the atmospheric electric field test. Compared with the original prediction, the results of the prediction of the series of the atmospheric electric field after cleaning are improved by 27.8% and 34.98% respectively in root mean square error and mean percentage error.

**Keywords:** atmospheric electric field; anomaly detection and correction; isolation forest; Chen-Liu algorithm

## 0 引言

大气电场是大气物理和空间物理领域的重要物理现象。大气中存在带正电荷的离子,地球表面携带负电荷,形成垂直向下的电场。实时监测大气电场时间序列对了解全球气候变化、反演雷暴云电荷结构、提高雷电预警效率具有

重要意义<sup>[1-3]</sup>。大气电场监测数据通常由大气电场仪测得,由于电场传感器对电场敏感,本身易受区域电磁环境以及气候等因素干扰<sup>[4]</sup>,同时采集的数据需要经历传输、存储和数据平台交互等环节,其中任何一个环节出现故障都会导致部分数据发生异常,对后续电场的挖掘与分析产生较大影响。为了更有效地利用监测数据,需要采用合理的方法

收稿日期:2022-06-13

\* 基金项目:国家重点研发计划政府间/港澳台重点专项(2021YFE0105500)、国家自然科学基金(41605121)项目资助

对大气电场数据进行预处理。

自 Malan 等<sup>[6]</sup>提出了经典的场磨式大气电场仪后,国内外学者主要通过优化电场传感器采集电路和校准实验来提高测量数据的精度<sup>[6-8]</sup>,或通过软件模拟仿真的方法对大气电场序列整体数值进行修正<sup>[9-11]</sup>。实际大气电场时间序列中还会包含一些与其他数据存在显著性差异的异常点,需要对这些异常点进行额外的检测和校正。国内外对时间序列异常点检测的研究起初是基于概率统计的方法,其核心思想是假设样本数据符合某种概率分布如高斯分布模型<sup>[12]</sup>,接着通过检验算法确定数据的异常情况。但在实际时间序列中,基本都表现出强随机性,难以将其归纳为某种概率模型。1999年,Knorr 等<sup>[13]</sup>提出基于最近邻的异常点检测方法,通过对所有数据进行距离计算确定异常点,这对于大量数据的异常检测而言计算过于庞大。2018年,Munir 等<sup>[14]</sup>提出的 DeepAnT 算法是一种基于深度学习的离群值检测方法,应用固定的卷积神经网络来预测未来的值,通过设置阈值来判定异常点,该方法属于有监督的机器学习方法,依赖于数据标签,同时转移到新的数据集需要重新训练。上述方法对于异常值的校正采用均值填充、众值填充等方法,忽视了时间序列内在的规律。Chen-Liu 迭代算法对时间序列进行建模分析,通过模型参数和异常点效应的联合估计能够准确检测异常点<sup>[15]</sup>,并对异常点进行有效地校正。但异常检测依赖于阈值 C 的选择且计算繁琐。

针对 Chen-Liu 迭代算法异常检测部分的不足,利用孤立森林自适应的优点<sup>[16]</sup>,将孤立森林与 Chen-Liu 迭代算法相结合,构建异常检测和校正模型。首先建立 ARIMA 时间序列模型得到残差序列。接着利用 Chen-Liu 迭代算法对检测到的异常点进行校正,构建大气电场序列异常检测和校正模型。通过实验仿真验证,两者结合可以准确地检测异常点并有效地进行校正,具有一定的可靠性和实用性。

## 1 大气电场时间序列的异常检测及校正原理

对大气电场时间序列进行数学建模分析,能够得到序列内在的规律和时序相关信息,可用于区分正常点与异常点。针对大气电场时间序列通常表现为非平稳状态,利用求和自回归移动平均模型 (autoregressive integrated moving average model, ARIMA)<sup>[17]</sup>,对原序列进行若干次差分使其成为平稳序列,对平稳序列建立自回归移动平均 (autoregressive moving average model, ARMA) 进行分析。

### 1.1 Chen-Liu 迭代算法

Chen-Liu 迭代算法是检测与校正 ARIMA 模型中异常点的有力方法。其核心环节是通过计算时间序列中每个时间点的异常效应标准统计量  $\hat{\tau}(n)$ , 并比较其绝对值与指定值 C 的大小,具体分析思路为:

假设  $x(n)$  为没有受到异常点影响的大气电场时间序列,且服从 ARIMA( $p, d, q$ ) 模型,可表示为:

$$x(n) = \frac{B(z)}{A(z)\nabla^d}\varepsilon(n) \quad (1)$$

$$A(z) = 1 + a_1z^{-1} + \dots + a_pz^{-p} \quad (2)$$

$$B(z) = 1 + b_1z^{-1} + \dots + b_qz^{-q} \quad (3)$$

式中:  $\varepsilon(n)$  为一离散白噪声序列,  $\nabla^d$  表示对序列做  $d$  阶差分,  $z^{-j}$  定义为后向移位算子:

$$z^{-j}x(n) = x(n-j) \quad (4)$$

在通常情况下,一个被观测的大气电场序列可能在多个的时间点上受异常值的影响,模型表达变为:

$$x(n)^* = x(n) + \sum_{j=1}^k \omega_j I_n(t_j) \quad (5)$$

$$I_n(t) = \begin{cases} 1, & n = t \\ 0, & n \neq t \end{cases} \quad (6)$$

式中:  $x(n)^*$  表示受到异常值影响的大气电场序列,  $I_n(t)$  为示性函数,  $k$  为异常点个数,  $\omega_j$  对应不同异常点的异常效应大小。因异常数据会影响时间序列拟合的精度,可通过对拟合残差的分析将异常数据的影响进行量化;单个异常点的处理方法如下,多个异常点的处理可以通过多次迭代的方式进行。

假设大气电场时间序列在  $t$  点收到异常值影响,序列的拟合残差为  $e(n)$ :

$$e(n) = \omega \pi(z) I_n(t) + \varepsilon(n) \quad (7)$$

$$\pi(z) = \frac{B(z)}{A(z)\nabla^d} = 1 - \pi_1 z^{-1} - \pi_2 z^{-2} - \dots \quad (8)$$

式中:  $\pi(z)$  表征残差影响的算子,  $\pi_1, \pi_2, \dots$  为  $\pi(z)$  相应的参数,对式(6)以矩阵方式扩展开来,有:

$$\begin{bmatrix} e(1) \\ \vdots \\ e(t-1) \\ e(t) \\ e(t+1) \\ \vdots \\ e(L) \end{bmatrix} = \omega \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ -\pi_1 \\ \vdots \\ -\pi_{L-t} \end{bmatrix} + \begin{bmatrix} \varepsilon(0) \\ \vdots \\ \varepsilon(t-1) \\ \varepsilon(t) \\ \varepsilon(t+1) \\ \vdots \\ \varepsilon(L) \end{bmatrix} \quad (9)$$

依据最小二乘估计理论,可得该点异常效应大小估计值  $\hat{\omega}$ , 为:

$$\hat{\omega} = \frac{e(t) - \sum_{i=1}^{L-t} \pi_i e(t+i)}{\sum_{i=0}^{L-t} \pi_i^2} \quad (10)$$

异常点检测方法是比异常点的标准检验统计量  $\hat{\tau}(t)$  的绝对值与预先给定的常数 C 的大小;若大于 C,则该点为异常点,异常效应大小为  $\hat{\omega}$ 。标准检验统计量表达式如下:

$$\hat{\tau}(t) = \frac{\hat{\omega}}{\sigma_\varepsilon} \sqrt{\sum_{i=t}^L \pi_i^2} \quad (11)$$

对于  $\sigma_\varepsilon$  的计算,为避免异常点存在导致计算值变大,采取 omit-one 法,除去该时刻数据,求残差序列的方差。

Chen-Liu 迭代算法主要参数  $C$  的选取往往取决于经验值且与序列的长度相关,会影响异常检测与校正的泛化能力;且对每个点计算标准检验统计量,会导致运算大、耗时长等问题。

1.2 孤立森林

孤立森林是一种基于集成思想的异常值检测算法,具有不需要预先设置参数(如距离或密度)、模型稳定性高、精度高等优点,且构建孤立森林属于无监督学习,无需设置标签,训练过程得到简化。孤立森林算法核心是对数据集样本进行子采样并构造孤立决策树,然后将多个决策树组成孤立森林以检测异常数据。构建孤立森林步骤如下:

- 1) 从待检测大气时间序列  $x(n)$  中随机选择  $\psi$  个样本点作为子样本集,放入树的根节点;
- 2) 随机选取子样本集的一个维度,在该维度中随机产生一个切割点  $p$  (切割点产生于指定的维度的最大值和最小值之间);
- 3) 用此切割点生成一个超平面,将当前节点的数据空间划分成两个子空间,把指定维度中小于  $p$  的数据点放在当前节点左子节点中,把大于等于  $p$  的数据点放在当前节点的右子节点;
- 4) 递归执行步骤 2) 和 3), 不断构造新的子空间节点,当数据不再可分或已达到树的最大深度  $\log_2 \psi$ , 停止递归;
- 5) 重复上述步骤,直到获得  $t$  棵孤立决策树后,孤立森林形成,训练过程结束。

对待测数据样本,令其遍历孤立森林的每一棵决策树,计算检测样本中的每一个值落在每棵决策树的层数,最后得出样本  $q$  在每颗树的平均深度。异常分值与样本在孤立决策树的深度有关,当样本在决策树的深度越小,则异常分值越高,即该样本为异常样本的概率越高。对  $n$  个数据样本,其平均搜索长度  $c(n)$  为:

$$c(n) = 2H(n-1) - \frac{2n-1}{n} \quad (11)$$

其中,  $H(i)$  为谐波数,表达式为:  $\ln(i) +$  欧拉常数。通过对所有样本点的长度进行归一化处理,可得到介于 0~1 之间的异常分值,记  $s(q, n)$  为异常指数,有:

$$s(q, n) = 2 \frac{E(h(q))}{c(n)} \quad (12)$$

$E(h(q))$  是某数据点在每棵孤立决策树的路径长度期望值。可以从上式推导出:当  $s$  趋于 0 时,它可能是正常值,而当  $s$  趋于 1 时,它可能是异常值。

1.3 异常检测与校正步骤

因异常点的存在会使序列模型参数的估计产生偏差,在针对异常点出现时刻与个数未知、预先不知道模型参数的情况下,使用迭代检验的方法对大气电场序列进行异常点检测与校正,具体流程如图 1 所示。

- 1) 数据采集与处理:利用大气电场仪采集地面大气电场。并进行数据标定,得大气电场信号  $x(n)$ 。
- 2) 建立模型:对大气电场序列  $x(n)$  建立 ARIMA 模

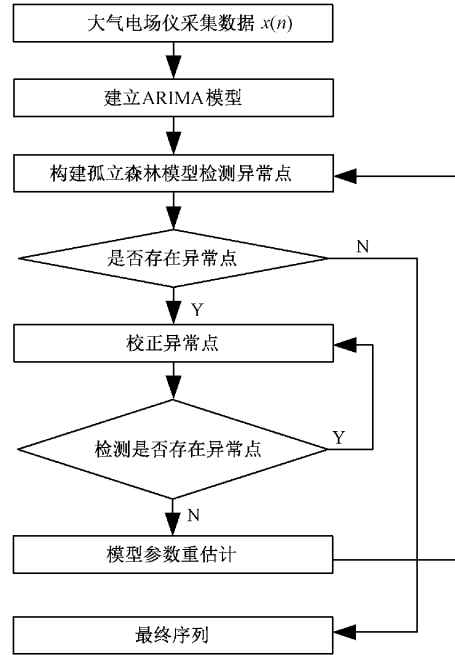


图 1 异常检测与校正流程

型,并依据估计的模型计算初始残差  $e(n)$ :

$$e(n) = \hat{\pi}(z)x(n) = \frac{\hat{\varphi}(z)\nabla^d}{\hat{\theta}(z)}x(n) \quad (13)$$

3) 外循环检测异常点:利用步骤 1) 得到的残差序列训练孤立森林模型;将训练好的模型去检测残差序列是否存在异常点。若存在异常点,进入内循环中校正异常值;否则,跳转到步骤 5)。

4) 在内循环中校正异常值:若  $t_1$  时刻为异常点,异常效应大小  $\omega$  由式(9)求得,通过修正可得新的残差序列:

$$\tilde{e}(n) = e(n) - \omega \hat{\pi}(z)I_n(t) \quad (14)$$

使用迭代的方法校正所有的异常点,在修正后的残差序列基础上再用训练好的孤立森林模型进行检测,若仍存在异常点,重复步骤 3); 否则内循环结束。

5) 修正大气电场序列:假设内循环结束检测出  $k$  个异常点,分别在  $t_1, \dots, t_k$  时刻,异常效应大小分别为  $\omega_1, \dots, \omega_k$ ; 通过校正,得到新的大气电场序列  $x^m(n)$  ( $x^0(n)$  表示第  $m$  次外循环迭代得到的序列,  $x^0(n)$  表示原序列):

$$x^m(n) = x^{m-1}(n) - \sum_{i=1}^k \frac{\hat{\varphi}(z)\nabla^d}{\hat{\theta}(z)}\omega_i I_n(t_i) \quad (15)$$

重新估计模型参数,得到新的残差序列;接着回到步骤 2) 重新进行异常检测。

6) 最终模型参数估计:最后一次异常检测后,对已经修正所有异常点的对大气电场序列进行联合估计,得到最终的时间序列模型;同时得到所有异常点的位置和异常效应大小。

2 实验和分析

为验证所提方法的有效性,进行两组实验。第 1 组实

验利用模拟时间序列,人为加入异常点,构建异常检测和校正模型,验证其有效性和可靠性。第2组实验采用大气电场时间序列作为待检测样本,通过比较清洗前后序列的预测效果,验证所提方法的实用性。

2.1 模拟序列实验

考虑如下模型:

$$x(n) = 1.5x(n-1) - 0.75x(n-2) + r(n) + 0.9r(n-1) \quad (16)$$

其中,  $x(n)$  为大气电场时间序列,  $r(n)$  为加入的白噪声,  $r(n)$  服从标准正态分布  $N(0,1)$ 。通过模拟产生符合该模型的时间序列,定义长度 800;在模拟序列中设置 8 个异常点,最终得到观测样本时间序列如图 2 所示,图 2 中横轴为数据序列号,纵轴为大气电场数值。大气电场幅值变化较大,肉眼难以判断所有设置的异常值。

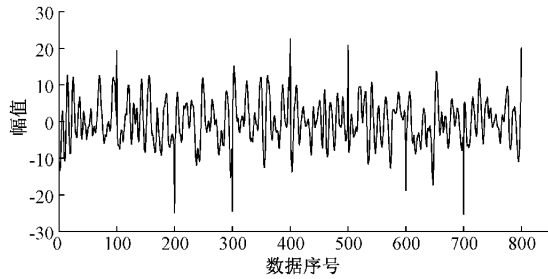


图2 模拟序列的时间序列图

检测结果如表 1 所示,最终拟合的模型表达式为:

$$x(n) = 1.5129x(n-1) - 0.7507x(n-2) + r(n) + 0.9363r(n-1) \quad (17)$$

由表 1 可知,异常点的位置能够被准确识别出来,且异常效应大小与实际值相差不大;与原时间序列模型相比,模型参数的估计值与实际值也是接近的,验证了该方法检测与校正异常数据的有效性。

表 2 所示为所提方法与 Chne-Liu 迭代算法的比较结果。评价指标为运行时间和均方根误差(RMSE),均方根误差计算方法如下:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (\hat{x}(t) - x(t))^2}{N}} \quad (18)$$

表3 不同阶数模型下测试结果统计

模型	误检率/%		漏检率/%		RMSE		运行时间/s	
	孤立森林结合 Chen-Liu 法	Chen-Liu 算法	孤立森林结合 Chen-Liu 法	Chen-Liu 算法	孤立森林结合 Chen-Liu 法	Chen-Liu 算法	孤立森林结合 Chen-Liu 法	Chen-Liu 算法
3 阶模型	5.8	8.2	0	0	0.15	0.15	2.58	7.23
4 阶模型	7.4	8.5	3.3	4.3	0.19	0.22	2.97	7.91
5 阶模型	8.7	10.2	4.1	4.8	0.28	0.57	3.60	9.37
6 阶模型	9.3	11.4	5.6	5.7	0.31	0.89	5.2	10.93

表1 模拟序列异常点检测结果

异常点位置	实际设置异常大小	孤立森林结合 Chen-Liu 检测 异常效应大小	Chen-Liu 算法检测异常 效应大小
100	20	20.29	20.22
200	18	-18.22	-17.97
300	24	-23.52	-24.31
400	22	20.64	22.03
500	26	25.92	26.00
600	-16	-16.11	-16.32
700	-26	-25.89	-26.09
800	12	12.33	10.69

表2 算法性能比较

算法	Chen-Liu 迭代算法	孤立森林结合 Chen-Liu 检测异常效应大小
RMSE	0.049	0.054
运行时间/s	6.95	1.88

式中:  $\hat{x}(t)$  表示修正后的序列,  $x(t)$  为实际序列,  $N$  为序列长度。

结合表 1 和 2,与 Chen-Liu 迭代算法相比,孤立森林结合 Chen-Liu 检测异常效应大小大部分异常点的校正效果略劣于 Chen-Liu 迭代算法,均方根误差 RMSE 增大 0.05,但总体相差不是很大;同时在运行时间上,孤立森林结合 Chen-Liu 检测方法有着显著的提升,提升了约 5 s,表明该方法时效性更强。

2.2 多序列实验

为进一步验证方法的可靠性,采用多组不同阶数的时间序列进行试验。固定 MA 阶数为 1,模拟 AR 阶数分别为 3、4、5、6 的时间序列,试验不同模型的适用程度;同时每个模型产生 10 组不同长度的序列,并在每个序列中随机加入 5% 的异常值。这里用 4 个统计量检测指标进行说明,分别为误检率、漏检率、RMSE 和运行时间(每个指标是不同阶数所有序列指标的平均值)。最终检测效果如表 3 所示。

由表 3 结果看出,对于不同阶数模型的序列,本文所提方法都能够得到很好的应用,得到较为准确的检测与校正结果;与 Chen-Liu 迭代算法相比,有着较低的误检率和漏检率,这是因为避免了参数阈值的选取,具有更好的泛化能力,能够适应不同模型的时间序列;同时在整体的均方根误差和运行时间上也优于 Chen-Liu 迭代算法,验证了所提方法的可靠性。

2.3 基于大气电场时间序列的异常检测与修复

上节对理想时间序列的异常点检测和校正的有效性 and 可靠性进行了验证,现将其应用于大气电场时间序列的数据清洗中。研制的大气电场仪观测现场如图 3 所示。选取 2018 年 1 月 2 日某段大气电场时间序列(采样间隔为 1 s)进行异常值检测和校正,检测结果如图 4 及表 4 所示。



图 3 大气电场观测实验

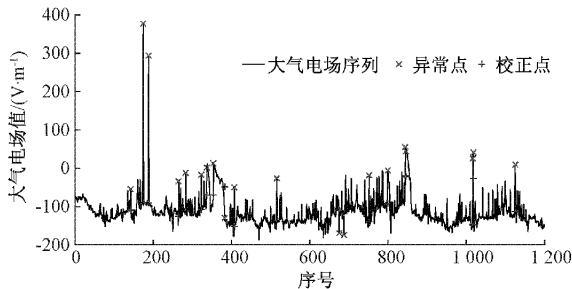


图 4 大气电场序列清洗结果

从结果可看出,某些明显的异常点能够被检测并校正,如时间序列的第 174 和 188 个观测点。考虑到大气电场序列的异常点和异常效应大小均是未知,采用文献[18]所提出的长短期记忆网络(long short-term memory, LSTM)方法进行大气电场值预测<sup>[18]</sup>,LSTM 拥有记忆功能,能够通过学习来挖掘时间序列之间隐藏的长期依赖关系<sup>[19-21]</sup>。通过分析清洗前后序列的预测结果来验证所提方法的实用性。具体步骤为分别采用原始序列前 1 000 点和清洗后序列的前 1 000 点预测未来 200 点的数据,其中待预测 200 点数据是经过清洗过后的数据,预测评价指标采用 RMSE 和 MAPE,RMSE 指均方根误差,MAPE 指平

表 4 大气电场序列异常位置及校正结果

位置	记录值/(V·m <sup>-1</sup> )	异常效应大小
142	-54.7	55.55
174	377.6	468.86
188	294.7	386.08
265	-34.12	91.77
283	-11.76	97.24
323	-16.47	91.19
337	1.75	72.36
354	13.53	84.03
382	-130	-81.3
407	-48.41	103.14
516	-25.88	112.63
675	-169.41	-59.95
687	-174.72	-56.1
752	-18.24	121.31
799	-6.47	85.69
844	55.88	77.59
845	43.53	64.31
1 017	25.88	-112.63
1 018	40.59	67.23
1 125	10	-122.41

均百分比误差,计算公式如下:

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (\hat{y}(t) - y(t))^2}{N}} \quad (19)$$

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{\hat{y}(t) - y(t)}{y(t)} \right| \quad (20)$$

式中:  $\hat{y}(t)$  表示预测序列,  $y(t)$  为实际序列,  $N$  为序列长度。预测结果如图 5 和表 5 所示。

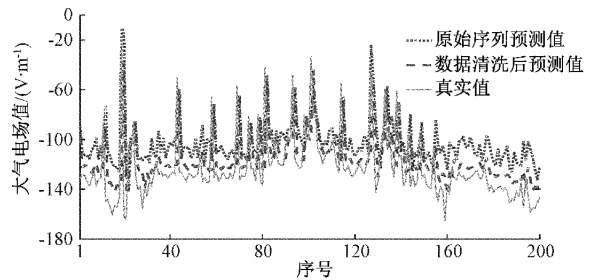


图 5 大气电场时间序列预测值与真实值

表 5 大气电场时间序列预测结果对比

序列	RMSE	MAPE
原始序列	31.85	27.27
清洗后序列	22.98	17.73

图 5 可以看出两种序列的预测结果在前 160 点相差不大,但从后 40 个数值的预测结果来看,清洗后的序列预测

效果明显优于原始序列的预测,与真实序列更为接近。由表5可知,清洗后的序列预测结果的均方根误差为22.98,平均百分比误差为17.73%,相对于原始序列,分别提高了27.8%和34.98%,表明对大气电场序列进行异常点检测和校正,有益于后面的进一步分析与挖掘,提供了更高质量的数据,验证了所提方法的实用性。

### 3 结 论

针对大气电场时间序列的非平稳性,利用ARIMA模型对大气电场序列进行建模,对大气电场时间序列进行拟合得到拟合残差,基于残差序列构建孤立森林异常检测模型,以确定异常点位置,弥补传统异常检测方法需要指定参数的不足;接着通过Chen-Liu迭代算法对检测到的异常点进行校正。

孤立森林结合Chen-Liu检测异常点的校正效果与Chen-Liu迭代算法接近,但在运行时间上,有着显著的提升,时效性更强。

对于不同阶数模型的序列,孤立森林结合Chen-Liu算法有着较低的误检率和漏检率,这是因为避免了参数阈值的选取,具有更好的泛化能力,能够适应不同模型的时间序列。在整体的均方根误差和运行时间上也优于Chen-Liu迭代算法。

对于原序列预测,采用孤立森林结合Chen-Liu清洗后大气电场序列预测结果在均方根误差和平均百分比误差都得到了明显改善。为后续的大气电场序列挖掘技术研究提供了更高质量的数据。

论文中主要检测和校正加性异常点(指某个孤立的异常点),在以后的实验中考虑扩展到连续的异常点。同时清洗后的大气电场序列对雷电预警和雷暴云电荷结构反演的优化效果,需要进一步研究。

### 参考文献

- [1] TAN Y, GUO X, ZHU J, et al. Influence on simulation accuracy of atmospheric electric field around a building by space resolution[J]. *Atmospheric Research*, 2014, 138: 301-307.
- [2] XU W, ZHANG C, JI X, et al. Inversion of a thunderstorm cloud charging model based on a 3D atmospheric electric field[J]. *Applied Sciences*, 2018, 8(12): 2642.
- [3] MENG Q, YAO W, XU L. Development of lightning nowcasting and warning technique and its application[J]. *Advances in Meteorology*, 2019, DOI: 10.1155/2019/2405936.
- [4] YANG P, CHEN B, WEN X, et al. A novel MEMS chip-based atmospheric electric field sensor for lightning hazard warning applications[C]. 2015 IEEE Sensors, IEEE, 2015: 1-4.
- [5] MALAN D J, SCHONLAND B F J. The electrical processes in the intervals between the strokes of a lightning discharge [J]. *Proceedings A*, 1951, 206(1085):145-163.
- [6] 徐伟,夏志祥,行鸿彦.基于集成经验模态分解和极端梯度提升的雷电预警方法[J]. *仪器仪表学报*, 2020, 41(8):235-243.
- [7] FORT A, MUGNAINI M, VIGNOLI V, et al. Design, modeling, and test of a system for atmospheric electric field measurement [J]. *IEEE Transactions on Instrumentation and Measurement*, 2011, 60(8):2778-2785.
- [8] KAINZ A, STEINER H, SCHALKO J, et al. Distortion-free measurement of electric field strength with a MEMS sensor[J]. *Nature Electronics*, 2018, 1(1): 68-73.
- [9] 罗福山,何渝晖,张华伟,等.电场的标定方法[J]. *空间科学学报*, 2007, 27(3):223-226.
- [10] ALESSANDRO F. On the optimum rod geometry for practical lightning protection systems[J]. *Journal of Electrostatics*, 2007, 65(2):113-121.
- [11] 周先春,邹宇,陈凯峰,等.建筑物不同参数对大气电场畸变效应的数值模拟研究[J]. *电瓷避雷器*, 2017(4):8-15.
- [12] HOETING J, RAFTERY A E, MADIGAN D. A method for simultaneous variable selection and outlier identification in linear regression[J]. *Computational Statistics & Data Analysis*, 1996, 22(3): 251-270.
- [13] KNORR E M, NG R T. Finding intensional knowledge of distance-based outliers[C]. *Vldb*, 1999, 99: 211-222.
- [14] MUNIR M, SIDDIQUI S A, DENGEL A, et al. DeepAnT: A deep learning approach for unsupervised anomaly detection in time series[J]. *IEEE Access*, 2018, 7: 1991-2005.
- [15] CHEN C, LIU L M. Joint estimation of model parameters and outlier effects in time series [J]. *Journal of the American Statistical Association*, 1993, 88(421): 284-297.
- [16] LIU F T, TING K M, ZHOU Z H. Isolation forest[C]. 2008 Eighth IEEE International Conference on Data Mining, IEEE, 2008: 413-422.
- [17] 刘士进,孙立华,郭鹏,等.基于EEDM-Arima算法的某地区电力系统负荷预测及算例分析[J]. *电子测量技术*, 2020, 43(7): 185-188.
- [18] GUO T, LIU R, YANG H, et al. Predict atmosphere

- electric field value with the LSTM neural network[J]. 2017 International Conference on Computer Systems, Electronics and Control(ICCSEC), IEEE, 2017: 263-266.
- [19] 李孟威,史元浩,杨彦茹,等. 融合 EMD 和 LSTM 的受热面积灰预测研究[J]. 电子测量与仪器学报, 2020, 34(11):166-172.
- [20] 陈安民,张春元,张泽林. 基于长短期记忆网络的无人机认知无线电频谱预测[J]. 国外电子测量技术, 2021,40(1): 37-43.
- [21] 谭峰,李成男,萧红,等. 基于 LSTM 循环神经网络的数控机床热误差预测方法[J]. 仪器仪表学报, 2020,

41(9): 79-87.

### 作者简介

夏志祥,硕士,主要研究方向为仪器仪表技术和信号检测与处理。

E-mail:1217937285@qq.com

李准,硕士,高级工程师,主要研究方向为雷电防护和雷电预警技术。

E-mail:zhun.lee@163.com

徐伟(通信作者),博士,硕士生导师,主要研究方向为气象观测方法及仪器。

E-mail:xw@nuist.edu.cn