

一种基于 PSO-VMD 和 LSTM 的复杂山地风电场 观测风速数据质量控制算法

熊雄¹, 姚润进², 程帅兵³, 李文龙⁴, 钱栋²

(1. 南京信息工程大学江苏省大气环境与装备技术协同创新中心, 信息与系统科学研究院, 南京 210044;

2. 中国电建集团江西省电力建设有限公司, 碳中和研究院, 南京 210018;

3. 江苏省海上风电叶片设计与制造技术重点实验室, 南京 210000;

4. 中国三峡新能源(集团)股份有限公司江西分公司, 南昌 330038)

摘要: 复杂山地风电场普遍存在观测风速数据质量差引起风资源评估误差大、风功率预测精度低的问题。而复杂山地风速呈现较强的间歇性、波动性和非平稳性, 导致常规质量控制方法无法有效提高数据质量。针对此, 提出一种基于粒子群改进变分模态分解和长短期记忆网络的集成学习算法(PVL), 并应用于复杂山地观测风速的质量控制以提高风速数据的质量。以广西某复杂山地风电场内5基观测塔2015—2016年逐10 min风速数据为案例进行PVL应用效果检验, 并与传统单站及空间回归法、反距离加权法进行对比。应用表明, PVL比传统方法具有更高的寻误率, 且在异地形、多风况上具有更强的适应性。

关键词: 风电场; 质量控制; 粒子群; 变分模态分解; 长短期记忆网络

中图分类号: TK89

文献标志码: A

0 引言

在全球气候变暖的大背景和“双碳”目标引导下, 全球能源结构低碳化转型在加速推进, 风电等可再生能源发展步伐加快。随着风力发电的大规模开发建设和并网, 在“竞价/平价”、电力市场化交易等机制下, 风能的“间歇性”问题得到凸显^[1]。一方面, 精准的风资源和发电量评估是风场投资开发阶段的迫切需要; 另一方面, 减少大规模风电并网爬坡事件、增强新型电力市场交易竞争力对在役风场风功率预测准确率的提升提出更高的要求。风速, 作为影响风能特征分布最主要、直接的因素, 具有明显的间歇性、波动性和非平稳性^[2], 这是风能呈现“间歇性”的内因, 尤其是复杂地形情况下。复杂地形情况下风电场的观测风数据是精细化评估的最主要参考依据, 其质量直接关系到投资收益率及并网冲击^[3]。根据 GB/T 18710—2002《风电场风能资源评估方法》(以下简称《标准》)规定, 用于风场产能评估的风速至少需一个完整年观测周期, 且数据有效完整率需达到 90% 以上^[4]。如何提高复杂地形情况下风速观测资料的质量, 是“双碳”目标进程中必须面对的挑战, 其核心问题是质量控制算法的构建。

风电场观测风速作为一种小尺度观测, 本质上是一种气

象观测资料。中国地面气象观测资料质量控制方法根据参考资料的差异分为单站和多站级质量控制技术路线。多站质量控制, 即融合周边观测数据构建相关模型进行控制。近年来国内外一些学者提出一些有效的方法。空间回归检查法(spatial regression test, SRT)^[5]、最佳内插法^[6]、反距离加权法(inverse distance weight IDW)^[7]通过观测值与理论值之间的差值来标记可疑数据; 贝叶斯质量控制法^[8]引入贝叶斯理论, 从观测资料概率统计的角度进行质量控制。熊雄等^[9]基于空间观测资料差异, 构建空间差异方程来进行质量控制; 叶小岭等^[10]以主成分分析优化选择参考站并通过随机森林学习完成地面气温的质量控制; 叶星瑜等^[11]引入时间卷积网络来对铁路沿线大风样本进行质量控制。然而, 这些方法都严重依赖周边观测资料。复杂山地风电场风速观测一般为小区域、短周期、单要素, 其观测时长和观测密度均较低, 且受微地形及小气候影响严重; 各观测塔数据存在仪器损坏、数据传输故障、结冰、倒塔等现象, 产生大量的随机误差、系统误差、微气象误差和粗大误差, 数据质量差, 且各观测塔之间数据相关性差^[12]。在这些观测密度低、相关性差的山地风电场, 想获得高质量的风速观测资料, 探索和研究有效的单站级质量控制方法显得尤为重要^[13]。

收稿日期: 2022-12-02

基金项目: 国家自然科学基金(42205150; 42275156); 江苏省自然科学基金(BK20210661); 中国电建集团江西省电力建设有限公司科技项目(JEPCC-KYXM-2023-002)

通信作者: 姚润进(1991—), 男, 硕士、助理研究员, 主要从事气象观测资料质量控制与风资源评估等方面的研究。runjin_1991@163.com

传统单站质量控制法包括格式检验、范围检验、极值检验和内部一致性检验的综合法^[14],是目标资料无有效周边参考信息时的有效性和准确性的基本保障。《标准》内对风电场观测风速的质量控制方法本质上属于传统级单站控制范畴,其能有效地标记出常规观测资料中的疑误数据,如物理极限值、冰冻“僵值”等,但其对特定条件下的异常值缺乏有效的检出,尤其在风况复杂多变的山地观测风场。为进一步提升复杂风况下的质量控制效果,本文从风速时间序列非平稳性及复杂度着手,提出一种 PSO 优化 VMD 和 LSTM 的集成算法,构造一种检错率高、适用于复杂山地风电场的观测风速的质量控制算法(PVL_QC)。

1 方 法

1.1 VMD和LSTM集成算法及优化

变分模态分解^[15](variational modal decomposition, VMD)是一种尺度可变化的信号处理方法,对信号噪声具有较好的鲁棒性。相较于经验模态分解方法,VMD 避免了端点效应和模态混叠问题,可降低复杂度高、非线性强的时间序列的复杂度^[16-17]。

长短时记忆网络(long short-term memory, LSTM)^[18]是一种时间递归神经网络,可对时间状态实现自动保存与更新,是循环神经网络(RNN)的变体,有效避免 RNN 的梯度消失或爆炸现象,多应用在时间序列预测、分类、文本挖掘等领域^[19-21],且呈现较好的效果。

考虑 VMD 和 LSTM 可降低时间序列的非平稳性和复杂度,以及挖掘重构的能力,本文集成 VMD 和 LSTM。即,首先通过 VMD 方法及其参数的粒子群(particle swarm optimization, PSO)寻优将时序风速信号分解成一系列异构模态分量,降低原风速序列的非平稳性和复杂度;然后,对各个模态分量均建立 LSTM 学习网络预测模型;最终将各子模态预测结果重构为时间序列预测值并引入到质量控制领域。

VMD 的核心参数主要有原始信号分解的模态参数 K 、惩罚因子 α 。其中,模态参数取值过大,即过度分解,会造成原始信号的丢失;模态参数取值过小,欠分解,各分解模态相互影响,达不到降低信号的复杂性;惩罚因子取值关系强干扰信号下模态分解的正确性。一般情况下, K 和 α 多根据实际情况进行选择,主观性和不确定性较大。

考虑 PSO 算法^[22]在复杂系统参数寻优上表现出稳健的全局收敛能力。因此,本文通过 PSO 最优化 K 和 α 。优化算法的主要思想是:将 VMD 的两个参数 K 和 α 看成 PSO 的两个粒子,将包络熵 E_p 作为粒子的适应度值^[23],如式(1)所示。改进算法流程图如图 1 所示。

$$\begin{cases} E_p = -\sum_i^N p_i \lg p_i \\ p_i = \frac{a(i)}{\sum_i^N a(i)} \end{cases} \quad (1)$$

式中: $a(i)$ ——原始信号 f 经希尔伯特变换解调后的包络信号; p_i —— $a(i)$ 的归一化表达。

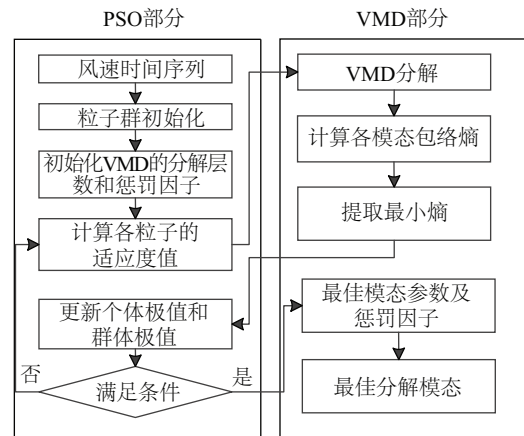


图 1 PSO 优化 VMD 算法流程图

Fig. 1 Flow chart of VMD improved with PSO algorithm

1.2 PVL 质量控制法

PVL_QC 具体步骤如下:

1) 记当前 t 时刻,取其前 n 长度的历史风速序列数据(本文取 10 d,即 $n = 1440$) $\{v(t-1), v(t-2), \dots, v[t-(n-1)]\}$ 。通过 PSO 最优化 K 和 α 进行 VMD,得到 K 个子模态序列,对各子模态进行重构样本向量,时间延迟取 1,可得到 $N(N = n - 1)$ 个 LSTM 样本。其中, $[N - 1]$ 个作为训练样本集和测试样本集,即 $\{(V_i^k(j), v_i^k(j+1)), j = 1, 2, \dots, N - 1\}$; 1 个为待估样本输入矢量 $V_i(j = N)$,输出即为 t 时刻子模态风速。

2) 根据 $[N - 1]$ 个训练、测试样本集,运用 LSTM 算法进行训练,得到各子模态时间序列模型,即 $v_i^k(t) = \text{lstm}_k(V_i^k(j), v_i^k(j+1))(j = 1, 2, \dots, N - 1)$,则各子模态在 t 时刻的预测值记为 $\hat{v}_i^k(t)$, $\hat{v}_i^k(t) = \text{lstm}_k(V_i^k(j), v_i^k(j+1))(j = N)$ 。

3) 对各子模态预测进行重构,可得 t 时刻风速预测值 $\hat{v}(t) = \sum_{k=1}^K \hat{v}_i^k(t)$ 。

4) 将当前 t 时刻的风速观测值 $v(t)$ 与步骤 2) 中的无偏估计进行对比,具体参照 1.3 节假设检验方案进行。

5) 重复步骤 1)~步骤 4),对目标观测站下一时刻风速时间序列进行检验。

1.3 假设检验方案

为了检验质量控制方法的可行性,本文采用 Hubbard 等^[5]在研究 SRT 方法时用到的假设检验方法。根据验证数据即经过严格质量控制后的风场原始风速观测数据,利用随机植入随机误差的方法模拟观测值,称为模拟观测数据;然后基于模型求得的预测值与模拟观测值的比对完成质量控制。具体地,模拟观测数据如式(2)所示。

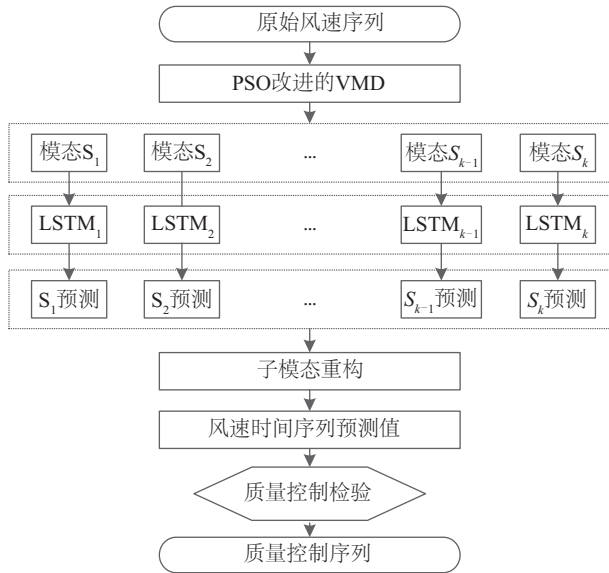


图2 PSO-VMD-LSTM集成学习算法流程图

Fig. 2 Flow chart of PSO-VMD-LSTM ensemble learning algorithm

$$\begin{cases} v_{\text{obs}} = v + v_c \\ (v_c)_t = s_t p_l \end{cases} \quad (2)$$

式中: v_{obs} ——模拟观测风速; v ——基准观测风速; v_c ——模拟人工误差; p ——服从区间 $[-3.5, 3.5]$, 均值为 0 的均匀分布的随机数; s ——被检测观测站风速时间序列的标准差; l ——误差插入的位置。

进一步地, 假设质量模型的无偏估计为 v_{est} , 则通过比较某时刻 v_{obs} 与 v_{est} , 如式(3)所示。若式(3)成立, 则认为该观测值存在疑误; 否则, 通过检验。

$$\begin{cases} |v_{\text{obs}}(t) - v_{\text{est}}(t)| \geq \theta, t = 1, 2, \dots, L \\ \theta = f \cdot \sigma \end{cases} \quad (3)$$

式中: θ ——质量控制阈值; f ——质控因子, 其值大小与质量控制后的资料质量有关; σ ——被检测观测站观测数据的标准差; L ——观测数据的采样长度。

1.4 质量控制算法评估指标

上述质量控制方案可概括为两个部分。一是根据基准参考资料, 通过质量控制算法获取被检站的预测值(预测资料); 二是根据预测资料对观测资料做出正确或疑误的判断。所以, 对一个质量控制算法的性能评估应从这两部分入手, 即预测模型的可靠性及质量控制模型的有效性。

可靠性是质量控制模型有效性的基础, 是对预测模型正确性能的判断。本文采用平均绝对误差(mean absolute error, MAE)、均方根误差(root mean squared error, RMSE)和纳什效率系数(Nash 等^[24]) (Nash-Sutcliffe model efficiency coefficient, NSC)评估。其中, NSC 的值介于 0~1, 数值越接近 1, 模型的性能越好; RMSE 和 MAE 为大于 0 的数, 越接近 0, 模型的性能越好。

有效性指标是衡量质量控制模型效果的主要指标, 本文

通过定义检错率来评估。分析本文假设检验方案, 存在 4 种情况, 见式(3): 即两种正确结论; 另外两种错误结论, 由质量控制模型的失效导致, 对应统计学中的“两类错误”。不失一般性, 记产生第一类错误和第二类错误的概率分别为 α 、 β , 研究表明^[25], 当样本容量固定时, α 的减小将引起 β 的增加。在质量控制领域研究中, 如何权衡这两类误差是质量控制领域不可避免的问题。屠其璞等^[26]认为, 在控制第一类错误的情况下, 应使第二类误差尽可能小。基于此, 本文提出最小绝对误差检错率概念, 给出衡量两类误差的定量指标。

由式(4)可知, 检错率是一个关于 α 、 β 以及 f 的函数, 则检错率可表述为:

$$\delta_{\text{ratio}} = \xi(\alpha(f), \beta(f)) \quad (4)$$

式中: $\xi(\cdot)$ ——某一函数映射关系; $\alpha(f)$ 、 $\beta(f)$ ——犯两类错误关于质控因子 f 的函数。

基于 Neymann-Pearson 检验原理, 取两类错误绝对误差最小时所对应 f 值为最佳质控因子, 记为 f_0 , 此时对应的检错率称为最小绝对误差检错率(记为 δ_{ratio_0})。 f_0 及 δ_{ratio_0} 可表示为:

$$\begin{cases} f_0 = \min |\alpha(f), \beta(f)| \\ \delta_{\text{ratio}_0} = \xi(\alpha(f_0), \beta(f_0)) \end{cases} \quad (5)$$

2 案例分析

2.1 案例数据概况

本文选取广西壮族自治区贺州市某山地风场为验证案例的数据。该风场所处区域北部为高山余脉, 东北偏北方向地势低洼, 东北方向为东北-西南走势山体, 东侧、东南地势开阔, 南部、西部均为山体, 如图 3(左侧)所示; 风场区内有 5 座观测塔, 即 58301#~58305#, 其相对位置示意图如图 3(右侧)所示。本文采用 5 座观测塔 2015—2016 年 70 m 高度、时间分辨率为 10 min 的风速作为实验验证数据。这些数据的质量已进行过控制, 已消除了明显离群值, 可正确反映该风电场区域风速的分布, 下文称为基准数据。

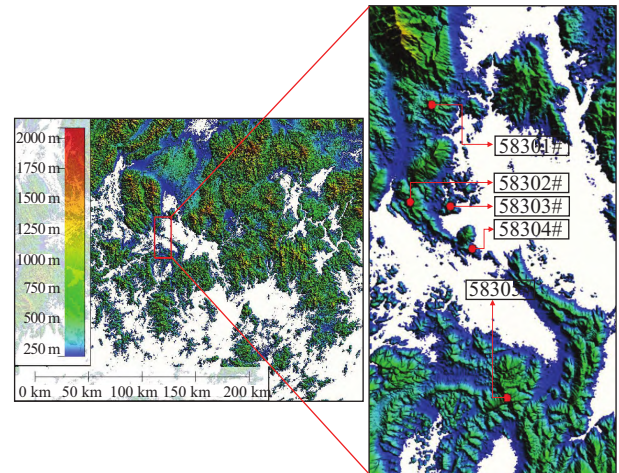


图3 广西某复杂山地风电场观测塔地形分布示意图

Fig. 3 Topographic distribution of observation towers of a complex mountain wind farm in Guangxi

对风速数据进行统计分析可知：一方面，5座观测塔2015—2016年各月风速整体变化趋势一致，但各观测塔风速差异较大，如图4所示；另一方面，5座观测塔的相关性差异较大，如图5所示。这种风速及相关性差异不仅源于海拔，还与周边观测地形、地貌相关。如58301#、58302#，前者比后者海拔低174m，但年均风速大0.474m/s，如表1所示；再如58301#、58305#风速与海拔整体呈正相关，但皮尔逊相关性系数仅为0.465，属于低相关性，如图5。

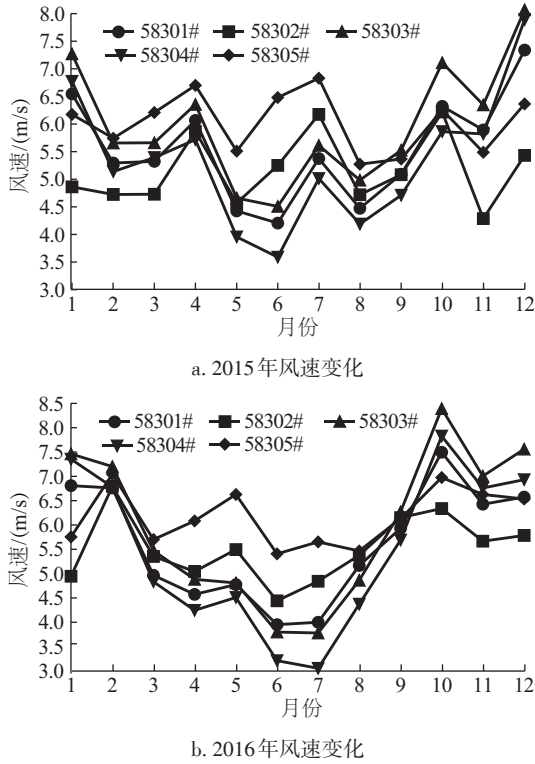


图4 5座观测塔各月风速变化趋势图

Fig. 4 Wind speed variation trend chart of five observation towers

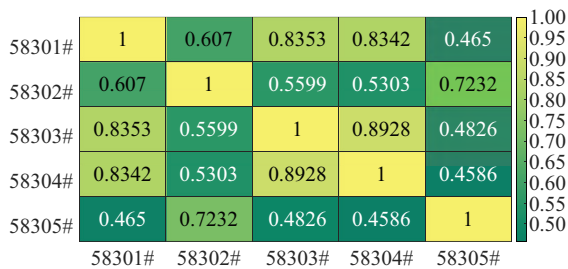


图5 5座观测塔10 min 风速数据相关性系数分布图

Fig. 5 Correlation coefficient of 10 min wind speed data of five observation towers

结合地形地貌对各观测塔进行分析：58301#观测塔正北、东北方向均为较高山体，北偏东方向为大型湖泊，形成“峡口”微地形^[27]；正北来流往北偏东偏转，并得到加速，风速威布尔分布呈明显“双峰”分布。58302#观测塔正北是较高南北走势山脉，处于“背风”坡地形；正北来流被阻挡，“风影

区”风速削弱明显，且呈现强烈的波动性。58203#观测塔西侧被较高山脉包围，正北有多个海拔矮小孤峰，东侧地势开阔，小地形复杂。58204#观测塔正北来流方向山峰遮挡，东北侧地势开阔，南依小型水库，小气候特征明显^[28]。58305#观测塔为东北-西南走势山脊的较高地，正北、正南方向开阔，气流稳定，风速相对稳定。

表1 广西某复杂山地风电场观测塔基本信息统计表

Table 1 Basic information of observation towers of a complex mountain wind farm in Guangxi

年份	塔号	海拔/m	样本点	风速/(m/s)
2015	58301#	533	52560	5.558
	58302#	707	52560	5.186
	58303#	466	52560	6.009
	58304#	327	52560	5.367
	58305#	777	52560	6.058
2016	58301#	533	52704	5.644
	58302#	707	52704	5.542
	58303#	466	52704	5.598
	58304#	327	52704	5.489
	58305#	777	52704	6.191

综上，该风场内5座观测塔属于复杂地形观测，观测风速复杂多变，且相关间呈现出不同相关性强度。

2.2 可靠性分析

模型参数选择参照表2。为验证PVL_QC的可靠性，即

表2 PVL_QC参数方案

Table 2 Parameter scheme of PVL_QC

参数	数值或方法
粒子位置规划方法	整数规划
粒子区间	[2, 30]; [100, 5000]
惯性权重	线性递减法 ^[29] ，初始值0.9，终止值0.4
最大进化代数	40
加速因子	1.49445
适应度函数	式(1)
输入层	1
隐含层	200
求解器	adam
梯度阈值、最大训练轮数	1, 300
初始学习率	0.005, 125轮训练后乘以因子0.2，降低学习率
最大训练轮数	300

较好的预测性能,基于上述案例数据设计 PVL_QC 方法的对比性实验。LSTM 指直接对风速时间序列进行预测;VMD-LSTM 指通过对风速时间序列分解后的各模态分别进行 LSTM 预测再重构;PSO-VMD-LSTM 为 VMD-LSTM 的参数优化模型。实验数据选取 58301#观测塔 2015 年逐 10 min 观测风速数据,运用 3 种方法进行预测并以月为单位进行结果统计分析。

图 6 是集成及改进前后算法各预测指标月变化趋势示意图,图 7 是集成及改进前后算法各预测指标年统计分布图。从图 6 可明显看出,PSO-VMD-LSTM 的预测性能最好,

VMD-LSTM 次之,LSTM 最差;结合图 7 可看出,PSO-VMD-LSTM 预测指标性能最高,且稳定性较好,VMD-LSTM 尤其 LSTM 波动性较大。

风速的分布直接影响风资源评估的结果,即要求风速预测序列的分布与背景场的偏差满足正态分布。风资源分析及业务中常用“双参数”威布尔模型来定量刻画风速分布^[30-31]。为进一步说明本文算法具有更好的风速分布还原性,分别对 3 种算法的风速序列预测值进行威布尔拟合,图 8 为不同模型 2015 年风速序列威布尔拟合示意图;同时对各月风速威布尔尺度参数 k 值与背景场进行误差分析,如图 9 所示。

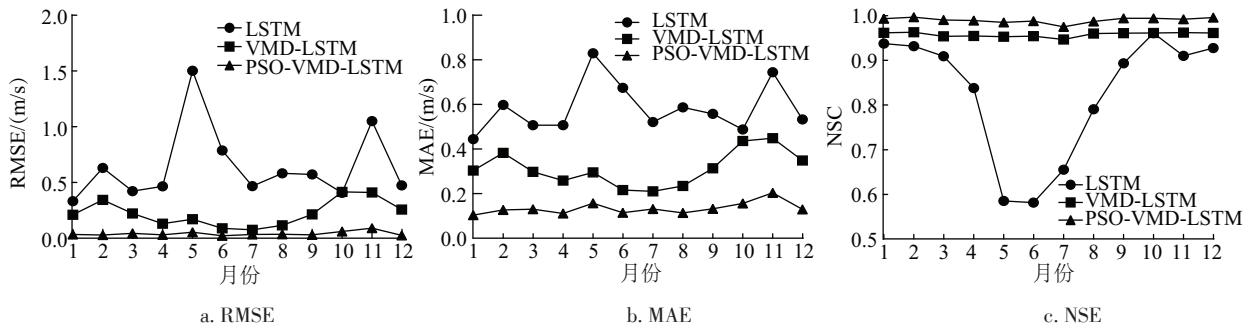


图 6 算法集成及优化前后 58301#观测塔各月 3 种预测指标对比

Fig. 6 Comparison of three prediction indexes of 58301# observation tower in each month before and after algorithm integration and optimization

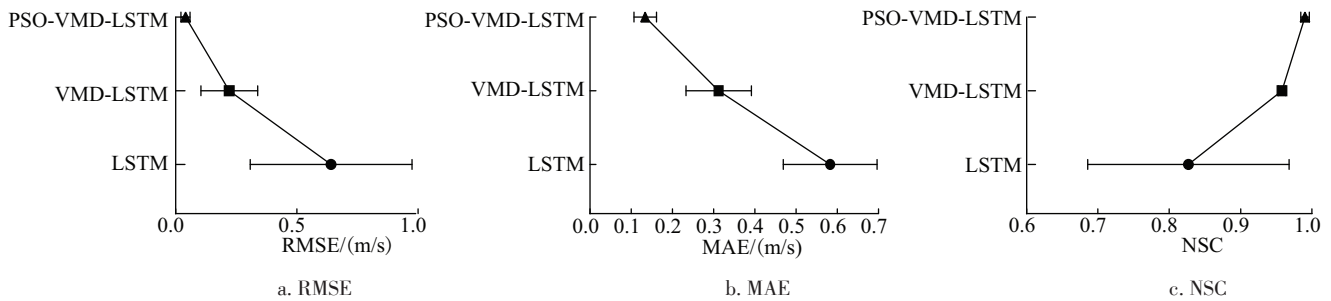


图 7 算法集成及优化前后在 58301#观测塔各月 3 种预测指标箱线图对比

Fig. 7 Box-plot of three prediction indexes in 58301# observation tower every month before and after algorithm integration and optimization

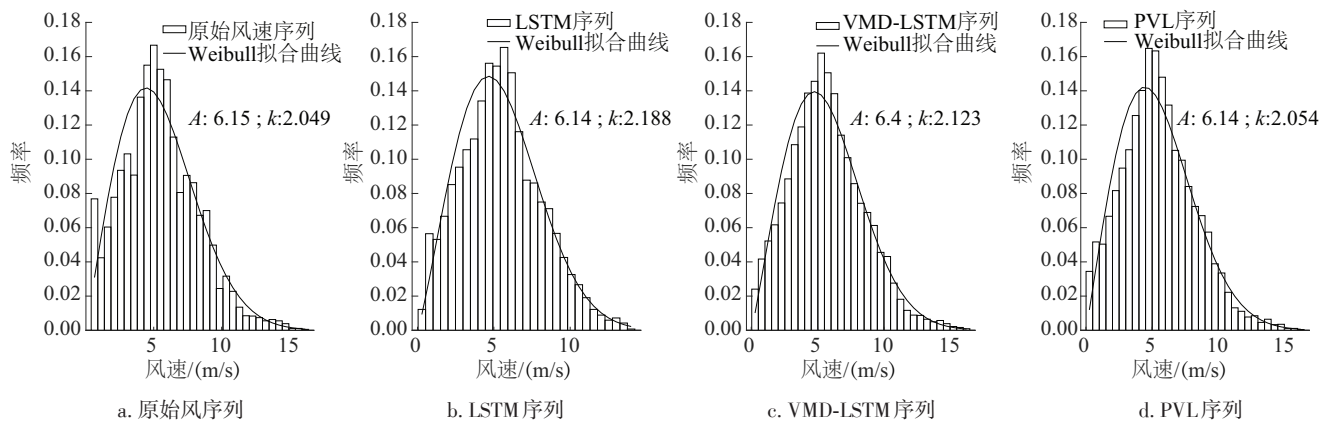


图 8 原始及不同算法预测风速序列威布尔拟合示意图

Fig. 8 Weibull fitting diagram of original and different algorithm prediction wind speed series

综合分析图 8 和图 9 可得出,PSO-VMD-LSTM 预测风速序列分布误差小,且波动性较小,平均误差 0.84%,较 VMD-LSTM 和 LSTM 分别提升 10.4%、3.5%;表明其具有较好且稳定的描述风速分布情况的性能。

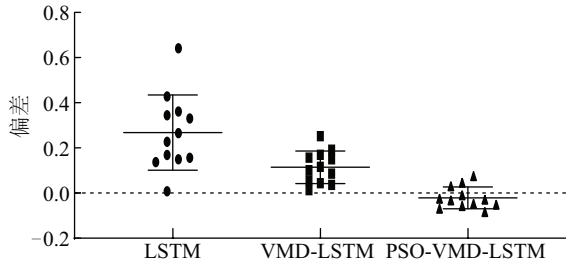


图 9 不同算法预测与观测风速序列拟合威布尔 k 值偏差箱线图

Fig. 9 Box-plot of Weibull k -value deviation of wind speed series observed and predicted by different algorithms

综上,PVL_QC 具有更好的风速预测性能,各预测性能指标较稳定;其预测风速与观测风速威布尔分布拟合参数误差较小,刻画风速分布更为吻合。

2.3 有效性分析

为验证 PVL_QC 的实际可行性,运用该方法对 58301#等 5 座观测塔 2015—2016 年的风速数据进行质量控制,并与《标准》内单站控制方法(记为 TS_QC)进行对比,表 3 是本文方法在不同观测点、不同年份的年平均检错率,图 10 是本文方法与 TS_QC 对比。从表 3 和图 10 可明显看出,相对于 TS_QC,本文方法具有明显优势,5 座观测站检错率水平均超过 88%,较 TS_QC 提升 60%以上,表明该方法检错率高,小气候和微地形适应性强;不同年份的同一观测点的检错率水平基本持平,偏差均值水平,表明 PVL 对同一气候特征下的观测数据具有相对的内在稳定性;不同观测点的检错率基本一致,但仍存在一定的差异性,这种差异主要可从微地形差异进行分析解释。

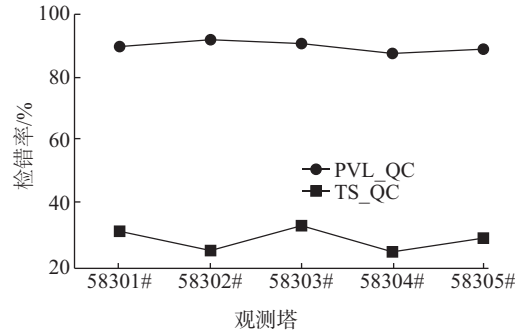
表 3 5 座观测塔 2015—2016 年 PVL_QC 的检验效果

Table 3 Testing results with QC method of PVL in five observation towers from 2015 to 2016

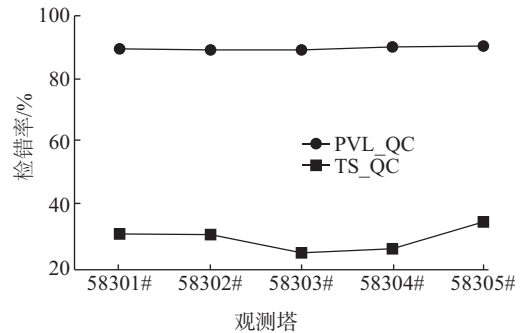
年份	检错率/%				
	58301#	58302#	58303#	58304#	58305#
2015	89.67	91.82	90.59	87.50	88.89
2016	89.51	89.20	89.19	90.12	90.42

为了说明 PVL_QC 较传统多站质量控制效果的优越性,分别使用 3 种方法对 58301#观测塔 2015—2016 年风速观测数据进行质量控制,SRT 和 IDW 参照文献[5,7]。图 11 为这 3 种方法在 58301#观测塔 2015—2016 年观测风速的检错效果对比。从图 11 可看出,3 种方法在 58301#观测塔不同年

份的观测数据的检错效果基本一致,但 PVL_QC 检错效果均明显好于 SRT 和 IDW 方法,平均提升 25.1%、28.8%。



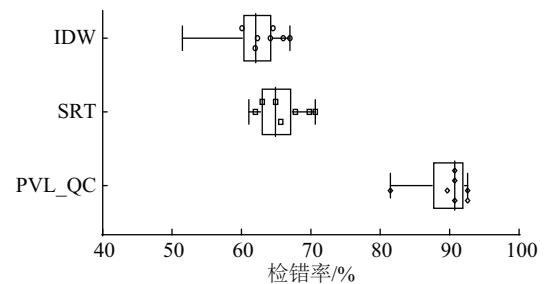
a. 2015 年检错率对比结果



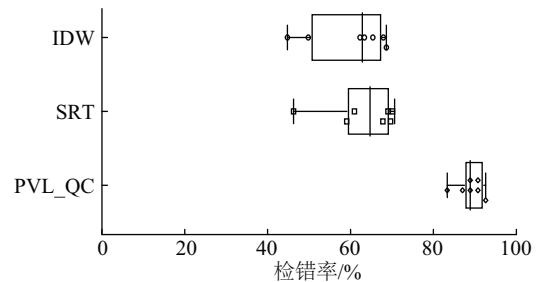
b. 2016 年检错率对比结果

图 10 PVL_QC 与传统单站在 58301#观测塔 2015—2016 年检错效果对比

Fig. 10 Comparison of detecting ratio between PVL_QC and TS_QC from 2015 to 2016 in 58301# observation tower



a. 2015 年不同方法检错率对比分析结果



b. 2016 年不同方法检错率对比分析结果

图 11 PVL_QC 与 SRT、IDW 在 58301#观测塔 2015—2016 年检错效果对比

Fig. 11 Comparison of detecting ratio between PVL_QC and SRT and IDW from 2015 to 2016 in 58301# observation tower

为进一步比较 PVL_QC 与 SRT 和 IDW 在不同观测点的检错效果,运用 3 种方法分别对 58301#等 5 基观测点 2015—2016 年的观测风速数据进行检错,图 12 是 3 种方法在不同观测点 2015—2016 年的检错效果。可明显看出,不同观测点 PVL_QC 的检错效果均明显优于 SRT 和 IDW 质量控制法,平均提升 26.3%、34.7%。

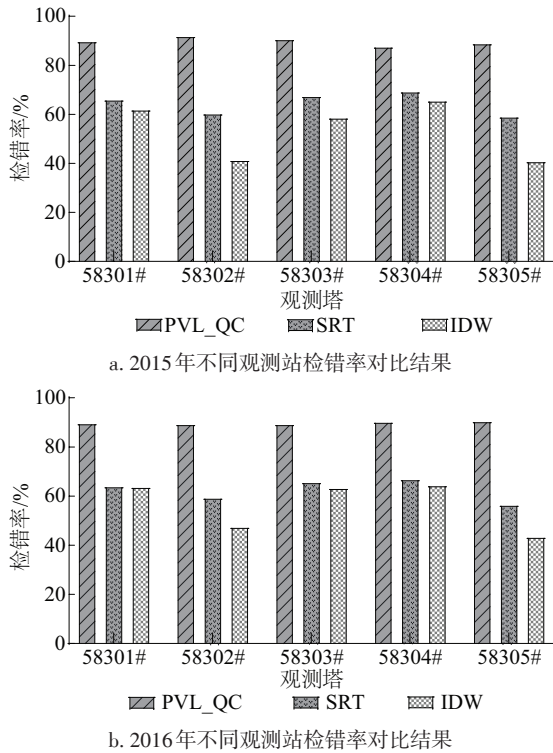


图 12 不同观测塔 2015—2016 年风速的 PVL_QC 与 SRT、IDW 检错效果对比

Fig. 12 Comparison of detecting ratio between PVL_QC and SRT, IDW in different observation towers from 2015 to 2016

2.4 不确定性分析

为进一步研究 PVL_QC 的质控因子参数的最佳取值 f_c 在不同地形、气候下的分布情况,对 58301#等 5 座观测塔 2015—2016 年风速观测数据以年和季度为单位进行 PVL_QC 分析。图 13 是各观测塔 2015—2016 年 f_c 值分布图。图 14 是各观测塔各年份、各季度的最佳 f_c 值分布图。

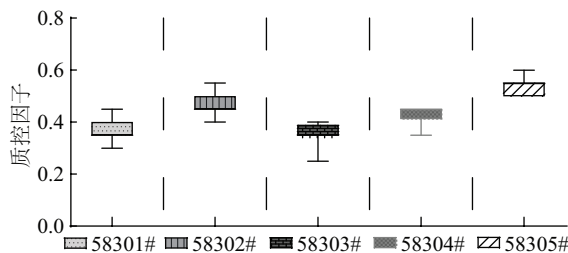


图 13 不同观测塔 2015—2016 年风速观测数据质控因子取值分布箱线图

Fig. 13 Box-plot of quality control factor value distribution of different observation towers

从图 13 可看出, PVL_QC 在各观测塔的取值具有一定的差异,但主要集中在 0.35~0.50 之间;从图 14 可看出,每个观测塔不同季度值也具有一定的差异,但同一观测塔不同年份同一季度基本一致。

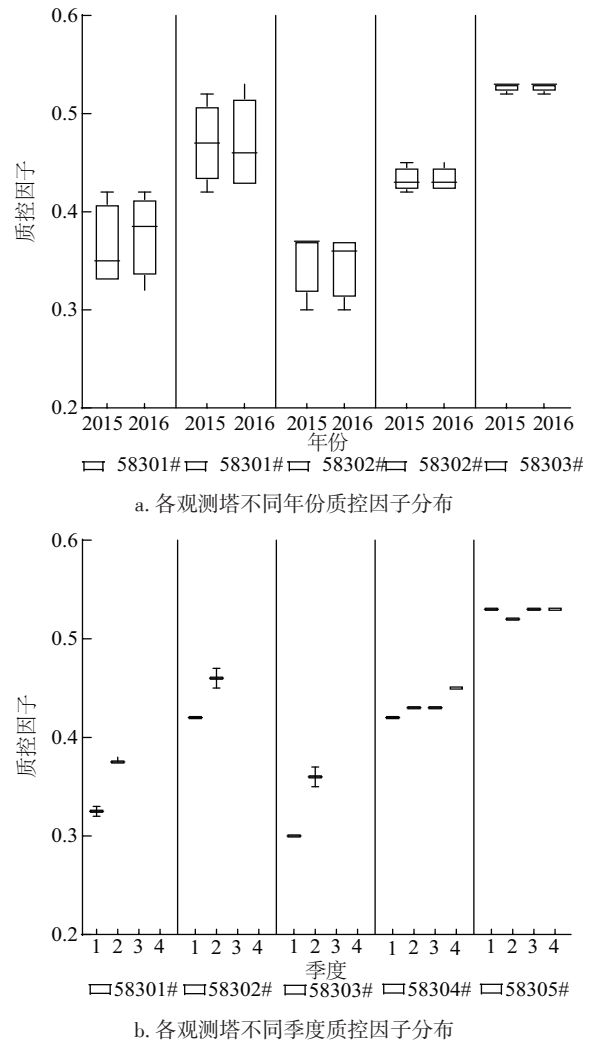


图 14 不同观测塔不同年份、季度观测风速质控因子取值分布图

Fig. 14 Box-plot of quality control factor value distribution of different observation towers in different years and quarters

对 PVL_QC 的 f_c 分析表明,通过构造最小绝对误差检错率可平衡两类统计学误差; f_c 的取值在同一位置、同一气候的观测具有一定的稳定性,同时在不同位置、不同气候的观测又具有差异性,可控及可移植性较好。

3 结 论

1) 本文从风速时间序列非平稳性和复杂度出发,集成变分模态分解技术(VMD)与长短时记忆网络(LSTM),并针对 VMD 中模态参数和惩罚因子的选取提出一种基于包络熵的优化选择方法,即 PVL 集成学习算法,并运用该学习算法对 58301#观测塔 2015—2016 年观测风速进行预测,通过多种

预测指标及风速分布进行验证,结果表明 PVL 集成学习算法具有较好的复杂山地观测风速时间序列预测能力。

2)PVL_QC 适用于复杂山地,可解决孤立观测或区域观测相关性较低、风速波动性较大的风速质量控制问题;其较传统质量控制法相比,对于复杂地形和小气候多变的观测风速均具有更好的疑误标识能力。

3)PVL_QC 核心是智能优化及学习重构算法,决定了其具有较强的地形、气候适应性。PVL_QC 在广西某复杂山地风场的 58301#等 5 座观测塔在不同微地形、小气候下表现出一定的差异性,在同地形和气候下具有一定的内部健壮性。

4)本文在平衡检错率与两类误差之间提出一种定量标准,使质控因子在不同位置和气候呈现一定的自适应性,增强了方法的可控及可移植性。

PVL_QC 在本文仅在单要素、固定高度的风速数据上简单应用,对如风向等其他观测要素的普适应还未分析,也未考虑时空约束耦合机理。在接下来的研究中,可从这些方面着手,构造更符合风速观测塔多要素、时空分布的模型,更好地提高数据质量,为风资源、发电量的精准评估和风功率预测提供一套基础入口数据。

[参考文献]

- [1] 欧阳志远, 史作廷, 石敏俊, 等. “碳达峰碳中和”: 挑战与对策[J]. 河北经贸大学学报, 2021, 42(5): 1-11.
OUYANG Z Y, SHI Z T, SHI M J, et al. Challenges and countermeasures of “carbon peak and carbon neutrality” [J]. Journal of Hebei University of Economics and Business, 2021, 42(5): 1-11.
- [2] FU W L, FANG P, WANG K, et al. Multi-step ahead short-term wind speed forecasting approach coupling variational mode decomposition, improved beetle antennae search algorithm-based synchronous optimization and Volterra series model[J]. Renewable energy, 2021, 179: 1122-1139.
- [3] SINGH S, BHATTI T S, KOTHARI D P. A review of wind-resource-assessment technology[J]. Journal of energy engineering, 2006, 132(1): 8-14.
- [4] GB/T 18710—2002, 风电场风能资源评估方法[S].
GB/T 18710—2002, Methodology of wind energy resource assessment for wind farm[S].
- [5] HUBBARD K G, YOU J S. Sensitivity analysis of quality assurance using the spatial regression approach- a case study of the maximum/minimum air temperature[J]. Journal of atmospheric and oceanic technology, 2005, 22(10): 1520-1530.
- [6] LORENC A C. A global three-dimensional multivariate statistical interpolation scheme[J]. Monthly weather review, 1981, 109(4): 701-721.
- [7] WADE C G. A quality control program for surface mesometeorological data [J]. Journal of atmospheric and oceanic technology, 1987, 4(3): 435-453.
- [8] LORENC A C, HAMMON O. Objective quality control of observations using Bayesian methods. Theory, and a practical implementation [J]. Quarterly journal of the royal meteorological society, 1988, 114(480): 515-543.
- [9] 熊雄, 叶小岭, 张颖超, 等. 基于空间观测差异的地面气温资料质量控制算法研究[J]. 地球物理学报, 2017, 60(3): 912-923.
XIONG X, YE X L, ZHANG Y C, et al. A quality control method for the surface temperature based on the spatial observation diversity [J]. Chinese journal of geophysics, 2017, 60(3): 912-923.
- [10] YE X L, YANG X, XIONG X, et al. A quality control method based on an improved random forest algorithm for surface air temperature observations [J]. Advances in meteorology, 2017, 2017: 1-15.
- [11] 叶星瑜, 叶小岭, 马伟叁, 等. 基于 LMD-TCN 的高铁沿线风速观测资料质量控制算法研究[J]. 铁道科学与工程学报, 2022, 19(4): 849-856.
YE X Y, YE X L, MA W S, et al. Research on the quality control method of wind speed observation data along the high-speed railway line based on LMD-TCN [J]. Journal of railway science and engineering, 2022, 19(4): 849-856.
- [12] 秦琼, 刘树洁, 赖旭, 等. GA 优化 ELM 神经网络的风电场测风数据插补[J]. 太阳能学报, 2018, 39(8): 2125-2132.
QIN Q, LIU S J, LAI X, et al. Interpolation of wind speed data in wind farm based on GA optimized ELM neural network [J]. Acta energiae solaris sinica, 2018, 39(8): 2125-2132.
- [13] 张颖超, 姚润进, 熊雄, 等. PSO-PSR-ELM 集成学习算法在地面气温观测资料质量控制中的应用[J]. 气候与环境研究, 2017, 22(1): 59-70.
ZHANG Y C, YAO R J, XIONG X, et al. Application of PSO-PSR-ELM-based ensemble learning algorithm in quality control of surface temperature observations [J]. Climatic and environmental research, 2017, 22(1): 59-70.
- [14] MEEK D W, HATFIELD J L. Data quality checking for single station meteorological databases [J]. Agricultural and forest meteorology, 1994, 69(1/2): 85-109.
- [15] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. IEEE transactions on signal processing, 2014, 62(3): 531-544.
- [16] WANG D Y, LUO H Y, GRUNDER O, et al. Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction [J]. Renewable energy,

- 2017, 113: 1345-1358.
- [17] SUN W, GAO Q. Short-term wind speed prediction based on variational mode decomposition and linear-nonlinear combination optimization model [J]. *Energies*, 2019, 12(12): 2322.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. *Neural computation*, 1997, 9(8): 1735-1780.
- [19] SHEN F R, CHAO J, ZHAO J X. Forecasting exchange rate using deep belief networks and conjugate gradient method [J]. *Neurocomputing*, 2015, 167: 243-253.
- [20] 邓三鸿, 傅余洋子, 王昊. 基于 LSTM 模型的中文图书多标签分类研究 [J]. *数据分析与知识发现*, 2017, 1(7): 52-60.
- DENG S H, FUYU Y Z, WANG H. Multi-label classification of Chinese books with LSTM model [J]. *Data analysis and knowledge discovery*, 2017, 1(7): 52-60.
- [21] 马远浩, 曾卫明, 石玉虎, 等. 基于加权词向量和 LSTM-CNN 的微博文本分类研究 [J]. *现代计算机(专业版)*, 2018(25): 18-22.
- MA Y H, ZENG W M, SHI Y H, et al. Research on text classification of weibo based on weighted word vectors and LSTM-CNN [J]. *Modern computer*, 2018(25): 18-22.
- [22] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory [C]//*Proceedings of the Sixth International Symposium on Micro Machine and Human Science*. Nagoya, Japan, 2002: 39-43.
- [23] 唐贵基, 王晓龙. 参数优化变分模态分解方法在滚动轴承早期故障诊断中的应用 [J]. *西安交通大学学报*, 2015, 49(5): 73-81.
- TANG G J, WANG X L. Parameter optimized variational mode decomposition method with application to incipient fault diagnosis of rolling bearing [J]. *Journal of Xi'an Jiaotong University*, 2015, 49(5): 73-81.
- [24] NASH J E, SUTCLIFFE J V. River flow forecasting through conceptual models part I-a discussion of principles [J]. *Journal of hydrology*, 1970, 10(3): 282-290.
- [25] ZHANG Y C, XIONG X, ZHANG Q D. An improved self-adaptive PSO algorithm with detection function for multimodal function optimization problems [J]. *Mathematical problems in engineering*, 2013, 2013: 1-8.
- [26] 屠其璞, 王俊德, 丁裕国, 等. 气象应用概率统计学 [M]. 北京: 气象出版社, 1984: 183-186.
- TU Q P, WANG J D, DING Y G, et al. *Meteorological application probability statistics* [M]. Beijing: China Meteorological Press, 1984: 183-186.
- [27] 席世平, 白凌霞, 赵培娟, 等. 2010年9月4日河南孟津罕见雷暴大风过程的分析和预报 [C]//第28届中国气象学会年会: S3 天气预报灾害天气研究与预报. 厦门, 中国, 2011: 919-927.
- XI S P, BAI L X, ZHAO P J, et al. Analysis and prediction of the rare thunderstorm Gale Process in Mengjin, Henan Province on September 4, 2010 [C]//*The 28th Annual Meeting of China Meteorological Society*, Xiamen, China, 2011: 919-927.
- [28] 华德尊, 李春艳, 蔡春苗. 二龙山水库流域不同程度生态破坏对小气候要素的影响 [J]. *环境科学研究*, 2002, 15(3): 16-18.
- HUA D Z, LI C Y, CAI C M. The influence of ecological destruction on microclimatic factors in erlongshan reservoir watershed [J]. *Research of environmental sciences*, 2002, 15(3): 16-18.
- [29] SHI Y H, EBERHART R C. Parameter selection in particle swarm optimization [M]. Berlin: Springer Berlin Heidelberg, 1998: 591-600.
- [30] PISHGAR-KOMLEH S H, KEYHANI A, SEFEEDPARI P. Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh County of Iran) [J]. *Renewable and sustainable energy reviews*, 2015, 42: 313-322.
- [31] 杨刚, 杜永贤, 陈鸣. 基于风频 Weibull 分布和风机功率特性求解风机发电量 [J]. *电力学报*, 2008, 23(4): 276-278, 300.
- YANG G, DU Y X, CHEN M. The calculation of energy provided by wind turbine based on its power characteristic and the wind frequency weibull distribution [J]. *Journal of electric power*, 2008, 23(4): 276-278, 300.

A QUALITY CONTROL ALGORITHM OF WIND SPEED OBSERVATIONS IN COMPLEX MOUNTAIN WIND FARM BASED ON PSO-VMD AND LSTM

Xiong Xiong¹, Yao Runjin², Cheng Shuaibing³, Li Wenlong⁴, Qian Dong¹

(1. *Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Information and Systems Science Institute, NUIST, Nanjing 210044, China;*

2. *Carbon Neutralization Research Institute, PowerChina Jiangxi Electric Power Construction Co., Ltd., Nanjing 210018, China;*

3. *Jiangsu Key Laboratory of Offshore Wind Power Blade Design and Manufacturing Technology, Nanjing 210000, China;*

4. *Jiangxi Branch, China Three Gorges New Energy (Group) Co., Ltd., Nanchang 330038, China)*

Abstract: There are many problems in complex mountain wind farms, such as large errors of wind resource evaluation and low accuracy of wind power prediction caused by poor quality of observed wind speed data. Because of the strong intermittent, fluctuating, and non-stationary characteristics presented by the wind speed in complex mountain wind farms, conventional quality control methods cannot effectively improve data quality. For this situation, an integrated learning algorithm (PVL) based on particle swarm optimization improved variational modal decomposition improved by particle swarm optimization and long short-term memory is proposed and applied to the quality control of wind speed observations in complex mountainous areas to improve the quality of wind speed data. In order to assess the feasibility and applicability of the proposed method, the 10 minutes wind speed observed in five observation tower of a complex mountain wind farm in Guangxi from 2015 to 2016 were examined. Otherwise, we compared this method to spatial regression test (SRT) and inverse distance weighting method (IDW). The results show that the method can more effectively flag suspicious data, and it also has the advantages of high identification accuracy, strong adaptability to different terrains and wind conditions.

Keywords: wind farm; quality control; particle swarm optimization; variational modal decomposition; long short-term memory