

基于新的组合模型的短期风速预测

胡文博, 袁柳洋, 马建清

武汉科技大学理学院, 湖北 武汉

收稿日期: 2024年1月2日; 录用日期: 2024年1月19日; 发布日期: 2024年2月29日

摘要

鉴于风速序列呈现高度随机和不稳定的特性, 本文构建了一个新的组合模型用于短期风速预测, 该模型由粒子群优化-时变滤波-经验模态分解(PSO-TVF-EMD)、排列熵(PE)、长短期记忆网络(LSTM)和自回归差分移动平均(ARIMA)组成。首先利用PSO-TVF-EMD算法将原始风速序列分解为若干模态分量, 以简化风速序列的复杂性; 其次, 使用PE把风速子模态分为高频率序列和低频率序列, 并对这两种序列分别构建LSTM和ARIMA预测模型; 最终, 将子序列预测结果叠加, 得出最终的风速预测值。试验结果表明, 新的组合预测模型一定程度上增加了预测的精度。

关键词

时变滤波, 粒子群优化, 排列熵模型, 长短期记忆网络, 短期风速预测

Short-Term Wind Speed Prediction Based on a New Combined Model

Wenbo Hu, Liuyang Yuan, Jianqing Ma

School of Science, Wuhan University of Science and Technology, Wuhan Hubei

Received: Jan. 2nd, 2024; accepted: Jan. 19th, 2024; published: Feb. 29th, 2024

Abstract

In view of the highly random and unstable characteristics of wind speed series, a new combined model is constructed for short-term wind speed prediction. The model consists of particle swarm optimization-time-varying filtering-empirical mode decomposition (PSO-TVF-EMD), permutation entropy (PE), long-short term memory network (LSTM) and auto regressive integrated moving average (ARIMA). Firstly, the original wind speed series are decomposed into several modal components by using PSO-TVF-EMD algorithm to simplify the complexity of the wind speed series, and secondly, the wind speed sub-modes are divided into high frequency series and low frequency se-

ries by using PE algorithm, the LSTM and ARIMA prediction models are constructed for the two series, and the wind speed prediction values are obtained by superimposing the results of the sub-series. The experimental results show that the new combined forecasting model increases the forecasting precision to some extent.

Keywords

Time-Varying Filtering, Particle Swarm Optimization, Permutation Entropy Model, Long-Term and Short-Term Memory Network, Short-Term Wind Speed Prediction

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在如今高速发展的社会形势下，能源的消耗速度在逐渐加速。近年来巨大的能源需求已经引发了资源耗尽和环境降级等问题。为了应对这种危机，全球各地已经开始专注于开发可再生能源，以推动能源的可持续发展。因此，寻求一个环保、可持续的能源是未来能源产业的发展趋势。就目前的生态形式和国家能源战略要求来看，作为新能源的风能是具有极大的发展潜能的。

风力发电的动力输入为大自然界的自然风，它具有不可控性、随机性、不连续性和波动性，导致了风力发电在维持电网电压频率稳定性方面比较差。这些特点同时也会加大风速高精度预测的难度。研究短期风速变化并进行风速预测是一项很有实际意义的工作。

针对短期风速预测存在的精度低，不可靠、鲁棒性差的问题，风电预测专家学者们已做出了很多工作，取得了很多值得关注的显著性成果。

初始的短期风速预测模型大都采用统计学的方法，比如灰色预测模型、马尔柯夫链预测模型，以及自回归预测模型[1]等。虽然这种统计学为基础的模型在预测上实现了部分效果，但是，由于其主要依赖于统计学原理和线性模型来进行模拟，因此对于风速的非线性和波动行为的精确预测，仍有改进之处。另一种是基于机器学习的模型，如支持向量机模型[2]和神经网络[3] [4]，通过学习风速的历史数据，能够更好地拟合风速变化。相比之下，基于机器学习的模型在处理风速的非线性回归上表现出色，并且预测的精确度也得到了提升。

尽管如此，但单独的预测模型在预测风速上仍存在准确度不足的问题，因此，目前研究人员开始将注意力放在混合模型上，其中采用“分解 - 预测 - 重构”策略的混合预测模型被广泛应用。相较于传统模型，其在提升预测精度上表现出更大的效果，并为预测风速提供了一种新的思路。该预测模型首先运用分解技术使得非线性非稳定的风速序列平稳，然后建立适合的数学模型进行预测。比如赵征等[5]觉察到风速序列强烈的波动性，他们提出了以 VMD 和 ARIMA 为基础的超短期风速预测模型。王贺等[6]在 CEEMDAN 和 LSTM 组合模型的基础上使用了布谷鸟算法对 LSTM 进行了优化，然后对风速时间序列进行了预测。向玲等[7]将 CEEMDAN 二次分解和 LSTM 进行结合进行了风速多步预测研究。王俊等[8]基于 VMD 和 LSTM 进行了超短期风速预测。

为了解决模态混叠问题并保持信号序列的时变性，Li 及其团队[9]引入了使用时变滤波器的 TVF-EMD 方法。以此为基础王文川[10]则利用 TVF-EMD 和 LSTM 用来进行月径流预测研究。

因此, 本文基于现有的研究成果, 引入了一个新的组合预测模型, 该模型运用了“分解 - 预测 - 重构”的方法, 该模型由粒子群优化 - 时变滤波 - 经验模态分解(PSO-TVF-EMD)、排列熵(PE)、长短期记忆网络(LSTM)和自回归差分移动平均(ARIMA)组成。首先利用 PSO-TVF-EMD 算法将原始风速序列分解为若干模态分量, 以简化风速序列的复杂性; 其次, 使用 PE 把风速子模态分为高序列和低序列, 并对这两种序列分别构建 LSTM 和 ARIMA 预测模型; 最终, 将子序列预测结果叠加, 得出最终的风速预测值。

2. 算法和模型

2.1. TVF-EMD

在 2017 年, 学者 Li 等[9]率先引入了时变滤波经验模态分解的概念, 对原先的 EMD 方法进行了调整和优化。这项技术创新地利用了时变滤波器, 有效地解决了模态混叠的困扰, 同时也保持了序列的时变性。不仅如此, 时变滤波经验模态分解还采用了瞬时幅度和瞬时频率, 以此自行调整局部截止频率, 以达到对输入序列进行时变滤波的效果, 在进一步划分出局部高频和局部低频成分后, 产生了固有模态函数(Intrinsic Mode Function, IMFs)。

TVF-EMD 方法的计算步骤为:

- 1) 利用 Hilbert 变换计算时间序列的瞬时幅值 $A(t)$ 和瞬时频率 $\varphi'(t)$:

$$A(t) = \sqrt{X(t)^2 + \hat{X}(t)^2}$$

$$\varphi'(t) = d\left(\arctan\left[\hat{X}(t)/X(t)\right]\right)/dt$$

本式中: $\hat{X}(t)$ 为时间序列 $X(t)$ 的 Hilbert 变换。

- 2) 确定瞬时幅值 $A(t)$ 的局部极大值序列和局部极小值序列, 分别表示为 $A(\{t_{\max}\})$ 和 $A(\{t_{\min}\})$;
- 3) 对 $A(\{t_{\max}\})$ 进行插值得到 $\beta_1(t)$, 以同样的方法对 $A(\{t_{\min}\})$ 进行插值得到 $\beta_2(t)$, 计算瞬时均值 $a_1(t)$ 和瞬时包络 $a_2(t)$:

$$a_1(t) = \frac{\beta_1(t) + \beta_2(t)}{2}$$

$$a_2(t) = \frac{\beta_1(t) - \beta_2(t)}{2}$$

- 4) 分别对 $A^2(\{t_{\min}\})\varphi'(\{t_{\min}\})$ 和 $A^2(\{t_{\max}\})\varphi'(\{t_{\max}\})$ 进行插值, 得到 $\eta_1(t)$ 和 $\eta_2(t)$, 并计算瞬时频率分量 $\varphi'_1(t)$ 和 $\varphi'_2(t)$;

- 5) 计算局部截止频率 φ'_{bis} :

$$\varphi'_{bis} = \frac{\varphi'_1(t) + \varphi'_2(t)}{2}$$

- 6) 为解决间歇问题, 重新对局部截止频率 φ'_{bis} 进行调整;
- 7) 计算信号 $h(t) = \cos\left[\int \varphi'_{bis}(t) dt\right]$, 并将 $h(t)$ 的极值点作为构造时变滤波器的节点, 采用样条插值对时间序列 $X(t)$ 进行逼近;
- 8) 计算停止准则 $\theta(t)$, 如果 $\theta(t) \leq \xi$, 则可确定 $X(t)$ 为一个 IMF, 不满足则令 $\theta(t)$ 重复执行(1)~(8); 分解后得到的所有模态分量 $X_i(t)$, 将全部的分量相加得到的值则为原时间序列 $X(t)$ 的值。

2.2. PSO 算法

粒子群优化(PSO)策略根据鸟儿觅食的行为获得启示, 常用于处理最优化问题。在 PSO 结构中, 各

个粒子象征着在解空间内移动的点，其速度各不相同。每个粒子都会根据与目标函数相匹配的适应度有不同的解，适应度函数是根据最优化目标设定的，可测量解的优劣。粒子们会追随现在最优的粒子在解空间里搜寻，通过不断的迭代最后找出最优解。

2.3. PE 算法

排列熵(PE)算法是一种基于信息论的度量方法。通过相邻值的比较，分析了时间序列数据的复杂性。PE 算法计算如下：

首先，假设时间序列 $X(t)(t=1,2,\dots,T)$ ，并重构其相空间：

$$x(t) = \{x_t, x_{t+\tau}, \dots, x_{t+(m-1)\tau}\}, t=1,2,\dots,T-(M-1)\tau$$

然后，按升序重新排列 $x(t)$ 中的元素：

$$x(t) = \{x_{t+(j_1-1)\tau}, x_{t+(j_2-1)\tau}, \dots, x_{t+(j_m-1)\tau}\}, x_{t+(j_1-1)\tau} \leq x_{t+(j_2-1)\tau} \leq \dots \leq x_{t+(j_m-1)\tau}$$

每个可能的排列都表示为 $w = \{j_1, j_2, \dots, j_m\}$ ， m 维序列存在 $m!$ 不同的排列。每个排列的概率是 P_1, P_2, \dots, P_g 。根据信息熵，置换熵的归一化可以定义为：

$$PE = \frac{PE(m)}{\ln[m!]} = \frac{-\sum_1^{m!} p_g \ln p_g}{\ln[m!]}$$

PE 的值介于[0, 1]之间，表示时间序列的复杂性。PE 值越接近 1，时间序列的复杂度越高；PE 值越接近 0，时间序列的复杂度越低。由于 ARIMA 模型对时间序列的平滑性要求较高，本文采用不同的置换熵值进行实验分析，然后确定置换熵的阈值为 0.2，大于 0.2 的是非平稳高频序列，小于 0.2 的是平稳低频序列。

2.4. LSTM

长短期记忆网络(LSTM)，作为递归神经网络(RNN)的升级版，引入了一种更佳的计算方法。它的模型能够自动提取时间序列数据中的特征，获取长距离关联信息，并有效避免在 RNN 训练过程中，梯度爆炸和梯度消失等问题的出现。LSTM 与 RNN 的主要不同之处，是在其门架构中增加了单元状态，这能记录历史数据状态信息。图 1 对 LSTM 模型的构造进行了展示，其核心由输入门、忘记门以及输出门构成。此外，Tanh 激活函数控制输入的更新， C_{t-1} 和 C_t 分别是 $t-1$ 和 t 时间的细胞状态， H_{t-1} 和 H_t 分别是 $t-1$ 和 t 时间的隐藏细胞状态。

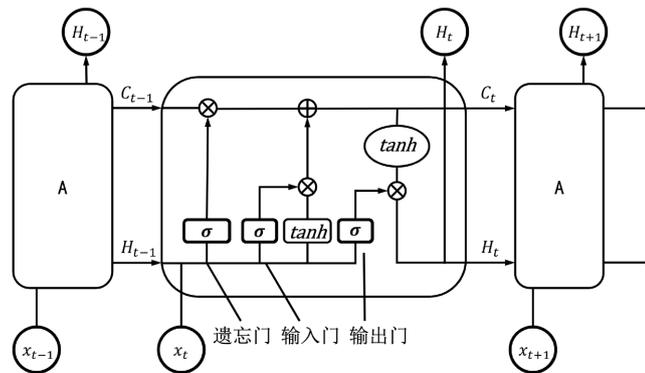


Figure 1. Long short-term memory network
图 1. 长短期记忆网络

2.5. ARIMA

自回归差分移动平均(ARIMA)是分析预测变量的未来值与历史值之间的线性函数关系的一种方法。ARIMA(P, D, Q)模型首先检验时间序列的平稳性, 如果平稳性不满足, 则对时间序列进行 d 阶差分将非平稳时间序列转化为平稳时间序列, 然后对平稳时间序列建立自回归移动平均模型(ARMA(p, q))。ARMA 方程为:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

3. 新的组合预测模型

3.1. PSO-TVF-EMD

参数设置是执行 TVF-EMD 分解的关键环节, 尤其是, 宽带阈值 ζ 和 B 样条级数 n 对分解效果有巨大影响, 因此, 如何选择这两个参数成为了问题的关键点。一般这两个参数都是根据经验, 人为选定, 但本文利用 PSO 算法对这两个参数进行优化, 粒子群优化算法是一类经典的集体智能优化方法, 它通过连续调整粒子的速率和位点, 同时保留了个体最优与全局最优, 经过持续的循环, 实现最优解的获得。

当用 PSO 算法来优化 TVF-EMD 分解的最优参数组合时, 必须设定一个目标函数。信噪比可以很有效地揭示信号分解降噪效果, 它的数值体现了降噪的效能, 所以本文选择信噪比作为 PSO 算法的目标函数。

具体步骤如下。首先用 TVF-EMD 对时间序列数据进行分解与处理, 如果重构出来的信噪比高, 这就意味着降噪效果优秀; 反之, 如果信噪比较低, 则降噪效果差。接着选取一组参数作为 PSO 算法的起点, 当第 i 个粒子位于某个位置, 便计算该位置下, 通过 TVF-EMD 分解和重构后的降噪数据信噪比, 把信噪比最大的标记为局部最佳值, 其对应的重构数据被视为最优重构数据。然而, 这只是局部的最优值, 为了在风速时间序列数据中找到降噪表现最卓越的重构数据, 目标函数被设为信噪比最大的情形, 并对参数配置进行优化。最后, 携带优化后的参数配置进行 TVF-EMD 分解, 预估出信噪比最大的重构数据, 从而获得预测结果。

算法步骤如下:

- 1) 初始化 PSO 算法的各项参数并将信噪比作为目标函数;
- 2) 初始化粒子种群, 随机产生一定数量的最佳影响参数组合 $[\zeta, n]$ 作为粒子的初始位置, 随机初始化粒子的速度;
- 3) 在不同粒子位置条件下对信号进行 TVF-EMD 分解, 计算每个粒子位置相应的信噪比;
- 4) 对比信噪比大小, 并更新个体局部极值和全局极值;
- 5) 更新粒子的速度和位置;
- 6) 循环迭代, 转至步骤(3), 直至迭代次数达到最大设定值, 并输出最大信噪比及粒子的位置;
- 7) 利用获得的参数组合 $[\zeta, n]$ 对信号进行 TVF-EMD 分解;
- 8) 对信噪比最大的 IMF 分量进行预测, 得出最终预测值。

3.2. 总体结构框架

由于风速数据存在着明显的非线性和非稳定特性, 本文首先使用了 PSO-TVF-EMD 算法对这些数据进行初步处理, 并将风速序列分解为不同频率的子序列。接下来, 运用 PE 算法计算出各子序列的排列熵, 并根据其复杂度选择出高频和低频子序列。对高频子序列, 利用了 LSTM 进行精确预测, 而对于低频子序列, 采用了 ARIMA 用以准确预测。收集到高频和低频子序列预测数据后, 将二者预测结果进行了结合, 以此获得了最终的风速时间序列预测值。

步骤如下:

利用 PSO-TVF-EMD 方法将原始风速序列分解为 n 个模态分量。

利用 PE 算法计算 n 个序列分量的置换熵,并根据时间序列的复杂性将模态分量分为高频部分和低频部分。

LSTM 用于预测高频序列, ARIMA 用于预测低频序列。

通过加入高频序列和低频序列的预测结果,得到风速的预测值。

风速预测框图如图 2 所示。

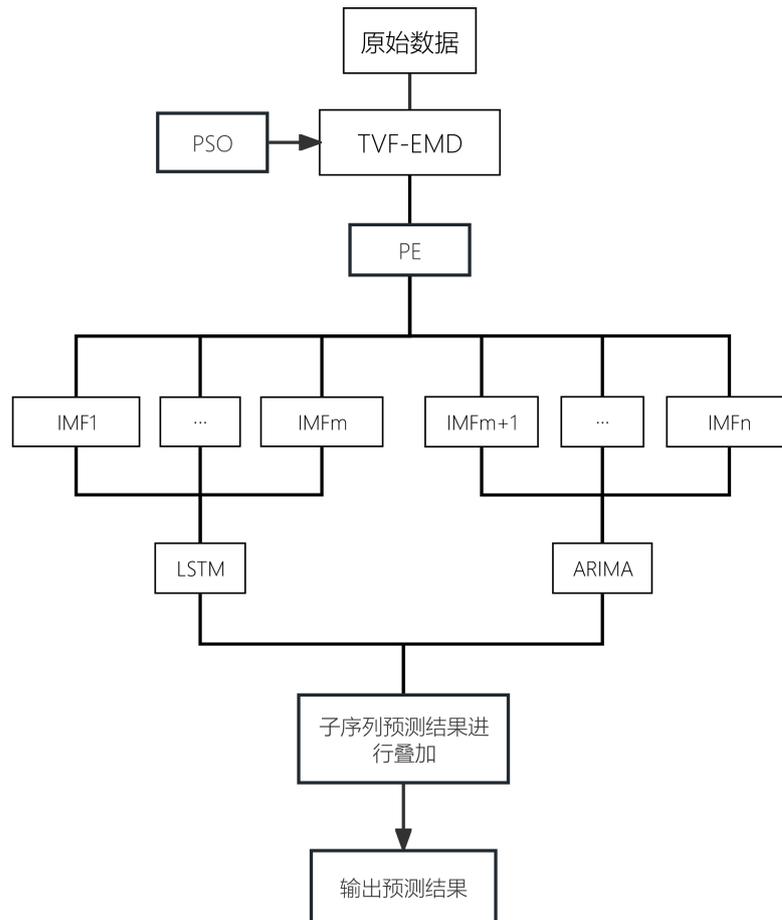


Figure 2. Structure block diagram
图 2. 结构框图

4. 实验及结果分析

4.1. 风速资料说明

本文所使用的数据来自于湖北省武汉市的气象台,这些数据的收集周期是在 2021 至 2022 年的某个时间段,每小时收集一次。共计有 5000 个数据点,从中挑选样本进行研究,其中风速数据的具体情况如图 3 所示。基于 9:1 的比例,本文从前 4501 个数据点中选择了训练样本,其余 499 个数据点被用作测试样本。如图 3 所示,风速数据展现出明显的非线性和不稳定性,增大了预测难度。然而,经过 TVF-EMD 分解处理,风速可以被拆为易于预测的时间子序列。此时,可以得到 26 个自高频到低频的模态函数。为

了量子序列频率的高低,用 PE 算法计算了序列的排列熵值。对 ARIMA 模型来说,排列熵值越小,其时间复杂度越低,预测精度也越高。其中,IMF1 到 IMF13 被视为高频序列,用于在 LSTM 网络模型中做预测,IMF13 到 IMF26 则被认为是低频序列,用于在 ARIMA 模型中做预测。

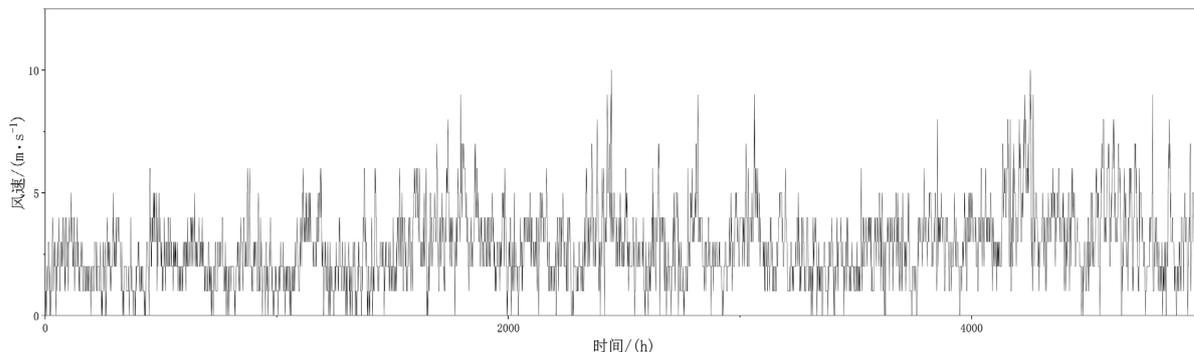


Figure 3. Raw data
图 3. 原始数据

4.2. 评价指标

本文运用了三种测量指标,即 RMSE (均方根误差)、MAE (平均绝对误差)和皮尔逊相关系数,来衡量与实际风速的差异及不同模型对于风速预测的准确度。这些指标的计算公式如下:

$$\text{RMSE}(X, h) = \sqrt{\frac{1}{m} \sum_{i=1}^m (h(x_i) - y_i)^2}$$

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$r_{xy} = \frac{\sum Z_x Z_y}{N}$$

4.3. 预测结果及对比分析

为了验证 PSO-TVF-EMD-LSTM-ARIMA 模型预测的效果,选择与 LSTM、EMD-LSTM、EMD-ARIMA 和 TVF-EMD-LSTM 这 4 种模型进行对比,这些模型均采用与目标模型相同的参数及变量设置。

由图 4 可以看出:① 单一的模型预测效果较差,只能显示序列的大概趋势,预测值误差较大。② 同一个预测模型采用不同方法对数据进行预处理,其预测精度都会有显著的差别。TVF-EMD-LSTM 模型预测精度从图中可以看出明显比 EMD-LSTM 模型要高。③ PSO-TVF-EMD-LSTM-ARIMA 模型预测精度最高,说明在数据的预处理上,PSO-TVF-EMD 模型有效的缓解了 EMD 存在的模态混叠问题和噪声问题,更适合处理非线性非平稳的序列,并且能取得较好的预测结果。

本文采用三个评价标准来测试目标模型预测的精度,其中 RMSE 与 MAE 的数值越小,且 r 系数越接近于 1,说明此模型的结果误差越小,精度越高。各个模型的误差评价指标的具体值如表 1 所示。由此可知:① LSTM 模型各项误差指标最大,随着采用 EMD 和 TVF-EMD 方法后,各模型的 MAE、RMSE 明显变小,如图 5 所示。RMSE 上,本文模型相对 LSTM、EMD-ARIMA、EMD-LSTM、TVF-EMD-LSTM 分别提升了 50.14%、26.44%、35.42%、12.60%。MAE 上,分别提升了 67.33%、54.62%、31.01%、15.74%。皮尔逊相关系数也更接近于 1,因此,PSO-TVF-EMD-LSTM-ARIMA 模型的预测精度更高。② 风速时间序列中普遍含有噪声,并且具有非平稳性和多频性,如果直接预测会使得 LSTM 模型的预测会产生较

大的误差。而 EMD 分解则会存在模态混叠的问题，高频序列可能会与低频序列混杂，对预测结果影响较大。TVF-EMD 就很好地解决了 EMD 的模态混叠和噪声问题，有效地提高了模型的预测精度，但是它对于参数有较高的要求，所以本文使用 PSO 对参数进行优化来提升模型的处理结果，从结果中可以发现，PSO-TVF-EMD 具有较好的分解效果，它优化了模型参数，解决了 EMD 存在的模态混叠和噪声的问题，使 PSO-TVF-EMD-LSTM-ARIM 模型预测结果最好。因此，PSO-TVF-EMD-LSTM-ARIM 模型在风速预测上具有较好的效果，适用于预测非平稳、非线性、复杂多频的风速时间序列。

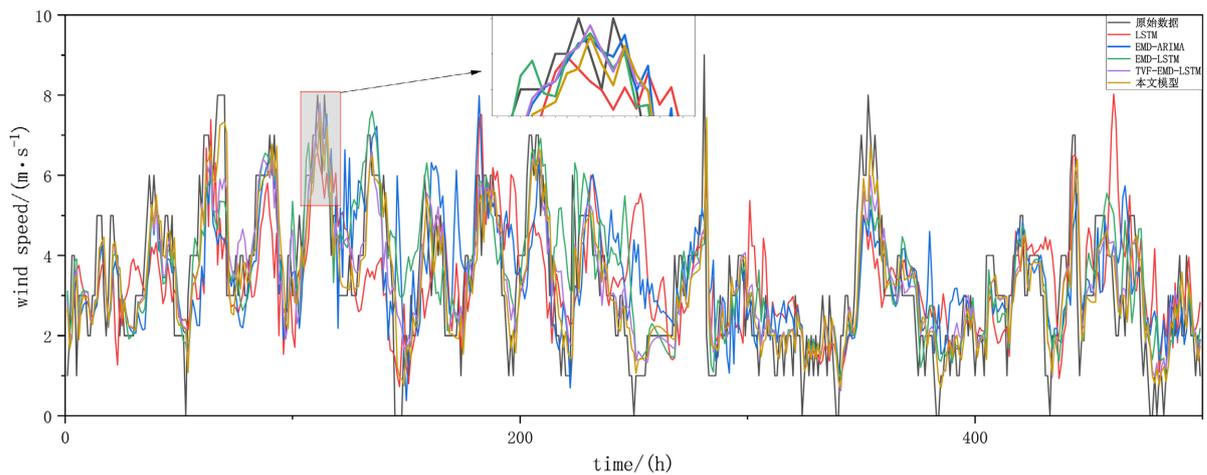


Figure 4. The predicted results of each model
图 4. 各模型的预测结果

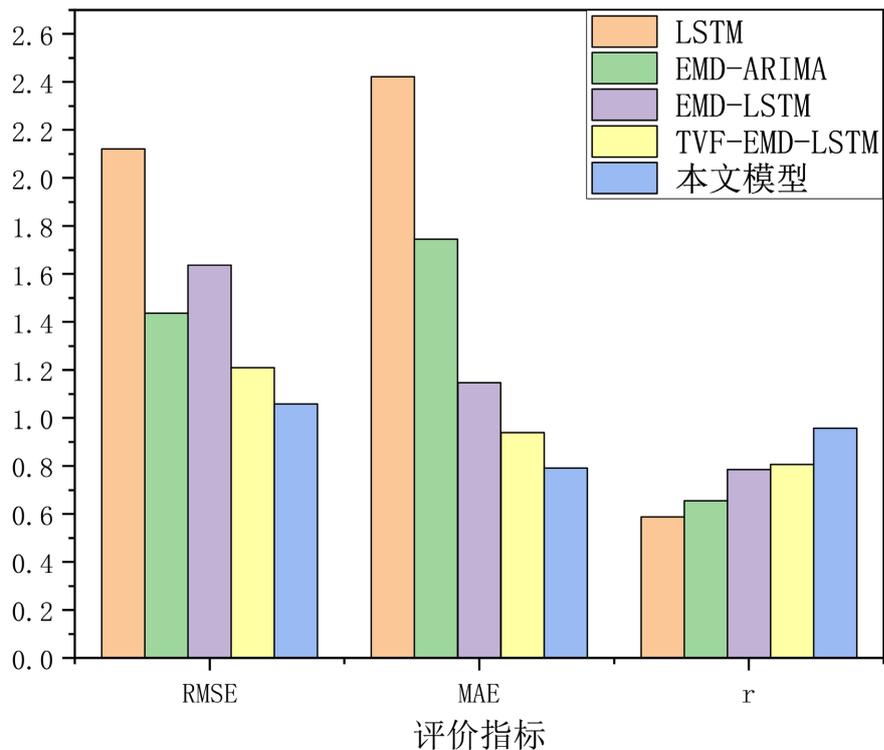


Figure 5. Evaluation indicators
图 5. 评价指标

Table 1. Error evaluation of each model
表 1. 各模型的误差评价

模型	RMSE	MAE	r
LSTM	2.1211	2.4215	0.5877
EMD-ARIMA	1.4378	1.7455	0.6544
EMD-LSTM	1.6376	1.1468	0.7852
TVF-EMD-LSTM	1.2101	0.9390	0.8056
本文模型	1.0576	0.7912	0.9576

5. 结论

本文运用了粒子群优化的 TVF-EMD 算法来分解风速序列，目的是解决模态分解的存在方法中分量的噪声和模态混叠问题，以提高模型预测的精准性。同时，采用 LSTM 和 ARIMA 模型分别针对高频和低频分量进行预测，这样做有助于对各种频率的模态分量做出有效预测，从而提升整体预测的准确性。以武汉每小时平均风速为例进行了风速预测对比及验证。并与其他模型进行对比，得到主要结论如下：

1) TVF-EMD 方法对于非线性和非平稳的风速时间序列有着更好的适用性。这种策略能够依据序列的特性作出相应的调整，并能通过使用时变滤波器有效地对抗模态混叠问题，同时保持序列的时变性。它也有助于大幅提高预测的精确度，为风速序列的预处理提出了新的对策。

2) 使用粒子群优化算法对 TVF-EMD 进行优化可以很好地提升 TVF-EMD 提取数据特征的能力，本文模型在各指标上预测精度均有明显提升，从数据中可看出提升显著。

不过，本文只考虑了短期风速预测，未来可以扩展到不同的时间尺度上进行风速预测，同时也可以考虑季节性变量或环境影响进行预测以确认此组合模型的通用性。

基金项目

国家自然科学基金项目(12361064, 52275504); 湖北省教育厅科学研究计划资助项目青年项目(Q20211111); 武汉科技大学冶金工业过程系统科学湖北省重点实验室开放基金项目(Y201905)。

参考文献

- [1] Sim, S.K., Maass, P. and Lind, P.G. (2018) Wind Speed Modeling by Nested ARIMA Processes. *Energies*, **12**, 69. <https://doi.org/10.3390/en12010069>
- [2] Tian, Z. and Chen, H. (2021) Multi-Step Short-Term Wind Speed Prediction Based on Integrated Multi-Model Fusion. *Applied Energy*, **298**, Article ID: 117248. <https://doi.org/10.1016/j.apenergy.2021.117248>
- [3] Zhang, Y., Chen, B., Zhao, Y., et al. (2018) Wind Speed Prediction of IPSO-BP Neural Network Based on Lorenz Disturbance. *IEEE Access*, **6**, 53168-53179. <https://doi.org/10.1109/ACCESS.2018.2869981>
- [4] 魏昱洲, 许西宁. 基于 LSTM 长短期记忆网络的超短期风速预测[J]. 电子测量与仪器学报, 2019, 33(2): 64-71.
- [5] 赵征, 汪向硕, 乔锦涛. 基于 VMD 和改进 ARIMA 模型的超短期风速预测[J]. 华北电力大学学报(自然科学版), 2019, 46(1): 54-59.
- [6] 王贺, 陈蕻峰, 熊敏, 等. 融合 CEEMDAN 和 ICS-LSTM 的短期风速预测建模[J]. 电子测量与仪器学报, 2022, 36(4): 17-23. <https://doi.org/10.13382/j.jemi.B2104894>
- [7] 向玲, 刘佳宁, 苏浩, 胡爱军, 朱泽宁. 基于 CEEMDAN 二次分解和 LSTM 的风速多步预测研究[J]. 太阳能学报, 2022, 43(8): 334-339.
- [8] 王俊, 李霞, 周昔东, 张凯. 基于 VMD 和 LSTM 的超短期风速预测[J]. 电力系统保护与控制, 2020, 48(11): 45-52.

-
- [9] Li, H., Li, Z. and Mo, W. (2017) A Time Varying Filter Approach for Empirical Mode Decomposition. *Signal Processing*, **138**, 146-158. <https://doi.org/10.1016/j.sigpro.2017.03.019>
- [10] 王文川, 高畅, 徐雷. 基于 TVF-EMD 与 LSTM 神经网络耦合的月径流预测研究[J]. 中国农村水利水电, 2022(2): 76-81+89.