

# 基于高斯过程回归和粒子滤波的短期风速预测

梁 智<sup>1</sup>, 孙国强<sup>1</sup>, 俞娜燕<sup>2</sup>, 倪晓宇<sup>3</sup>, 沈海平<sup>2</sup>, 卫志农<sup>1</sup>

(1. 河海大学能源与电气学院, 南京 210098; 2. 国网江苏省电力公司无锡供电公司, 无锡 214000;

3. 无锡扬晟科技股份有限公司, 无锡 214106)

**摘 要:** 建立高斯过程回归和粒子滤波相结合的短期风速预测模型, 实现对历史风速序列异常值的在线动态检测与修正。首先, 在训练样本集中通过高斯过程回归建立状态空间方程, 采用粒子滤波算法对当前量测值进行状态估计, 对估计值和量测值的残差进行分析, 并根据“ $3\sigma$ ”原则判断异常值。其次, 修正异常值, 并对修正后的风速序列重新建立高斯过程回归预测模型。在进行提前 15 分钟风速预测时, 同样采用粒子滤波算法对最新量测值进行状态估计, 实现了异常值在线检测并修正。算例分析结果表明, 粒子滤波算法能够有效检测出异常风速值, 降低了风速预测误差, 提前 15 分钟风速预测时平均绝对百分比误差和均方根误差分别降至 8.92% 和 0.5826 m/s。

**关键词:** 高斯过程回归; 粒子滤波; 异常值检测与修正; 短期风速预测

**中图分类号:** TK81

**文献标志码:** A

## 0 引 言

中国大力发展风能、太阳能等可再生能源, 在一定程度上缓解了能源紧张、环境恶化的局面, 但可再生能源电力的波动性和随机性给电网安全稳定运行带来了极大挑战。近年来, 风能在电网中的渗透率逐年上升, 2019 年全国风电新增装机 2574 万 kW, 截至 2019 年底, 全国风电累计装机 2.1 亿 kW。准确预测未来一定时间内的风速, 进而实现风力输出功率的短期预测对调度部门安排日前计划具有极其重要的意义。

目前用于风速预测主要有数值天气预报和统计模型两类方法<sup>[1]</sup>。数值天气预报需要建立物理模型, 通过微观气象学理论和计算流体力学得到风速、风向、温湿度等信息。统计模型方法主要采用数学统计的思想, 通过挖掘数据间存在的内在规律进行预测。此类方法主要有时间序列、神经网络、支持向量机、卡尔曼滤波等<sup>[2-3]</sup>。文献[4]建立了遗传算法优化相关向量机参数的短期风速预测模型, 同时利用 ARMA 拟合误差序列, 实现对风速预测值的校正。文献[5]引入 Elman 动态神经网络风速预测模型, 缓解了传统静态前馈神经网络易陷入局部最优解和动态性能不足问题。相关向量机和神经网络方法具有较好的预测精度, 但过程基于“黑箱”原理, 难于建立显性数学表达。文献[6]通过时

间序列分析建立风速的线性数学模型, 推导出物理意义明确的卡尔曼滤波状态方程和观测方程, 由卡尔曼递推方程进行预测, 取得了理想的预测精度。但卡尔曼滤波方法适用于线性数学模型, 对于非线性过程处理能力弱。文献[7]采用支持向量机 (support vector machines, SVM) 建立非线性状态空间模型, 用无迹卡尔曼滤波动态更新变换 Sigma 点获得均值和协方差, 完成预测值动态修正, 这种方法适用于高阶非线性系统, 获得了较好单步和多步预测效果。文献[8]基于高斯过程回归 (Gaussian process regression, GPR) 输出具有概率分布特点, 建立短期风速预测模型, 并对不同协方差函数对预测精度影响。与 SVM 和 BP 神经网络模型相比, GPR 具有更好预测精度。

短期风速预测往往存在 2 个方面的重要问题:

1) 历史风速序列混杂的噪声影响了预测模型精度。风速序列在采集、传输、存储等过程中不可避免地会受到各种噪声因素的影响, 如量测设备采集错误、数据传输过程中数据丢失等。在训练预测模型时, 这些异常值将导致预测值偏离真实值, 数据缺失严重情况下甚至难以给出预测结果。同时, 模型参数的不准确估计也降低了预测精度。文献[9-10]分别建立基于小波去噪和经验模态分解的异常值-检测和修正方法, 能够有效检测出异常值。此外, 统计学方法指出: 在训练

收稿日期: 2017-09-08

通信作者: 孙国强 (1978—), 男, 博士、副教授, 主要从事电力系统运行分析与控制、可再生能源发电系统方面的研究。

hhusunguoqiang@163.com

过程中,拟合值远远偏离真实值的样本点即为异常值点<sup>[11]</sup>。由此,产生基于偏差的异常值检测方法。该方法主要过程为:首先利用已知数据建立数学模型,根据拟合数据和真实值间的残差来判断数据是否异常。其假设条件是拟合残差服从均值为零的高斯分布<sup>[12]</sup>。文献[13]提出基于 SVM 的异常值检测方法,并将修正后的序列用于中期风速预测。该方法认为,若拟合残差绝对值大于给定阈值,则该点风速值异常,并用拟合值修正该异常值。文献[14]提出多元线性回归和最小二乘法异常值检测方法,该方法同样根据残差达到异常值检测目的,并将修正后序列用于短期风速预测建模。

2) 由于统计模型方法机理自身存在的滞后性,导致预测值变化趋势落后于风速真实值,尤其在风速发生突变时刻,预测模型输出值往往难以反映真实值,需要对预测值进行修正。

针对风速预测过程中存在的混杂噪声影响预测模型精度的问题,本文采用基于偏差的异常值检测方法。在利用 GPR 拟合真实值过程中,由于风速的强随机性和不确定性、预测模型自身的滞后性,使得拟合值和真实值有较大偏差,从而“淹没”了异常值处的偏差,导致异常值检测困难或造成误检、漏检。为消除正常值处较大的偏差,本文引入非线性非高斯滤波方法-粒子滤波(particle filter, PF),从而提出基于高斯过程回归和粒子滤波(GPR-PF)相结合的短期风速预测模型,实现了异常值在线动态检测与修正。通过算例验证了本文模型能够有效地对异常值进行检测。

## 1 基于GPR的风速预测模型

GPR 以贝叶斯理论和统计学习理论为基础,在处理高维数、小样本和非线性等复杂回归问题时具有易编程实现、超参数自适应获取以及输出概率分布等优点<sup>[15]</sup>,在时间序列分析、动态系统模型辨识等多领域获得广泛应用。

建立 GPR 短期风速预测模型,假设训练样本集合  $D = \{(x_i, y_i) | i = 1, 2, 3, \dots, n\} = (X, y)$ , 其中:  $x_i \in \mathbf{R}^m$  为  $m$  维输入向量,  $m \times n$  维输入矩阵则可表示为  $X = [x_1, x_2, \dots, x_n]$ ,  $n$  表示训练样本点数量,  $y_i \in \mathbf{R}$  为对应于  $x_i$  的输出标量。用数学语言描述风速预测过程为<sup>[16]</sup>:

定义函数空间  $f(x) = \Phi(x)^T \omega, f(x^{(1)}), f(x^{(2)}), \dots, f(x^{(n)})$  构成随机变量的一个集合,且服从联合高斯分布,高斯过程模型就可表示为:

$$y \sim GP[m(x), k(x, x') + \sigma_n^2 \delta_{ij}] \quad (1)$$

式中,权重向量  $\omega$  服从先验分布  $\omega \sim N(0, \Sigma_p)$  满足  $E(\omega) = 0$  且  $E(\omega\omega^T) = \Sigma_p$ ; 独立高斯白噪声服从均值为 0, 方差为  $\sigma^2$  的高斯分布,记做  $\varepsilon \sim N(0, \sigma^2)$ ;  $\delta_{ij}$  为 Kronecker delta 函数,当  $i=j$  时,函数  $\delta_{ij} = 1$ ;  $m(x)$  为有限维分布族的均值函数,描述风速均值输出结果;  $k(x, x')$  为协方差函数,刻画风速方差大小。

为简化推导,风速均值  $m(x)$  进行数据预处理使之 0。GPR 预测模型在  $n$  维训练集  $D$  内建立先验分布,在  $n_*$  维测试集  $D_* = \{(x_i, y_i) | i = n+1, \dots, n+n_*\}$  下转变为后验分布,则训练样本观测值  $y$  和测试数据的输出向量  $f_*$  之间构成联合高斯分布,即

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left\{ 0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X, X_*) & K(X_*, X_*) \end{bmatrix} \right\} \quad (2)$$

其中,  $K(X, X) = K_n$  表示  $n \times n$  的核矩阵,其元素  $K_{ij} = k(x_i, x_j)$ ;  $K(X, X_*) = K(X_*, X)^T$  为测试数据  $X_*$  与训练集的输入  $X$  之间的协方差矩阵;  $K(X_*, X_*)$  为  $X_*$  自身的协方差。

由此得出预测值  $f_*$  后验分布为:

$$f_* | X, y, x_* \sim N(\bar{f}_*, \text{cov}(f_*)) \quad (3)$$

其中

$$\bar{f}_* = K(x_*, X)[K(X, X) + \sigma_n^2 I_n]^{-1} y \quad (4)$$

$$\text{cov}(f_*) = k(x_*, x_*) - K(x_*, X)[K(X, X) + \sigma_n^2 I_n]^{-1} K(X, x_*) \quad (5)$$

均值向量  $\bar{f}_*$  为 GPR 模型风速预测均值,对应于点预测输出,  $\hat{\sigma}_{f_*}^2 = \text{cov}(f_*)$  为对应于  $\bar{f}_*$  的方差,可获得概率分布意义下风速区间预测结果。

## 2 粒子滤波基本原理

PF 算法具有良好的非线性非高斯系统状态滤波能力,随机量无需满足高斯分布制约条件,因而在信号处理、通信等领域获得应用<sup>[17]</sup>。本文选择粒子滤波算法对风速序列异常值进行处理,提出基于 GPR-PF 的短期风速预测模型。

PF 基本思想是用一组粒子近似表示系统的后验概率分布,然后用这一近似的表示估计非线性系统的状态。用非线性系统动态状态空间模型描述粒子滤波过程<sup>[18]</sup>为:

$$\begin{cases} x_k = f(x_{k-1}) + w_{k-1} \\ z_k = h(x_k) + v_k \end{cases} \quad (6)$$

式中,  $x_k, z_k$  ——  $k$  时刻系统状态向量和量测值;  $f(\cdot)$ 、 $h(\cdot)$  —— 系统状态转移函数和测量模型函数;  $w_{k-1}$ 、 $v_k$  —— 系统过程噪声和观测噪声。

从重要性密度函数  $q(x_k | x_{1:k-1}, z_k)$  抽样得到  $N$  个样

本,并将这些样本表示为:  $\{x_{k-1}^i\}_{i=1}^N$ 。状态  $k$  时刻的预测值为  $x_k^i$ , 即:

$$x_k^i = f_{k-1}(x_{k-1}^i) + w_{k-1}^i \quad (7)$$

式中,  $w_{k-1}^i$  ——系统已知噪声分布中采样得到的对应于粒子  $i$  的独立样本。完成预测阶段的所有粒子就构成  $k$  时刻的先验概率样本, 表示为  $\{x_k^i\}_{i=1}^N$ , 也就得到先验概率密度  $p(x_k|Z_{k-1})$ 。在获得新的观测值  $z_k$  后, 各个粒子  $x_k^i$  根据权值公式进行更新:

$$w_k^i = w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^{i-1})}{q(x_k^i|x_{k-1}^{i-1}, z_k)} \quad (8)$$

对权值进行归一化处理:

$$\bar{w}_k^i = \frac{w_k^i}{\sum_{j=1}^N w_k^j} \quad (9)$$

利用重采样方法克服粒子退化现象<sup>[19]</sup>。根据每个粒子归一化权值对样本重新采样, 复制权值较大的粒子, 删除权值较小的粒子, 得到等权值粒子集, 即可得到后验分布密度函数:

$$\hat{p}(x_k|z_k) = \frac{1}{N} \sum_{i=1}^N \delta(x_k - x_k^i) \quad (10)$$

式中,  $\delta(\cdot)$  ——狄拉克 Kronecker delta 函数, 一个循环迭代过程结束。

最小均方误差准则意义下状态最优估计为:

$$\hat{x}_k = \frac{1}{N} \sum_{i=1}^N x_k^i \quad (11)$$

### 3 短期风速预测的 GPR-PF 模型

为准确估计模型参数, 降低异常值对预测精度的影响, 本文提出基于 GPR-PF 的短期风速预测模型, 实现对异常值的在线动态检测与修正。应用 PF 算法对风速序列进行异常值检测时, 关键过程是确定模型的状态向量及状态方程。状态方程通过在训练样本集中建立 GPR 非线性回归方程得到, 因而状态向量即为 GPR 模型的输入变量。

图 1 为建立 GPR-PF 短期风速预测模型流程图, 主要过程为: 首先, 将偏自相关函数用于原始风速序列<sup>[20]</sup>, 确定状态向量及构建合适的训练样本集合。其次, 建立 GPR 状态转移函数并求解超参数, 采用 PF 算法对当前量测值进行状态估计, 获得估计值和量测值间的残差, 并根据“ $3\sigma$ ”准则判断异常值, 从而实现了异常值的检测。然后, 修正异常风速值, 并对修正后的风速序列重新建立 GPR 预测模型。最终, 在进行提前 15 分钟的短期风速预测时, 同样采用 PF 算法对最新量测值

进行状态估计, 实现异常值的在线动态检测与修正。

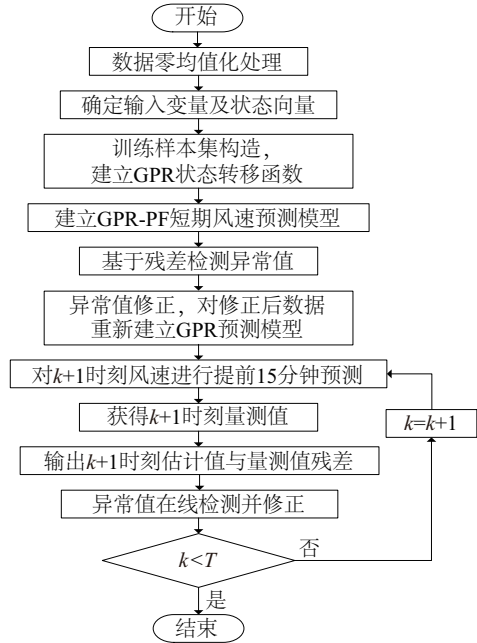


图1 GPR-PF 短期风速预测流程图

Fig. 1 Flow chart of short-term wind speed forecasting based on GPR-PF model

#### 3.1 确定状态向量

自相关函数和偏自相关函数在识别模型类型及估计阶数过程中具有重要意义<sup>[21]</sup>。本文根据自相关函数和偏自相关函数衡量  $X_k$  与  $X_{k-\tau}$  间的相关关系, 从而分析出时间延迟, 确定输入变量集合及状态向量。其中,  $\tau$  为延迟阶数。

将江苏省某风电场实测风速(时间间隔为 15 min)用于 GPR-PF 短期风速预测建模, 图 2 为时间序列分析自相关函数图和偏自相关函数图。从图 2 中可看出, 自相关函数具有拖尾特征, 而偏自相关函数截尾, 从而风速序列满足 AR 模型。结合偏自相关函数图, 本文选取 4 个输入变量, 即预测  $k+1$  时刻的风速  $X_{k+1}$  时, 将

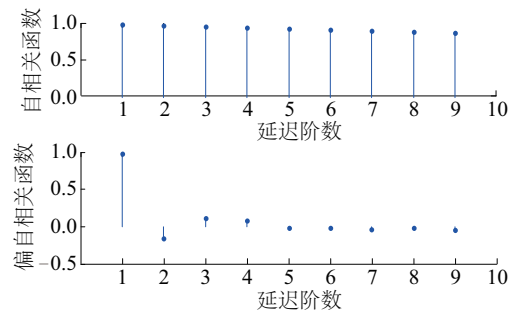


图2 自相关和偏自相关函数图

Fig. 2 Figure of autocorrelation function and partial autocorrelation function



$k, k-1, k-2, k-3$  时刻的风速  $X(k), X(k-1), X(k-2), X(k-3)$  作为输入变量。令  $X_1(k)=X(k), X_2(k)=X(k-1), X_3(k)=X(k-2), X_4(k)=X(k-3)$  [22], 则  $k$  时刻状态向量为  $X(k)=[X_1(k)X_2(k)X_3(k)X_4(k)]^T$ 。

### 3.2 建立 GPR-PF 状态空间方程

在采用粒子滤波进行风速序列异常值检测与修正时,首先采用历史风速值预测当前时刻风速状态,然后根据当前时刻风速量测值对预测值进行修正,从而获得此时刻量测值的最优估计,并获得风速估计值与量测值间的残差。

结合 3.1 节确定的状态向量,本文的 GPR-PF 短期风速预测状态空间模型为:

$$\begin{cases} X_{k+1} = GP(X_k) + W_k \\ y_{k+1} = HX_{k+1} + V_{k+1} \end{cases} \quad (12)$$

式中,  $H=[1 \ 0 \ 0 \ 0]$ ; 非线性状态转移函数  $GP(\cdot)$  反映由  $k$  时刻状态  $X_k$  得到  $k+1$  时刻风速预测值,该函数由训练样本集合通过训练 GPR 模型参数得到。

粒子滤波通过对  $k$  时刻的状态来估计得到  $k+1$  时刻风速估计值,并根据  $k+1$  时刻量测值修正估计值,从而获得了滤波后风速最优估计值,其滤波过程具体实现步骤可参考文献[23]。

### 3.3 异常值判别方法

本文采用基于偏差的数据异常值检测方法检测并修正原始风速序列中存在的异常值,通过分析 GPR-PF 风速估计值和量测值之间存在的残差  $r$ ,进而根据“ $3\sigma$ ”准则来判断异常值。

对于测试数据  $r_1, r_2, \dots, r_n$ , 取其算术平均值:

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i \quad (13)$$

及剩余误差值  $v_i = r_i - \bar{r}$ , 求出其均方根偏差为:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n v_i^2} \quad (14)$$

则异常值判别依据为:若  $|r_i - \bar{r}| > 3\sigma$ , 该值为异常数据;若  $|r_i - \bar{r}| \leq 3\sigma$ , 则  $r_i$  为正常数据。

### 3.4 评价指标

本文选择平均绝对百分比误差(mean absolute percentage error, MAPE)和均方根误差(root mean square error, RMSE)作为模型预测效果评价指标,计算公式分别为:

$$MAPE = \frac{1}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \quad (15)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2} \quad (16)$$

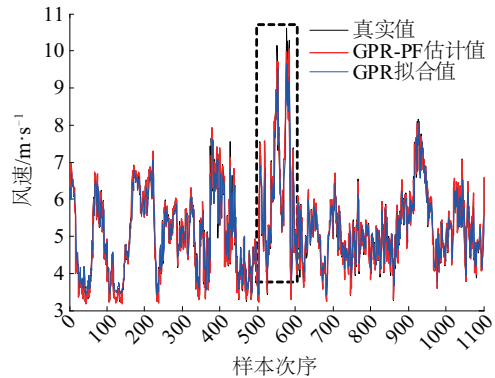
式中,  $T$ ——预测点数量;  $y_i$ ——第  $i$  个预测点风速真实值;  $\hat{y}_i$ ——第  $i$  个预测点模型预测值。

## 4 算例分析

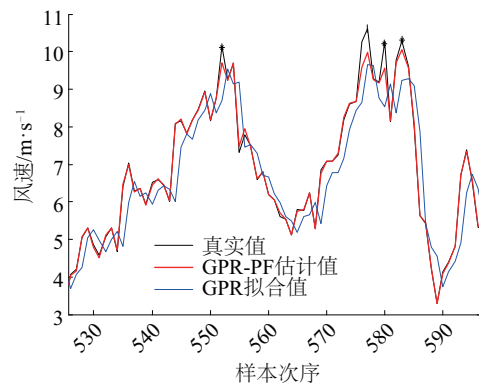
本文采用某风电场 2008 年 5 月 14 日 12:00~5 月 25 日 23:45 共 1104 个实测风速值作为训练样本序列,数据采样时间间隔为 15 min,建立 GPR-PF 风速预测模型,对 5 月 26 日的 96 个风速值做提前 1 步(即提前 15 分钟)预测。

首先,由训练样本集建立 GPR 短期风速预测模型并求解模型超参数,建立状态空间方程,分析模型拟合值与量测值间的残差,从而采用基于偏差的方法进行异常值检测。

图 3a 为训练过程中采用 GPR 与 GPR-PF 2 种方法得到的预测结果。从图 3a 中可看出,采用 GPR 模型进行短期风速预测时,由于风速强波动性与随机性,使得 GPR 模型难以很好地跟踪风速变化趋势,存在一定的滞后性,从而产生较大的偏差。而 GPR-PF 模型采用当前时刻量测值对风速估计值进行动态更新,能够很好地估计出风速的真实状态。从图 3b 局部放大图可更



a. 训练过程风速拟合结果



b. 风速拟合结果局部放大图

图3 风速拟合结果及局部放大图

Fig. 3 Wind speed fitting results and partial magnification

细致比较 2 种模型拟合效果。对风速预测值与实际量测值进行残差分析,图 4 分别为采用 GPR 和 GPR-PF 模型进行拟合时残差结果。单独采用 GPR 模型时,残差分布较分散,不利于异常值分析与检测。而采用 GPR-PF 混合方法时,残差分布较集中,可较易采用“ $3\sigma$ ”准则判断异常值。

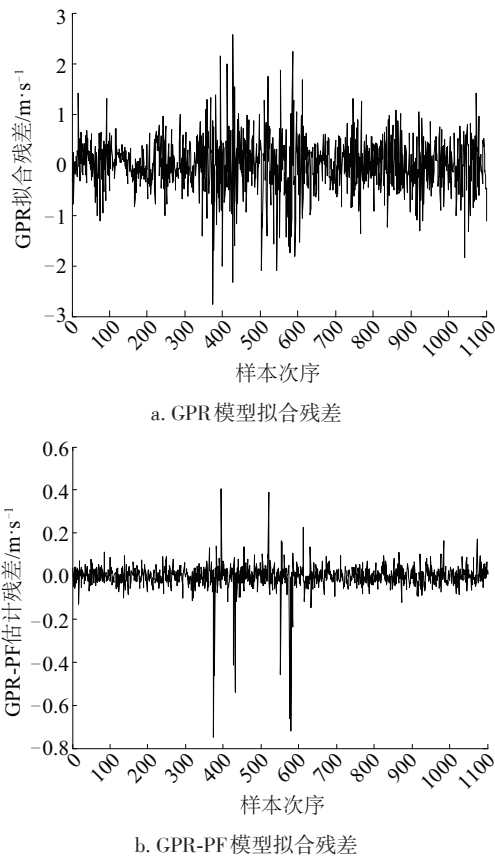


图 4 不同模型拟合残差

Fig. 4 Fitting residuals of different model

通过对残差分布的分析及“ $3\sigma$ ”准则,对异常值检测与修正结果如表 1 所示。

对修正后的风速序列重新建立 GPR 预测模型,并进行提前 15 分钟的短期风速预测。图 5 为采用 BP 神经网络、SVM、GPR 和 GPR-PF 这 4 种模型的风速预测结果。预测模型定量评价指标结果如表 2 所示。可看出,相对于 BP 神经网络、SVM 这 2 种预测模型,GPR 可给出较好的预测精度。通过采用粒子滤波算法对异常值进行修正后,本文的 GPR-PF 风速预测模型削弱了噪声影响,从而获得最优预测结果。

为进一步验证模型有效性,本文按照相同步骤将 GPR-PF 模型用于风速的提前 30 分钟预测。表 3 为提前 30 分钟风速预测结果。

表 1 速度异常值检测与修正结果

Table 1 Abnormal value detection and correction results

序号	日期	时刻	异常值	修正值
1	2008-05-18	10:30	7.42	4.66
2	2008-05-18	15:30	4.95	7.09
3	2008-05-18	23:30	7.54	5.22
4	2008-05-19	00:30	6.63	5.01
5	2008-05-20	06:45	10.15	8.70
6	2008-05-20	12:45	10.25	8.64
7	2008-05-20	13:45	10.28	8.54
8	2008-05-20	14:30	10.29	9.23
9	2008-05-20	21:45	3.90	5.57

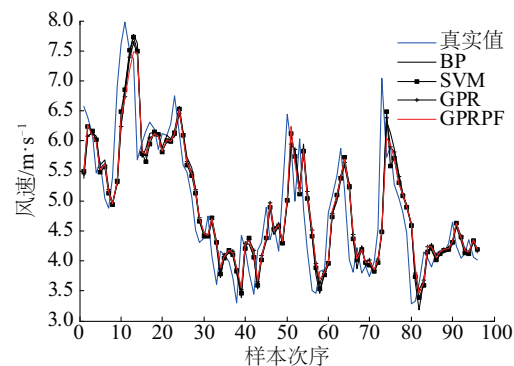


图 5 提前 15 分钟风速预测结果

Fig. 5 Wind speed forecasting results ahead 15 mins

表 2 提前 15 分钟风速预测误差

Table 2 Error results of wind speed forecasting ahead 15 mins

评价指标	BP	SVM	GPR	GPR-PF
MAPE/%	9.94	9.74	9.85	8.92
RMSE/ $\text{m}\cdot\text{s}^{-1}$	0.6525	0.6294	0.6460	0.5826

表 3 提前 30 分钟风速预测误差

Table 3 Error results of wind speed forecasting ahead 30 mins

评价指标	BP	SVM	GPR	GPR-PF
MAPE/%	11.54	11.32	11.37	10.46
RMSE/ $\text{m}\cdot\text{s}^{-1}$	0.7322	0.7056	0.7043	0.6410

## 5 结 论

针对风速具有的强非线性、随机性以及量测值混杂噪声特点,本文提出基于高斯过程回归和粒子滤波的短期风速预测模型,实现了异常值的在线动态检测与修正。算例结果表明,通过结合粒子滤波过程,可有

效检测出风速序列中含有的异常值,对修正后的风速序列重新建模,降低了风速预测误差,提前 15 分别占该风速预测时平均绝对百分比误差和均方根误差分别降至 8.92% 和 0.5826 m/s。下一步将研究高斯过程回归优化模型,以进一步提高模型预测精度及鲁棒性。

### [参考文献]

- [1] YAN J, LIU Y Q, HAN S, et al. Reviews on uncertainty analysis of wind power forecasting[J]. Renewable and sustainable energy reviews, 2015, 52: 1322-1330.
- [2] LOUKA P, GALANIS G, SIEBERT N, et al. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering[J]. Journal of wind engineering and industrial aerodynamics, 2008, 96: 2348-2362.
- [3] 杜颖, 卢继平, 李青, 等. 基于最小二乘支持向量机的风电场短期风速预测[J]. 电网技术, 2008, 32(15): 62-66.  
DU Y, LU JP, LI Q, et al. Short-term wind speed forecasting of wind farm based on least square-support vector machine[J]. Power system technology, 2008, 32(15): 62-66.
- [4] 孙国强, 卫志农, 翟玮星. 基于RVM与ARMA误差校正的短期风速预测[J]. 电工技术学报, 2012, 27(8): 187-193.  
SUN G Q, WEI Z N, ZHAI W X. Short term wind speed forecasting based on RVM and ARMA error correcting[J]. Transactions of China Electrotechnical Society, 2012, 27(8): 187-193.
- [5] 尹东阳, 盛义发, 蒋明洁, 等. 基于粗糙集理论-主成分分析的Elman神经网络短期风速预测[J]. 电力系统保护与控制, 2014, 42(11): 46-51.  
YIN D Y, SHENG Y F, JIANG M J, et al. Short-term wind speed forecasting using Elman neural network based on rough set theory and principal components analysis[J]. Power system protection and control, 2014, 42(11): 46-51.
- [6] 潘迪夫, 刘辉, 李燕飞. 基于时间序列分析和卡尔曼滤波算法的风电场风速预测优化模型[J]. 电网技术, 2008, 32(7): 82-86.  
PAN D F, LIU H, LI Y F. A wind speed forecasting optimization model for wind farms based on time series analysis and Kalman filter algorithm[J]. Power system technology, 2008, 32(7): 82-86.
- [7] CHEN K L, YU J. Short-term wind speed prediction using an unscented Kalman filter based state-space support vector regression approach[J]. Applied energy, 2014, 113: 690-705.
- [8] 孙斌, 姚海涛, 刘婷. 基于高斯过程回归的短期风速预测[J]. 中国电机工程学报, 2012, 32(29): 104-109.  
SUN B, YAO H T, LIU T. Short-term wind speed forecasting based on Gaussian process regression model[J]. Proceedings of the CSEE, 2012, 32(29): 104-109.
- [9] LIU H, TIAN H Q, CHEN C, et al. A hybrid statistical method to predict wind speed and wind power[J]. Renewable energy, 2010, 35(8): 1857-1861.
- [10] GUO Z, ZHAO W, LU H, et al. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model[J]. Renewable energy, 2012, 37(1): 241-249.
- [11] TAN P, STEINBACH M, KUMAR V. Introduction to data mining[M]. Boston, MA, USA: Pearson Addison Wesley, 2006, 651 - 683.
- [12] 苏卫星, 朱云龙, 胡琨元, 等. 基于模型的过程工业时间序列异常值检测方法[J]. 仪器仪表学报, 2012, 33(9): 2080-2087.  
SU W X, ZHU Y L, HU K Y, et al. Model-based outlier detection method for time series of process industry[J]. Chinese journal of scientific instrument, 2012, 33(9): 2080-2087.
- [13] WANG J, QIN S, ZHOU Q, et al. Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China[J]. Renewable energy, 2015, 76: 91-101.
- [14] WANG J, XIONG S. A hybrid forecasting model based on outlier detection and fuzzy time series-A case study on Hainan wind farm of China[J]. Energy, 2014, 76: 526-541.
- [15] RASMUSSEN C E, WILLIAMS C K I. Gaussian processes for machine learning[M]. USA: MIT Press, 2006.
- [16] KO J, FOX D. GP- Bayes Filters: Bayesian filtering using Gaussian process prediction and observation models[J]. Autonomous robot, 2009, 27(1): 75-90.
- [17] 王法胜, 鲁明羽, 赵清杰, 等. 粒子滤波算法[J]. 计算机学报, 2014, 37(8): 1679-1694.  
WANG F S, LU M Y, ZHAO Q J, et al. Particle filtering algorithm[J]. Chinese journal of computers, 2014, 37(8): 1679-1694.
- [18] ZHAO W, TAO T, DING Z S, et al. A dynamic particle filter-support vector regression method for reliability prediction[J]. Reliability engineering and system

- safety, 2013, 119: 109-116.
- [19] 胡士强, 敬忠良. 粒子滤波算法综述[J]. 控制与决策, 2005, 20(4): 361-365.
- HU S Q, JING Z L. Overview of particle filter algorithm [J]. Control and decision, 2005, 20(4): 361-365.
- [20] LIU D, NIU D, WANG H, et al. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm[J]. Renewable energy, 2014, 62(3): 592-597.
- [21] 范剑青, 姚琦伟, 陈敏. 非线性时间序列: 建模预报及应用[M]. 北京: 高等教育出版社, 2005: 28-46.
- FAN J Q, YAO Q W, CHEN M. Nonlinear time series: modeling, forecasting and application [M]. Beijing: Higher Education Press, 2005: 28-46.
- [22] LIU H, TIAN H Q, LI Y F. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction[J]. Applied energy, 2012, 98(1): 415-424.
- [23] EMAMI K, FERNANDO T, IU H H-C, et al. Particle filter approach to dynamic state estimation of generators in power systems[J]. IEEE transactions on power systems, 2015, 30(5): 2665-2675.

## SHORT-TERM WIND SPEED FORECASTING BASED ON GAUSSIAN PROCESS REGRESSION AND PARTICLE FILTER

Liang Zhi<sup>1</sup>, Sun Guoqiang<sup>1</sup>, Yu Nayan<sup>2</sup>, Ni Xiaoyu<sup>3</sup>, Shen Haiping<sup>2</sup>, Wei Zhinong<sup>1</sup>

(1. College of Energy and Electrical Engineering, Hohai University, Nanjing 210098, China;

2. State Grid Jiangsu Electric Power Company Wuxi Power Supply Company, Wuxi 214000, China

3. Wuxi Yang Sheng Technology Co., Ltd., Wuxi 214106, China)

**Abstract:** Improving the wind speed prediction accuracy of wind farm will help enhance the power grid stability and economy. Noise or data loss often appears in historical wind speed sequences. These abnormal values will lead to inaccurate estimation of model parameters. Therefore, the affecting the prediction accuracy. Therefore, the detection and correction of abnormal value is the prerequisite and necessary measure to effectively analyze the law of wind speed. In this paper, a short-term wind speed forecasting model combining Gaussian process regression and particle filter is established, which realizes online dynamic detection and correction of outliers. Firstly, the state space equation is established by the Gaussian process regression in the training sample set. The particle filter algorithm is then used to estimate the current measurement value. The residuals between the estimated and measured values are analyzed and the anomalous values are detected according to the principle of “ $3\sigma$ ”. Secondly, after the anomaly being corrected, the Gaussian process regression forecasting model is reconstructed. The particle filter algorithm is repeatedly used to estimate the latest measurement value during the process of 15 mins ahead wind speed forecasting, realizing the online dynamic detection and correction of outliers. The case study show that the particle filter algorithm can effectively detect the abnormal values and reduce the wind speed prediction error, the average absolute percentage error and root mean square error are reduced to 8.92% and 0.5826 m/s respectively when the wind speed is predicted 15 minutes ahead.

**Keywords:** Gaussian process regression; particle filter; detection and correction of outliers; short-term wind speed forecasting