

机电技术

基于时序长记忆模型的风电场短期功率预测

卢锦玲¹ 王 阳¹ 杨 月¹ 何振民¹ 项 丽² 李笑宁²

(华北电力大学电气与电子工程学院¹,保定 071003; 国网宁夏电力公司²,银川 750001)

摘 要 随着风电的飞速发展,风电场接入电网的规模日益扩大,随之而来的是对风电功率预测准确性要求的提高,准确的风电功率预测可以更好地利用风能资源,减小风电并网对电网的不利影响。为了提高风电预测的精度,采用最大期望算法(expectation maximization algorithm, EM 算法)对风电场功率历史数据进行处理,填补缺失值,替换错误数据,然后采用修正重标极差分析法即修正 R/S 分析方法分析数据的长记忆性,采用时间序列 ARFIMA 模型,然后根据预测时刻之前的功率数据,通过贝叶斯统计推断对模型参数进行估计,生成预测模型方程,进而对风电场输出功率进行预测。

关键词 风电功率预测 EM 算法 时间序列分析法 修正 R/S 分析法 ARFIMA 模型

中图分类号 TM614; **文献标志码** A

准确的风电场功率预测有利于电网调度合理地安排调度计划并更有效利用风电资源,提高风电上网小时数,有利于电网的安全稳定运行,减轻风电对整个电网的冲击,并且有利于正确制定电能交换计划、减少运行备用容量从而提高了运行的经济性^[1,2]。

风速预测方法有物理预测方法^[3]、统计学方法(时间序列法^[4]、灰色预测法^[5]、卡尔曼滤波法等^[6])、人工智能预测方法(神经网络法^[7]、模糊逻辑等^[8])和综合预测方法^[9]。其中,时间序列法是通过分析风电场功率历史数据考虑时间序列的相关性,确定模型类型和阶数,对参数进行估计,最后进行模型检验建立预测模型,从而利用预测模型对风电场功率进行预测。常用的随机线性时间序列模型有自回归模型(autoressive, AR)、滑动平均模型(moving average, MA)、自回归滑动平均模型(autoressive and moving average, ARMA),这些模型属于短记忆模型,只考虑了时间序列的短记忆性,但是时间序列不仅具有短记忆性许多时间序列中存在着长记忆现象。

准确的预测模型是建立在准确的历史数据的基础上的,但是在生产实际中数据的采集过程可能出现由测量系统故障造成数据异常或者丢失以及由数据传输通道故障及拥堵造成的数据错误及不真实,若直接使用这些不良数据对其输出功率进行预测,不利于预测的准确度的提高^[10]。因此,为了提高功

率预测的精度,采用了一种基于最大期望算法(expectation maximization algorithm, EM 算法)的修正和补全不良数据方法。

1 EM 算法

1.1 EM 算法的基本原理

EM 算法是一种可以用于含缺失或不良数据的数据分析中的一种迭代算法,此算法是以一种极大似然估计的方法处理不完全数据^[11]。EM 算法是分析数据找到缺失数据和不良数据用较接近真实值的估计值代替这些值,替换数据时选用条件期望值进行替换效果较好,但是由于是不完全数据,数据的参数估计并不知道,无法得到不良数据处的条件期望,这就需要首先在 E 步先用初始估计值替换缺失数据和不良数据,然后在 M 步进行一次参数估计,重新估计分布参数,以使得数据的似然性最大,用新参数计算出的新的缺失数据和不良数据估计值代回到 E 步,两个步骤交替进行计算一直进行到满足收敛条件为止。如果第 n 次迭代开始时参数 θ 的估计值记做 $\theta_{(n)}$,则在第 $n+1$ 次迭代中的期望步和最大似然步为:

E 步(期望步),前一次迭代所得对数似然函数的条件期望

$$Q(\theta | \theta_{(n)}) = \hat{E}_Z[\ln f(Y | \theta) \theta_{(n)}, Z]。$$

M 步(最大似然步),最大化对数似然函数以确定参数的值,并用于下步的迭代

$$Q(\theta_{(n+1)} | \theta_{(n)}) = \max_{\theta} Q(\theta | \theta_{(n)})。$$

如此形成了一次迭代由 $\theta_{(n)}$ 得到 $\theta_{(n+1)}$ 。重复 E 步和 M 步直至 $\|\theta_{(n+1)} - \theta_{(n)}\|$ 满足充分小条件时

2015 年 7 月 21 日收到

第一作者简介:卢锦玲(1971—),女,博士,副教授。研究方向:电力系统运行、分析与控制。E-mail:hd_lujingling@126.com。

迭代停止。

1.2 基于 EM 算法的不完全数据处理

采用 EM 算法对含缺失或不良数据的风电场输出功率历史数据进行处理的步骤,具体如下:

步骤一 分析数据,做出初始划分;

步骤二 初始化:对数据集密度分布待估计的参数 Θ 初值进行设置,包括各类别的比例 $\alpha^{(0)}j$ 、均值向量 $e^{(0)}j$ 和协方差矩阵 Σ_j ;将数据中的有效数据分为均服从高斯分布的 n 组,给定每组的初始权重,即 $\alpha_j = 1/n, j = 1, 2, 3, 4, \dots, n$;

步骤三 计算 E 步:计算隐含变量数据的期望值,此处,只需求数据的条件概率 $P(j|x)$ 即可。

采用随机变量 C 指示数据成分,则概率 $P_{ij} = P(C = i | x_j)$ 表示数据 x_j 由成分 i 产生的概率,也即由第 j 个高斯分布产生的概率。

由贝叶斯公式得,

$P_{ij} = P(x_j | C = i)P(C = i)$, 其中 $P(x_j | C = i)$, 即 x_j 在第 i 个高斯分布中的概率, $P(C = i)$ 是第 i 个高斯分布的权重参数。按照混合高斯模型的思想,表示如下:

$$P(j|x) = \frac{\alpha_j G(x, e_j^i, \Sigma_j)}{P(x, \Theta)}, j = 1, 2, 3, \dots, n$$

步骤四 计算 M 步:通过求解对数似然方程,计算期望值到达极大值点时新的均值 e_j 和权重 α_j , 其中:

$$\begin{cases} e_j^{(i+1)} = \frac{\sum_{i=1}^n P(j|x_i) x_i}{\sum_{i=1}^n P(j|x_i)} \\ \alpha_j^{(i+1)} = \frac{1}{n} \sum_{i=1}^n \frac{\alpha_j^{(i)} G(x, e_j^i, \Sigma_j)}{\sum_{i=1}^n \alpha_j^{(i)} G(x, e_j^i, \Sigma_j)} = \frac{1}{n} \sum_{i=1}^n P(j|x_i) \end{cases}$$

步骤五 测试是否满足初始设置的循环迭代条件 $|\alpha(i+1)j - \alpha(i)j| \leq \varepsilon$, 如果不满足条件,则转至步骤三;如果满足条件,则迭代停止,输出最终参数 $\alpha_j, e_j, j = 1, 2, 3, 4, \dots, n$, 此时 $e_j, j = 1, 2, 3, 4, \dots, n$ 即作为数据集中缺失点的插补值。

流程如图 1 所示。

2 长记忆时间序列预测

2.1 长记忆时间序列定义

当分析时间序列数据时会发现相距较远的数据值的相关性虽然已经比较小但并不足以被忽视,并且分析这种时间序列的自相关函数会发现它的自相关函数呈双曲率缓慢下降,这种特性就是时间序列

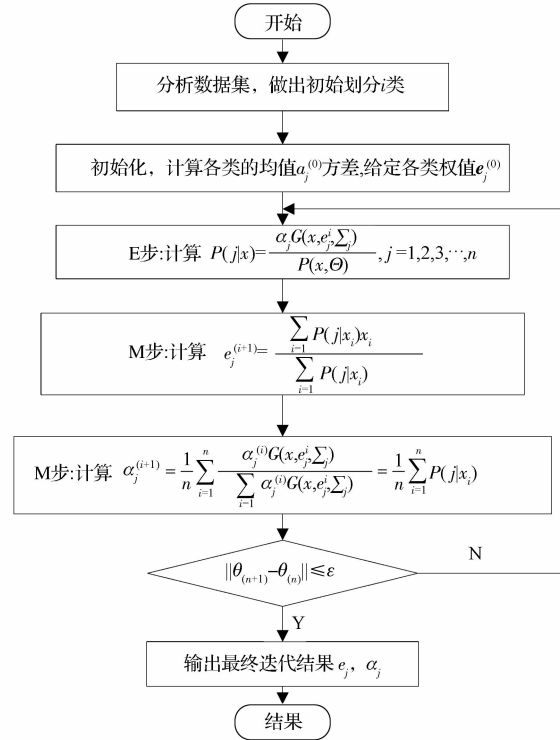


图 1 基于 EM 算法的不完全数据处理流程图

Fig. 1 Incomplete data processing flow chart based on EM algorithm

的长记忆性。如在时域中分析时间序列的长记忆性,记 ρ_τ 为平稳时间序列 $\{x_t\}$ 的自相关函数,若 ρ_τ 满足 $\rho_\tau \sim c\tau^{2d-1}, \tau \rightarrow \infty$, 其中, c 为常数,“ \sim ”表示收敛速度相同,当 $0 < d < 0.5$ 时, $\lim_{n \rightarrow \infty} \sum_{\tau=-n}^{\tau=n} |\rho_\tau| \rightarrow \infty$, 此时称时间序列 $\{x_t\}$ 为长记忆序列^[12]。

2.2 时间序列长记忆性分析

对于时间序列 $\{x_t\}$, 为了研究时间序列的长记忆性,可以取 n 个序列观测值,通过重标极差分析法,即 R/S 分析方法,计算 Hurst 指数,进而分析其长记忆性。R/S 统计量定义为: $Q_n = R_n/S_n$, 式中

$$R_n = \max_{1 \leq k \leq n} \sum_{j=1}^k (x_j - \bar{x}_n) - \min_{1 \leq k \leq n} \sum_{j=1}^k (x_j - \bar{x}_n), S_n = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2}, \text{ 其中 } \bar{x}_n = \frac{1}{n} \sum_{j=1}^n x_j. \text{ 可以证得}$$

$$Plim(n^{-H} Q_n) = c.$$

式中 c 为常数, H 为 Hurst 指数,则 H 的近似估计为 $H = \ln Q_n / \ln n$ 。当 $H \leq 0.5$ 时,时间序列为短记忆;而当 $H > 0.5$ 时,时间序列为长记忆。

R/S 分析法可以分析任何时间序列的分形性质,确定其长期记忆特征,但 R/S 分析方法有一定的缺点, R/S 统计量在短期记忆和异方差情况缺乏稳健性,所以对 R/S 统计量进行改进,使其不必太

顾忌短记忆的形式,可以引入修正的 R/S 统计量 Q 它具有如下形式:

$$\tilde{Q}_n = R_n / \sigma_n(q)。$$

式中,

$$\sigma_n(q) = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x}_n)^2 + \frac{2}{n} \sum_{j=1}^q w_j(q) \left[\sum_{i=j+1}^n (x_i - \bar{x}_n)(x_{ij} - \bar{x}_n) \right]} = \sqrt{\hat{\sigma}_x^2 + 2 \sum_{j=1}^q w_j(q) \hat{\gamma}_j}。$$

式中, $\hat{\sigma}_x^2$ 和 $\hat{\gamma}_j$ 是 $\{x_i\}$ 的样本方差和样本协方差, $w_j(j)$ 是 Bartlett 窗权重:

$$w_j(q) = 1 - \frac{j}{q+1}, \text{ 对于 } q < n。$$

在 $\{x_i\}$ 是非长记忆序列的零假设下, V 统计量 $V_n(q) = \tilde{Q}_n / \sqrt{n}$ 的极限分布为 $F(v) = 1 + 2 \sum_{k=1}^{\infty} (1 - 4k^2 v^2) e^{2(kv)^2}$ 。式中, v 为分位数; $F(v) = P(x < v)$ 。可得任意显著性水平下检验的临界值,通过检验 $V_n(q)$ 的显著性就能反映序列是否为长记忆时间序列。

2.3 长记忆时间序列模型

描述长记忆时间序列主要是分整自回归移动平均模型,即 ARFIMA(p, η, d, q) 模型与常用的随机线性时间序列模型 AR 模型、MA 模型、ARMA 模型、ARIMA 模型一般称为短记忆模型有本质性的区别^[13]。对于时间序列 $\{x_i\}$, ARFIMA(p, η, d, q) 模型的一般形式为

$$\Phi(L)(1-L)^\eta(1-L)^d x_i = \theta(L)\mu_i。$$

式中, L 为滞后算子,即 $L^k x_i = x_{i-k}$; 随机误差项 $\mu_1, \mu_2, \dots, \mu_i$ 相互独立,并且均服从正态分布 $N(0, \tau^{-1})$, $\tau > 0$ 为模型误差项 μ_i 的精度,即方差的 σ^2 的逆; $(1-L)^\eta$ 为整数差分算子; $(1-L)^d$ 为分数差分算子^[14],其具体表达式如下

$$(1-L)^d = 1 - dL + \frac{d(d-1)L^2}{2} - \frac{d(d-1)(d-2)L^3}{3!} + \dots = \sum_{k=0}^{\infty} (-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)} L^k。$$

式中, $\sum_{k=0}^{\infty} (-1)^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}$ 是一超几何函数,当 d 确定后此函数是 k 的函数,将其记为 $g(k)$ 。 d 为模型分数差分阶数, $-0.5 < d < 0.5$; $\Phi(L) = \Phi_1 L + \Phi_2 L^2 + \dots + \Phi_p L^p$ 为 p 阶平稳的自回归算子, $\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$ 为 q 阶可逆的移动平均算子,其所有特征根都在单位圆外,

则称 $\{x_i\}$ 满足分整自回归移动平均模型,记为 ARFIMA(p, η, d, q)。由以上分析可知 ARFIMA(p, η, d, q) 模型用 $p+q$ 个参数描述过程的短记忆性,以参数 d 反映过程的长记忆特征。当 $d = 0$ 时, ARFIMA(p, η, d, q) 模型就是 ARIMA(p, η, q) 模型;当 $d = \eta = q = 0$ 时, ARFIMA(p, η, d, q) 模型简化为 AR(p) 模型: $x_i = \Phi_1 x_{i-1} + \Phi_2 x_{i-2} + \dots + \Phi_p x_{i-p} + \mu_i$;当 $p = d = \eta = 0$ 时, ARFIMA(p, η, d, q) 模型即 MA(q) 模型: $x_i = \mu_i - \theta_1 \mu_{i-1} - \theta_2 \mu_{i-2} - \dots - \theta_q \mu_{i-q}$;当 $p = q = 0$ 且 $\eta = 0$ 时,则 ARFIMA($0, d, 0$) 模型即 FDN 模型: $(1-L)^d x_i = \mu_i$ 。因此, ARFIMA(p, η, d, q) 模型综合考虑到过程的长、短记忆特征,它既优于单独描述短记忆的 ARIMA 模型,又优于单独描述长记忆的 FDN 模型^[15]。

3 算例验证与分析

3.1 数据选取与数据挖掘

选取宁夏地区某风电基地号风电场的实测功率数据,选取时间点 1 至时间点 16 448 的风功率实测数据,每个时间点间隔 5 min。将数据分为两部分,前 1~16 400 个时间点的数据用来建立模型,后边 48 个时间点的数据用来验证预测的效果。本文通过前面所述的 EM 算法对历史数据进行处理,修正或补全这些不良数据,得到风功率时间序列 $\{y_i\}$ 。

3.2 长记忆性检验

分析时间序列长记忆性前需检验时间序列的平稳性,采用 ADF 单位根检验的方法,利用 eviews 软件对时间序列 $\{y_i\}$ 进行 ADF 单位根检验,发现原序列具有一个单位根,所以原时间序列是非平稳序列。为将原时间序列变为平稳时间序列,对原序列进行一阶差分发现,一阶差分后的序列的自相关函数随滞后阶数增加迅速下降到零,说明一阶差分序列为平稳序列。

从图 2 可以清楚地看到,风功率一阶差分时间序列 $\{y_i\}$ 绝对值的自相关函数呈缓慢下降趋势,滞后 80 阶后自相关函数才逐渐接近于零。这说明一阶差分后的风功率时间序列 $\{y_i\}$ 存在长记忆性。

3.3 建立模型

首先,通过修正 R/S 分析法计算 hurst 指数。由 hurst 指数可以确定 ARFIMA 模型的分数差分 d 值, $d = H - 0.5$,其中 H 为 hurst 指数。利用 MATLAB 软件编写修正 R/S 分析程序求得风功率时间序列的 hurst 指数 H 为 0.817 5。 $0.5 < H < 1$ 并且 H 较接近于 1 表明相关性较强,进一步表明了风功率时间序列 $\{y_i\}$ 存在长记忆性。分数差分系数 d 可

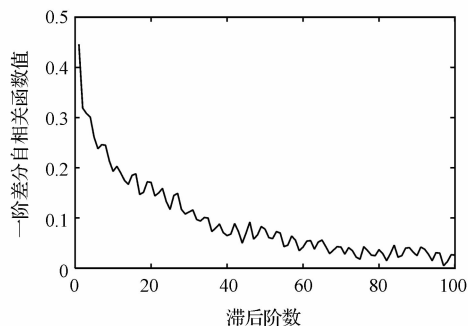


图2 一阶差分自相关函数图

Fig. 2 Auto-correlation function of first order difference

以由 $d = H - 0.5$ 的关系得到为 0.317 5。通过 eviews 软件对风功率时间序列 $\{y_t\}$ 进行分数差分, 除去其长记忆性, 得到短记忆的时间序列。然后通过 IBM SPSS Statistics 软件里的专家建模功能, 使用 ARMA 模型对分数差分后的短记忆时间序列进行分析建模, 得到参数 $p = 1, q = 6$ 。所以风功率时间序列 $\{y_t\}$ 长记忆模型为 ARFIMA(1, 1, 0.317 5, 6)。通过 eviews 得到模型的参数估计如表 1, 则模型 ARFIMA(1, 1, 0.405 6, 1) 表示为

$$(1 + 0.082\ 806)(1 - L)(1 - L)^{0.317\ 5}y_t = (1 - 0.428\ 061L)(1 - 0.122\ 610L)(1 - 0.047\ 214L)(1 - 0.024\ 542L)(1 - 0.010\ 491L)(1 + 0.017\ 889L)\mu_1。$$

表1 模型参数统计量

Table 1 Parameter statistics of the model

模型参数	系数值	标准差	T 统计量	P 值
C	0.000 476	0.016 306	4.577 021	0.000 0
AR(1)	-0.082 806	0.403 424	2.773 550	0.006 5
MA(1)	0.428 061	0.403 361	1.995 952	0.048 4
MA(2)	0.122 610	0.139 636	3.743 210	0.000 3
MA(3)	0.047 214	0.039 059	3.231 963	0.001 6
MA(4)	0.024 542	0.018 364	-2.307 063	0.022 9
MA(5)	0.010 491	0.012 745	3.365 392	0.001 1
MA(6)	-0.017 889	0.010 333	-2.307 063	0.022 9

3.4 模型评价

表 2 是模型的拟合统计表, 从中可以看出, 模型的拟合效果比较理想, 决定系数 (R^2) 达到了 0.980, 说明模型可解释原序列 98% 的信息, BIC 值也比较小, 并且 $\text{Sig} > 0.05$, Ljung-Box 统计量的值也是显著的, 这些都说明模型 ARFIMA(1, 1, 0.317 5, 6) 能很好地拟合风功率时间序列 $\{y_t\}$ 的数据。图 3 是风功率时间序列 $\{y_t\}$ 的观测值, 图 4 是 ARFIMA(1, 1, 0.317 5, 6) 模型的拟合值, 由图 5 可以直观地看出 ARFIMA(1, 1, 0.317 5, 6) 模型拟合效果比较理想。

表2 模型的拟合统计量

Table 2 Fitting statistics of the model

模型拟合统计量		均值
R^2		0.980
正态化的 BIC		0.676
统计量		22.623
Ljung-Box Q(18)	DF	15
	Sig	0.682

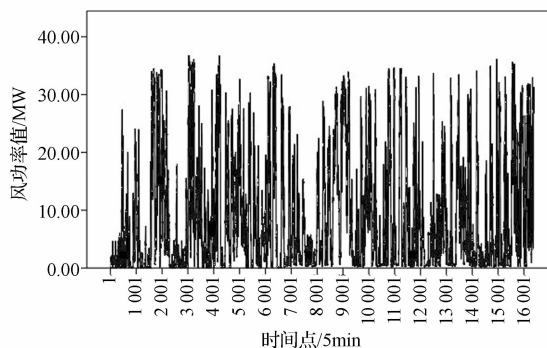
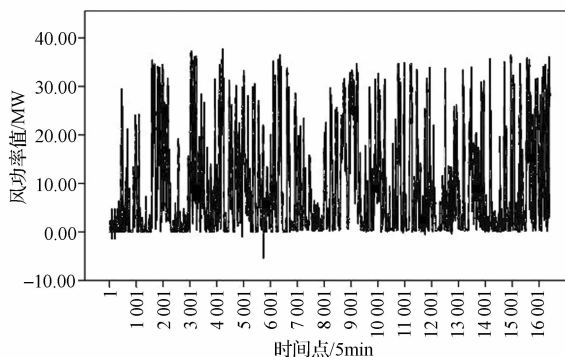
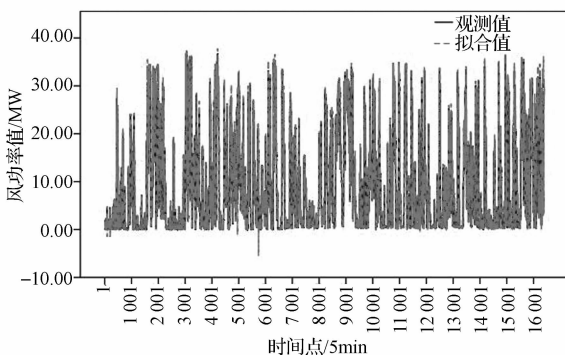
图3 风功率时间序列 $\{y_t\}$ 的观测值Fig. 3 Observation value of the $\{y_t\}$ 

图4 ARFIMA 模型的拟合值

Fig. 4 Fitting values of the ARFIMA model

图5 风功率时间序列 $\{y_t\}$ 的

观测值与拟合值对比图

Fig. 5 The comparison between the observation value and the fitting values

3.5 功率预测

通过 ARFIMA(1, 1, 0.317 5, 6) 模型对第二部

分 4 h 风功率数据进行预测,即后 48 个时间点。为了便于研究验证长记忆模型的预测效果,选择风功率预测中较常用的 ARIMA 模型对风功率时间序列进行预测并将其预测效果与 ARFIMA 模型做比较,采用与 ARFIMA 模型相同的数据估计 ARIMA 模型并预测第二部分的 48 个时间点的风电场输出功率,然后对两种模型的预测效果进行比较。图 6、图 7 分别为 ARIMA 模型和 ARFIMA 模型对 48 个时间点的风电场输出功率预测结果,通过图可以直观看出 ARFIMA 模型对数据预测的更加准确。为了更客观的评价模型预测性能,采用最大绝对百分误差 MRPE 和平均绝对百分比误差 MAPE 这两种统计量作为模型的灵敏度指标,对模型的预测结果进行评估。

$$MRPE = \max_{i=1}^N \left\{ \frac{|\hat{y}_i - y_i|}{y_i} \times 100\% \right\};$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|\hat{y}_i - y_i|}{y_i} \times 100\%。$$

式中 \hat{y}_i 是预测值, y_i 是实际值, N 为样本数量。通过计算 ARIMA 模型的最大百分误差为 14.247%, 平均绝对百分比误差为 8.673%; ARFIMA 模型的最大百分误差为 8.347%, 平均绝对百分比误差为 3.452%。由此可见 ARFIMA 模型的预测效果好于 ARIMA 模型。

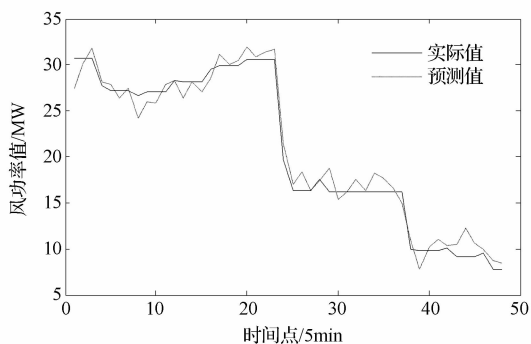


图 6 ARIMA 模型对 48 个时间点的预测结果

Fig. 6 Forecast results of the ARIMA model for 48 time points

4 结论

对基于时序长记忆模型的风电场短期功率预方法进行了阐述。根据传统时间序列 ARIMA 模型预测风速的不足,提出了基于 ARFIMA 长记忆模型的风电场功率预测;由于风电场实测风功率存在缺失和错误数据,采用 EM 算法处理不完全数据。通过实际风电场功率预测比较,计算预测结果统计量。算例结果表明长记忆模型的风电场短期功率预测提

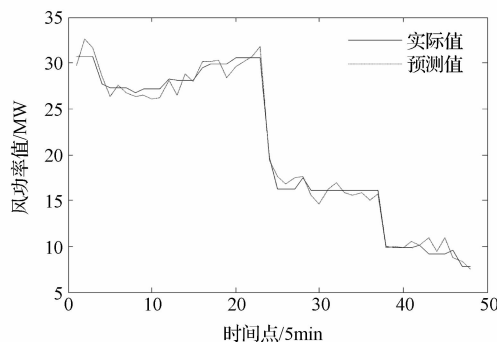


图 7 ARFIMA 模型对 48 个时间点的预测结果

Fig. 7 Forecast results of the ARFIMA model for 48 time points

高了风电场功率预测的准确性,具有实际意义。

参考文献

- 1 Kariniotakis G, Mayer D, Moussafir J. ANEMOS: development of a next generation wind power forecasting system for the large-scale integration of onshore & offshore wind farms. European Wind Energy Conference & Exhibition, Madrid, Spain, 2003
- 2 Giebel G, Landberg L, Kariniotakis G. State-of-the-art on methods and software tools for short-term, prediction of wind energy production. European Wind Energy Conference & Exhibition, Madrid, Spain, 2003
- 3 冯双磊, 王伟胜, 刘纯, 等. 风电场功率预测物理方法研究. 中国电机工程学报, 2010; 30(2): 1—6
Feng S L, Wang W S, Liu C, et al. Study on the physical approach to wind power prediction. Proceedings of the CSEE, 2010; 30(2): 1—6
- 4 陈新娜. 基于时间序列的超短期功率预测方法研究. 长沙: 长沙理工大学, 2013
Chen X N. Study on Wind Power Prediction Method Based on Time Series Analysis. Changsha: Changsha University of Science and Technology, 2013
- 5 王子赞, 纪志成. 基于灰色—辨识模型的风电功率短期预测. 电力系统保护与控制, 2013; 41(12): 79—85
Wang Z Y, Ji Z C. Grey-identification model based wind power generation short-term prediction. Power System Protection and Control, 2013; 41(12): 79—85
- 6 赵攀, 戴义平, 夏俊荣, 等. 卡尔曼滤波修正的风电场短期功率预测模型. 西安交通大学学报, 2011; 45(5): 47—51
Zhao P, Dai Y P, Xia J R, et al. A kalman filter based correction model for short-term wind power prediction. Journal of Xi'an Jiaotong University, 2011; 45(5): 47—51
- 7 范高峰, 王伟胜, 刘纯, 等. 基于神经网络的风电功率预测. 中国电机工程学报, 2008; 28(34): 118—123
Fan G F, Wang W S, Liu C, et al. Wind power prediction based on artificial neural network. Proceedings of the CSEE, 2008; 28(34): 118—123
- 8 吴桂联. 短期风电功率预测技术研究. 天津: 天津大学, 2012
Wu G L. Study of short-term wind power forecasting method. Tianjin: Tianjing University, 2012

- 9 刘 纯, 范高峰, 王伟胜, 等. 风电场输出功率的组合预测模型. 电网技术, 2009; 33(13):74—79
Liu C, Fan G F, Wang W S, *et al.* A combination forecasting model for wind farm output power. Power System Technology, 2009; 33(13):74—79
- 10 刘兴杰. 风电输出功率预测方法与系统. 保定:华北电力大学, 2011
Liu X J. Wind power prediction approach and system. Baoding: North China Electric Power University, 2011
- 11 Graham C G, Juan C A. Approximate EM algorithms for parameter and state estimation in nonlinear stochastic models. Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, 2005:12—15
- 12 朱慧明. 时间序列 ARFIMA 模型的贝叶斯预测分析. 统计与决策, 2006; 2(4):4—6
Zhu H M. Bayesian analysis of prediction base on ARFIMA model of the time series. Statistics & Decision, 2006; 2(4):4—6
- 13 Ellis C, Wilson P. Another look at the forecast performance of ARFIMA models. International Review of Financial Analysis, 2004; 13(1):63—8
- 14 金 秀, 姚 瑾, 庄新田. 基于分数阶差分的 ARFIMA 模型及预测效果研究. 数理统计与管理, 2007; 26(5):896—907
Jin X, Yao J, Zhuang X T. Study of ARFIMA model and its forecast performance based on fractional differencing. Journal of Applied of Statistics and Management, 2007; 26(5):896—907
- 15 Granger C W J, Ding Zhuanxin. Varieties of long memory models. Journal of Econometrics, 1996; 73:61—78

Wind Power Short-term Prediction Based on the Long-term Memory Model the Time Series Analysis Method

LU Jin-ling¹, WANG Yang¹,

YANG Yue¹, HE Zhen-min¹, XIANG Li², LI Xiao-ning²

(School of Electrical and Electronic Engineering, North China Electric Power University¹, Baoding 071003, P. R. China;

State Grid Ningxia Electric Power Company², Yinchuan 750001, P. R. China)

[**Abstract**] With the rapid growth of wind power, more wind power plants have been connected to the power grid, followed by the requirement of increasing accuracy of wind power prediction. Accurate prediction is indispensable for the better use of wind energy resources, and reducing the negative impact on the power grid. In order to improve the accuracy of prediction of wind power, expectation maximization algorithm (EM) was applied for processing the history data of wind power, filling missing values, and replace erroneous data. Modified rescaled range analysis was used to analyze the long memory of data. Then autoregressive integrated moving average model (ARIMA) was used. Based on the previous prediction data of the power, a prediction model equations can be made after the model parameters are estimated by Bayesian inference. This model can be used for the prediction of wind power.

[**Key words**] wind power prediction expectation maximization algorithm the time series analysis method
modified rescaled range analysis ARFIMA model