

基于ARIMA模型的自动站风速预测

蒋金良, 林广明

(华南理工大学电力学院, 广东 广州 510640)

摘要: 对风速预测进行了研究, 提出了基于ARIMA模型的风速预测模型, 为了检验ARIMA模型的有效性, 综合考虑可决系数和AIC(最小信息量)准则, 利用历史150天数据进行ARIMA建模, 对某自动站后一天的风速进行预测, 经过多次仿真计算, 结果表明该方法是有用的。

关键词: ARIMA模型; 风速; 预测

中图分类号: TM641 **文献标识码:** A

Automatic station wind speed forecasting based on ARIMA model

JIANG Jin-liang, LIN Guang-ming

(Electric Power College, South China University of Technology, Guangzhou Guangdong 510640, China)

Abstract: A wind speed forecasting model based on ARIMA (autoregressive integrated moving average) model is proposed in this paper. In order to test the effectiveness of the ARIMA model, the identified coefficient and AIC (Akaike information criterion) criterion are comprehensively dealt with to establish the ARIMA model. By using the historic data of the last 150 days, this model forecasts the next one day's wind speed in one automatic station. The validity of the proposed method is verified by the results of the simulation.

Key words: ARIMA model; wind speed; forecast

1 引言(Introduction)

随着我国能源结构的调整, 可再生能源特别是风能的开发利用已得到高度重视, 从技术成熟性和经济可行性看, 风电在可再生能源中具有很好的前景^[1]. 随着风能的加速发展, 越来越多的大型风电场将纳入统调电网, 风电在电网的比重越来越大, 风电的强随机性增大了电力调度的难度. 风速的准确预测可以减少电力系统运行成本和旋转备用, 提高风电穿透功率极限, 有利于调度部门及时调整计划, 从而减轻风电对电网的影响^[2]. 目前, 风速预测的方法有卡尔曼滤波法(Kalman filters)^[3]、时间序列法(time series method)^[4]、人工神经网络法(ANN)^[5]、模糊逻辑法(fuzzy logic)^[6]等. 卡尔曼滤波法是在假定噪声的统计特性已知的前提下得出, 事实上估计噪声的统计特性是该方法应用的难点所在; 人工神经网络方法则具有难以科学确定网络结构、学习速度慢、存在局部极小点、记忆具有不稳定性等固有缺陷; 模糊逻辑法预测学习能力较弱, 模糊系统的辨识还未形成完善的理论, 在预测系统中选择模糊系统的结构尚需作进一步的研究. 随机时间序列法建模所需信息少, 运算方便, 应用广泛. 它分为4种

不同的模型: 自回归(AR)模型, 滑动平均(MA)模型, 自回归-滑动平均(ARMA)模型; 累积式自回归-滑动平均(ARIMA)模型. 目前, 该方法使用最多的是ARIMA模型, 因为该模型引入了差分方法来处理非平稳数据.

本文以某自动站数据为例, 应用ARIMA模型对自动站风速进行预测, 在经过模型识别, 参数估计和模型定阶, 并通过检验后应用该模型进行风电场风速预测. 对自动站风速数据的预测值和实际值的比较说明该方法适合风速的预测.

2 预测模型建立(Establishment of forecasting model)

ARMA模型可以近似表示为^[7]

$$y_t = \varphi_1 y_{t-1} - \cdots - \varphi_p y_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \quad (1)$$

式中: $\varphi_1, \varphi_2, \cdots, \varphi_p$ 和 $\theta_1, \cdots, \theta_q$ 都是常数, a_t 是白噪声序列, 则称时间序列 $\{y_t\}$ 服从 (p, q) 阶自回归滑动平均模型, 简记为 ARMA (p, q) . 若 $\{y_t\}$ 是一个非平稳序列, 经 d 次差分后, $\{\nabla^d y_t\}$ 是一个平稳的 ARMA 序列, 则称

$$\varphi_p(B)\nabla^d y_t = \theta_q(B)a_t \quad (2)$$

为 (p, d, q) 阶的累积式自回归—滑动平均模型, 记作ARIMA (p, d, q) . 其中 y_t 是原始序列, B 为后移算子.

2.1 模型的识别与定阶(Model identification and order determination)

对时间序列数据平稳化处理后, 首先计算原始序列自相关函数ACF, 偏相关函数PACF, 对于平稳时间序列 y_t , 有

自协方差:

$$\hat{\gamma}_k = \frac{1}{N} \sum_{j=1}^{N-k} y_j y_{t+k} \quad (3)$$

自相关函数:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (4)$$

偏相关函数为

$$\begin{cases} \hat{\alpha}_{11} = \hat{\rho}_1, \\ \hat{\alpha}_{k+1,k+1} = (\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\rho}_{k+1-j} \hat{\alpha}_{kj}) \times \\ \quad (1 - \sum_{j=1}^k \hat{\rho}_j \hat{\alpha}_{kj})^{-1}, \\ \hat{\alpha}_{k+1,j} = \hat{\alpha}_{kj} - \hat{\alpha}_{k+1,k+1} \times \hat{\alpha}_{k,k-j+1}. \end{cases} \quad (5)$$

可通过判断 $\hat{\rho}_k, \hat{\alpha}_k$ 的截尾性来初步确定模型的阶数, 其结构判定的基本准则见表1所示.

表1 平稳时间序列模型结构识别准则
Table 1 Stationary time series model structure identification rule

类别	模型识别		
	AR(p)	MA(q)	ARMA(p, q)
自相关函数	拖尾	截尾	拖尾
偏相关函数	截尾	拖尾	拖尾

2.2 参数估计(Parameter estimation)

时间序列模型的参数辨识采用最小二乘估计, 即估计参数 $\varphi_1, \varphi_2, \dots, \varphi_p, \theta_1, \dots, \theta_q$ 使残差平方和

$$\sum_{t=1}^N a_t^2 = \sum_{t=1}^N (\theta_q^{-1}(B)\varphi_p(B)\nabla^d y_t)^2 \quad (6)$$

达到最小值.

2.3 模型检验(Model validation)

1) 残差序列白噪声检验: 残差序列白噪声检验就是要检验序列的自相关函数 $\rho_k(\hat{a})$ 与零无显著差异, 构造box-pierce的修正统计量 Q :

$$Q = (N - D - \max(p, q)) \sum_{k=1}^m \rho_k^2(\hat{a}). \quad (7)$$

式中: D 为差分后损失的数据个数, 给定显著性水平 α . 根据卡方表临界值 $\chi_{\alpha}^2(m-p-q)$ 判别准则, 若 $Q \leq \chi_{\alpha}^2(m-p-q)$, 则认为模型是合适的, 否则就

认为模型拟合不好.

2) AIC(最小信息量)准则: AIC准则一般形式:

$$AIC = -\frac{2L}{n} + \frac{2k}{n} = -\frac{2L}{n} + \frac{2(p+q+1)}{n}. \quad (8)$$

其中: L 是对数似然值, n 是观测值数目, k 是被估计的参数个数. AIC准则要求AIC的取值越小越好.

3) 可决系数(又称决定系数)是判断回归模型拟合程度优劣最常用的数量指标.

$$R^2 = \frac{\sum (\hat{y}_t - \bar{y})^2}{\sum (y_t - \bar{y})^2}. \quad (9)$$

式中的取值范围是 $[0, 1]$. R^2 值越大模型拟合程度愈高.

3 算例及结果分析(Computing example and result analysis)

本文以某自动站实测数据为例, 选取被预测日前150个点的风速数据建立ARIMA (p, d, q) 模型, 对2005年9月15日自动站每隔15 min风速 v 进行预测. 首先对原始数据线性插值, 得到当日96个点的数据(按全天1440 min计). 对2005年9月15日96个点每一点分别运用ARIMA (p, d, q) 模型进行预测, 综合考虑可决系数和AIC准则, 分别确定每一点的ARIMA (p, d, q) 模型. 得到原始序列 y_t 的预测值 $\hat{y}_t$, 结果如图1所示.

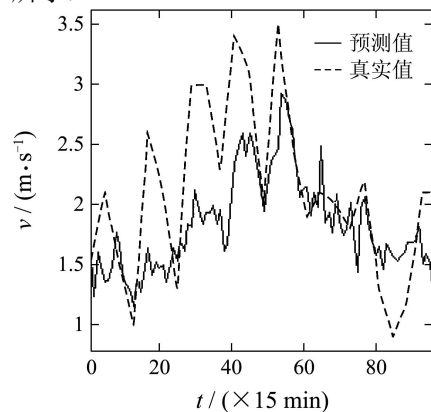


图1 预测值与实际值比较

Fig. 1 Comparison between forecasting value and real value

根据持续预测法原理, 与 $y_t, y_{t-1}, y_{t-2}, y_{t-3}$ 相关性很大, 因此作如下修正:

$$\hat{y}_t = \frac{(\hat{y}_t + y_{t-1} + y_{t-2} + y_{t-3})}{4}. \quad (10)$$

修正后预测结果如图2所示, 从图2可以看出, 经过修正后预测效果得到了一定的改善, 但预测在风速时间序列转折点处并不理想, 这是由于风速的随机性很强, 而ARIMA模型又很难对转折点发生的时刻进行预测. 误差评价指标定义包括:

1) 误差均方根:

$$N^{-1}E\|\hat{y} - y\|^2 = N^{-1} \sum_{i=0}^{N-1} E(\hat{y}(\frac{i}{N}) - y(\frac{i}{N}))^2. \quad (11)$$

2) 绝对误差:

$$E = |\hat{y} - y|. \quad (12)$$

求得误差均方根为0.5236 m/s, 图3为预测误差绝对值E.

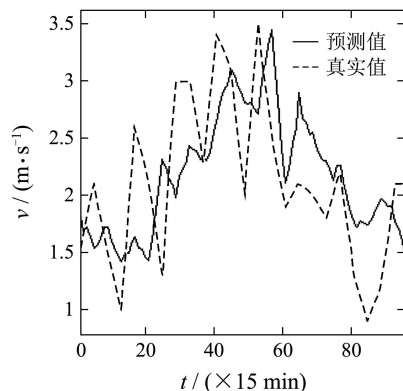


图2 经过修正后的预测值与实际值比较

Fig. 2 Comparison between forecasting value and real value after revised

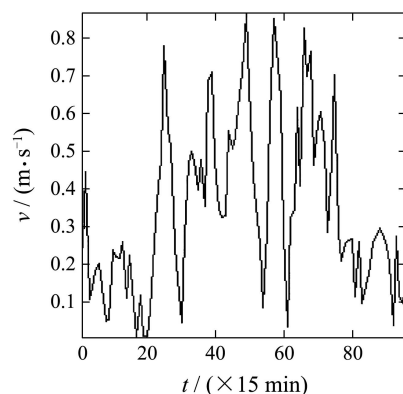


图3 预测绝对误差

Fig. 3 Forecasting absolute error

4 结论(Conclusion)

本文应用随机时间序列理论对经过插值后的风速数据进行建模, 得到各预测点的相对最佳ARIMA模型后对自动站某日风速进行预测, 取得了较好的效果, 预测结果表明, 基于改进的时间序列模型(ARIMA)适用于自动站风速预测. 但由于原始数据缺乏导致样本长度不足, 以及预测未考虑温度、湿度、气压等气象环境因素. 这些方面的欠缺都致使预测精度受到一定的限制.

参考文献(References):

- [1] 丁明, 张立军, 吴义纯. 基于时间序列分析的风电场风速预测模型[J]. 电力系统自动化设备, 2005, 25(8): 32-34.
(DING Ming, ZHANG Lijun, WU Yichun. Windspeed forecasting model on wind farm based on time series analysis[J]. *Electric Power Automation Equipment*, 2005, 25(8): 32-34.)
- [2] 杨秀媛, 肖洋, 陈树勇. 风电场风速和发电功率预测研究[J]. 中国机电工程学报, 2005, 25(11): 1-5.
(YANG Xiuyuan, XIAO Yang, CHEN Shuyong. Wind speed and generated power forecasting in wind farm[J]. *Proceedings of the Chinese Society for Electrical Engineering*, 2005, 25(11): 1-5.)
- [3] BOSSANYI E A. Short-term wind prediction using Kalman[J]. *Wind Engineering*, 1985, 9(1): 1-8.
- [4] WANG P, BILLINTON R. Time-sequential simulation technique for rural distribution system reliability cost/worth evaluation including wind generation as alternative supply[J]. *IEEE Proceedings: Generation, Transmission and Distribution*, 2001, 148(4): 355-360.
- [5] LI Shuhui, WUNSCH D C, O'HAIR E A, et al. Using neural networks to estimate wind turbine power generation[J]. *IEEE Transactions on Energy Conversion*, 2001, 16(3): 977.
- [6] 张昊, 吴捷, 郁滨. 电力负荷的模糊预测方法[J]. 北京: 电力系统自动化, 1997, 22(7): 14-17.
(ZHANG Hao, WU Jie, YU Bin. Electric load forecasting using fuzzy method[J]. *China Electric Power Press*, 1997, 22(7): 14-17.)
- [7] 牛东晓, 曹树华. 电力系统负荷预测技术及其应用[M]. 北京: 中国电力出版社, 1998.
(NIU Dongxiao, CAO Shuhua. *Power System Load Forecasting Technology and Its Application*[M]. Beijing: China Electric Power Press, 1998.)