

# 基于时间序列的电力负荷数据分析

袁硕, 陈礼定, 孙国鹏 (东南大学数学系, 南京, 211189) \*

2016年5月发表于《应用数学进展》汉斯出版社。

指导老师: 林金官

**摘要:** 时间序列分析方法是电力负荷预测领域的重要工具之一, 它主要通过建立相关模型描述历史数据随时间动态变化的规律以预测未来的值。本文采用温特线性与指数平滑法和季节乘积 ARIMA 模型对电力负荷实测数据进行建模, 然后分别使用平均相对误差 MAPE 衡量预测精度。研究结果表明: 两种方法均有较高的拟合与预测精度。

**关键词:** 电力负荷; ARIMA; 时间序列; 预测; 指数平滑

**Abstract:** Time series analysis method is one of the important tools in the field of power load forecasting, it mainly describes the law of the historical data dynamic change over time to predict the future value by establishing a relevant model. In this paper, Winter's exponential smoothing method and seasonal ARIMA product model are applied to model estimating on the the power load data, and use the Mean Absolute Percentage Error (MAPE) to measure accuracy. The results prove that both of them have high fitting and forecasting precision.

**Key words:** electric load; ARIMA; time series; forecasting; exponential smoothing

## 引言

电力负荷预测是能量管理系统的重要组成部分, 作为电力行业中的一个热点问题一直倍受关注, 已经成为实现电力系统管理的重要内容之一。电力负荷预测是根据电力负荷的历史数据分析其变化规律, 综合考虑影响负荷变化的因素, 在某些要求下, 使用合适的预测模型和方法, 以未来的经济形势、气候条件等因素为依据, 预测未来某时段电力负荷数据的过程, 其主要目标是预测未来不同时间和空间上的电力负荷, 将预测结果加以应用, 以便在未来时刻可以合理地规划和调度电力负荷, 既不产生短缺也不造成浪费。

电力市场的改革使负荷预测的重要性凸显, 各电力公司对电力负荷预测精度的要求也越来越高。由于影响因素的不确定性以及负荷数据的复杂性的存在, 负荷的预测总会存在偏差。随着现代科学技术和学术理论的进步, 学科交叉日益频繁, 理论研究逐步深入, 以专家系统理论、神经网络等为代表的新兴交叉学科理论的出现, 也为负荷预测的快速发展提供了强大的理论基础。如果选择预测模型的标准是追求预测精度的极大化, 则最好选择时间序列模型。

在实际情况下, 绝大部分序列都是非平稳的, 因而对非平稳序列的分析更具有普适性, Gramer 定理说明: 任何一个序列的波动都可以视为同时受到了确定性影响和随机性影响的共同作用<sup>[2]</sup>。一般来讲, 非平稳时间序列的分析方法分为确定性时间序列分析和随机性时间序列分析<sup>[4]</sup>。社会经济现象存在着不确定性, 无论预测方法是否得当, 预测模型对历史数据的拟合程度多高, 观察值与预测值之间仍会存在偏差。使得误差降到最低是研究预测方法的重要任务。

## 一. 模型介绍

### (一) 温特线性与季节指数平滑法

温特方法有三个基础公式和一个预测方程组成<sup>[5, 12]</sup>, 这些公式的推导过程在此不详述。总体平滑公式:

---

\*作者简介: 袁硕 (1993-) 男 河北省衡水市人 应用统计学专业

$$S_t = \alpha \frac{x_t}{I_t} + (1 - \alpha)(S_{t-1} + b_{t-1}),$$

趋势平滑公式:

$$b_t = \gamma(S_t - S_{t-1}) + (1 - \gamma)b_{t-1},$$

季节的平滑公式:

$$I_t = \beta \frac{x_t}{S_t} + (1 - \beta)I_{t-L},$$

预测方程:

$$F_{t+m} = (S_t + b_t m)I_{t+m-L},$$

式中,  $L$ 是季节性的长度。 $I$ 是季节指数, 由于时间序列的原始数据 $x_t$ 里包含了季节性、趋势性和随机性,

## (二) ARIMA 模型

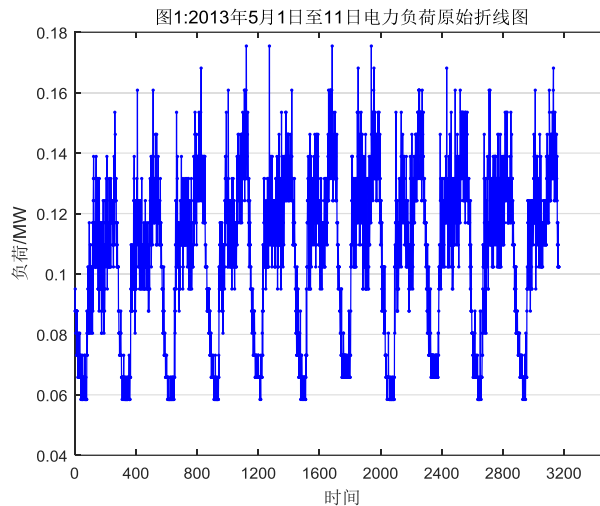
在实际情况下, 有时需要通过差分使得数据变平稳, 若 $Y_t = (1 - B)^d X_t$ 是一个 ARMA ( $p, q$ ) 的序列, 则称 $\{X_t\}$ 是 ARIMA ( $p, d, q$ ) 序列。

针对具有季节性的时间序列, 周期记为 $s$ , 将每个周期段内的同一时刻点位置的数据单独看成一个时间序列, 总共产生 $s$ 个子序列。若每个子序列都可以用 ARMA ( $p, q$ ) 模型拟合, 则有以下模型:  $A(B^s)X_t = B(B^s)\varepsilon_t$ ,  $t = 1, 2, 3 \dots$ , 其中 $\{\varepsilon_t\}$ 不可假设为白噪声序列, 故不能否认各子列之间的相关性。但是在每个子列里面可以假设为白噪声, 为了解决这个问题, 对 $\{\varepsilon_t\}$ 拟合 ARMA ( $p, q$ ) 模型:  $\alpha(B)\varepsilon_t = \beta(B)a_t$ , 其中,  $\{a_t\}$ 为白噪声 $WN(0, \sigma^2)$ , 可得模型:  $\alpha(B)A(B^s)X_t = \beta(B)B(B^s)a_t$ , 这就是季节乘积ARMA( $p, q$ )  $\times$  ( $P, Q$ )模型。

若每一个子列需要经过差分 $Y_t = (1 - B)^d (1 - B^s)^D X_t$ , 之后才能建立 ARMA 模型, 故模型公式变成  $\alpha(B)A(B^s)(1 - B)^d (1 - B^s)^D X_t = \beta(B)B(B^s)a_t$ , 该模型称为季节乘积ARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ )<sub>288</sub>模型。

## 二. 电力负荷数据案例分析

本文采用 2013 年 5 月 1 日到 11 日江苏省南京市新街口金鹰国际购物中心某层用电功率的实测数据 (单位: 兆瓦) 作为电力负荷建模预测的原始数据, 由南京市供电局提供。用 SAS 画出原始数据图如图 1 所示。以 1 日至 10 日的数据进行建模, 记为 $\{X_t\}$ , 将 2013 年 5 月 11 日的数据用以检验模型的预测精度。



### (一) 基于温特线性与季节性指数平滑法的探究性分析

使用温特线性方法时,  $S_t$ 的初值为:  $S_{L+1} = x_{L+1}$ , 式中 $L$ 为季节长度。季节指数初始值 $I$ 为:  $I_1 = \frac{\bar{x}_1}{\bar{x}}, I_2 =$

$$\frac{\bar{x}_2}{\bar{x}}, \dots, I_L = \frac{\bar{x}_L}{\bar{x}}, \bar{x}_L \text{ 为不同周期同一时刻的平均值, } \bar{x} \text{ 为总体平均值。} b_{L+1} \text{ 的初始值为: } b_{L+1} = \frac{\sum_{i=1}^L x_{L+i} - \sum_{j=1}^L x_j}{L^2}。$$

采用穷举法选取平滑系数, 即考虑三个系数的所有组合, 选取 MSE 最小的组合值。通过 MATLAB 编程

以[0.01, 0.99]为区间, 步长为 0.01 进行反复计算, 选择 MSE 最小的组合作为最佳平滑系数<sup>[5]</sup>。得到最优平滑系数 $\alpha = 0.12$ ,  $\beta = 0.01$ ,  $\gamma = 0.01$ ,  $\min\text{MSE} = 9.7152\text{E} - 5$ , 95%的数据点相对误差为 0.006~17.2 (%) , 使用该模型对 11 日的负荷值进行预测, 模型的 MAPE=7.4873<10, 认为其预测精度较高。

## (二) 基于 ARMA 模型的探究性分析

观察原始数据及折线图的变化趋势, 可以认为该序列没有缺失值和异常值, 且数据围绕一条水平线上下波动, 初步判断为平稳序列, 其自相关系数呈现出围绕零轴正弦波动的现象, 这也说明序列具有周期性。

考虑对原数据先进行季节差分(季节周期为 288), 季节差分后的序列记为 $\{\nabla_{288}X_t\}$ , 由于自相关系数迅速衰减为零, 在滞后 288 处再次出现显著大于零的值, 偏自相关系数具有截尾特征。用 Ljung-Box 统计量来检验序列 $\{\nabla_{288}X_t\}$ 是否为白噪声序列<sup>[1]</sup>, P 值小于 0.05, 序列显然是非白噪声。采用 ADF 统计量检验是否含有单位根<sup>[6]</sup>, P 值都小于 0.05, 认为序列不存在单位根, 故继续以 ARMA 模型进行拟合。进一步采用最小信息准则<sup>[1, 6, 9, 10, 13]</sup>选择模型, 选择 BIC 最小(-8.52573)的 ARMA(1, 1)模型。采用最小二乘法<sup>[1, 9]</sup>进行参数估计, 得到拟合模型公式如下:

$$(1 - 0.94112B)(1 - B^{288})X_t = (1 + 0.84690B)\varepsilon_t$$

表 1: 残差检验

滞后数	$\chi^2$ 值	P 值
6	4.27	0.3704
12	8.07	0.6217

为了检验该模型是否适合于原序列, 采用 Ljung-Box 统计量<sup>[1]</sup>对残差进行白噪声检验, 从表 1 的检验结果看出, 延迟 6、12 期的 P 值均大于 0.05, 残差序列显著为白噪声, 认为残差序列之间不存在显著的自相关性, 序列中的相关信息已经被充分提取。利用建立的最佳 ARMA 模型, 比较拟合值与真实值, 发现只有极少数点处于 95%置信区间外, 95%的数据点的相对误差为 0.02~25.7 (%)。对比 11 日预测值与真实值, 计算 MAPE=10.5925, 预测效果不够理想。可能是趋势性的存在所导致的, 继续采用季节乘积 ARIMA 模型进行建模。

## (三) 基于乘积季节 ARIMA 模型的探究性分析

现对原序列进行一阶差分再做 288 步季节差分, 希望能消除季节效应和趋势效应, 差分后的序列记作 $\{\nabla_{288}\nabla x_t\}$ , 采用 ADF 单位根检验, 检验结果显示序列不存在单位根。ACF 图只有滞后一阶时超出了 2 倍标准差范围, 可以判断为 1 阶截尾的, 而偏自相关系数下降较缓慢, 初步判断为 MA(1)模型<sup>[1, 6, 10]</sup>, 尝试拟合季节乘积 ARIMA 模型<sup>[8, 11]</sup>, 模型拟合的 AIC、SBC 及残差白噪声检验结果见表 2。

表 2: 比较各季节乘积模型

季节乘积模型	AIC	SBC	残差白噪声检验
SARIMA(0,1,1) × (0,1,0) <sub>288</sub>	-14710.1	-14698.4	**
SARIMA(0,1,1) × (0,1,1) <sub>288</sub>	-15645.7	-15628.1	*
SARIMA(0,1,1) × (0,1,2) <sub>288</sub>	-15643.7	-15620.3	*
SARIMA(0,1,1) × (1,1,0) <sub>288</sub>	-15296.6	-15279	**
SARIMA(0,1,1) × (1,1,1) <sub>288</sub>	-15643.7	-15620.3	*
SARIMA(0,1,1) × (1,1,2) <sub>288</sub>	-15643.3	-15614	---
SARIMA(0,1,1) × (2,1,0) <sub>288</sub>	-15296.6	-15279	**
SARIMA(0,1,1) × (2,1,1) <sub>288</sub>	-15643.7	-15620.3	*
SARIMA(0,1,1) × (2,1,2) <sub>288</sub>	-15643.3	-15614	---

从表中看出(0,1,1) × (0,1,1)<sub>288</sub>的 AIC 和 SBC 均最小, 所以选择该模型为最佳的拟合模型, 采用条件最小二乘法得到参数估计并剔除掉常数项, 得到拟合模型公式为:

$$(1 - B)(1 - B^{288})X_t = (1 - 0.88665B)(1 - 0.70596B^{288})\varepsilon_t$$

采用已得季节乘积 ARIMA 模型对 5 月 1 日至 10 日的数据进行拟合, 95%的数据点的相对误差为 0~20.6

(%)，对比来看，进一步用该模型对 11 日的负荷值进行预测，计算 MAPE=7.4897，认为模型是合理的。

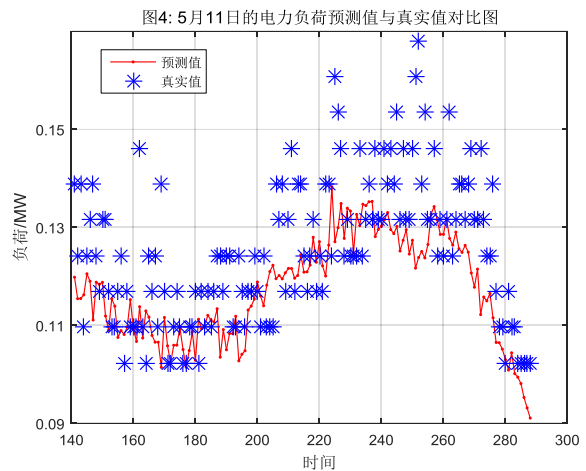
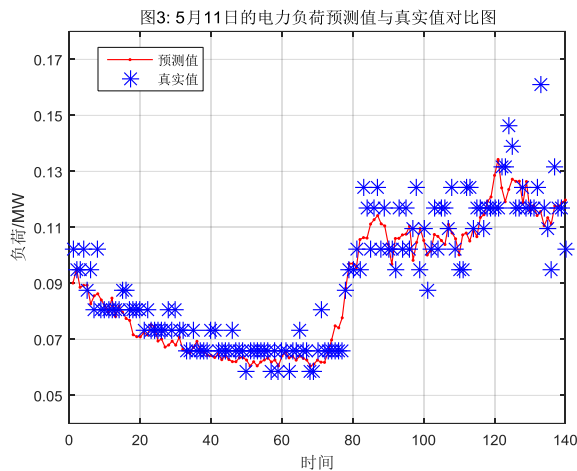
### 三. 结论

预测精度是衡量预测方法是否适合预测对象的重要指标。本文采用的预测精度为平均相对误差：

$$Mape = \frac{1}{n} \sum_{t=1}^n \left| \frac{(x_t - \hat{x}_t)}{x_t} \right|$$

如果 MAPE 越小，说明模型的预测精度越高。若 MAPE<10，说明模型的预测结果比较好。

三个模型的 MAPE 分别为：7.4873、10.5925、7.4897，所以温特线性方法预测效果最好，其预测图与真实值的对比图见图 3 和图 4 两幅图(x: 预测值, y: 真实值)，预测是合理的，可以将温特线性方法用于该样本数据的实际预测。从图中看，前段为深夜时分，用电量较低，走势较平稳，偏差较小；后段出现较大波动，这可能是由于白天用电较多，设备的启用使用电功率产生不规则的波动，整体看预测值比实测值偏小一些。这在一定程度上说明温特线性方法对趋势的处理需要改进。猜测其原因，一方面，其线性处理不适合于数据；另一方面，短期内外界因素导致的波动对实际数据产生影响<sup>[7]</sup>。温特线性方法操作简单，精确度高，该方法在今后的一般线性预测中将会发挥很重要的作用，但它所暴露的缺点更需进一步完善。



### 参考文献

- [1] 王燕. 应用时间序列分析[M]. 中国人民大学出版社, 2012:41-110.
- [2] 邓祖新. 数据分析方法和 SAS 系统[M]. 上海财经大学出版社, 2006:390-443
- [3] 牛东晓. 电力负荷预测技术及其应用[M]. 中国电力出版社, 1998:40-54. 59-88. 95-134. 185-101
- [4] 吴仁群. 经济预测与决策[M]. 中国人民大学出版社, 2015:34-36. 92-140
- [5] 梁慧玲, 赵嘉阳, 林玉蕊. 温特线性与季节指数平滑法在气温中的应用[J]. 科技和产业, 2015(9):145-148.
- [6] 孙国红, 沈跃, 徐应明, 等. 基于 Box-Jenkins 方法的黄河水质时间序列分析与预测[J]. 农业环境科学学报, 2011, 30(9):1888-1895.
- [7] 叶舟, 陈康民. 温斯特线性与季节性指数平滑法在电力负荷预测中的应用及改进[J]. 电力建设, 2000, 21(8):13-16.
- [8] 邓盛川, 于德亮, 齐维贵. 基于乘积季节 ARIMA 模型的供热负荷预报[J]. 沈阳工业大学学报, 2011(3):321-325.
- [9] Brockwell P J, Davis R A. 时间序列: 理论与方法[M]. 田铮译. 北京: 高等教育出版社. 2001. 200-202. 214-254.
- [10] Kumar S V, Vanajakshi L. Short-term traffic flow prediction using seasonal ARIMA model with limited input data[J]. European Transport Research Review, 2015, 7(3):1-9.
- [11] Eni D, Adeyeye F J. Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria[J]. Journal of Geoscience & Environment Protection, 2015(3):91-98.
- [12] Hanzák T. Holt-Winters method with general seasonality[J]. Kybernetika -Praha-, 2012, 1(1):1-15.
- [13] Box G E P. 著. 时间序列分析: 预测与控制[M]. 王成璋等译. 北京: 机械工业出版社. 2011:116-125.