

基于K-L信息量和ARIMA误差修正的月度电量预测

陈明帆¹, 宁光涛¹, 李琳玮¹, 何礼鹏¹, 刘丽新²

(1. 海南电网有限公司, 海口 570204; 2. 北京清软创新科技股份有限公司, 北京 100085)

Monthly electricity forecasting based on K-L information and ARIMA error correction

CHEN Mingfan¹, NING Guangtao¹, LI Linwei¹, HE Lipeng¹, LIU Lixin²

(1. Hainan Power Grid Co., Ltd., Haikou 570204, China;

2. Beijing Tsingsoft Technology Co., Ltd., Beijing 100085, China)

摘要:为考虑多因素对电量的影响及提高月电量预测准确率,提出基于K-L信息量法和ARIMA误差修正的月度电量预测方法。在筛选相关性强的指标基础上,利用相关分析法对影响指标与电量进行回归建模,计算拟合误差并构建新的非平稳时间序列,结合ARIMA模型对此序列进行修正,进而获得准确度性更佳的月度电量预测值,具有较高的应用价值。

关键词:K-L信息量;ARIMA;月度电量预测;误差修正

Abstract: In order to consider the influence of multiple factors on electricity and improve the accuracy of monthly electricity forecasting. A monthly electricity forecasting method is proposed based on the K-L information method and ARIMA error correction. Based on screening relevant indicators, the correlation analysis method is used to make regression modeling on the influence indicators and electricity, calculate of fitting errors, and construct a new non-stationary time series, combined with the ARIMA model to modify this series. Monthly electricity predictions obtained with better accuracy, have higher application value.

Key words: K-L information method; ARIMA; monthly electricity consumption forecast; error correction

0 引言

随着电力体制改革的不断深入,城乡居民暂不参与电力市场,需要电网公司对其承担保底供电责任。所以电网公司越来越重视城乡居民月度用电量的预测,以便于合理安排月度计划、更好履行保底供电责任。

城乡居民用电特性受气象、经济等因素影响显著。随着经济快速发展,城乡居民的生活用电量迅速增加;由于家用电器的普及率越来越高,特别是空调负荷已经成为城乡居民用电的重要组成部分,并且从根本上改变了居民生活用电特性。若要提高城乡居民月度电量的预测精度,不能单纯依靠历史序列进行外推,需要结合经济、气象等多因素的影响,筛选出相关性强的指标,再利用相关因素分析法建立回归模型进行预测。传统只依赖相关因素分析法建立回归模型的预测误差往往较大,如文献[1]通过基于粗糙集理论找到相关因素后进行回归分析的预测方法。自回归差分移动平均模型(autoregressive integrated moving average model, ARIMA)对预测非平

稳时间序列具有很大优势,文献[2]在ARIMA模型的基础上先消除季节周期分量影响再对月度电量进行预测;文献[3]提出一种结合X12季节分解方法的ARIMA月度电量预测方法。但这些研究多集中在应用ARIMA模型进行月度电量预测,忽视了ARIMA模型是可以对预测误差进行修正的价值。

本文在考虑相关因素建立回归模型进行预测的基础上,充分利用ARIMA模型对误差修正的优势,创新性地提出基于K-L信息量法和ARIMA误差修正的月度电量预测法。在筛选相关性强的指标基础上,利用相关分析法对影响因素与电量进行回归建模,计算拟合误差并构建新的非平稳时间序列,结合ARIMA模型对此序列进行修正,提升月电量的预测准确率。

1 城乡居民用电特性分析

城乡居民用电呈现出变化趋势相对稳定、空调用电明显的特征。如图1所示某地城乡居民历年月电量,一般表现为6月份开始电量快速上升,8月达到顶峰,9月开始回落,每年11月至来年5月是全年用电低谷。

收稿日期:2020-09-03;修回日期:2020-12-15

基金项目:海南电网有限责任公司科技项目(070000 KK52160001)

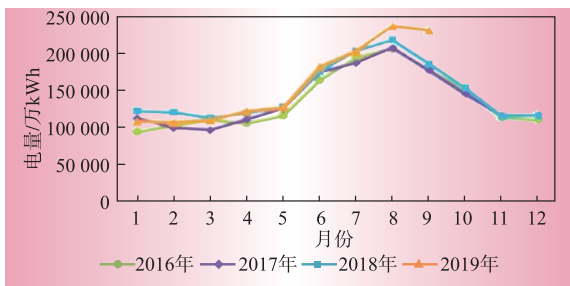


图1 2016—2019年城乡居民生活用电情况

Fig. 1 Domestic electricity consumption of urban and rural residents in 2016—2019

2 相关因素分析预测法

根据城乡居民用电特性,如历年变化趋势相对稳定,空调用电特性明显等特点,采用K-L信息量法从多种影响因素中筛选出最相关因素建立回归模型进行初步预测。

2.1 基于K-L信息量的相关因素筛选法

K-L信息量法是对指标总体概率分布的拟合程度进行评价的一种方法^[4]。对于带有随机性的现象,可以认为它是服从某一概率分布的随机变量的一些实际值。如果已知真正的概率分布,通过估计选择的概率模型与实际的概率分布相近的程度来判断模型的好坏。具体计算公式如下。

(1) 设一个基准指标 $y = \{y_1, y_2, \dots, y_n\}$, 该指标的概率分布列为 $p = \{p_1, p_2, \dots, p_n\}$, 则标准化处理后的基准变量序列 p_i 为

$$p_i = y_i / \sum_{j=1}^n y_j \quad t = 1, 2, \dots, n \quad (1)$$

式中: y_j 为基准随机变量; p_i 为对 y_i 进行标准化处理后的序列。

(2) 设另一个指标为被选指标为 $x = \{x_1, x_2, \dots, x_n\}$, 该备选指标概率分布序列为 $q = \{q_1, q_2, \dots, q_n\}$, 则标准化处理后的指标序列 q_i 为

$$q_i = x_i / \sum_{j=1}^n x_j \quad t = 1, 2, \dots, n \quad (2)$$

式中: x_j 为备选指标随机变量; q_i 为对 x_i 进行标准化处理后的序列。

(3) 定义分布列 q 关于分布列 p 的K-L信息量 k_i 为

$$k_i = \sum_{i=1}^n p_i \ln(p_i/q_i) \quad (3)$$

以某地城乡居民月电量为历史数据,选取尽可能丰富的外部影响指标,应用K-L信息量法计算出的各个指标的K-L信息量,从表1可以看出,与城乡居民月度电量最相关的影响因素为最高气温平均值。

表1 月度电量与各个指标的K-L信息量

Table 1 Monthly electric quantity and K-L information of each indicator

外部影响指标	K-L信息量	外部影响指标	K-L信息量
城市居民消费价格总指数	419	降雨量	77
商品零售价格总指数	85	家用电器产量	613
商品进出口总值	181	房间空气调节器产量	63
商品进口总值	69	电子计算机产量	111
商品出口总值	90	彩色电视机产量	85
第三产业固定资产投资额	552	房地产开发施工面积	77
最高气温	162	房地产开发竣工面积	97
最低气温	98	社会消费品零售总额	346
最高气温平均值	48		

2.2 基于相关因素的回归模型

相比于时间序列模型,相关因素回归模型不再只针对历史电量数据进行分析,而是把时间项换为相关因素,建立电力需求与一个或多个相关变量的回归方程式。回归分析法对相关因素的选择要求比较高,使用的相关因素必须与电力需求之间具有长期的稳定的相关关系。

常用的回归模型有线性模型、指数模型、对数模型、幂函数模型、抛物线模型、多元回归模型。

根据城乡居民的月度电量与当地月最高气温平均值之间的拟合误差最小建立回归方程,得到历史月拟合误差时间序列和预测月预测误差。但只依赖相关因素分析法建立回归模型的预测误差往往较大,需要对预测误差进行修正后再预测才能更好地提高预测精度。

3 基于ARIMA模型的误差修正

ARIMA模型对预测非平稳时间序列具有很大优势。利用上述回归模型得到的历史月电量误差值刚好构成一个非平稳时间序列,所以可以采用ARIMA模型进行修正。

3.1 ARIMA模型原理

ARIMA模型是采用差分运算或对数运算后将非平稳时间序列转换成对平稳时间序列后,利用自相关和偏自相关函数建立的一种时间序列预测方法^[3],该模型已成熟应用在时间序列分析领域。ARMA是对平稳时间序列的预测模型,ARIMA是对非平稳时间序列先进行平稳化检验与处理后再利用ARMA模型进行预测的模型。

针对非平稳时间序列 X_t 的ARIMA(m, d, n)模

型表达式为

$$X_t = \left\{ \begin{aligned} &(1-B)^d X_t \\ &\varphi_1 X_{t-1} + \cdots + \varphi_m X_{t-m} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \cdots - \omega_n \varepsilon_{t-n} \end{aligned} \right\} \quad (4)$$

式中: d 为差分次数; B 为滞后算子; X_t 为经过 d 次差分后的平稳时间序列; m 和 n 分别为自回归阶数和滑动平均阶数; $\varphi_1, \varphi_2, \cdots, \varphi_n$ 分别为自回归系数; $\omega_1, \omega_2, \cdots, \omega_n$ 分别为滑动平均系数; $\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_{t-n}$ 为零均值白噪声序列。

3.2 ARIMA 模型的预测步骤

ARIMA 模型预测步骤如下:

(1) 首先,获取时间序列,检查该序列数据是否属于平稳时间序列。常用的方法是时序图法,如果不平稳,需要对非平稳时间序列进行 d 次差分直到平稳。一般 $d \leq 2$ 。

(2) 其次,用 Box-Jenkins 法进行模型识别,即计算自相关函数和偏自相关函数,利用两个函数的截尾和拖尾特性判断模型阶数 p 和 q 。

(3) 再次,用最小二乘法计算自回归系数和滑动平均系数。

(4) 最后,进行模型验证。主要检查 ε_t 是否为白噪声序列,检查方法是利用 Barlett 定理构造检验统计量^[1]。若为白噪声序列,模型通过,若不是白噪声序列,需要重新进行模式识别。

4 基于 K-L 信息量和 ARIMA 误差修正的月度电量预测

设历史月度电量序列为 $y = \{y_1, y_2, \cdots, y_t\}$ 。结合 K-L 信息量和 ARIMA 误差修正的月度电量预测流程步骤如下,流程如图 2 所示。

(1) 采用 K-L 信息量筛选指标:首先采用 K-L 信息量计算月度电量与多指标的分布函数,通过 K-L 信息量的计算公式得到各个指标与月度电量的 K-L 信息量。选取 K-L 信息量在 50 以内的一个或多个影响指标作为月度电量的相关因素建立回归方程。

(2) 相关因素分析法建立回归方程:若为一个影响指标,则建立一元回归方程;若为多个影响指标,则建立多元回归方程。

(3) 获取误差时间序列:通过回归方程,计算拟合曲线与历史真实值之间的误差值。构建拟合误差时间序列 $D = \{D_1, D_2, \cdots, D_t\}$ 。

(4) 利用 ARIMA 模型进行误差修正^[5]:运用 R 语言中的 ARIMA 工具包^[6]对误差时间序列 D_t 直接建模,通过 ARIMA 预测模型实现对误差值的修正。

(5) 基于修正误差预测月度电量:将相关因素

分析法的预测值与修正后的预测误差值进行加减预算,得到修正后的月度电量预测值。

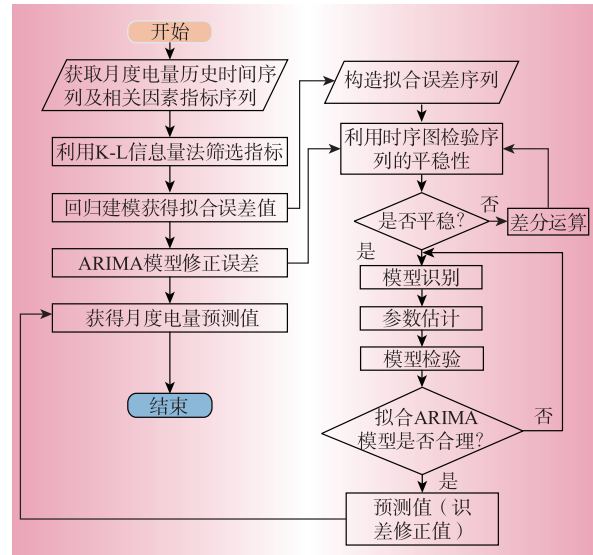


图 2 基于 K-L 信息量和 ARIMA 误差修正的预测流程

Fig. 2 Prediction process based on K-L information and ARIMA error correction

5 算例验证

为验证本文所提出的基于 K-L 信息量和 ARIMA 误差修正的月度电量预测模型的可行性,选取某地 2016 年 1 月至 2019 年 6 月的城乡居民月度电量作为历史数据序列,以 2019 年 7、8、9 月作为预测目标,预测具体步骤如下。

(1) 为对比多元回归模型与一元回归模型的预测精度,结合 K-L 信息量计算结果,首先选取对电量影响较大的前四种影响因素,选择多元回归模型,建立月度电量与最高气温平均值 a 、房间空气调节器产量 b 、降雨量 c 、房地产开发施工面积 d 的月度电量值 y 多元回归模型为

$$y = 0.80a + 0.14b + 0.03c + 0.03d \quad (5)$$

(2) 在 K-L 信息量得出温度对电量影响最大基础上,选择一元回归方程的抛物线模型,建立月度电量 y 与最高气温平均值 x 的回归模型为

$$y = 651.9x^2 - 28583x + 416774 \quad (6)$$

(3) 利用回归模型获取误差时间序列,得到修正前的预测误差精度如表 2 所示。

表 2 修正前拟合误差

Table 2 Fitting error before correction

	%		
误差类型	7月	8月	9月
对多元回归方程的拟合误差	9.01	7.08	6.32
对一元回归的拟合误差	8.21	6.78	5.74

(4) 利用 ARIMA 模型修正误差

选取历史误差值为历史值构建训练样本集, 2019年7月至9月为预测月, 结合R语言的ARIMA工具包, 将上述误差时间序列逐次进行平稳化处理、模型识别、参数估计、模型检验等, 最后输出误差修正值。修正后的预测误差如表3所示。

表3 修正后预测误差

Table 3 Fitting error after correction

误差类型	%		
	7月	8月	9月
对多元回归方程的预测误差	5.52	4.32	5.30
对一元方程的预测误差	4.51	3.22	4.20

(5) 基于修正误差的月度电量预测

由于一元方程的预测误差更小, 因此采用一元回归方程结合ARIMA修正方法对月度电量进行预测, 即将相关因素分析中的预测值与ARIMA修正后的预测误差值进行加减运算, 得到修正后的预测值见表4。

表4 预测误差修正值

Table 4 Modification of prediction error

时间	万 kWh		
	7月	8月	9月
预测结果	193 911	229 546	221 746

最后, 单独采用相关分析法及文献[7]中的ARIMA预测法两种预测方法进行计算, 与基于K-L信息量与ARIMA误差修正预测法进行对比, 并进一步计算3种结果预测误差, 结果如表5, 误差越小, 预测精度越高。由此可见, 本文所提的基于K-L信息量和ARIMA误差修正的月度电量预测方法对于城乡居民用电具有较好的预测效果。

表5 3种预测方案对比

Table 5 Comparison of three forecasting schemes

预测方案	7月/ 万 kWh	8月/ 万 kWh	9月/ 万 kWh	平均误差/%
实际值	203 069	237 183	231 468	
一元相关分析法	186 398	221 102	218 182	6.91
ARIMA 预测法	192 693	222 193.2	221 469	5.25
本文所提方法	193 911	229 546	221 746	3.98

6 结束语

将回归方程的拟合误差构建成新的非平稳时间序列, 利用ARIMA对非平稳时间序列的预测优势, 修正预测误差, 为提高预测精度提供思路。

基于K-L信息量和ARIMA误差修正的月度电

量预测方法既考虑了对城乡居民用电量影响较大的相关因素, 又修正了相关因素分析法的预测误差, 提高了预测精度。D

参考文献:

[1] 夏丽. 基于ARIMA模型及回归分析的区域用电量预测方法研究[D]. 南京: 南京理工大学, 2013.
XIA Li. Research on regional electricity consumption forecasting method based on ARIMA model and regression analysis [D]. Nanjing: Nanjing University of Science & Technology, 2013.

[2] 崔和瑞, 王娣. 基于季节ARIMA模型的华北电网售电量预测研究[J]. 华东电力, 2009, 37(1): 70-72.
CUI Herui, WANG Di. Electricity sales forecast of north China power grid based on seasonal ARIMA model [J]. East China Electric Power, 2009, 37(1): 70-72.

[3] 颜伟, 程超, 薛斌, 等. 结合X12乘法模型和ARIMA模型的月售电量预测方法[J]. 电力系统及其自动化学报, 2016, 28(5): 74-80.
YAN Wei, CHENG Chao, XUE Bin, et al. Monthly electricity sales forecasting method based on X12 multiplication model and ARIMA model [J]. Proceedings of the CSU-EPSA, 2016, 28(5): 74-80.

[4] 石雪梅, 葛斐, 肖夕林. 基于K-L信息量法的安徽省工业用电量预测[J]. 电网与清洁能源, 2015, 31(11): 58-62.
SHI Xuemei, GE Fei, XIAO Xilin. Prediction of industrial power consumption in Anhui Province based on K-L information method [J]. Power System and Clean Energy, 2015, 31(11): 58-62.

[5] 田中大, 李树江, 王艳红, 等. 高斯过程回归补偿ARIMA的网络流量预测[J]. 北京邮电大学学报, 2017, 40(6): 65-73.
TIAN Zhongda, LI Shujiang, WANG Yanhong, et al. Network traffic prediction based on Gaussian process regression compensation ARIMA [J]. Journal of Beijing University of Posts and Telecommunications, 2017, 40(6): 65-73.

[6] 麦洪坤, 肖坚红. 基于R语言的负荷预测ARIMA模型并行化研究[J]. 电网技术, 2015, 39(11): 3 216-3 220.
MAI Hongkun, XIAO Jianhong. Research on parallelization of ARIMA model for load forecasting based on R language [J]. Power System Technology, 2015, 39(11): 3 216-3 220.

[7] 孙颖. 基于季节性ARIMA模型的全社会用电量预测研究[J]. 黄冈师范学院学报, 2017, 37(6): 52-56.
SUN Ying. Study on the prediction of electricity consumption of the whole society based on seasonal ARIMA model [J]. Journal of Huanggang Normal University, 2017, 37(6): 52-56.

(责任编辑 郝 洁)