

基于回归分析和虚拟变量的短期用电量预测管理模型

为适应电力市场化交易,实现短期用电量的准确预测,调研了影响短期用电量的主要因素,重点分析了气温与用电量数据的关联关系,并利用用电量在时间分布上具有惯性的特征,提出了考虑气象条件的因变量滞后回归模型。同时对比了工作日、休息日、固定假日和移动假日用电量的差异,针对不同情况获得了虚拟化变量的系数,在此基础上建立了短期用电量预测模型,并对模型进行了静态预测和动态预测的实证分析,模型的建立为地区电网项目建设起到重要的理论支撑作用。

李 凯 赵滨滨 曹占峰 刘展展/国网天津市电力公司
曹云峰/内蒙古电子信息职业技术学院

0 引言

2015年3月,国务院印发了《关于进一步深化电力体制改革的若干意见》(中发〔2015〕9号),启动了我国新一轮电力体制改革。此次深化改革坚持市场化方向,以建立健全电力市场机制为主要目标,着力推动电力交易机构组建和规范运行,充分发挥市场在资源配置中的决定性作用。随着电力交易和运行机制的广泛推行,项目所产生的效益一般短期就能得到体现。因此,发电企业、售电企业、电力交易机构以及电网公司在项目管理方面更加注重短期用电量预测,使之成为电力市场和交易管

理中的研究重点和热点。

目前我国对用电需求的预测主要以中长期为主^[1-4],对于短期用电量预测,特别是未来数天日用电量预测研究相对较少。本文以短期用电量预测为主要研究对象,分析了影响用电量的主要影响因素。以天津市历史实际气象数据和用电量数据为基础,探索了气温与用电量的相关关系及其敏感度,并考虑每周的社会经济特性以及节假日用电特征,建立了天津市短期用电量预测模型。最后采用静态预测和动态预测方法进行了实证分析,验证了该模型具有较高的准确性和实用性^[5-6]。

1 用电量管理的主要影响因素分析

在以往的中长期电力需求预测中,一般认为全社会用电量主要受到社会经济因素影响,包括人口、城镇化率和地区生产总值等。如表1所示,近5年来天津市全社会用电量与常住人口、城镇化率和地区生产总值同步发展。但对于短期预测来说,社会经济因素变化难以精准量化。其中,难以获得一个地区每天的常住人口变动指标,也很难得到每天的第一产业、第二产业和第三产业增加



李 凯/高级工程师

关键词/Keywords

回归模型·
敏感度·
虚拟变量·
滞后期·
静态预测·
动态预测·

表1 天津市常住人口、地区生产总值与全社会用电量

时间	常住人口/万人	城镇化率(%)	地区生产总值/亿元	全社会用电量/亿kW·h
2010年	1 299.3	79.6	9 224.5	645.7
2011年	1 354.6	80.5	11 307.3	695.2
2012年	1 413.2	81.5	12 893.9	722.5
2013年	1 472.2	82.0	14 442.0	774.5
2014年	1 516.8	82.5	15 722.5	794.4

值指标。因此，从社会经济因素变化角度对地区短期日用电量进行预测和分析存在很大难度。

为此，从用电量的构成角度进行深入分析，以2014年天津市全社会用电量为例，第一产业占比仅为2%，第二产业占比达72%，第三产业以及居民用电量占比分别为16%和10%。可见，第一产业用电量占比较小，全社会用电量主要受到第二、第三产业以及居民用电的影响较大。对于短期用电量来说，第二产业用电量相对稳定，主要受经济状况、工作周期和节假日的影响；对于第三产业、居民用电量而言，主要受到工作周期、节假日以及气象因素影响。

气象因素影响主要是由于随着社会经济的发展，人民生活的提高，受到空调器的拥有率、使用率等因素影响。特别在大城市空调器已成为大众商品，夏天降温设备与冬天取暖设备的大量使用，气温成为高负荷、高用电量的主要原因。图1为2012年1月~2015年7月天津市用电量时间序列图，从图1中可见每年夏天的7、8月，冬天的12、1月是一年中用电量最高峰，4月和10月相对来说是用电量的低谷，之所以产生这种周期性的波动，究其原因就是气象因素。

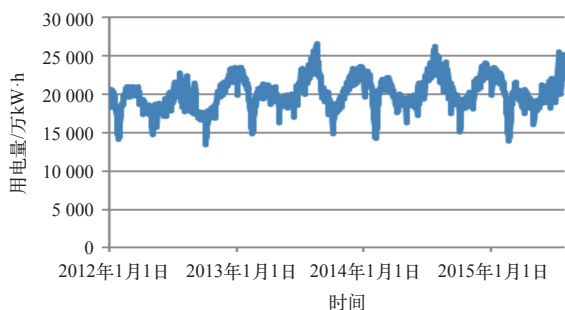


图1 用电量时间序列图

2 气温与用电量回归管理模型

2.1 数据组织与收集

为实现短期用电量的预测，首先需要利用近年来气象、用电量等数据为基础，定量探讨气温和用电量之间的内在关联关系，探索其规律性。随后考虑工作周期以及节假日等因素对用电量的影响和作用，用于开展短期用电量预测，供相关

人员参考。

气象指标，包括每天的最高温度、最低温度、平均温度、湿度、气压和降雨量等，这些指标可以从气象局得到，日用电量从天津市电力公司得到。

下面需要讨论的是模型数据的组织，是用1年数据建模，还是用3年数据建模，还是用5年数据建模。天津市用电量和最大负荷数据与各省电力公司数据相同，最大特点是趋势性和周期波动性（见图1），用最小二乘法建模得到的是离散点的平均趋势。选择几年的数据，反映较长时间的平均趋势，模型相对稳定，但拟合度不一定高，预测效果也不一定好。如果选择1年的数据，这1年数据特征正好与预测年特征一致，那么所建模型可以得到最佳的预测效果，但对未来经济情况、气象特点存在很多未知数，也很难选择。从另一角度来看，时间序列数据具有惯性，相近年份的数据内部结构是最相近的，故在一般情况下选择上一年的数据作为建模数据会得到相对较好的预测结果。天津公司提供了2012年1月~2015年8月数据，选取2014年8月~2015年7月数据作为模型数据，选取2015年8月数据作为模型验证数据。下面就模型数据的分布特征进行进一步的探索。

2.2 气温与用电量数据分析

一年中夏季随着温度的上升用电量也会上升，而在冬季却相反，用电量是随着温度的下降而上升的。春秋季气温相对舒适，气温与用电量不存在确定的关系。通过对数据的考察，将一年中11月、12月、1月、2月和3月份定义为冬季数据，4月、10月定义为春秋数据，5月、6月、7月、8月和9月定义为夏季数据。表2、表3和表4分别为气温与用电量相关系数表（注：季节划分非气象意义上的划分，而是根据人体舒适程度用电量需求进行划分。）。从相关系数表可以得出以下几点结论。

1) 冬季气温与用电量相关系数为负，相关系数为 $-0.33 \sim 0.37$ ，即温度越低用电量越高，与平均温度相关系数最大，其次最低温度，再其次最

高温度。

2) 春秋季气温与用电量相关系数为负, 相关系数为0~0.39, 为微弱相关, 主要由于天津4月上旬和10月下旬最低温度仍较低。

3) 夏季气温与用电量相关系数为正, 相关系数为0.66~0.78, 为显著相关, 相关系数大小依次为最低温度、平均温度和最高温度。

4) 气压、降水量和相对湿度与用电量的相关系数表现得比较复杂, 气压冬季为正相关, 夏季为负相关; 降水量均为负相关, 湿度冬季为负相关, 夏季为正负相关, 这与人体的感觉是一致的。

5) 综合考虑, 平均温度与用电量之间的关系更具有代表性, 因此选用平均温度作为模型中温度自变量。

表2 冬季气温与用电量相关系数

	最高温度	最低温度	平均温度	气压	降水量	相对湿度
用电量	-0.333 9	-0.353	-0.364 9	0.366	-0.272 3	-0.127 6

表3 春秋季气温与用电量相关系数

	最高温度	最低温度	平均温度	气压	降水量	相对湿度
用电量	0	-0.394 9	-0.256 7	0.054 4	-0.153 2	-0.230 5

表4 夏季气温与用电量相关系数

	最高温度	最低温度	平均温度	气压	降水量	相对湿度
用电量	0.668 8	0.779 8	0.765 2	-0.352 4	-0.075 8	0.183 4

图2为不同季节下用电量数据的盒装图, 表5为用电量数据均值分析情况。从中可以看出:

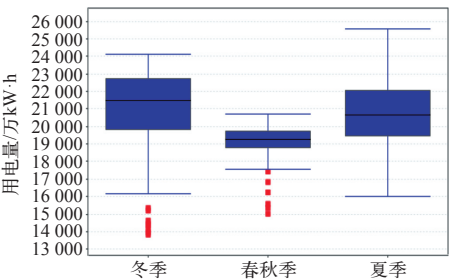


图2 盒装图分析

1) 春秋季日中位数用电量在19 000万kW·h, 代表了天津市常态下的用电量, 夏季日中位数用电量为20 800万kW·h, 平均每天要高出春秋1 800万kW·h, 冬季日中位数用电量为21 000万kW·h, 高出春秋2 000万kW·h。

表5 气温与用电量数据均值分析

季节	有效记录数	最小值 /万kW·h	最大值 /万kW·h	平均值 /万kW·h	标准差
冬季	151	13 831.07	24 135.16	20 953.85	2 274.72
春秋	61	15 021.58	20 695.47	19 046.65	1 208.65
夏季	153	15 991.53	25 581.41	20 813.89	1 866.47
总数	365	13 831.07	25 581.41	20 576.44	2 074.32

2) 方差分析。春秋季用电量分布相对集中, 方差小, 夏季和冬季气温波动大, 故方差也大, 夏季日最大用电量要高出春秋季最大用电量5 000万kW·h, 冬季日最大用电量高出春秋季4 000万kW·h。

3) 冬季和春秋季存在较多的异常点, 这主要是春节、国庆节假期, 企业放长假而引起的用电量下降的结果。

用电量与平均温度的散点关系图如图3所示。从图3中可以看出:

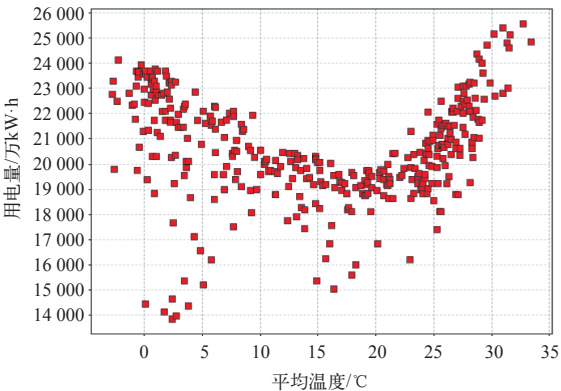


图3 用电量与平均温度散点图

1) 用电量与平均温度散点图总体上呈向上开口的弧形分布, 春节和国庆节用电量最低, 构成图2中的异常点用电量(也可认为是长假期用电量), 除去春节和国庆节用电量外, 用电量最低点在平均温度15~16℃, 也就是从气象特征来认识, 平均温度15~16℃左右是人体感觉最为舒适的, 大于该温度或小于该温度, 用电量都将随温度的变化而上升。

2) 由于用电量除了受平均温度影响外, 也受到其他因素的影响: 如最低温度、气压、湿度、工作日、休息日和节假日等, 所以用电量在形态上表现为宽体开口向上的弧形。

3) 春节和国庆节, 由于大多数企业休息, 故春节和国庆节用电量都较低, 特别是春节用电量为全年最低, 而且影响到春节前后的若干天, 这些异常点的用电量分析将在后续讨论。

4) 用电量与平均温度之间呈向上弧形分布, 故回归模型应为一元二次方程。

2.3 敏感度分析

所谓温度对用电量敏感度分析就是定量研究由于温度的变动而导致用电量变动情况, 为了排除异常数据的干扰, 首先将用电量均值+2倍标准差和均值-2倍标准差的数据剔除, 然后对样本进行回归分析, 式(1)为回归方程, 该方程各项系数全部通过检验

$$S = 22\,727.8 - 487.279T + 16.1361T^2 \tag{1}$$

式中, S为日用电量; T为平均气温。

根据式(1)可以得出以下几点结论:

- 1) 当平均温度为15.1℃时用电量取极小值, 估值为19 049万kW·h。
- 2) 当平均温度高于15.1℃或低于15.1℃时, 用电量的加速变化率为32.27万kW·h, 也就是用电量对平均温度的敏感度不是一个常数, 而是以15.1℃为基准, 平均温度每增高1℃或降低1℃用电量的敏感度会以32.27万kW·h/℃²的加速变化率上升。

平均温度对用电量敏感度分析总结如表6所示。

表6 平均温度对用电量敏感度分析

指标	极小值/℃	敏感度(加速度)	95%置信区间
用电量	15.1	32.27	29.08—35.46

注: 敏感度指平均温度增加或减少1℃, 用电量的加速变化率, 用电量单位: 万kW·h/℃²。

上述结论是在仅考虑平均温度一个变量得到的回归估计值, 实际情况要复杂得多, 每天的用电量除受平均温度的影响外, 还与其他多因素有关, 如湿度、降雨量和经济活动等。下面进一步讨论气温与用电量模型。

3 短期用电量预测模型

3.1 气温与用电量滞后模型

在时间序列模型中, 社会经济变量具有明显

的惯性, 所以由社会经济变量决定的用电量也具有明显的惯性, 为确实最佳回归模型, 利用AIC、BC系数的测定决定模型的滞后期。

前面讨论已经知道用电量对于平均温度而言是一元二次方程。在此基础上, 分别加入应变量的滞后一期、滞后二期和滞后三期, 回归结果如表7所示。从表7中可见, 对于用电量回归方程, 滞后二期, 修正后R²较大, AIC和SC相对较小; 滞后三期, 尽管修正后R²略大, AIC略小, 但SC比滞后二期大了, 所以对于用电量回归模型应该是滞后二期为最佳模型^[7-8]。

表7 回归模型滞后期讨论

指标	修正后R ²	AIC	SC
滞后一期	0.870 9	16.058 2	16.122 7
滞后二期	0.787 8	10.662 2	10.715 9
滞后三期	0.786 4	10.665 4	10.729 9

3.2 节假日的虚拟化处理

从图1中可以看到节假日用电量较低, 特别是国庆节和春节要比常态下的用电量减少1/3, 其次工作日和休息日用电量不同, 休息日用电量明显小于工作日, 再其次每个工作日内部也存在差异, 需要用虚拟化变量来解决这些问题。

星期虚拟变量, 以星期日为基准, 星期一为1, 其他为0, 星期二为1其他为0, 其他仿此设定, 结果如表8所示。

表8 星期虚拟变量设定

星期一	星期二	星期三	星期四	星期五	星期六
1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0
0	0	0	0	0	1

节假日虚拟变量设定, 节假日包括固定假日和移动假日, 元旦、清明、五一、端午和中秋为休息3天, 对用电量和最大负荷有影响, 但相对有限, 假日赋值为1, 国庆节和春节休息7天, 对用电量和最大负荷的影响相对较大, 国庆节虚拟变量从9月30日~10月7日。根据序列观测, 其中10

月1日、2日和3日电量和最大负荷下降较快,赋值为2,其余赋值为1。春节对用电量影响不仅仅是春节7天,还会波及到节前和节后。根据序列观测节前5天和节后3天也会受到春节的影响,当然节前、节中和节后受影响的程度是不一致的,节中影响最大。故从年三十~初七赋值为4,其他日子为1,如表9所示。

表9 国定假日、移动假日虚拟变量设定

元旦	春节	清明	五一	端午	中秋	国庆
1	1	1	1	1	1	1
1	1	1	1	1	1	2
1	1	1	1	1	1	2
0	1	0	0	0	0	2
0	4 (年三十)	0	0	0	0	1
0	4 (初一)	0	0	0	0	1
0	4 (初二)	0	0	0	0	1
0	4 (初三)	0	0	0	0	1
0	4 (初四)	0	0	0	0	0
0	4 (初五)	0	0	0	0	0
0	4 (初六)	0	0	0	0	0
0	4 (初七)	0	0	0	0	0
0	1	0	0	0	0	0
0	1	0	0	0	0	0
0	1	0	0	0	0	0

3.3 短期用电量回归模型

在 $\alpha = 0.05$ 显著性条件下,气压和降水量被剔除,用电量回归模型如下:短期用电量回归规模设定为

$$S = 6658.31 + 0.84 \times S(-1) - 0.16 \times S(-2) - 183.53 \times T + 5.96T^2 + 3.99A + 1416.97 \times D(1) + 870.0 \times D(2) + 756.1 \times D(3) + 667.68 \times D(4) + 671.33 \times D(5) + 223 \times D(6) - 630.43 \times J \quad (2)$$

式中, S 为日用电量; $S(-1)$ 为日用电量滞后一期; $S(-2)$ 为日用电量滞后二期; T 为平均气温; A 为相对湿度; $D(i)$ 为星期虚拟变量; i 分别为1~6; J 为节假日虚拟变量。

修正 $R^2 = 0.9354$,说明模型的拟合度相当高, $F = 437.6$, $P = 0.0000$,说明在 $\alpha = 0.05$ 显著性条件下,模型是显著的, $D.W = 1.6072$,说明模型不存在自相关。模型的各项系数在 $\alpha = 0.05$ 显著性条件下全部通过了检验,模型系数进行解释

如下。

常数项=6 658.3,即其他各项系数全为0时,天津市用电量为6 658.3万kW·h。

用电量(-1)=0.838 5,说明滞后一期用电量的85.29%对本期有影响。

用电量(-2)=-0.160 2,说明滞后二期用电量的-16.02%对本期的影响。

平均温度一次项系数=-183.53,二次项系数=5.962 3,由于平均温度与用电量之间是抛物线关系,可以计算当平均温度为15.39℃时,用电量取极小值;当平均温度大于或小于15.39℃时,用电量以11.92万kW·h/℃的加速变化率上升,这与前面散点图探索和敏感度分析的结论基本一致。

相对湿度系数=3.99,说明当相对湿度增加1度时,用电量将增加3.99万kW·h。

星期一系数=1 417,说明以星期日为基准,星期一的用电量比星期日高出1 417万kW·h,星期二要高出870万kW·h,星期三要高出756万kW·h,星期四要高出667.7万kW·h,星期五要高出671.3万kW·h,星期六要高出223万kW·h。

节假日系数=-630.4,平均而言节假日用电量将减少630.4万kW·h。

气温与用电量回归模型拟合图如图4所示。根据回归模型残差数据测算,模型的平均绝对误差为361.87万kW·h,相对误差为1.79%。

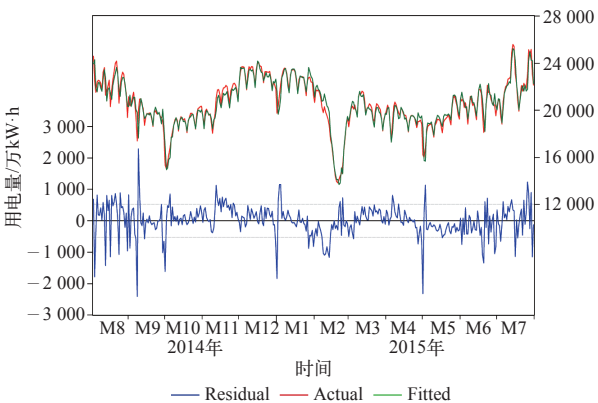


图4 气温与用电量回归模型拟合

4 预测分析及对项目优化

建立模型的目的一就是用来预测,发现事

物的发展规律对未来的工作进行指导。为验证上述模型的准确性和可用性，本文对2015年8月份数据进行实证分析。

4.1 模型评估方法

采用平均绝对误差（MAE）和平均相对误差（MPE）来评估模型的准确性和可用性。一般认为，平均相对误差越低，表示预测模型准确性和实用性越大。计算公式为

MAE = \frac{1}{h} \sum_{t=T+1}^{T+h} |\hat{y}_t - y_t| \tag{3}

MPE = \frac{1}{h} \sum_{t=T+1}^{T+h} \left| \frac{\hat{y}_t - y_t}{y_t} \right| \tag{4}

式中， h 为预测期数； $\hat{y}$ 和 y 分别为预测值和实际值。

4.2 预测方法

由于模型中包含因变量的滞后变量，可以进行两种方法的预测：一种为静态预测；另一种为动态预测。

静态预测：在其他自变量已知的条件下，因变量滞后变量用实际值代替，静态预测只能做一期向后预测。

动态预测：在其他自变量已知的条件下，因变量滞后变量用预测值代替，故可以进行多期向后预测，第一期的预测可用实际值代替。

4.3 预测结果与误差分析

(1) 静态预测

对2015年8月3日~8月9日（星期一~星期日）的用电量进行静态预测，结果如表10所示。静态预测结果，用电量的平均绝对误差为502.43kW·h，平均相对误差为2.3%，其中最大相对误差为3.59%，最小相对误差仅0.89%。

表10 用电量静态预测评估

日期	用电量 /万kW·h	用电量预测值 /万kW·h	用电量 绝对误差	用电量 相对误差
8月1日	20 750.05			
8月2日	20 550.79			
8月3日	21 193.97	21 676.99	483.03	0.022 8
8月4日	21 087.87	21 751.46	663.58	0.031 5
8月5日	22 376.67	21 578.00	798.68	0.035 7
8月6日	22 364.15	22 562.80	198.65	0.008 9

8月7日	22 658.23	22 342.51	315.72	0.013 9
8月8日	21 780.61	22 153.11	372.50	0.017 1
8月9日	22 175.76	21 490.92	684.84	0.030 9
平均值	21 948.18	21 676.99	502.43	0.023

根据静态测试评估可见测试区与模型区结果基本一致，平均绝对误差和平均相对误差多比较小，说明根据上一年度数据建立的模型完全可以用在未来的预测中。

(2) 动态预测评估

采用动态预测模型就上述静态测试数据再进行动态预测评估，表11为用电量动态预测结果。用电量的平均绝对误差为366kW·h，平均相对误差为1.7%，其中最大相对误差为5%，最小相对误差仅0.1%。

表11 用电量动态预测评估

日期	用电量 /万kW·h	用电量预测值 /万kW·h	用电量 绝对误差	用电量 相对误差
8月3日	21 193.97	21 676.99	483.03	0.022 8
8月4日	21 087.87	22 157.20	1 069.33	0.050 7
8月5日	22 376.67	22 398.95	22.27	0.001 0
8月6日	22 364.15	22 410.42	46.26	0.002 1
8月7日	22 658.23	22 377.80	280.42	0.012 4
8月8日	21 780.61	21 910.16	129.54	0.005 9
8月9日	22 175.76	21 644.61	531.16	0.024 0
平均值	21 948.18	22 082.30	366.00	0.017 0

由于表10静态预测与表11动态预测用的是同一数据，所以可以对预测结果进行比较，从上述比较结果可看出，动态预测要好于静态预测，这只不过是偶然现象，大多数情况下静态预测要好于动态预测。在进行动态预测时要注意：动态预测是将预测结果作为下一期预测的滞后变量值，如果前面预测已产生较大误差，误差就会传递下去，使误差越来越大，所以动态预测可以进行多期预测，但不易做过长时间的预测。

4.4 预测模型对电网项目投资优化

用电量预测模型引入中国以来，已经在国民经济发展过程中得到了广泛的普及和应用，在此基础上的短期用电量预测模型必将有助于我国电网项目投资优化管理。

短期用电量模型的好处在于能够迅速捕捉电

力市场短期内变化动态,通过可量化指标参数体现市场短期变化规划,避免项目投资的盲目性,提高投资效率,增强投资效益。

5 结束语

力求所建模型考虑各种影响因素尽可能周到,所建模型尽可能完善,然而周到和完美是相对的。

本文分析提出影响短期用电量的主要因素除了经济社会因素外,还包括气象条件和节假日等,并发现了气温与用电量存在二次方的关联关系。随后利用用电量在时间分布上具有惯性的特征,提出了气温与用电量的滞后回归模型,针对工作日、休息日、固定假日和移动假日用电量的差异进行了节假日虚拟化变量处理。综合获得了短期用电量预测模型,最后针对该模型进行了动态预测和静态预测方法的实证分析。对比预测用电量和实际用电量数据评估结果,平均相对误差较低,说明该模型具有较高的准确性和可用性。

在今后的模型的应用过程中,可进一步增加相关重要的指标(如空调器拥有比例),逐渐完善模型。另外为增强模型数据组织的时效性,可采用与时俱进的方式每年更新一次模型,进一步提高模型的准确性和实用性,为地区项目投资管理提供重要的技术支撑;同时进一步考虑各地区

差异化影响因素,全面完善模型,使之适用于共性的同时兼顾特性。

参考文献

- [1] 毛李帆,江岳春,龙瑞华,等.基于偏最小二乘回归分析的中长期电力负荷预测[J].电网技术,2008(19):71-77.
- [2] 康重庆,夏清,张伯明.电力系统负荷预测研究综述与发展方向的探讨[J].电力系统自动化,2004(17):1-11.
- [3] 傅守强,汪莹,刘素伊,等.基于市场的中长期电力消费需求预测[J].电网与清洁能源,2015,31(7):74-80.
- [4] 龙瑞华,毛弋,毛李帆,等.基于诱导有序加权平均算子和马尔可夫链的中长期电力负荷组合预测模型[J].电网技术,2010(3):150-156.
- [5] 高铁梅.计量经济分析方法与建模:EvIEWS应用及实例[M].北京:清华大学出版社,2009.
- [6] 古扎拉蒂.计量经济学第[M].3版.林少宫,译.北京:中国人民大学出版社,1999.
- [7] 黄晖,李鸿琪.马克威统计分析与数据挖掘应用案例[M].北京:中国统计出版社,2012.
- [8] 毛玉凤.基于时间序列分析的电力需求预测及季节调整模型的研究[D].北京:北京工业大学,2013.

(收稿日期:2016-10-15) **EA**

(上接第24页)

灯两端的电压的大小,达到了降低路灯的功率以节约电能的目的,相对于该市目前的节能技术,有着非常明显的优势。

两组现场试验的结果发现,该新型节能器能在保证路面照度达标的条件下,达到超过35%的节能率。同时对该市城市路灯线路适应性良好,能在实际运行过程中,达到约30%左右的节能率。

参考文献

- [1] 郝书堂.路灯[M].北京:中国电力出版社,

2008.

- [2] GB/T 12325—2008 电能质量供电电压允许偏差[S].
- [3] 鲍民.国内照明节能控制设备的性能比较[J].道路照明,2004,8(1):41-42.
- [4] CJJ 45—2015 城市道路照明设计标准[S].
- [5] 周剑,陈汝义,王长友.交流斩波调压技术在路灯节能上的应用[J].江苏电器,2008(12):45-47,64.

(收稿日期:2015-11-19) **EA**