

# Big Data Analysis of Mid-Long Term Load Forecasting in Power System

Xiaoping Wang, Yang Li, Jun Yong, Hao Zhang, Bing He, Tao Zheng, Qianshan Cheng, Hualiang Fang

Mianyang Power Supply Company, State Grid Sichuan Electric Power Company, Mianyang  
Email: [qianshan\\_ch@163.com](mailto:qianshan_ch@163.com)

Received: Oct. 31<sup>st</sup>, 2014; revised: Nov. 19<sup>th</sup>, 2014; accepted: Nov. 26<sup>th</sup>, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

A mid-long term load forecasting method based on big data theory is proposed because of the present complicated forecasting models and bad adaptability. The method focuses on the data and the load in large area is partitioned to different levels and classified to different types. Then the relationships among data are researched. Based on the characteristic of each partition and analysis of corresponding data, the load forecasting model is established, and thus the elaborate load forecasting is realized. In the end, the validity of the method is proved by practical example in power system.

## Keywords

Mid-Long Term Load Forecasting, Big Data, Partition and Classification of Load, Elaborate Load Forecasting Model

---

# 电力系统中长期负荷预测的大数据分析

王小平, 李阳, 雍军, 张浩, 何冰, 郑涛, 程潜善, 方华亮

国网四川省电力公司绵阳供电公司, 绵阳

Email: [qianshan\\_ch@163.com](mailto:qianshan_ch@163.com)

收稿日期: 2014年10月31日; 修回日期: 2014年11月19日; 录用日期: 2014年11月26日

## 摘要

针对当前电力系统中长期预测模型复杂、适应性差等不足，提出基于大数据的中长期负荷预测方法。该方法以数据为中心，将大区域的负荷进行分层分区及分类，并研究数据之间的关联性。针对各区间负荷结构等特点，在数据分析基础上，建立与数据相适应的负荷预测模型，实现大区域内负荷的精细化预测。最后以某地区电网为例，验证了该方法的有效性。

## 关键词

中长期负荷预测，大数据，负荷分区分类，精细化预测

## 1. 引言

中长期负荷预测是对未来五到十年及以上的负荷进行预测，为电网规划和建设提供参考。预测时间周期长、影响因子多，除了经济水平、产业结构、相关政策外，还受到新型负荷如电动汽车等的迅速发展，节能环保政策，以及气候的异常变化等的影响，中长期负荷预测是一个十分复杂的多维、非线性的不确定性问题。很多学者对中长期负荷预测进行了研究，主要有回归分析法[1]、灰色系统理论预测法[2][3]、以及人工智能预测方法如神经网络预测法[4][5]、证据理论预测法[6]、数据挖掘技术[7]及支持向量机预测法[8]。这些预测方法主要是以模型或算法为中心，基于某种数学理论，建立负荷预测模型，并根据实际情况进行相应的改进，计算得到负荷预测结果。预测结果严重依赖于数学模型，对于基础数据类型多样，模型的适应性较差。

中长期负荷预测受到诸多随机因素的影响，没有明确的数学机理，数学模型无法准确地描述负荷变化规律[9]。根本上决定负荷变化的是各类型数据，而模型只是数据的支撑。所以精准的负荷预测只有尽可能搜集、发掘影响负荷变化因子的数据，全面考虑各种不同因素对负荷的影响。而实际上，电网在运营过程中已收集到大量的负荷信息，如各地负荷的类型与分布情况，另外各相关部门提供的诸如经济、气候等也会随技术革新而愈加全面、准确。根据这种趋势，本文提出了基于大数据思想的中长期负荷预测方法，是在数据的分析基础上建模预测。

大数据技术[10][11]是解决预测问题强有力的工具，为负荷预测提供了一种全新的思路，它以数据的分析处理为中心，从数据中挖掘出隐含未来发展的信息。中长期负荷预测是典型的数据决定问题，利用大数据方法，对负荷大数据进行细致的分析处理，能很好地发掘出负荷的分布情况和变化趋势。本文首先从整体上构建大数据思想下中长期负荷框架，然后就框架中数据分析处理和建模预测两个核心部分进行了详细的分析，通过数据的分区分层与分类、建立与之相适应的预测模型，最后通过实例验证了方法的有效性。

## 2. 大数据思想负荷预测框架

负荷预测一般包括确定预测目标、负荷数据收集、数据分析处理、建模预测和误差分析等环节。基于大数据思想的负荷预测对数据的要求更高，要对数据源进行全面分析，且更依赖通信网络和微机技术来传送汇集存储数据，以形成大数据平台，如图1所示。

数据源分析与采集：中长期负荷预测的影响因素多，相关数据不仅体量大、类型多，相互影响与层次关系也非常复杂。在收集之前，对数据进行全面分析，确定其种类和影响负荷的方式。在此基础上，

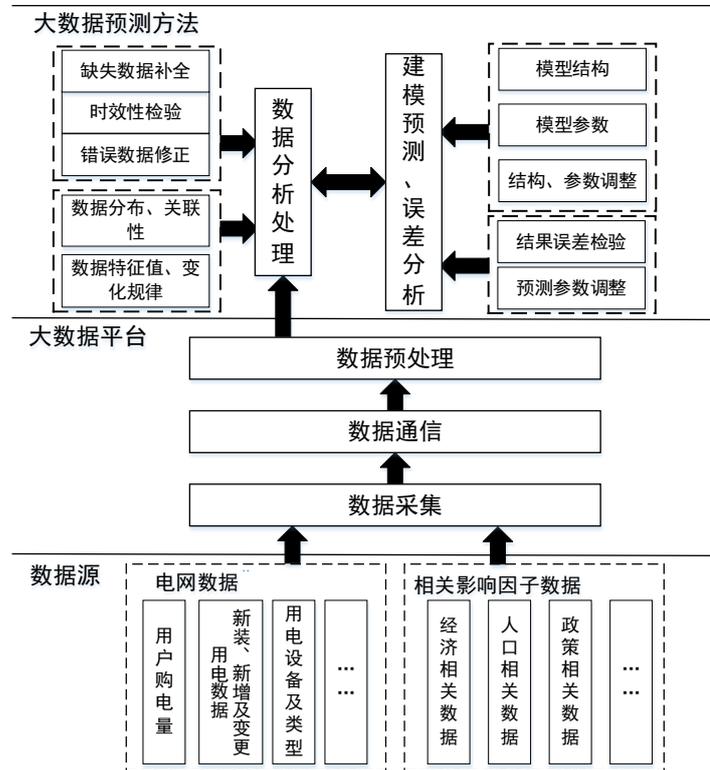


Figure 1. Frame diagram of big data load forecasting

图 1. 大数据负荷预测框架图

结合数据分布情况，通过各种数据采集终端、政府信息发布平台等渠道收集这些数据。

数据通信与预处理：预测基础数据是分布在广域空间上的，各采集终端所获得数据必须通过通信传输汇集起来，通过格式转换、整理归类等预处理操作存储起来，形成负荷预测大数据信息平台。

数据处理与分析：在大数据平台下，对数据的分析处理分为两步。首先针对数据的不足进行的处理，包括时效性检验、缺失数据补全、错误数据修正等，剔除错误数据，补足重要数据。其次利用大数据分析技术，对各类型数据进行详细的分析，提取特征参数和变化规律，挖掘出数据中隐含信息。

建模预测与误差分析：基于大数据的负荷预测建模是针对数据的建模，是根据数据的分布规律与变化规律建立的与数据相适应的模型，能针对数据的改变做出相应调整，最后通过误差分析进一步修正模型中的参数。

在预测框架中，数据源分析、数据分析处理及建模预测为其核心部分，决定预测精度。大数据思想正是通过数据源分析，大量掌握负荷相关数据，利用大数据分析技术全面分析负荷大数据特点，建立以数据为中心的预测模型，最终实现大区域内负荷的精细化预测。在预测体系中，负荷数据对最终结果起着很大作用。实际中不可避免出现数据不足的情况，在部分数据严重缺失的情况下某些参数会无法提取，会使得模型过于简单，预测结果较为粗糙，所以数据源的分析 and 数据的收集在大数据负荷预测中至关重要。

### 3. 负荷大数据分析

中长期负荷预测除了需搜集大量的影响因素数据外，电网本身也会有大量的负荷数据产生。这些数据不仅体量大，类型多，而且相互关系复杂，在建模预测前必须进行详细的数据分析，构建起数据间层

次结构体系，这便需要对负荷进行分区分层与分类处理。

负荷分区分类是细致考虑底层负荷个性，实现精细化预测的手段，能针对不同地域负荷的分布、类型等特点作出相应的预测处理，更加贴合实际。负荷分层分区与分类要与实际分布情况相统一，能反映出负荷密度、负荷类型等特点，图2是按照我国电网管理方法的开放性负荷分类方法。

这种分类方法结合电网管理，实现对区域负荷多层次的划分，能自由根据所收集数据选择划分的层级，当数据足够充分时，甚至能精确到某个具体的负荷。对划分到底层的负荷，依次进行独立的分析预测，但在实际中，由于产业链等因素，负荷之间存在复杂的关联关系，即某一类型负荷的变化会造成相关负荷的改变，特别是关联紧密产业间的大负荷。

大负荷间的关联关系与产业间关系相对应，主要有横向关联与纵向关联。横向关联指同一级别的产业负荷间相互影响关系，既有竞争又有共生，纵向关联指不同级别，即存在上下游产业负荷间作用关系，只有共生关系，如汽车生产企业负荷间的横纵关联关系如图3所示。

对于有横纵关联关系的大负荷，可将其分别从原有的各分区内提取出来，组成独立的虚拟分区，对虚拟分区内的负荷进行单独预测，整体修正。单独预测是对虚拟分区内的负荷依据各自的变化趋势和影响因素分别做出预测，整体修正是根据负荷间的关联类别和关联程度对各负荷的预测结果作出整体上的修正。

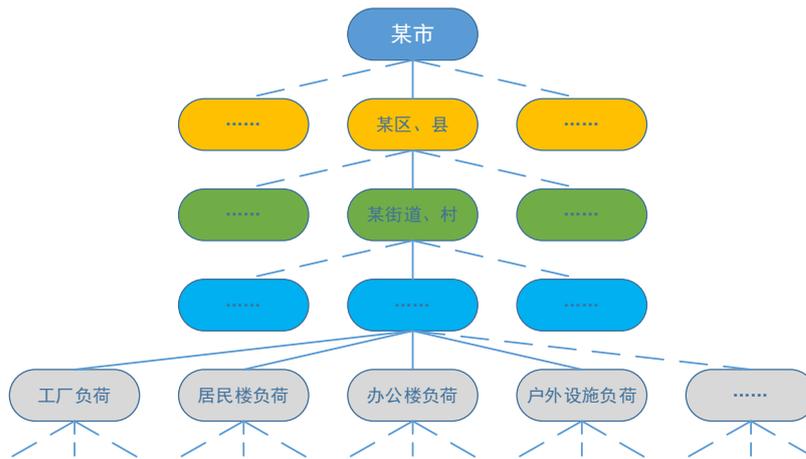


Figure 2. Open hierarchy chart of load partition and classification  
图2. 开放性负荷分区分类层次图

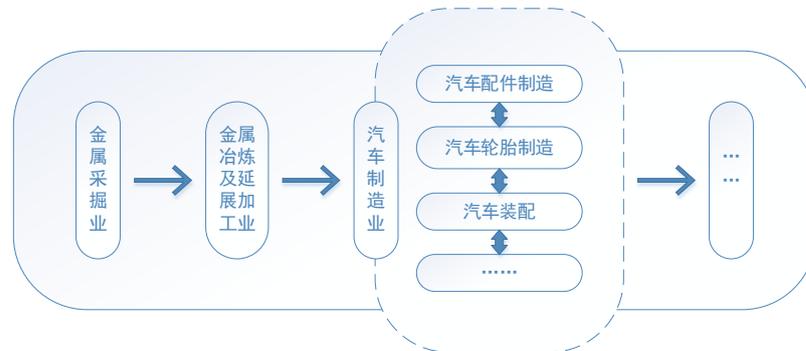


Figure 3. Horizontal and vertical relationship diagram among big load  
图3. 大负荷间横、纵关联关系图

## 4. 大数据预测建模

### 4.1. 分区模型

分区建模要依据分区的层次，从一级分区开始，逐层深入。发展不同、负荷密度不同的区域，分区级数会有所区别。对某一分区用  $I_n^m$  表示，其中  $m$  表示分区的级数， $n$  表示每个分区层级中的不同分区，若对于子分区  $I_n^m$  有  $s$  个下级子分区，则有

$$I_n^m = \sum_{k=1}^s I_{n,k}^{m+1} \quad (1)$$

那么整个预测区域  $I$  可表示为

$$I = \sum_{i=1}^n I_i^1 = \sum_{j=1}^p I_{1,j}^2 + \sum_{j=1}^q I_{2,j}^2 + \sum_{j=1}^r I_{3,j}^2 + \dots + I_k^1 + \dots = \dots \quad (2)$$

对应上式， $I_1^1$ 、 $I_2^1$ 、 $I_3^1$  分别表示具有  $p$ 、 $q$ 、 $r$  个下级子分区的分区， $I_k^1$  表示没有子分区的分区。可以看出，不同子分区下可以有不同个数、不同层级的下级子分区，在实际情况下，根据数据的详实程度一直确定到细化程度最高的级别。

对于有纵横关联性的产业大负荷，将这些负荷从各分区内提取出来，组成虚拟分区  $I_v$ ，虚拟分区的处理同其他分区的处理方法类似，即根据分区特点独立进行预测。新建负荷的虚拟分区后，整个预测区域  $I$  可表示为

$$I = \sum_{j=1}^p I_{1,j}^2 + \sum_{j=1}^q I_{2,j}^2 + \sum_{j=1}^r I_{3,j}^2 + \dots + I_k^1 + \dots + I_{v1}^1 + I_{v2}^1 + \dots \quad (3)$$

### 4.2. 分类模型

负荷分类要在分区的前提下进行，针对每个分区内不同类型，不同作用因素的负荷进行区分，具体分类要根据分区所深入的程度，当分区深入到基层负荷时，那负荷可划分为办公楼负荷，居民楼负荷，户外设施负荷等具体类型负荷，当分区层级比较少，也可划分为三大产业的负荷以及居民用电负荷等类型。负荷分类要在分区内进行，对分区  $I_n^m$ ，若共有  $L$  类型负荷，则有

$$P_{I_n^m} = \sum_{i=1}^L P_{i,I_n^m} \quad (4)$$

### 4.3. 预测模型

负荷分区分类的最终目的是实现不同类型负荷的精细化预测。在分区分类的基础上，建立反映不同分区负荷特性的预测模型。

对于包含两类子分区的  $r$  级预测分区  $I_k^r$ ，即有  $p$  个包含子分区的分区和  $l$  个不包含子分区的分区，详细的建模过程如图 4。

对于不包含子分区的  $l$  个分区内的某一类型负荷  $P_i$ ，假设其受到  $n$  种影响因素的影响，写成向量即  $X_i = (x_{i,1}, x_{i,2}, x_{i,3}, \dots, x_{i,n})$ ，对  $P_i$  的预测，首先根据  $P_i$  变化的整体变化趋势确定基准值  $P_{i,t,\text{base}}$ ，基准值的确定可采取多种方法，选用弹性系数法有

$$P_{i,t,\text{base}} = P_{i0} (1 + k_i * r_{\Delta G_i})^t \quad (5)$$

$P_{i0}$  为基值参考值，可选择某一年负荷量或多年均值等作为参考， $k_i$  为弹性系数， $r_{\Delta G_j}$  为经济的增长

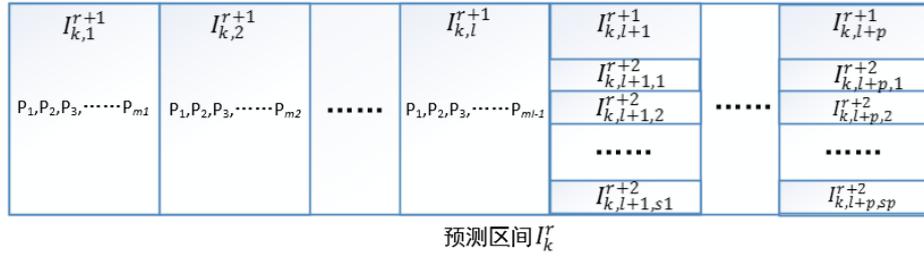


Figure 4. Model structure diagram of load partition

图 4. 预测分区模型结构图

率， $t$  为相对基值参考值经过的年限。

然后根据每种影响因子对负荷的作用程度对预测基值进行修正，有

$$P_{i,t} = P_{i,t,\text{base}} + \sum_{i=1}^n \alpha_{i,t} * \Delta x_{i,t} \quad (6)$$

$P_{i,t}$  为相应预测分区经过  $t$  年后的预测值， $\Delta x_{i,t}$  为各种影响因素  $t$  年后的预测数据变化向量， $\alpha_{i,t}$  为相应变化对  $P_{i,t}$  的影响系数。

对预测分区  $I_{k,1}^{r+1}$ ， $t$  年后负荷的预测值  $P_{I_{k,1}^{r+1},t}$  为

$$\begin{aligned} P_{I_{k,1}^{r+1},t} &= P_{I_{k,1}^{r+1},t,\text{base}} + \sum_{j=1}^{m_1} (\alpha_{j,t} * \Delta P_{j,t}) \\ &= P_{I_{k,1}^{r+1},t,\text{base}} + \sum_{j=1}^{m_1} (\alpha_{j,t} * (P_{j,t} - P_{j,t,\text{base}})) \\ &= P_{I_{k,1}^{r+1},t,\text{base}} + \sum_{j=1}^{m_1} \left[ \alpha_{j,t} * \left( \sum_{i=1}^n \alpha_{i,t} * \Delta X_{i,t} \right) \right] \end{aligned} \quad (7)$$

对于包含子分区的  $p$  个分区，同样从最底层的负荷出发，逐级向上级分区分步计算，对于预测分区  $I_{k,l}^{r+1}$  有

$$P_{I_{k,l}^{r+1},t} = P_{I_{k,l}^{r+1},t,\text{base}} + \sum_{j=1}^{s_l} (\alpha_{j,t} * \Delta P_{I_{k,l,j}^{r+2},t}) = P_{I_{k,l}^{r+1},t,\text{base}} + \sum_{j=1}^{s_l} (\alpha_{j,t} * (P_{I_{k,l,j}^{r+2},t} - P_{I_{k,l,j}^{r+2},t,\text{base}})) \quad (8)$$

综合两种子分区，最后得出预测分区  $I_k^r$  的预测结果为

$$\begin{aligned} P_{I_k^r,t} &= P_{I_k^r,t,\text{base}} + \sum_{j=1}^p (\alpha_{j,t} * \Delta P_{I_{k,j}^{r+1},t}) = P_{I_k^r,t,\text{base}} + \sum_{j=1}^{l-1} \left\{ \alpha_{j,t} * \sum_{q=1}^{m_1} \left[ \alpha_{q,t} * \left( \sum_{i=1}^n \alpha_{i,t} * \Delta X_{i,t} \right) \right] \right\} \\ &\quad + \sum_{j=l}^p \left\{ \alpha_{j,t} * \sum_{j=1}^{s_l} (\alpha_{j,t} * (P_{I_{k,l,j}^{r+2},t} - P_{I_{k,l,j}^{r+2},t,\text{base}})) \right\}. \end{aligned} \quad (9)$$

## 5. 算例分析

利用大数据分析方法，对某市的用电量进行分析预测。该市的用电量如表 1 所示。

由数据可将该市分为 7 个分区，每个分区内的负荷分为第一产业、第二产业、第三产业以及居民生活用电 4 种类型。在横向上，首先对该市负荷做整体分析，确定关键分区以及各类型负荷的重点分区(图 5)。

从图 5 中看出，该市以第二产业用电为主，超过全市负荷的一半，而第一产业负荷非常少。全市负荷主要集中在 A、B、C 三个区域，尤其是 A、C 两个分区，拥有全市接近 70% 的负荷。各类型负荷，也

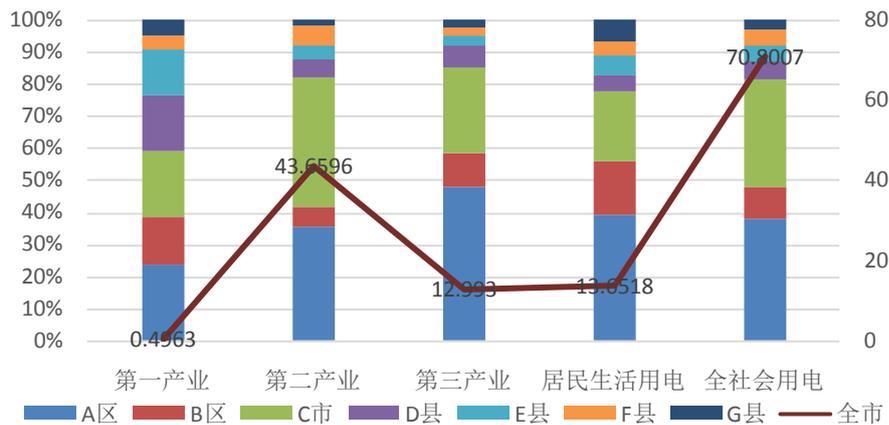
都是 A、B、C 三个分区占据比重最大，所以要准确的进行该市负荷预测，需将这三分区作为重点分区考虑。

然后对各分区内的负荷进行分析，从下面的百分比图可以看出，包括重点分区 A、B、C 在内的所有分区均以第二产业负荷为主，特别是 C 区，达到 70% 多，因各分区产业结构的不同，各类型负荷的比例均不相同(图 6)。

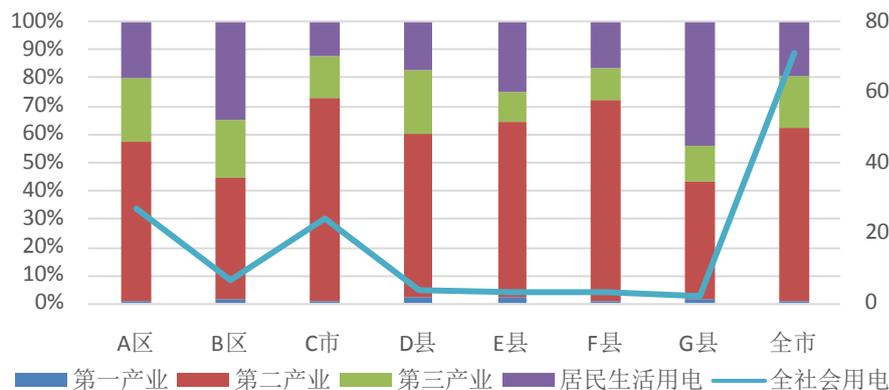
在纵向上，分析各分区内各类型负荷比例变化。随着地区的发展，不同产业经济增速是不一样的，

**Table 1. Basic data of load forecasting**  
**表 1. 负荷预测基础数据**

	A 区	B 区	C 市	D 县	E 县	F 县	G 县	全市
第一产业	0.1182	0.0743	0.1012	0.0846	0.0718	0.0204	0.0258	0.4963
第二产业	15.361	2.8948	17.4885	2.3998	2.0737	2.5125	0.9293	43.6596
第三产业	6.1988	1.3627	3.4767	0.9308	0.3496	0.3919	0.2825	12.993
居民生活用电	5.3296	2.3346	2.9259	0.692	0.8385	0.5686	0.9626	13.6518
全社会用电	27.0076	6.6664	23.9923	4.1072	3.3336	3.4934	2.2002	70.8007



**Figure 5. Load types analysis in whole city**  
**图 5. 全市负荷类型分析**



**Figure 6. Load analysis diagram in each partition**  
**图 6. 分区类型负荷分析图**



Figure 7. Structure changing of the load in partition A  
图 7. A 区负荷结构变化图

Table 2. Forecasting result of each kind of load  
表 2. 分类型负荷预测结果

年份	第一产业		第二产业		第三产业		居民生活用电		全社会用电	
	实际值	预测值	实际值	预测值	实际值	预测值	实际值	预测值	实际值	预测值
2005	1.11	1.01	28.51	30.98	5.94	5.85	5.81	5.57	41.37	41.22
2006	0.96	0.95	28.29	28.40	6.99	6.97	7.34	6.30	43.58	45.21
2007	0.89	0.84	27.7	30.83	7.80	7.69	7.88	8.16	44.27	44.71
2008	0.85	0.80	24.08	27.20	7.62	8.60	7.50	7.50	40.05	41.08
2009	0.74	0.73	28.58	26.05	8.77	8.85	9.18	8.11	47.26	46.21
2010	0.62	0.66	35.24	32.40	9.84	9.51	10.24	10.22	55.95	51.28
2011	0.54	0.52	35.26	36.10	11.87	10.70	13.46	11.55	61.13	63.38
2012	0.57	0.47	33.46	34.29	13.16	12.95	14.39	14.37	61.58	61.08

造成各类型负荷的比重会有改变。以 A 分区为例，如图 7 所示，可以看出，当各类型负荷比例保持相对稳定时，总体负荷的增长越平稳，规律性较强；在经济动荡时候会出现负荷的较大波动。

通过上述全市及分区各类型负荷的分析，分别确定相应负荷的基准值和影响系数等参数，尤其是在重点分区内，对精度要求会更高。对各类型负荷用相应的经济、人均收入等影响因素对基值进行相应的修正，逐分区层级向上，得出最终的预测结果如表 2 所示。

从预测结果来看，均与实际值较为接近。误差主要的来源是基值的误差和修正量的误差，基值误差由负荷的分析引起的；修正量的误差由各影响因素数据统计误差和影响系数引起的，这两种误差均会随着数据的丰富，精度得到提高。

## 6. 结论

本文提出了基于大数据思想的中长期负荷预测方法，并建立了大数据思想下的预测框架体系，建立体系中分区分类和相应预测模型，能针对负荷类型等特点，综合考虑影响负荷变化的各类因素，体现数据对预测结果的决定性作用。实例证明，本文提出的方法体系预测结果能反映负荷实际情况，较传统方法更加精确。

## 基金项目

国家自然科学基金(51107090)。

## 参考文献 (References)

- [1] 毛李帆, 江岳春, 姚建刚 (2009) 采用正交信号修正法与偏最小二乘回归的中长期负荷预测. *中国电机工程学报*, **16**, 82-88.
- [2] 张伏生, 刘芳, 赵文彬 (2003) 灰色 Verhuist 模型在中长期负荷预测中的应用. *电网技术*, **5**, 37-39.
- [3] 周德强 (2009) 改进的灰色 Verhuist 模型在中长期负荷预测中的应用. *电网技术*, **18**, 124-127.
- [4] Hsu, C.C. and Chen, C.Y. (2003) Regional load forecasting in Taiwan-application of artificial neural networks. *Energy Conversion and Management*, **44**, 1941-1949.
- [5] 李春祥, 牛东晓, 孟丽敏 (2009) 基于层次分析法和径向基函数神经网络的中长期负荷预测综合模型. *电网技术*, **2**, 99-104.
- [6] 倪明, 高晓萍, 单渊达 (1997) 证据理论在中期负荷预测中的应用. *中国电机工程学报*, **3**, 199-203.
- [7] 崔旻, 顾洁 (2004) 基于数据挖掘的电力系统中长期负荷预测新方法. *电力自动化设备*, **6**, 18-21.
- [8] 李瑾, 刘金朋, 王建军 (2011) 采用支持向量机和模拟退火算法的中长期负荷预测方法. *中国电机工程学报*, **16**, 63-66.
- [9] 郑志杰, 李磊, 赵兰明 (2011) 考虑数据不确定性的中长期电力负荷预测. *电力系统保护与控制*, **7**, 123-126.
- [10] 赵云山, 刘焕焕 (2014) 大数据技术在电力行业的应用. *研究电信科学*, **1**, 57-62.
- [11] 宋亚奇, 周国亮, 朱永利 (2013) 智能电网大数据处理技术现状与挑战. *电网技术*, **4**, 927-935.