

报名序号：1254

论文题目：电力系统短期负荷预测

	姓 名	班 级	有效联系电话
参赛队员 1			
参赛队员 2			
参赛队员 3			

指导教师：

参赛学校：

证书邮寄地址、邮编、收件人：

报名序号：

	阅卷专家 1	阅卷专家 2	阅卷专家 3
论文等级			

摘要

提高负荷预测进度是保障电力系统优化决策科学性的重要手段。根据已有电力负荷数据及气象因素数据,文章主要建立了4个模型来解决关于短期负荷预测方面的问题。

针对问题一,建立日最高负荷量模型、日最低负荷量模型、日峰谷差模型、日平均负荷量模型以及日负荷率模型。利用 Excel 软件可将两地区 014 年各个负荷量的统计值求出(详见附件 1),其中地区二 2014 年 1 月 1 日的日最高负荷量、日最低负荷量、日峰谷差、日平均负荷量以及日负荷率分别为 6765.5、3748.48、3017.05、5138.23 和 0.76。通过观察两地 2014 年负荷数据变化曲线图,考虑数据的波动性等因素可得出地区二更准确的预测结果的结论。

针对问题二,构建多元线性回归模型,利用 SPSS 软件对日最高负荷、日最低负荷、日平均负荷与各气象因素进行回归分析。通过观察标准化残差图(详见图 4),认为没有趋势性,回归模型有效。用同样的方法可得出两地区各个因变量的回归方程(详见表 5)。对多元线性方程做回归误差分析,认为将不重要的气象因素剔除可减小误差。利用逐步回归法可进行更合理的回归分析,得出优先推荐平均温度来提高负荷预测精度。

针对问题三,构建 ARIMA 预测模型,对数据进行预处理,取每年春季的负荷量作为参照数据,消除了季节成分的影响。通过自相关方面的分析,确定模型为 ARIMA(1,1,1),利用 SPSS 软件可得出所需的预测结果。例如地区一在时间点 T0000 的负荷量预测模型为 $\Delta x_t = 0.928\Delta x_{t-1} + \varepsilon_t - 0.999$ 。模型拟合的可决系数都在 0.8 以上,说明预测结果精度比较高。

针对问题四,构建基于 BP 神经网络算法的多元非线性系统模型,确定模型为 $y = ANN(x_1, x_2, x_3, x_4, x_5)$,利用 Matlab 编程可训练出相应的神经网络结构,得出预测结果。通过参照数据、模型原理这两个方面,论证了计及气象因素影响的负荷预测结果的精度得到了改善这一结论。

针对问题五,提取两地区日负荷率作为待处理数据,分别对两地区日负荷率进行正态拟合、T 分布拟合、Logistic 拟合,做出拟合曲线并对各个拟合进行拟合曲线广义似然比检验。得出地区二的数据比地区一的数据更有规律的结论。

关键词: 短期负荷预测; 多元线性回归; ARIMA 预测模型; BP 神经网络; 拟合

1. 问题的重述

短期负荷预测是电力系统运行与分析的基础，对机组组合、经济调度、安全校核等具有重要意义。提高负荷预测精度，是保障电力系统优化决策科学性的重要手段。现代电力系统中，构成电力负荷的用电器种类繁多，空调等受气象条件影响的负荷占比持续增高，气象因素（温度、湿度、降雨量等）对电力系统负荷的影响愈显突出。考虑气象因素成为调度中心进一步改进负荷预测精度的主要手段之一。

已知地区 1、地区 2 从 2009 年 1 月 1 日至 2015 年 1 月 10 日的电力负荷数据（每 15 min 一个采样点，每日 96 点，量纲为 MW）以及 2012 年 1 月 1 日至 2015 年 1 月 17 日的气象因素数据（日最高温度、日最低温度、日平均温度、日相对湿度以及日降雨量）。

具体要求如下：

1.请分析两个地区 2014 年 1 月 1 日—2014 年 12 月 31 日的负荷数据，统计各地区全年的日最高负荷、日最低负荷、日峰谷差、日负荷率指标的分布情况，并绘制两地区 2014 年全年的负荷持续曲线；结合上述结果，分析两地区负荷变化的主要差异；初步预判哪个地区的负荷可以获得更准确的预测结果，说明你的理由。

2.根据 2012 年 1 月 1 日至 2014 年 12 月 31 日的的数据，分别对日最高负荷、日最低负荷、日平均负荷与各气象因素的关系进行回归分析，分析回归误差；如果要用气象因素来提高负荷预测精度，在诸气象因素中，你优先推荐哪个（或哪几个）？简要说明理由。

3.请根据已知负荷数据，构建预测方法，对两个地区 2015 年 1 月 11 日至 17 日共 7 天的电力负荷进行预测（间隔 15min），给出负荷预测结果；在不知道实际负荷数据的条件下，你对预测结果的准确度有何推断，请说明理由。

4.如果已获得 2015 年 1 月 11 日至 17 日的气象因素数据，你能否构建计及气象因素的负荷预测方法，对两个地区 2015 年 1 月 11 日至 17 日共 7 天的电力负荷再次进行预测（间隔 15min），给出预测结果；与原有的预测结果相比，你认为计及气象因素影响的负荷预测结果精度得到改善了吗？有何证据？请说明理由。

5.综合上述计算结果，你如何评价两地区负荷规律性的优劣？你还有什么证据可以佐证两地区负荷整体规律性优劣的判断？

2. 1 对于问题一的分析

问题一要求分析两个地区二 014 年的负荷量数据的一些统计量，全年的日最高负荷、日最低负荷、日峰谷差、日负荷率指标的分布情况。可以直接建立最大量最小量模型以及一些简单算数模型来解决，利用 Excel 软件可以很快求出答案。题目还要求绘制出两地区二 014 年全年的负荷数据变化曲线，可以利用 Matlab 的绘图工具来绘制出想要的结果。最后对所得统计量以及两地区二 014 年全年的负荷数据变化曲线进行分析，可以初步预判哪个地区的负荷可以获得更准确的预测结果。

2. 2 对于问题二的分析

问题二要求对日最高负荷、日最低负荷与各气象因素的关系进行回归分析，分析回归误差，还要求用推荐哪个（或哪几个）气象因素，来提高负荷预测精度。可利用统计学知识分别对日最高负荷、日最低负荷与各气象因素的关系进行回归分析，并通过回归分析所得的一些统计学数据来进行回归误差分析以及选出推荐的气象因素。

2. 3 对于问题三的分析

该问题要求根据一致负荷数据，构建预测方法，对两个地区二 015 年 1 月 11 日至 17 日共 7 天的电力负荷进行预测。此问题没有提及气象因素对负荷的影响，说明要求我们通过负荷数据本身进行预测，这是个时间序列预测问题，可建立 ARIMA 模型就可预测出指定 7 日的负荷量。

2. 4 对于问题四的分析

该问题要求构建计及气象因素的负荷预测方法，并给出预测结果。气象因素对负荷影响是很大的，我们可以尝试构建人工建神经网络的模型，通过训练网络可以比较准确地找到各气象因素与负荷之间的关系，进而预测出指定 7 日的负荷量。该问题还要求将通过气象因素预测出的结果与问题 3 的预测结果进行比较，可以从多个方面比较预测结果的精度。

2. 5 对于问题五的分析

该问题要求对两地区负荷规律性的优劣进行评价，既然是考虑规律性，我们可以将两地区的负荷数据进行正态拟合、Logistic 拟合以及 T 分布拟合，比较两个地区负荷的拟合效果，就可以得出哪个地区的规律性更好。

3. 模型的假设与符号说明

3. 1 模型的假设

- (1) 假设2009年1月1日至2015年1月10日的电力负荷数据均为真实有效数据；
- (2) 神经网络训练期间，“坏数据”带来的训练误差；不会使网络不能收敛到理想误差。

3. 2 符号说明

M	隐层节点数
F	权值输入端连接的神经节点数
X_{ijk}	第 <i>i</i> 个地区第 <i>j</i> 天第 <i>k</i> 个时刻所测量的负荷数据
a_{ij}	第 <i>i</i> 个地区第 <i>j</i> 天的日最高负荷量
b_{ij}	第 <i>i</i> 个地区第 <i>j</i> 天的日最低负荷量
c_{ij}	第 <i>i</i> 个地区第 <i>j</i> 天的日峰谷差
d_{ij}	第 <i>i</i> 个地区第 <i>j</i> 天的日平均负荷，
e_{ij}	第 <i>i</i> 个地区第 <i>j</i> 天的日负荷率
Y	日最高负荷、日最低负荷、日平均负荷中的一种变量
ANN	非线性函数
X_1	最高温度
X_2	最低温度
X_3	平均温度
X_4	相对湿度
X_5	降雨量

4. 模型的准备

4.1 回归分析法基本原理

回归分析法是根据历史数据的变化规律和影响负荷变化的因素，寻找自变量与因变量之间的相关关系及回归方程式，确定模型参数，据此推断将来时刻的负荷值。

回归分析法的优点是计算原理和结构形式简单，预测速度快，外推性能好，对于历史上没有出现的情况有较好的预测。

4.2 针对问题三对原始数据进行预处理

在解决问题三的过程中，利用 ARIMA 预测模型，首先运用 SPSS 软件将地区一的原始负荷数据导入，对时间点 T0000 构建如下的序列图。

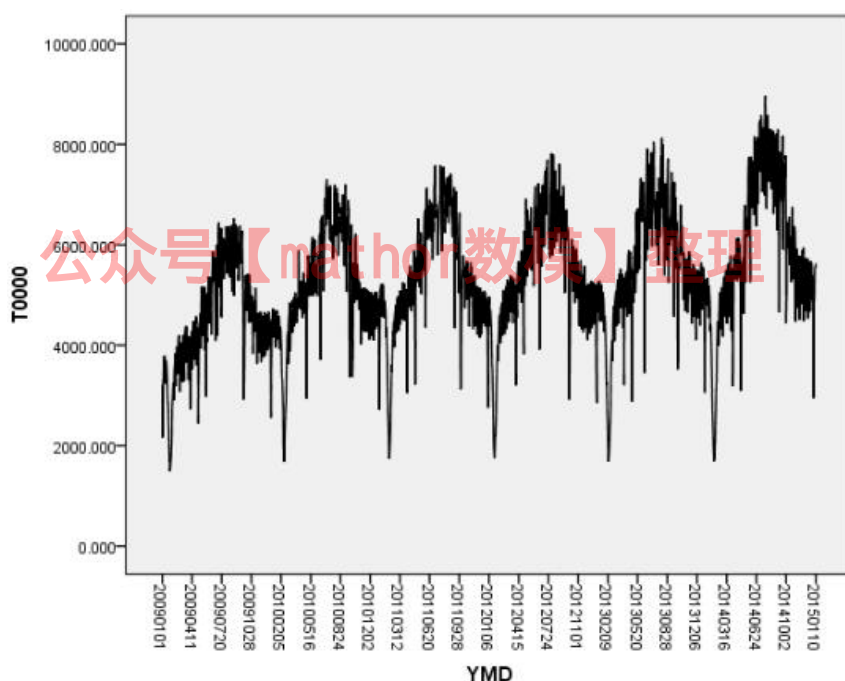


图1 数据处理前地区一 T0000 时间点序列图

图中有明显的季节成分，因此需要做季节分解。题目要求预测两个地区二015年1月11日至17日共7天的电力负荷，都属于春季。因此只需提取每年的前三个月的负荷数据作为输入的数据。分解后，序列图如下。

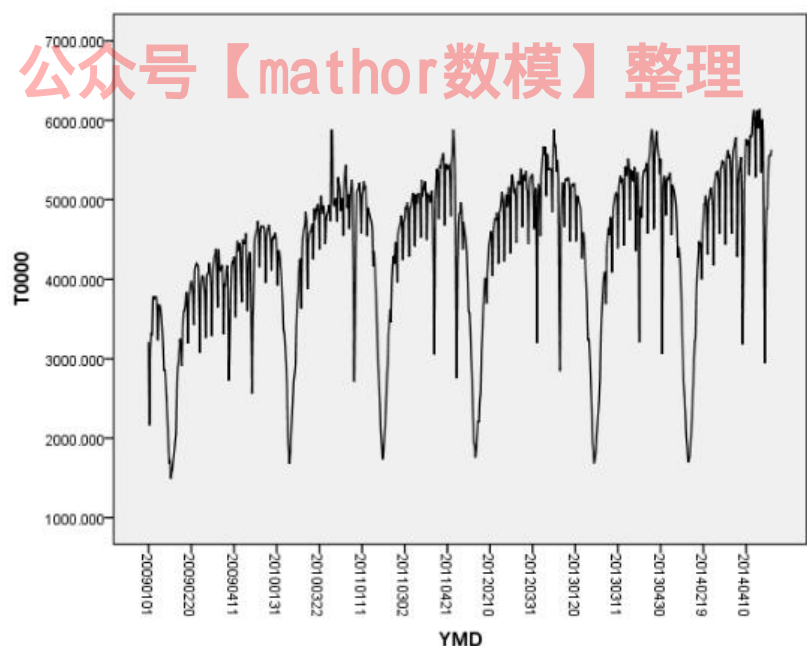


图 2 数据处理后地区一 T0000 时间点序列图

从上图可知，排除了季节成分。所做的预测将会更精准，同时计算的复杂程度将会降低。

4. 3 BP 神经网络基本原理概述

4. 3. 1 BP 神经网络基本原理

BP 网络模型处理信息的基本原理是：学习过程由信号的正向传播和误差的反向传播两个过程组成。正向传播时，输入信号通过中间层作用于输出层，经过非线性变换，产生输出信号;若输出层的实际输出与期望输出不符，则转向误差的反向传播阶段。误差的反向传播是将输出误差以某种形式通过中间层向输入层逐层反转，并将误差分摊给各层的所有单元，从而获得各层的误差信号作为修正各单元权值的依据。此过程周而复始，直到输出的误差降到可以接受的程度。此时经过训练的神经网络即能对类似样本的输入信息自行处理，进而输出误差最小的经过非线性转换的信息，然后可通过检验神经网络的有效性。

运用 BP 神经网络处理实际问题时分为两个步骤即网络训练和网络应用。第一步网络训练采用有监督的学习，有监督的学习是指每一个训练样本都对应一个代表环境信息的教师信号作为期望输出，训练时计算实际输出与期望输出之间的误差，根据误差的大小和方向反复调整网络连接权值，直到误差达到预订的精度为止。

4. 3. 2 BP 神经网络的结构

BP 神经网络是一种多层前馈网络，其神经元连接权值的调整规则采用误差反传算法即 BP 算法。BP 神经网络又是一个多层感知器，多层次感知器强调神经网络在结构上由输入层、隐含层、输出层等多层构成，BP 网络则强调层间连接权值通过误差反传算法进行调整。

BP 神经网络的特点是:网络由多层次构成，包括输入层、隐含层（单层或多层）和输出层;层与层之间全连接，同层神经元之间无连接;传递函数必须可微，

常用的有 Sigmoid 型的对数、正切函数或线性函数;采用误差反传算法进行学习,逐层向前修正网络连接权值。

BP 神经网络结构在设计时主要包括以下方面:

(1) 网络层数

BP 神经网络至少包括一个输入层和一个输出层,可以包含一个或多个隐含层,所以网络层数的决定问题即是隐含层层数的决定问题。理论上已经证明,单个隐层可以通过适当增加神经元节点数达到任意的非线性映射,因此大多数情况单隐层结构的神经网络足以满足需求。在样本较多的情况下,增加一个隐层可以有效减小网络规模。

(2) 输入层节点数

输入层节点数取决于输入向量维数,具体可根据实际问题 and 数据类型确定。如果输入数据为模型信号波形,则可根据波形的采样点数目决定输入向量维数;如果输入数据为时间序列数据,则输入节点为时间点数;如果输入为图像,则输入单元可以为图像像素或经处理的图像特征。

(3) 隐含层节点数

隐含层节点数在很大程度上影响着 BP 神经网络的性能。对此一个非常重要的定理表述为对任何一个在闭区间内的连续函数都可以用三层即单隐层 BP 神经网络逼近,因而单隐层 BP 网络可以完成任意的 n 维到 m 维的映射。一般而言,隐含层较多节点可使网络达到更好的性能,但可能导致较长的收敛时间。实践中,通常采用以下经验公式选择最佳节点数:

第一种: $\sum_{i=0}^n C_m^i > k$, 其中 k 为样本数, M 为隐层节点数。如果 $i > M$, 规定 $C_m^i = 0$ 。

第二种: $M = \sqrt{n+m} + a$, 其中 n 为输入节点数, m 为输出节点数。 a 是 $[0,10]$ 之间的常数。

第三种: $M = \log_2 n$, n 为输入节点数。

(4) 输出层节点数

输出层节点数需要根据实际问题的抽象模型进行确定。例如在利用神经网络解决模式

分类问题中,如果共有 n 个类别,则输出节点数为 n 或 $\lceil \log_2 n \rceil$, $\lceil x \rceil$ 表述不小于 x 的最小整数。

(5) 传递函数

根据研究经验,一般情况下输入层和隐层的传递函数选用 S 型函数

$$f(x) = \frac{1}{1+e^{-x}}, (0,1)$$

或正切 S 型函数

$$f(x) = \frac{1-e^{-x}}{1+e^{-x}}, (-1,1)$$

输出层选用线性函数作为传递函数,用 purelin 表示。

(6) 训练方法

BP 神经网络采用迭代调整的方式进行权值确定,因此在训练之前需要确定初始值作为迭代调整的起点。初始值的大小会影响网络的性能,通常情况将初始

值定为较小的非零随机值，经验值为 $(-\frac{2.4}{\sqrt{F}}, \frac{2.4}{\sqrt{F}})$ 或 $(-\frac{3}{\sqrt{F}}, \frac{3}{\sqrt{F}})$ 之间，其中 F 为权值输入端连接的神经节点数。

5. 模型的建立与求解

5. 1 问题一的模型建立与求解

对于第一问，设 X_{ijk} 为第 i 个地区第 j 天第 k 个时刻所测量的负荷数据，可建立日最高负荷量的数学模型：

$$a_{ij} = \max \{X_{ijk}\} (k = 0000, 0015, 0030, \dots, 2345)$$

该模型中 a_{ij} 表示第 i 个地区第 j 天的日最高负荷量。

同样可建立最日低负荷量的数学模型：

$$b_{ij} = \min \{X_{ijk}\} (k = 0000, 0015, 0030, \dots, 2345)$$

该模型中 b_{ij} 表示第 i 个地区第 j 天的日最低负荷量。

对于日峰谷差，可建立如下模型：

$$c_{ij} = \max \{X_{ijk}\} - \min \{X_{ijk}\} (k = 0000, 0015, 0030, \dots, 2345)$$

该模型中 c_{ij} 表示第 i 个地区第 j 天的日峰谷差。

日负荷率为日平均负荷与日最大负荷的比值，可建立如下模型：

$$\begin{cases} d_{ij} = \frac{\sum_{k=0000}^{k=2345} X_{ijk}}{96} \\ e_{ij} = \frac{d_{ij}}{a_{ij}} \end{cases}$$

其中 d_{ij} 为第 i 个地区第 j 天的日平均负荷， e_{ij} 表示第 i 个地区第 j 天的日负荷率。

依据上述模型可利用 Excel 软件求出部分下列结果如下（详见附件 1）：

表 1 2014 年地区二负荷量的统计量结果

日期	最高负荷	最低负荷	日峰谷差	日平均负荷	日负荷率
20140101	6765.53551	3748.481751	3017.053759	5138.225299	0.759470598
20140102	8464.700806	3278.47475	5186.226056	6232.420477	0.736283611
20140103	8642.119917	4141.438257	4500.68166	6703.741343	0.775705661
20140104	8350.459638	4269.594754	4080.864884	6550.929709	0.784499296
.....
20141228	8480.166777	4356.945	4123.221777	6557.451155	0.773269126
20141229	9010.524753	4238.837121	4771.687632	6936.880418	0.769864199
20141230	8780.473733	4455.129564	4325.344169	6898.469963	0.785660338
20141231	8059.246529	4297.719613	3761.526916	6494.801036	0.805881916

利用 Matlab 软件，将数据导入后利用输入相应代码（详见附录 1），可得出如下负荷持续曲线图：

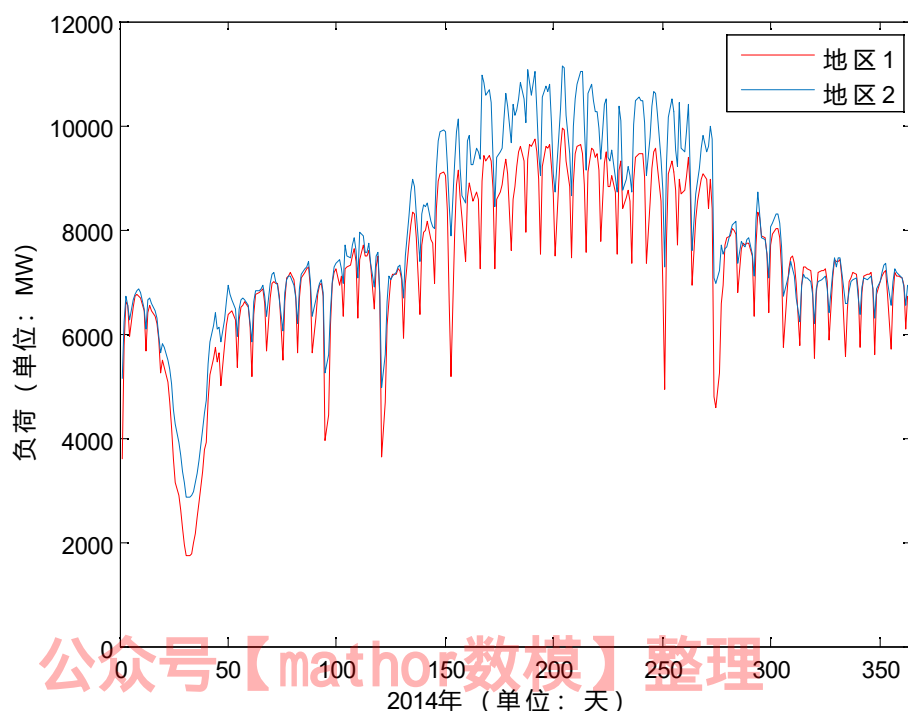


图 3 两地 2014 年负荷持续曲线图

通过结合上述结果，分析两地区负荷变化的主要差异，初步预判地区二的负荷可获得更准确的预测结果。原因是通过对附件 1 的统计量结果的分析，地区二的日峰谷差更小，通过图 1 也可以明显看出负荷持续波动更小，因此地区二可获得更准确的预测结果。

5. 2 问题二的模型建立与求解

5. 2. 1 多元线性回归模型的建立

变量 Y 和变量 X_1, X_2, X_3, X_4, X_5 的关系：

$$Y = f(X_1, X_2, X_3, X_4, X_5) + \varepsilon$$

其中 X_1, X_2, X_3, X_4, X_5 分别代表最高温度、最低温度、平均温度、相对湿度以及降雨量， Y 代表日最高负荷、日最低负荷、日平均负荷中的一种变量。 ε 为均值为 0 的随机变量。 f 的函数为线性的，即整个线性模型为：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \varepsilon。$$

为了得到回归参数的估计值，就要对变量进行观测，对变量的 $n = 1096$ 次独立观测数据为： $\{(y_i, x_{i1}, x_{i2}, \dots, x_{im}), i = 1, \dots, n\}$ ，则这些观测数据应满足式，即有

$$\begin{cases} y_1 = b_0 + b_1x_{11} + b_2x_{12} + b_3x_{13} + b_4x_{14} + b_5x_{15} + \varepsilon_1 \\ y_2 = b_0 + b_1x_{21} + b_2x_{22} + b_3x_{23} + b_4x_{24} + b_5x_{25} + \varepsilon_2 \\ \dots\dots\dots \\ y_n = b_0 + b_1x_{n1} + b_2x_{n2} + b_3x_{n3} + b_4x_{n4} + b_5x_{n5} + \varepsilon_n \end{cases}$$

其中 $E(\varepsilon_i) = 0, Cov(\varepsilon_i, \varepsilon_j) = \delta_{ij} \sigma^2, (i, j = 1, \dots, n)$,
若记 $Y = (y_1, y_2, \dots, y_n)^T, \beta = (b_0, b_1, \dots, b_m)^T, \varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}_{n \times (m+1)}$$

则多元线性回归的数学模型式 (4-6) 可以写成矩阵形式

$$Y = X\beta + \varepsilon$$

其中 $E(\varepsilon) = 0, Var(\varepsilon) = \sigma^2 I_n$ 。

为了获得参 β 的估计, 我们采用最小二乘法, 即选择 β , 使

$$Q(\beta) = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (Y - X\beta)^T (Y - X\beta) \quad (4-8)$$

达到最小。

将 $Q(\beta)$ 对 β 求导数并令其为零, 得

$$\frac{\partial Q}{\partial \beta} = -2X^T(Y - X\beta) = 0$$

即 $X^T X\beta = X^T Y$ 。记 $L = X^T X$, 则

$$L\beta = X^T Y$$

上述方程称为正规方程, 其中 X 为 $n \times (m+1)$ 阶矩阵, 一般假定 $rank(X) = m+1$, 由线性代数理论可知, $L = X^T X$ 为满秩矩阵, 它的秩 $rank(L) = m+1$, 则正规方程有唯一解, 记作

$$\hat{\beta} = L^{-1} X^T Y$$

我们来证明上式中 $\hat{\beta}$ 为参数向量 β 的最小二乘法估计量, 现用矩阵形式来叙述其证明步骤。对任意的 β 有 $Q = (Y - X\beta)^T (Y - X\beta)$ 则有

$$\begin{aligned} (Y - X\beta)^T (Y - X\beta) &= [(Y - X\hat{\beta}) + X(\hat{\beta} - \beta)]^T [(Y - X\hat{\beta}) + X(\hat{\beta} - \beta)] \\ &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) + (\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta) + (Y - X\hat{\beta})^T X(\hat{\beta} - \beta) + \\ &\quad (\hat{\beta} - \beta)^T X^T (Y - X\hat{\beta}) \geq (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \end{aligned}$$

上述证明过程中应用了如下结果:

$$(\hat{\beta} - \beta)^T X^T X(\hat{\beta} - \beta) = [X(\hat{\beta} - \beta)]^T [X(\hat{\beta} - \beta)] \geq 0$$

$$(Y - X\hat{\beta})^T X(\hat{\beta} - \beta) = (Y^T X - \hat{\beta}^T X^T X)(\hat{\beta} - \beta) = (Y^T X - Y^T X)(\hat{\beta} - \beta) = 0$$

至此, 在 $|L| \neq 0$ 时, 证明了正规方程中的 $\hat{\beta}$ 是 β 的最小二乘法估计量。

在实际工作中, 常称 $\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \dots + \hat{b}_m x_m$ 为经验线性回归方程。

5. 2. 2 多元线性回归模型的求解

首先本文利用问题一中所给模型，求出 2012 年 1 月 1 日至 2014 年 12 月 31 日的日最高负荷、日最低负荷、日平均负荷，部分结果如下表（详见附件 2）：

表 2 2012 年到 2014 年地区一统计量结果

日期	最高负荷	最低负荷	日平均负荷
20120101	3967.259968	2674.310752	3193.704021
20120102	6203.327488	2102.91472	4518.020952
20120103	7362.322144	3413.132512	5688.646099
20120104	7654.747168	3972.14128	6055.78345
20120105	7772.622784	4126.712512	6158.695814
20120106	7635.090112	4130.5792	6103.478207
20120107	7478.730112	4114.814656	5928.308079
20120108	6682.430272	3930.446176	5384.340865
20120109	7316.607328	3658.456096	5716.865955
20120110	7267.226368	3895.22992	5743.61441
.....
20141224	9041.502784	4830.504064	7095.792178
20141225	9055.009408	4861.555456	7122.141599
20141226	8960.655616	4805.517856	7073.389654
20141227	8899.329088	4826.893792	6912.876737
20141228	7501.186048	4616.671072	6078.493122
20141229	8738.955616	4224.25984	6707.0431
20141230	8479.017856	4578.106048	6627.303336
20141231	7797.769888	4313.017216	5880.430551

根据多元线性回归模型，利用 SPSS 软件，可对日最高负荷、日最低负荷、日平均负荷与各气象因素的关系进行回归分析。将数据导入软件后，设置回归分析方法为进入法，分别将日最高负荷、日最低负荷、日平均负荷作为因变量，进行回归分析。例如，对地区一日最高负荷与各气象因素的关系进行回归分析，可得以下分析结果：

表 3 地区一最高负荷与各气象因素回归分析的模型汇总

模型	R	R 方	调整 R	标准估计	更改统计量				
			方	的误差	R 方改	F 更改	df1	df2	Sig. F 改
1	.623 ^a	.388	.385	1517.28958	.388	137.693	5	1088	.000

从上表看出可决系数为 0.388，其模型的拟合程度最好，但还是很一般。

表 4 地区一最高负荷与各气象因素回归分析的系数

模型	非标准化系数		标 系 试用 版	t	Sig.	相 关 性			共线性统计量	
	B	标误				零阶	偏	部分	容差	VIF
(常量)	5604.140	402.913		13.909	.000					
最高温度	-33.500	30.155	-.113	-1.111	.267	.573	-.034	-.026	.055	18.306
最低温度	130.059	60.378	.412	2.154	.031	.614	.065	.051	.015	64.995
平均温度	105.834	79.770	.334	1.327	.185	.615	.040	.031	.009	112.876
相对湿度	-12.906	4.419	-.091	-2.921	.004	.112	-.088	-.069	.582	1.718
降雨量	5.856	3.235	.046	1.810	.071	.074	.055	.043	.858	1.165

上表给出了各个自变量的回归系数，但在这得出结论之前，必须要观察以下标准化残差图：

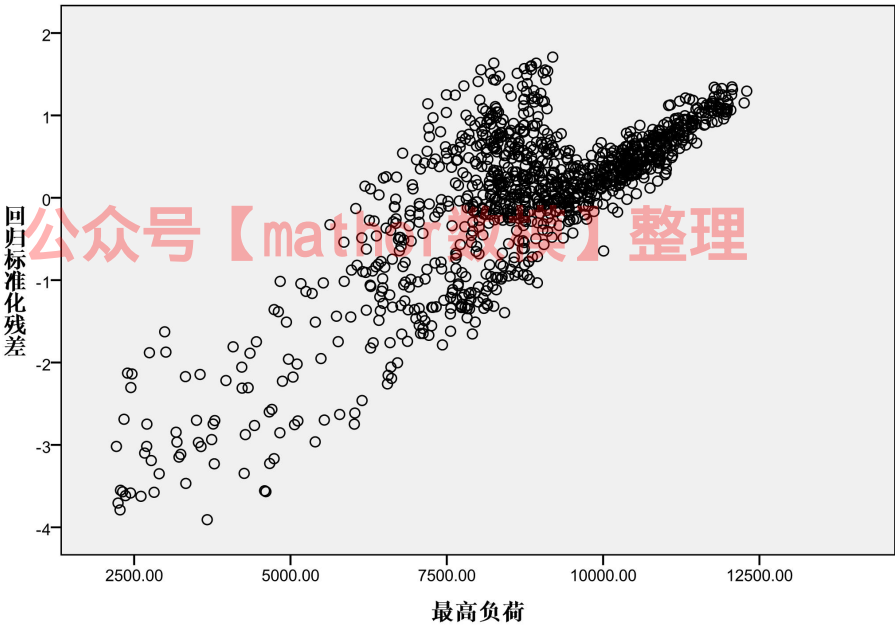


图 4 地区一最高负荷与各气象因素回归分析标准化残差图

从图中可以看出，残差图中的分布是随机的，可以看作没有出现趋势性，所以回归模型是有效的。最终的回归模型为：

$$y = 5604.140 - 33.5X_1 + 130.059X_2 + 105.834X_3 - 12.906X_4 + 5.856X_5。$$

用同样的分析过程可得两个地区各个因变量的回归分析，结果如下表：

表 5 各个回归方程汇总表

地区一	日最高	$y = 5604.140 - 33.5X_1 + 130.059X_2 + 105.834X_3 - 12.906X_4 + 5.856X_5$
	日最低	$y = 2886.322 - 16.266X_1 + 90.362X_2 + 53.659X_3 - 8.75X_4 + 3.956X_5$
	日平均	$y = 4401.141 - 24.384X_1 + 109.575X_2 + 78.188X_3 - 11.018X_4 + 4.37X_5$
地区二	日最高	$y = 6300.062 - 19.918X_1 + 18.774X_2 + 219.213X_3 - 21.968X_4 + 12.607X_5$
	日最低	$y = 2891.563 - 8.736X_1 + 19.088X_2 + 149.577X_3 - 17.684X_4 + 9.364X_5$
	日平均	$y = 4808.043 - 16.877X_1 + 18.501X_2 + 195.402X_3 - 20.913X_4 + 9.279X_5$

总的来说回归方程的有效性还是可以的，气候因素确实对负荷有影响。

5. 2. 3 多元线性回归误差的分析

本文将地区二的日平均负荷作为实例进行误差分析。我们知道两个因素之间的相关性可作为两个因素的相互影响程度的衡量标准，因此可以通过下表来得出一些结论：

表 6 地区二的日平均负荷与各因素的相关系数表

		日均负荷	最高温度	最低温度	平均温度	相对湿度	降雨量
Pearson 相关性	日均负荷	1.000	.715	.656	.740	.123	.119
	最高温度	.715	1.000	.795	.962	.138	.032
	最低温度	.656	.795	1.000	.878	.398	.177
	平均温度	.740	.962	.878	1.000	.278	.115
	相对湿度	.123	.138	.398	.278	1.000	.411
	降雨量	.119	.032	.177	.115	.411	1.000

从上表可以看出，相对湿度与日平均负荷的相关性为 0.123，降雨量与日平均负荷的相关性为 0.119。这两个相关系数都比较低，说明相对湿度和降雨量对日平均负荷的影响很少。如果将相对湿度与降雨量强行作为自变量的话，就会加大误差。因此如果将相对湿度与降雨量这两个因素从自变量中排除，可减小回归误差。可以对回归分析模型的汇总进行比较。

表 7 地区二日平均负荷与各气象因素回归分析的模型汇总

模型	R	R 方	调整 R2	标准估计误差	更改统计量				
					R2 改	F 更改	df1	df2	Sig. F 改
1	.750 ^a	.563	.561	1054.322281	.563	280.634	5	1089	.000

表 8 地区二日平均负荷与部分象因素回归分析的模型汇总

模型	R	R 方	调整 R2	标准估计误差	更改统计量				
					R2 改	F 更改	df1	df2	Sig.F 改
1	.741 ^a	.549	.547	1070.475794	.549	442.276	3	1092	.000

虽然最高的 R^2 即可决系数在去掉两个自变量后减小了一点为 0.549，但因为原始数据的减小，我们任然可以认为降雨量与相对湿度是造成误差加大的一个比较重要的原因。

5. 2. 4 为提高负荷预测精度对气象因素的选择

在 SPSS 软件中，有多种回归方法可供选择，现将回归方法改为逐步回归法。以地区二日最高负荷与各气象因素的回归分析为例，结果如下：

整理

公众号

表 9 地区二日最高负荷与部分象因素回归分析的模型汇总

模型	R	调整 R 方			标准估计误差	更改统计量				Durbin-Watson
		R 方	调整 R2			R2 更改	F 更改	df1	df2	Sig. F 改
1	.706	.498	.498		1315.65	.498	1086.35	1	1093	.000
2	.709	.503	.502		1309.989	.005	10.482	1	1092	.001
3	.715	.511	.510		1300.499	.008	16.999	1	1091	.000

由上表知模型 3 的可决系数为 0.511，但相互差别不大。模型模型拟合程度最高，DW 值为 0.459，通过检验，说明残差项不存在一阶自相关。

表 9 地区二日最高负荷与部分象因素回归分析的方差分析表

模型		平方和	df	均方	F	Sig.
1	回归	1.880E9	1	1.880E9	1086.358	.000 ^a
	残差	1.892E9	1093	1730958.153		
	总计	3.772E9	1094			
2	回归	1.898E9	2	9.492E8	553.132	.000 ^b
	残差	1.874E9	1092	1716071.308		
	总计	3.772E9	1094			
3	回归	1.927E9	3	6.424E8	379.824	.000 ^c
	残差	1.845E9	1091	1691292.379		
	总计	3.772E9	1094			

整理

公众号

上表中可明显看出模型1的F值最大，说明模型1的回归效果最显著。

表 10 地区二的日最高负荷与各因素的相关系数表

模型	非标准化系数		系数	t	Sig.	相关性			共线性统计量	
	B	标误	试用			零阶	偏	部分	容差	VIF
(常量)	4670.460	144.119		32.407	.000					
平均温度	209.531	6.357	.706	32.960	.000	.706	.706	.706	1.000	1.000
(常量)	5486.218	289.964		18.920	.000					
平均湿度	215.463	6.590	.726	32.697	.000	.706	.703	.697	.923	1.084
相对湿度	-11.995	3.705	-.072	-3.238	.001	.130	-.098	-.069	.923	1.084
(常量)	5932.160	307.513		19.291	.000					
平均温度	215.453	6.542	.726	32.935	.000	.706	.706	.697	.923	1.084
相对湿度	-18.564	4.008	-.111	-4.631	.000	.130	-.139	-.098	.777	1.287
降雨量	13.053	3.166	.096	4.123	.000	.133	.124	.087	.831	1.203

因为模型 1 的回归效果最显著，因此可以认为最好的回归方程为

$$y = 4670.460 + 209.531X_3。$$

同理，可得出其他经过筛选后的回归方程，结果如下表：

表 11 对气象因素筛选后各个回归方程汇总表

地区一	日最高负荷	$y = 4361.94 + 194.674X_3$
	日最低负荷	$y = 3008.140 + 90.362X_2 - 9.85X_4 + 4.129X_5$
	日平均负荷	$y = 3357.264 + 156.729X_3$
地区二	日最高负荷	$y = 4670.460 + 209.531X_3$
	日最低负荷	$y = 282.286 + 20.228X_2 + 139.249X_3 - 17.13X_4 + 9.484X_5$
	日平均负荷	$y = 3263.269 + 188.247X_3$

综上，可认为在诸气象因素中，优先推荐平均温度。

5.3 问题三的模型建立与求解

5.3.1 ARIMA 预测模型的建立

一个时间序列通常存在长期趋势变动、季节变动、周期变动和不规则变动因素。时间序列的目的就是逐一分解和测定时间序列中各项因素的变动程度和变动规律，然后将其重新综合起来，预测统计指标今后综合的变化的发展情况。

采用 ARIMA 模型对现有的数据进行建模，首要问题是确定模型的阶数，即相应的 p, q 值，对于 ARIMA 模型的识别主要是通过序列的自相关函数和偏自相关函数进行的。

序列 y_t 的自相关函数度量了 y_t 与 y_{t-k} 之间的线性相关程度，用 β_k 表示，定义如下：

$$\beta_k = \frac{r_k}{r_0}$$

式中： $r_k = \text{cov}(y_t, y_{t-k})$; $r_0 = \text{cov}(y_t, y_t)$ 表示序列的方差。

自相关函数刻画的是 y_t 与 y_{t-k} 分别与它们的中间部分 $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ 之间存在关系，如果在给定 $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ 之间的前提下，对 y_t 与 y_{t-k} 之间的条件相关关系进行刻画，则要通过偏自相关函数 φ_{kk} 进行，偏自相关函数可由下面的递推公式得到：

$$\begin{aligned} \varphi_{11} &= \beta_1 \\ \varphi_{kk} &= \frac{\beta_k - \sum_{j=1}^{k-1} \varphi_{k-1,j} \beta_{k-j}}{1 - \sum_{j=1}^{k-1} \varphi_{k-1,j} \beta_j} \\ \varphi_{k,j} &= \varphi_{k-1,j} - \varphi_{kk} \varphi_{k-1,k-j}, j = 1, 2, \dots, k-1 \end{aligned}$$

AIC 准则既考虑拟合模型对数据底接近程度，也考虑模型中所含待定参数的个数。

关于 ARIM(p, q)，对其定义 AIC 函数如下：

$$\text{AIC}(p, q) = n \ln(\hat{\sigma}^2) + 2(p + q)$$

其中 $\hat{\sigma}^2$ 是拟合 ARIM(p, q) 模型时残差的方差，它是 (p, q) 的函数。如果模型中含有常数项，则 $p+q$ 被 $p+q+1$ 代替。AIC 定阶的方法就是选择 ARIM(p, q) 最小的 (p, q) 作为相应的模型阶数。

模型阶数确定后，就可以估计模型。主要方法有三种估计方法:据估计，极大似然估计和最小二乘估计。最小二乘估计和极大似然估计的精度比较高，因而一般称为模型参数的精估计。

5. 3. 2 ARIMA 预测模型的求解

在数据处理的基础上，同样以地区一在时间点 T0000 的数据为例，做自相关分析，结果如下：

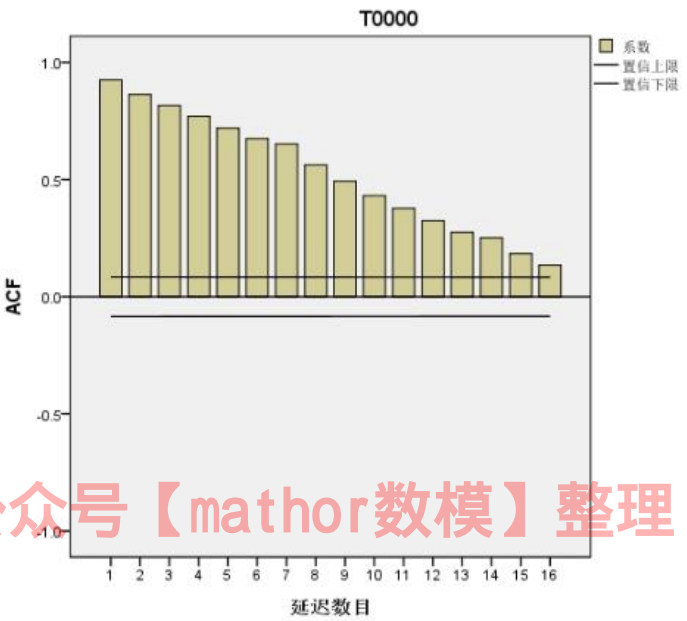


图 5 地区一 T0000 的 ACF 图

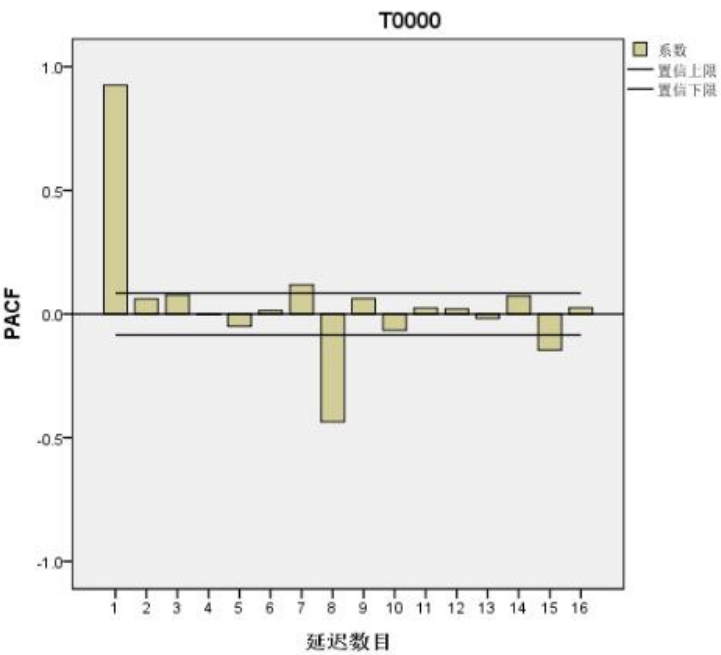


图 6 地区一 T0000 的 PACF 图

从图中可以看出，序列的自相关图（ACF）和偏自相关图（PACF）都是拖尾的，说明序列是非平稳的。数据序列通常不是平稳序列，但一般一阶差分都是平稳的，因此可以通过差分做进一步分析。

将差分设为 1，绘制差分序列的序列图如下：

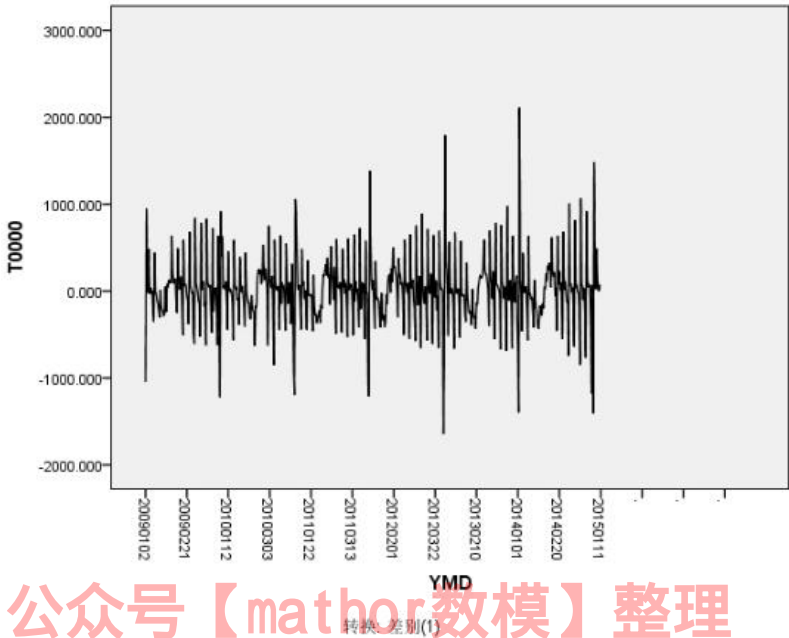


图 7 地区一 T0000 的差分序列图

由图可以知道，差分序列基本均匀分布在 0 刻度线上下两侧，因此可以认为差分序列是平稳的。

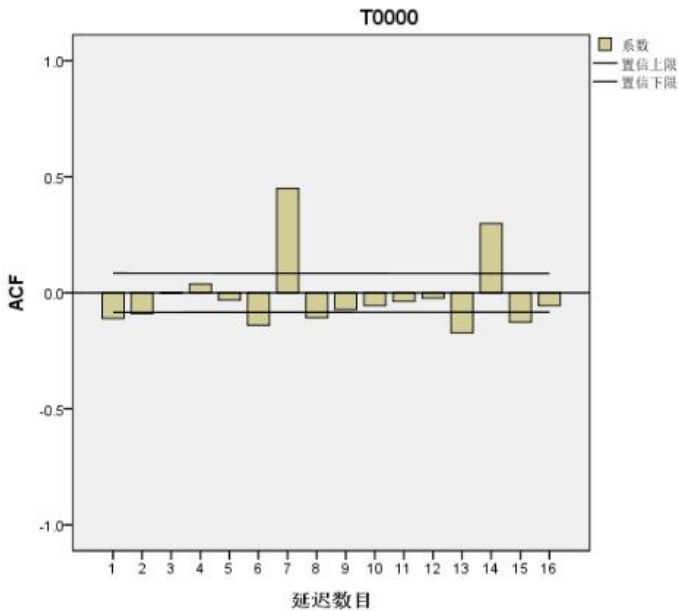


图 8 调整后地区一 T0000 的 ACF 图

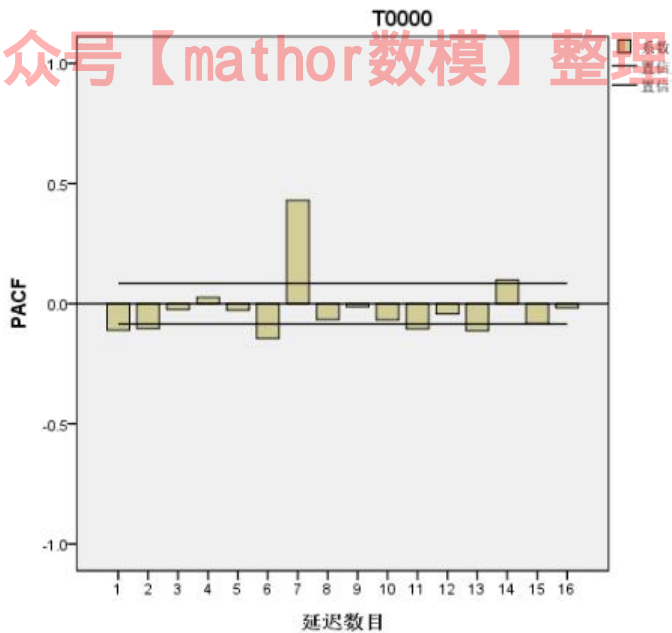


图 9 调整后地区一 T0000 的 PACF 图

由图可知，差分序列的 ACF 和 PACF 都是拖尾的，因此，可对序列建立 ARIMA (p,1,q) 模型。经过反复试验，确定模型为 ARIMA (1,1,1)，模型运行如下：依次点击“分析”，“预测”，“创建模型”，弹出时间序列建模器。可求出最后所需的结果，下表给出了地区一预测模型的部分统计量（详见附件 3、附件 4）：

表 12 地区一预测模型统计量

模型	预测变量数	模型拟合统计量		Ljung-Box Q (18)			离群值数
		平稳 R 方	R 方	统计量	DF	Sig.	
T0000-模型_1	1	.035	.859	216.160	16	.000	0
T0015-模型_2	1	.035	.861	208.964	16	.000	0
T0030-模型_3	1	.017	.858	186.099	16	.000	0
T0045-模型_4	1	.035	.861	200.881	16	.000	0
T0100-模型_5	1	.014	.858	179.677	16	.000	0
.....
T2245-模型_92	1	.044	.840	218.546	16	.000	0
T2300-模型_93	1	.048	.842	211.686	16	.000	0
T2315-模型_94	1	.049	.843	201.502	16	.000	0
T2330-模型_95	1	.050	.844	193.489	16	.000	0
T2345-模型_96	1	.051	.845	186.130	16	.000	0

从上表可看出 R^2 都在 0.8 以上，可证明拟合的结果比较科学。结果中给出了各个 p,q 的值，如下表所示：

表 13 地区一 ARIMA 预测模型参数

公众号【mathor数模】整理								
			估计		SE	t	Sig.	
T0000-模型_1	T0000	无转换	常数		4.113	4480.697	.001	.999
			AR	滞后1	.928	.023	40.350	.000
			差分		1			
			MA	滞后1	.999	.091	10.969	.000
	YMD	无转换	分子	滞后0	-7.771E-8	.000	.000	1.000
T0015-模型_2	T0015	无转换	常数		3.432	3987.218	.001	.999
			AR	滞后1	.923	.022	41.114	.000
			差分		1			
			MA	滞后1	1.000	.127	7.845	.000
	YMD	无转换	分子	滞后0	-4.978E-8	.000	.000	1.000
公众号【mathor数模】整理								
T2330-模型_95	T2330	无转换	常数		1401.964	16126.571	.087	.931
			AR	滞后1	.092	.179	.515	.607
			差分		1			
			MA	滞后1	.323	.170	1.900	.058
	YMD	无转换	分子	滞后0	-6.940E-5	.001	-.087	.931
T2345-模型_96	T2345	无转换	常数		1655.127	15410.761	.107	.915
			AR	滞后1	.067	.179	.376	.707
			差分		1			
			MA	滞后1	.302	.171	1.767	.078

同样拿地区一的 T0000 时间点举例，可得其预测模型如下：

$$\Delta y_t = 0.928\Delta y_{t-1} + \varepsilon_t - 0.999$$

用同样的方法可预测出地区二的指定七天的负荷量，部分结果如下（详见附件 Q3-Area1-Load、附件 Q3-Area2-Load）：

表 13 地区二 ARIMA 预测结果

YMD	T0000	T0015	T0030	T0045	T2300	T2315	T2330	T2345
20150111	6198.557	6007.202	5863.55	5700.1	7091.34	6820.8	6546.5	6245.16
20150112	6205.14	6010.508	5871.95	5704.86	7088.19	6821.1	6548.83	6246.7
20150113	6209.693	6015.623	5876.24	5709.49	7096.02	6828.13	6554.84	6252.43
20150114	6214.616	6020.377	5881.05	5714.13	7101.93	6834	6560.52	6257.86
20150115	6219.471	6025.203	5885.8	5718.76	7108.17	6840.06	6566.23	6263.31
20150116	6224.338	6030.015	5890.55	5723.39	7114.36	6846.1	6571.93	6268.76
20150117	6229.203	6034.829	5895.3	5728.02	7120.56	6852.14	6577.64	6274.21

5. 4 问题 4 的模型建立与求解

5. 4. 1 多元非线性模型

当有 q 个应变变量 $y = (y_1, \dots, y_q)'$ 时, 而 $x = (x_0, x_1, \dots, x_{p-1})'$, $x_0 \equiv 1$, 的是:

$$Y = XB + U$$

其中 $\varepsilon(U) = 0$, $\text{Var}(\text{vec}(U)) = I_n \otimes \Sigma$, $\Sigma > 0$ 式中 $Y: n \times q$ 是应变量的 n 组随机独立

抽样的观察值矩阵, $X: n \times p$ 是对应于 Y 的自变量的已知的观察值矩阵, $B: p \times q$ 是未知的回归系数矩阵, $U: n \times q$ 是未知的随机误差矩阵, 一般称为残差阵。

与一元的线性模型一样, 多元方差分析及多元协方差分析。

一般, 在线性模型中多假设有下分布:

$$U \sim N_{n \times q}(0, I_n \otimes \Sigma), \Sigma > 0$$

与上假设等价的是

$$Y \sim N_{n \times q}(XB, I_n \otimes \Sigma)$$

5. 4. 2 基于 BP 神经网络算法的多元非线性系统模型的建立

在科学研究和生产实践中, 对具有表现系统特征或运行状态的离散数据进行建模, 用于系统预测、评价等, 是科学决策和决策系统建立的重要基础。由于大多数研究对象普遍具有多变量且依从高度非线性关系等特征, 因此多元非线性系统建模极其重要。

人工神经网络是由大量简单的处理单元(神经元)广泛地互相连接形成的复杂非线性系统。它不需要任何先验公式, 可直接从训练样本(离散数据)中自动归纳规则, 提取离散数据之间复杂的依从关系(可以是高度非线性关系), 储存于网络权重之中, 从而建立研究问题的神经网络模型。其中由 Rumelhart 提出的多层前馈神经网络, 由于采用误差反传的学习算法, 被称为 BP 网络, 其应用非常广泛。在理论上已经证明具有三层结构(一个隐含层)的 BP 网络能够逼近任何有理函数。

题目中给出了 5 个自变量、1 个因变量。有三层 BP 神经网络模型逼近存在于样本数据间的函数关系, 其模型为 $y = ANN(x_1, x_2, x_3, x_4, x_5)$, 这是一个非线性函数。此模型为隐含表达式, 即不能用通常数学公式表示, 故称为“知识库”。

5. 4. 3 基于 BP 神经网络算法的多元非线性系统模型的求解

根据这个多元非线性系统模型，利用 Matlab 编程可训练出相应的神经网络结构。首先还是考虑负荷的季节性很明显，为排除季节因素对负荷的影响，将数据预处理。只留下两地区每年春季的负荷量数据，以及两地区每年春季的各气候因素的数据作为预测的原始数据。

利用 Matlab 编程（详见附录二、附录三），可预测出指定七天的负荷量，地区一的负荷量预测结果如下（详见附件 Q4-Area1-Load、附件 Q4-Area2-Load）：

表 14 地区一负荷量预测结果

YMD	T0000	T0015	T0030	T0045	……	T2300	T2315	T2330	T2345
20150111	4455.22	4332.89	4313.60	4319.68	……	5182.73	4969.39	4766.26	4609.26
20150112	4692.49	4202.82	4318.70	4399.88	……	5659.89	5488.15	4961.61	5189.43
20150113	4692.49	4202.82	4318.70	4399.88	……	5659.89	5393.79	4961.61	5189.43
20150114	4455.22	4332.89	4318.70	3862.89	……	5035.62	4969.39	4766.26	4373.57
20150115	4455.22	4332.89	4318.70	4149.47	……	5182.73	4969.39	4766.26	4373.57
20150116	4455.22	4332.89	4318.70	4298.76	……	5182.73	4969.39	4766.26	4609.26
20150117	4399.20	4332.89	4306.45	4218.56	……	5288.23	4969.39	4766.26	4609.26

5. 4. 4 预测结果与原有的预测结果进行对比

本模型得出的预测结果与原有预测结果进行比较，得出的结论是计及气象因素影响的负荷预测结果精度得到了改善，我们可以从以下几个方面来论证：

1. 从预测参照的数据进行论证

在解决第三问时，参照的数据是往年的负荷量；而在第四问中参照的数据是往年的负荷量以及各个气象因素的参数。我们知道气象因素对负荷量的影响是比较大的，如果单纯考虑往年负荷量的变化是比较片面的，计及气象因素影响的负荷量预测的结果精度会更高。

2. 从模型的原理进行论证

ARIMA 模型是基于考虑一个时间序列通常存在长期趋势变动、季节变动、周期变动和不规则变动因素。依赖时间的变动较强，但假设将要预测的日期中气候条件突然骤变，可想而知负荷量也会变化比较大，这时 ARIMA 模型所预测出的结果的可信度就不会很高。如果计及气象因素影响进行负荷量预测，而且是通过 MATLAB 训练所形成的神经网络而得来的预测，其精度明显高于前者。

5. 4 问题 5 的模型建立与求解

首先，确定用日负荷率作为观察数据，因为日负荷率是一个综合的指标，其变化能准确的反映负荷量。利用 Matlab 软件做出负荷率变化曲线如下：

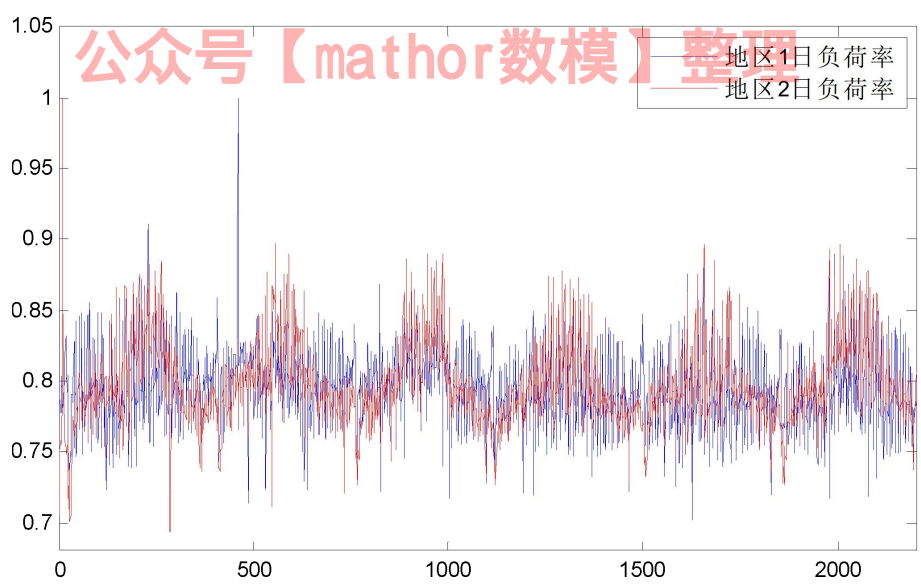


图 10 两地区日负荷率变化曲线

由上图可知：地区二日负荷率分布更加集中，波动性较地区一小，初步判定地区二的负荷量更加规律。

接下来利用 Matlab 软件做出地区一的负荷率分布直方图并作正态拟合、Logistic 拟合以及 T 分布拟合，结果如下：

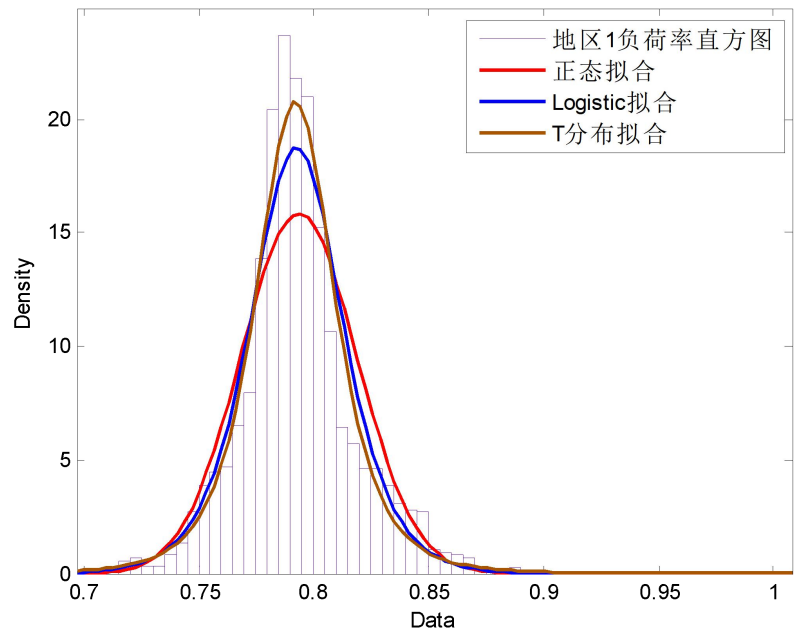


图 11 地区一各个拟合结果

用同样的方法对地区二的负荷率分布直方图并作正态拟合、Logistic 拟合以及 T 分布拟合，结果如下：

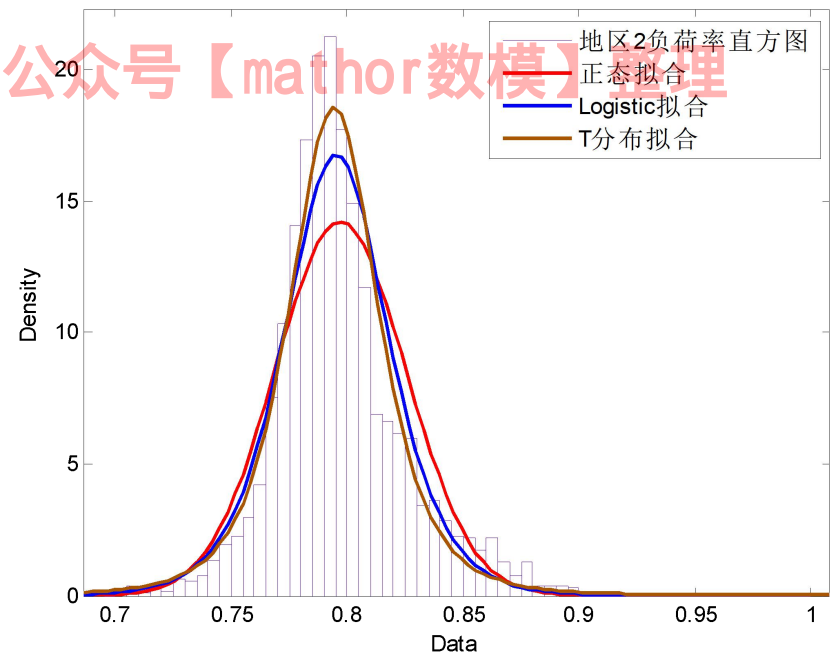


图 12 地区二各个拟合结果

对地区一、地区二日负荷率的拟合结果进行拟合曲线的广义似然比检验，结果如下：

表 15 两地区拟合曲线广义似然比检验

	正态拟合	Logistic 拟合	T 分布拟合
地区 1 负荷率	4981.55	5066.27	5080.16
地区 2 负荷率	4740.42	4818.75	4832.94

这里需要说明的是，一般认为似然比越大拟合效果越好，但在实际求解中将最大似然比乘以了-2，所以所得的结果越小越好。地区二的广义似然比检验都比地区一的小，所以可以得出结论：地区二的数据比地区一的数据更有规律，更适合研究。

6. 模型结果的分析与检验

6.1 对线性回归预测模型的显著性检验

在对线性模型求解之前必须对于该问题是否具有线性回归模型进行显著性检验。由回归模型 $y = \beta_0 + \beta_1 x_i + \varepsilon, \varepsilon \sim (0, \delta^2)$ 可知, 当 $\beta_1 = 0$ 时, 就认为 Y 与 x 之间不存在线性回归关系, 故检如下假设:

$$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$$

为了检验假设 H_0 , 先分析样本 y_1, y_2, \dots, y_n 的差异, 它可以用总的偏差平方和来度量, 记为

$$S_{\text{总}} = \sum_{i=1}^n (y_i - \bar{y})^2$$

由正规方程组, 有

$$S_{\text{总}} = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

利用 F 检验发, 当 H_0 为真时, 取统计量

$$F = \frac{S_{\text{回}}}{S_{\text{剩}}(n-2)} \sim F(1, n-2)$$

由给定显著性水平 α , 查表得 $F_{\alpha}(1, n-2)$, 根据数据计算 F 值, 若 $F > F_{\alpha}(1, n-2)$ 时拒绝 H_0 , 表明回归效果显著; 若 $F \leq F_{\alpha}(1, n-2)$ 时, 接受 H_0 , 此时回归效果不显著。

对地区二日平均负荷量与气象因素的关系作为举例, 利用 SPSS 软件对原始数据进行线性显著性检验, 结果如下:

表 14 地区二线性显著性检验结果

模型		平方和	df	均方	F	Sig.
4	回归	1.559E9	4	3.898E8	350.744	.000 ^d
	残差	1.211E9	1090	1111234.492		
	总计	2.770E9	1094			

所求 F 值为 350.744, Sig 值为 0.000, 所以, 其显著性概率远小于 0.01, 所一显著的拒绝总体回归系数为 0 的假设。

6.1 对 ARIMA 预测模型的显著性检验

利用 SPSS 软件能比较容易得求出 ARIMA 预测模型的显著性等各参数的检验, 用地区一的负荷量预测模型举例, 显著性检验结果如下:

表 15 地区一 ARIMA 预测模型显著性检验结果

模型	预测变量数	模型拟合统计量 Ljung-Box Q (18)					离群值数
		平稳 R 方	R 方	统计量	DF	Sig.	
T0000-模型_1	1	.035	.859	216.160	16	.000	0
T0015-模型_2	1	.035	.861	208.964	16	.000	0
T0030-模型_3	1	.017	.858	186.099	16	.000	0
T0045-模型_4	1	.035	.861	200.881	16	.000	0
.....
T2300-模型_93	1	.048	.842	211.686	16	.000	0
T2315-模型_94	1	.049	.843	201.502	16	.000	0
T2330-模型_95	1	.050	.844	193.489	16	.000	0
T2345-模型_96	1	.051	.845	186.130	16	.000	0

Sig 值都为 0.000，其显著性概率远远小于 0.01，所以该模型具有显著性。

7. 模型的推广与改进方向

7. 1 模型的推广

由于大多数研究对象普遍具有多变量且依从高度非线性关系等特征，因此多元非线性系统建模极其重要。人工神经网络是由大量简单的处理单元（神经元）广泛地互相连接形成的复杂非线性系统。它不需要任何先验公式，可直接从训练样本（离散数据）中自动归纳规则，提取离散数据之间复杂的依从关系，储存于网络权重之中，从而可以建立研究问题的神经网络模型。

7. 2 模型的改进

BP 神经网络也存在识别精度依赖于系统的训练数据、训练方法和训练精度，神经网络的拓扑结构只有经过相当的尝试后才能最终确定，没有筛选主导因子的能力等缺点。因此，加强对网络权重，阈值的调整，优化网络结构或和多元回归联合使用，不失为一个新的研究方向。

8. 模型的优点

- 1.在第一问中，建立日最高负荷量模型、日最低负荷量模型、日峰谷差模型、日平均负荷量模型以及日负荷率模型，简单直接并且准确地解决了问题；
- 2.在问题三中，在构造 ARIMA 预测模型之前先进行季节划分，排除季节性影响，使预测更准确，计算的复杂度也会降低
- 3.在问题四中，结合 BP 神经网络算法构建多元非线性系统模型，计及气象因素进行负荷量预测，精度更高
- 4.在问题五中，将日负荷率作为参照数据准确且全面，将两地日负荷率进行多个分布种类拟合，得出结论可信度高。

参考文献

- [1] 韩中庚. 数学建模方法及其应用 (第二版) [M]. 北京: 高等教育出版社, 2009.
- [2] 刁在筠. 运筹学[M]. 北京: 高等教育出版社, 2007.
- [3] 傅家良. 运筹学方法与模型 (第二版) [M]. 上海: 复旦大学出版社, 2014.
- [4] 张杰, 郭丽杰, 周硕等. 运筹学模型及其应用[M]. 北京: 清华大学出版社, 2012.
- [5] 卓金武. MATLAB 在数学建模中的应用 [M]. 北京: 北京航空航天大学出版社, 2014.
- [6] 杜栋, 庞庆华, 吴炎. 现代综合评价方法与案例精选[M]. 北京: 清华大学出版社, 2014.
- [7] 张学敏. MATLAB 基础及应用 (第二版) [M]. 北京: 中国电力出版社, 2012.

附录

附录1 两地区2014年全年的负荷持续曲线matlab代码

```
function createfigure1 (YMatrix1)
figure1 = figure ('PaperSize',[20.98404194812 29.67743169791]);
axes1 = axes ('Parent',figure1);
xlim (axes1,[0 365]);
box (axes1,'on');
hold (axes1,'all');
% 使用 plot 的矩阵输入创建多行
plot1 = plot (YMatrix1,'Parent',axes1);
set (plot1 (1), 'Color',[1 0 0],'DisplayName','地区1');
set (plot1 (2), 'Color',[0 0.447058826684952 0.74117648601532],...
'DisplayName','地区2');
% 创建 xlabel
xlabel ('2014年 (单位: 天)');
% 创建 ylabel
ylabel ('负荷 (单位: MW)');
% 创建 legend
legend (axes1,'show');
end

clc,clear;
load ('2014年平均负荷变化数据.mat')
createfigure1 (average2014);
saveas (gcf,'2014年平均负荷变化图');
```

附录2 基于BP神经网络算法的多元非线性系统模型地区一Matlab程序

```
Y = zeros (7,96);
for i = 1 :96
load ('地区1数据.mat');
p = [T_max1 (1:281);T_min1 (1:281);T_average1 (1:281);Shidu1 (1:281);Rain1 (1:281)
'];
t = Area1_4 (:,i);
s1 = 9; s2 = 1;
w = [min (T_max1 (1:281)),max (T_max1 (1:281));min (T_min1 (1:281)),max (T_min1 (1:281));min (T_average1 (1:281)),max (T_average1 (1:281));min (Shidu1 (1:281)),max (Shidu1 (1:281));min (Rain1 (1:281)),max (Rain1 (1:281))];
net = newff (w,[s1,s2],{'tansig','purelin'},'trainlm');
net.trainParam.show = 50;
net.trainParam.lr = 0.05;
net.trainParam.epochs = 500;
net.trainParam.goal = 1e-6;
net = train (net,p,t);
pp = [T_max1 (282:288);T_min1 (282:288);T_average1 (282:288);Shidu1 (282:288);Rain1 (282:288)
'];
a = sim (net,pp);
```



```
Y(:,i) = a;
end
```

附录3 基于BP神经网络算法的多元非线性系统模型地区二Matlab程序

```
Y = zeros (7,96) ;
for i = 1 :96
load ('地区 2 数据.mat') ;
p = [T_max2 (1:281)';T_min2 (1:281)';T_average2 (1:281)';Shidu2 (1:281)';Rain2 (1:281)
'];
t = Area2_4 (:,i)';
s1 = 9; s2 = 1;
w = [min (T_max2 (1:281)) ,max (T_max2 (1:281)) ;min (T_min2 (1:281)) ,max (T_min2
(1:281)) ;min (T_average2 (1:281)) ,max (T_average2 (1:281)) ;min (Shidu2 (1:281)) ,max
(Shidu2 (1:281)) ;min (Rain2 (1:281)) ,max (Rain2 (1:281)) ];
net = newff (w,[s1,s2],{'tansig','purelin'},'trainlm') ;
net.trainParam.show = 50;
net.trainParam.lr = 0.05;
net.trainParam.epochs = 500;
net.trainParam.goal = 1e-6;
net = train (net,p,t) ;
pp = [T_max2 (282:288)';T_min2 (282:288)';T_average2 (282:288)';Shidu2 (282:288)';Rain2
(282:288)'];
a = sim (net,pp)';
Y(:,i) = a;
End
```

附录4 两地区日负荷率变化曲线Matlab程序

```
function createfigure (YMatrix1)
figure1 = figure;

axes1 = axes ('Parent',figure1) ;
%% 取消注释以下行以保留轴的 X 极限
xlim (axes1,[0 2201]) ;
%% 取消注释以下行以保留轴的 Y 极限
ylim (axes1,[0.68 1.05]) ;
box (axes1,'on') ;
hold (axes1,'all') ;

% 使用 plot 的矩阵输入创建多行
plot1 = plot (YMatrix1,'Parent',axes1) ;
set (plot1 (1) ,'DisplayName','地区 1 日负荷率') ;
set (plot1 (2) ,'Color',[1 0 0],'DisplayName','地区 2 日负荷率') ;

% 创建 legend
legend (axes1,'show') ;

load ('负荷率.mat') ;
YMatrix1=[fuhelv1,fuhelv2];
createfigure (YMatrix1) ;
```

附录5 两地区日负荷率拟合 Matlab 程序

```
function [pd1,pd2,pd3]=NiHe (fuhelv2,i)
t = strcat ('地区',num2str (i));
t = strcat (t,'负荷率直方图');
fuhelv2 = fuhelv2 (:);
clf;
hold on;
LegHandles = []; LegText = {};
[CdfF,CdfX] = ecdf (fuhelv2,'Function','cdf'); % compute empirical cdf
BinInfo.rule = 1;
[~,BinEdge] = internal.stats.histbins (fuhelv2,[],[],BinInfo,CdfF,CdfX);
[BinHeight,BinCenter] = ecdfhist (CdfF,CdfX,'edges',BinEdge);
hLine = bar (BinCenter,BinHeight,'hist');
set (hLine,'FaceColor','none','EdgeColor',[0.333333 0 0.666667],...
    'LineStyle','-','LineWidth',1);
xlabel ('Data');
ylabel ('Density');
LegHandles (end+1) = hLine;
LegText{end+1} = t;
XLim = get (gca,'XLim');
XLim = XLim + [-1 1] * 0.01 * diff (XLim);
XGrid = linspace (XLim (1),XLim (2),100);
% --- Create fit "正态拟合"
pd1 = fitdist (fuhelv2,'normal');
YPlot = pdf (pd1,XGrid);
hLine = plot (XGrid,YPlot,'Color',[1 0 0],...
    'LineStyle','-','LineWidth',2,...
    'Marker','none','MarkerSize',6);
LegHandles (end+1) = hLine;
LegText{end+1} = '正态拟合';
% --- Create fit "Logistic 拟合"
pd2 = fitdist (fuhelv2,'logistic');
YPlot = pdf (pd2,XGrid);
hLine = plot (XGrid,YPlot,'Color',[0 0 1],...
    'LineStyle','-','LineWidth',2,...
    'Marker','none','MarkerSize',6);
LegHandles (end+1) = hLine;
LegText{end+1} = 'Logistic 拟合';
% --- Create fit "T 分布拟合"
pd3 = fitdist (fuhelv2,'tlocationscale');
YPlot = pdf (pd3,XGrid);
hLine = plot (XGrid,YPlot,'Color',[0.666667 0.333333 0],...
    'LineStyle','-','LineWidth',2,...
    'Marker','none','MarkerSize',6);
LegHandles (end+1) = hLine;
LegText{end+1} = 'T 分布拟合';
box on;
hold off;
hLegend = legend (LegHandles,LegText,'Orientation','vertical','Location','NorthEast');
set (hLegend,'Interpreter','none');
```

```
load('负荷率.mat');
[pd1_1,pd1_2,pd1_3]=NiHe(fuhelv1,1);
saveas(gcf,'拟合结果图/地区 1 拟合图');
[pd2_1,pd2_2,pd2_3]=NiHe(fuhelv2,2);
saveas(gcf,'拟合结果图/地区 2 拟合图');
```

附录 6 地区一日负荷率各个拟合结果

地区 1 负荷率正态拟合结果

Distribution: Normal
Log likelihood: 4981.55
Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.793598
Variance: 0.00063363

Parameter	Estimate	Std. Err.
mu	0.793598	0.000536547
sigma	0.025172	0.000379525

Estimated covariance of parameter estimates:

	mu	sigma
mu	2.87883e-07	-5.50596e-21
sigma	-5.50596e-21	1.4404e-07

地区 1 负荷率 Logistic 拟合结果

Distribution: Logistic
Log likelihood: 5066.27
Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.792592
Variance: 0.000584075

Parameter	Estimate	Std. Err.
mu	0.792592	0.000485426
sigma	0.0133243	0.000241861

Estimated covariance of parameter estimates:

	mu	sigma
mu	2.35638e-07	3.40325e-09
sigma	3.40325e-09	5.8497e-08

地区 1 负荷率 t 分布拟合结果

Distribution: t Location-Scale
Log likelihood: 5080.16
Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.792086
Variance: 0.00072776

Parameter	Estimate	Std. Err.
-----------	----------	-----------

mu	0.792086	0.00046106
sigma	0.0179131	0.000582481
nu	3.57726	0.35821

Estimated covariance of parameter estimates:

	mu	sigma	nu
mu	2.12576e-07	4.29803e-08	2.79561e-05
sigma	4.29803e-08	3.39284e-07	0.00016002
nu	2.79561e-05	0.00016002	0.128315

附录 7 地区二日负荷率各个拟合结果

地区 2 负荷率正态拟合结果

Distribution: Normal
Log likelihood: 4740.42
Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.797101
Variance: 0.000788849

Parameter	Estimate	Std. Err.
mu	0.797101	0.000598669
sigma	0.0280864	0.000423467

Estimated covariance of parameter estimates:

	mu	sigma
mu	3.58405e-07	-1.02676e-20
sigma	-1.02676e-20	1.79325e-07

地区 2 负荷率 Logistic 拟合结果

Distribution: Logistic
Log likelihood: 4818.75
Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.795282
Variance: 0.000732158

Parameter	Estimate	Std. Err.
mu	0.795282	0.000544638
sigma	0.0149181	0.000270353

Estimated covariance of parameter estimates:

	mu	sigma
mu	2.96631e-07	7.81813e-09
sigma	7.81813e-09	7.30907e-08

地区 2 负荷率 T 分布拟合结果

Distribution: t Location-Scale
Log likelihood: 4832.94

Domain: $-\text{Inf} < y < \text{Inf}$
Mean: 0.794274
Variance: 0.000910133

Parameter	Estimate	Std. Err.
mu	0.794274	0.000537099
sigma	0.0200522	0.000637353
nu	3.5829	0.350257

Estimated covariance of parameter estimates:

	mu	sigma	nu
mu	2.88475e-07	9.89018e-08	5.7654e-05
sigma	9.89018e-08	4.06219e-07	0.000168696
nu	5.7654e-05	0.000168696	0.12268