

西安电子科技大学 2024 年数学建模校内赛

承 诺 与 产 权 转 让 书

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们同意将参赛论文以及支撑材料中的所建模型、算法以及程序产权归属西安电子科技大学所有。数学与统计实训基地可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等）。

我们参赛选择的题号是（从 A/B/C 中选择一项填写）： C

参赛报名队号为 005

队长所属学院（书院）（请填写完整的全名）： 丁香书院

参赛队员姓名与学号（此处附电子签名）：

1. 雷毅东 23009200417

2. 郭为峰 23009100137

3. 袁思达 23009290033

日期： 2024 年 5 月 4 日

西安电子科技大学 2024 年大学生数学建模校内赛

评 阅 专 用 页

	评阅人 1	评阅人 2	评阅人 3	总评
成绩				

对架空输电线路动态增容系统的计量预测与宏观分析

摘 要

随着近几年国民经济的迅速发展,各行业对电力的需求量大大增加。输电线路作为电力系统的重要组成部分,输电线路的输电效率亟待提高,以满足一天之内各行业的电力需求。

针对问题一的第一问,考虑用 **BP 神经网络** 预测 2020 年 8 月的动态载流量需求。对于 **BP 神经网络** 的各个变量,首先建立**时间序列预测**模型,考虑关键影响因素 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]。为确保数据为完整的连续数据,利用 Python 处理 DEV 1 的 2019 年 7 月至 2020 年 7 月时间段内的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据。接着使用插值算法,将样本数据转换为**连续数据**。以每小时各样本数据为原始数据,预测 DEV 1 的 2020 年 8 月的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据。最后利用 **BP 神经网络算法**预测 2020 年 8 月的动态载流量需求。结果可得在 2020-8-12 12:00 之前,Line Current[A]振幅较小。在 2020-8-12 12:00 之后,振幅出现了明显的增大趋势,最高峰值 Line Current[A]达到了 340[A]左右,这与夏季用电高峰有关。

针对问题一的第二问,使用**傅立叶变换分析法**对数据进行处理。首先用使用**三次 Hermite 插值**对缺失的载流量数据进行补充。然后使用 Python 对数据进行**傅立叶变换**,得到**傅立叶变换**后的频率图像。分析频率的峰值可获知数据的周期性变化情况。结果可得载流量的变化有明显的周期性,期中 24h 的周期性最为显著。

针对问题二,首先根据变量之间的 **Spearman 相关系数**,选择与载流量相关性较强的变量。然后利用 Stata 进行多元线性回归分析。接着对回归方程进行联合相关性检验和 White 检验,根据检验结果,判断回归方程的回归效果,最后根据回归方程分析温度、弧垂角度对载流量的定量影响。

针对问题三,对于附件给出的多组应力数据,使用**因子分析**提取公共因子,简化模型分析。首先使用 SPSS 软件对应力数据进行**因子分析**。结果分别从两侧的应力数据中提取了**两对公共因子**。对于与多组自变量相关的两组因变量之间关系的分析,使用**典型相关分析**,得到两组因变量之间的相关关系。使用 SPSS 对两侧应力公共因子变量与载流量进行**典型相关分析**,得到两侧应力与载流量之间的关系,提取了一对**典型变量**。对于题目中给出的弧垂角与应力成线性变化的条件,最后使用**一元线性回归**,建立一元线性回归方程。

针对问题四,题目要求给出动态增容方案并给出最大增容量。对于含有较多位置信息的变量,使用**灰色预测模型**进行预测。利用 Python 对提取的具有和预测变量具有类似性质的 13 个变量进行**灰色预测**,得到 2020 年 8 月的最大载流量预测值,进而计算出最大增容量。最后依据动态增容的工程实例,结合相关文献与实际背景,给出**动态增容方案**。

关键词: BP 神经网络、傅里叶变换、多元线性回归、因子分析、典型相关分析、灰色预测

一、问题重述

1.1 问题背景

随着近几年国民经济的迅速发展，各行业对电力的需求量大大增加。输电线路作为电力系统的重要组成部分，输电线路的输电效率亟待提高，以满足一天之内各行业的电力需求。

架空输电线路(Overhead transmission line, OHTL)由线路杆塔、导线等装置构成。电流的热效应会导致电线温度过高，电线温度直接决定弧垂的大小。因此，高架空输电线路不可避免出现弧垂降低、导线变形等问题，若时间过长，架空输电线路易发生电力运行故障。为了解决安全隐患，考虑架空输电线路的极限温度和弧垂极限很有必要。导线的温度与环境温度、导线内电流的大小、弧垂角度有关。在输电线路上安装温度在线检测装置(Overhead Temperature Line Monitoring, OTLM)是一种监控电流的技术。通过该方法，可以获取温度、弧垂角度等数据，进而可以获知输电效率与时间和耗电之间的关系。

动态增容技术(dynamic line rating, DLR)是目前解决该问题的主要方法。该方法由早期发展的热方程理论所确定，根据环境变化随时间动态调整输电线路容量。它包括环境调节式 DLR 和实时监测式 DLR。通过研究该技术，可以解决电力供不应求的问题。

1.2 问题要求

附件给出了在线监测装置 OTLM1 与 OTLM2 设备采集的数据，记录了 2019 年 7 月至 2020 年 7 月一年内等时段间隔的电流负载状态。为了满足一天之内所有行业耗电需求，特别是高峰时段用电需求，现需结合实际情况与所给信息建立数学模型，分析以下问题：

问题一：结合相关文献，依据 OTLM1 设备的采样数据，整理相关数据，建立数学模型预测 2020 年 8 月的动态载流量需求，并分析不同时间间隔载流量的变化情况，分析电流的周期性特征。

问题二：综合考虑 OTLM1 和 OTLM2 采集的数据，分析温度、弧垂角度对载流量的定量影响。

问题三：附件中 OTLM 设备的输出数据记录了具有 L 型切面四脚钢结构的架空输电线路温度以及振动数据。在 L 型切面脚钢结构中，应力反映了导线内张力的变化，而张力与导线温度有关，张力与 OTLM1 与 OTLM2 设备监测到的弧垂角度呈线性关系。结合图 5 和图 6 的相关信息，建立数学模型来讨论导线电流变化与钢架腿部每侧应力变化之间的关系。

问题四：已知 2020 年 8 月 15 日电用量有大幅度增加，整合分析附件中的相关数据，给出相应的增容方案，并给出最大增容量。

二、问题分析

2.1 问题一的分析

针对问题一的第一问，考虑用 BP 神经网络预测 2020 年 8 月的动态载流量需求。对于 BP 神经网络的各个变量，首先建立时间序列预测模型，由题可知，导线温度直接决定弧垂的大小。根据数据侧写中的 Spearman 相关系数判断，现考虑关键影响因素 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]。为确保数据为完整连续数据，利用 Python 处理 DEV 1 的 2019 年 7 月至 2020 年 7 月时间段内的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据，接着使用三次

Hermite 插值和**三次样条插值**，将样本数据转换为**连续数据**。以每小时各样本数据为原始数据，预测 DEV 1 的 2020 年 8 月的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据。最后运用预测的数据，利用 **BP 神经网络算法**，设置合理的神经网络的结构，预测 2020 年 8 月的动态载流量需求。针对问题一的第二问，对于周期性变化的数据，考虑使用**傅立叶变换分析法**对数据进行处理，从而检验数据的周期性。首先用使用**三次 Hermite 插值**对缺失的载流量数据进行补充。然后使用 Python 对数据进行傅立叶变换，得到**傅立叶变换**后的频率图像。分析频率的峰值可获知数据的周期性变化情况。

2.2 问题二的分析

针对问题二，首先根据变量之间的 **Spearman 相关系数**，选择与载流量相关性较强的变量，然后利用 Stata 进行**多元线性回归分析**，得到载流量与其他因变量之间的回归方程，接着对回归方程进行联合相关性检验和 White 检验，根据检验结果，判断回归方程的回归效果，最后根据回归方程分析温度、弧垂角度对载流量的**定量影响**。

2.3 问题三的分析

针对问题三，对于附件给出的多组应力数据，考虑使用**因子分析**提取公共因子，简化模型分析。首先使用 SPSS 软件对应力数据进行**因子分析**，得到两侧应力的公共因子变量。对于与多组自变量相关的两组因变量之间关系的分析，通常使用**典型相关分析**，得到两组因变量之间的相关关系。使用 SPSS 对两侧应力公共因子变量与载流量进行典型相关分析，得到两侧应力与载流量之间的关系。对于题目中给出的**弧垂角与应力成线性变化**的条件，最后使用**一元线性回归**，建立一元线性回归方程，阐释弧垂角与两侧应力之间的关系。

2.4 问题四的分析

针对问题四，题目要求给出动态增容方案并给出最大增容量。对于含有较多位置信息的变量，常常使用**灰色预测模型**进行预测。利用 Python 对提取的具有和预测变量具有类似性质的 13 个变量进行灰色预测，得到 2020 年 8 月的最大载流量预测值，进而计算出最大增容量。最后依据动态增容的工程实例，结合相关文献与实际背景，给出动态增容方案。

三、模型假设

- 1. 假设 Stress (MPa) 与 Line Current[A]满足联合正态分布排除生活中的小概率事件(例如黑天鹅事件、非正常情况)。
- 2. 假设前 10 个月载流量峰值由用电量大幅度增加引起使用模型中要求的假设。
- 3. 假设 2020 年 8 月 15 日最大载流量为包含较多未知信息的灰色预测预测量。

四、符号说明

符号	说明	单位
Line Temperature	导线温度(LT)	° C
Ambient Temperature OTLM	环境温度(AT)	° C
Line Current	载流量(LC)	A
Humidity	湿度(H)	%

Angle	弧垂角(AG)	°
Altitude	海拔(AL)	M
Vbat	电压(V)	V
Stress	应力(ST)	MPa

五、数据侧写

鉴于数据较为**复杂和庞大**，为更好地理解和分析数据，需在进行问题求解前提前进行数据分析。根据题目要求，综合附件中全部信息和数据，并利用 Excel、Python、MATLAB 和 SPSS 等工具，对数据进行**多角度、多标准的整理与观察**，实现初步了解所有数据的大致情况的目的，便于后续对问题的解答

5.1 利用 Python 和 Excel 处理数据

首先，附件中数据存在小数格式表示错误，数字格式为货币等问题，利用 Excel 的宏功能，在 Excel 中定义函数，对所有数据格式进行处理，纠正了数据格式错误的问题。

其次，数据并非完全间隔 5 分钟，为了便于后续问题的解决，利用 Python 编写脚本，对部分数据进行处理，以每小时内所有数据的平均值作为该小时的数据，从而**减小数据不连续的影响**。

5.2 利用插值法补充样本数据

数据经过 Python 脚本处理后，仍存在部分小时数据缺失的问题。由于数据的复杂性，现利用**三次 Hermite 插值法**和**三次样条插值法**对处理后样本数据进行插值，保证样本数据为连续数据，便于后续求解问题。

5.2.1 三次 Hermite 插值

三次 Hermite 插值是通过给定的数据点处的函数值和导数值，进而构造出一个**三次多项式**逼近原函数。该多项式满足函数值与给定数据点函数值相等，且给定数据点处的导数与给定的导数值一致。

假设已知函数 $f(x)$ 在插值区间 $[p, q]$ 上的 $n+1$ 个互不相同的节点 $x_i (i = 0, 1, \dots, n)$ 处满足 $f(x_i) = f_i$ 及 $f'(x_i) = f'_i (i = 0, 1, 2, \dots, n)$ ，如果函数 $G(x)$ 存在且满足下列条件：

- ① $G(x)$ 在每个小区间上的多项式次数为 3；
- ② $G(x) \in C^1[a, b]$ ；
- ③ $G(x_i) = f(x_i)$ ， $G'(x_i) = f'(x_i)$ ， $i = (0, 1, 2, \dots, n)$

就称 $G(x)$ 是 $f(x)$ 在 $n+1$ 个节点 x_i 上的**分段三次 Hermite 插值多项式**。

$$\begin{aligned}
 \text{所以, } G(x) &= h_k y_k(x) + h_{k+1} y_{k+1}(x) + H_k(x) y'_k + H_{k+1}(x) y'_{k+1} \\
 &= \left(1 + 2 \frac{x-x_k}{x_{k+1}-x_k}\right) \left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2 y_k + \left(1 + 2 \frac{x-x_{k+1}}{x_{k+1}-x_k}\right) \left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2 y_{k+1} \\
 &\quad + (x-x_k) \left(\frac{x-x_{k+1}}{x_k-x_{k+1}}\right)^2 y'_k + (x-x_{k+1}) \left(\frac{x-x_k}{x_{k+1}-x_k}\right)^2 y'_{k+1}
 \end{aligned}$$

三次 Hermite 插值具有对数据点的**高度适应性和平滑性**的优点，具有处理高次导数值可能引入振荡的缺点。

5.2.2 三次样条插值

三次样条插值是将给定的数据点分段连接，且在相邻数据点之间采用三次多项式进行插值，这些多项式函数在相邻数据点处具有连续的一阶和二阶导数，进而保证插值曲线的平滑性和连续性。

对于给定区间 $[p, q]$ 上 $p = x_0 < x_1 < \dots < x_n = q$ 这 $n + 1$ 个节点和在这些点上的函数值 $f(x_i) = y_i (i = 0, 1, \dots, n)$ ，若函数 $g(x)$ 满足：

- ① 在每个子区间 $[x_{i-1}, x_i] (i = 1, 2, \dots, n)$ 上，多项式 $g(x)$ 的次数不超过 3；
- ② $g(x)$ ， $g'(x)$ ， $g''(x)$ 在 $[p, q]$ 上连续；
- ③ 满足 $g(x_i) = y_i (i = 0, 1, 2, \dots, n)$ 的插值条件

则 $g(x)$ 是函数 $f(x)$ 关于 n 个节点 x_i 处的三次样条插值函数。

三次样条插值具有插值稳定的优点，具有在数据点之间距离较大或数据不平滑的情况下，插值出现振荡的缺点。

5.3 利用矩阵散点图判断线性关系

矩阵散点图是一种多变量数据可视化技术，用于同时展示多个变量之间的线性关系。矩阵散点图是散点图的一种扩展，适用于多个变量的情况。

分别研究 Line Current [A] DEV1 和 Line Current [A] DEV2 与其他变量之间的关系，利用 SPSS 作矩阵三点图如下：

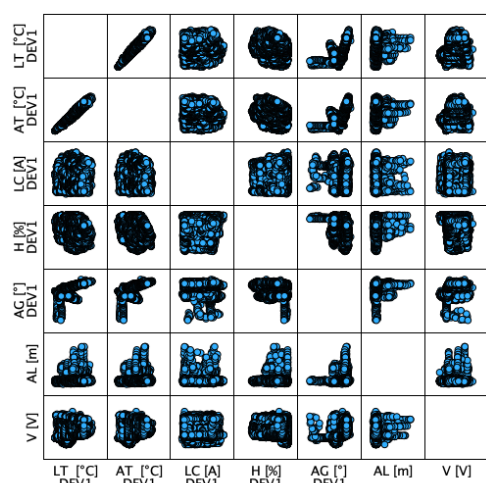


图 1 Line Current [A] DEV1 的矩阵散点图

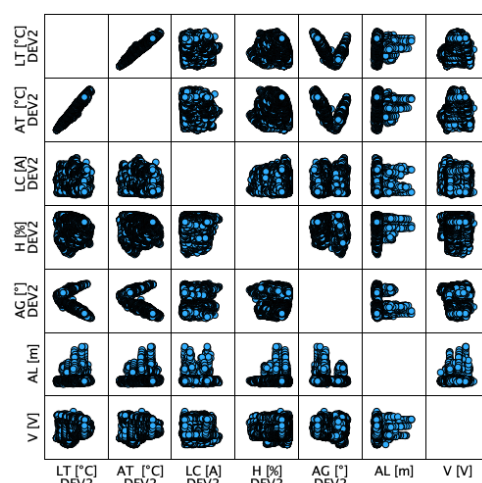


图 2 Line Current [A] DEV2 的矩阵散点图

利用 SPSS 软件，绘制 OTLM1 与 OTLM2 采集的各组数据之间的矩阵散点图。由图可知，各组变量之间有一定的线性相关性，Line Temperature [°C]与 Ambient Temperature OTLM [°C]呈现出较强的线性相关性。为更精确地描述各组数据之间的关系，先对各组数据进行正态分布检验，以判断其是否满足 Pearson 相关系数假设检验的条件。

5.4 利用 Quantile-Quantile plot 鉴别样本数据

统计学中，Quantile-Quantile plot 是一种通过比较两个概率分布的分位数对这两个概率分布进行比较的概率图方法。

首先选定分位数的对应概率区间集合，在此概率区间上，点 (x, y) 对应于第一个分布的一个分位数 x 和第二个分布在和 x 相同概率区间上相同的分位数。对于该问题中的数据，选择正态分布和要检验的随机变量，并对其做出 QQ 图。

选择对 Line Temperature [°C] DEV1、Ambient Temperature OTLM [°C] DEV1、Line Current [A] DEV1、Humidity [%] DEV1、Angle [°] DEV1、Line Temperature [°C] DEV2、Ambient Temperature OTLM [°C] DEV2、Line Current [A] DEV2、Humidity [%] DEV12、Angle [°] DEV2、Altitude [m]、Vbat [V]数据进行 QQ 图鉴别。

利用 MATLAB 作样本数据的 QQ 图。Line Current [A] DEV1 与 Line Current [A] DEV2 的 QQ 图如下图所示（由于篇幅原因，只展示两个样本数据的 QQ 图，其余结果见附件）：

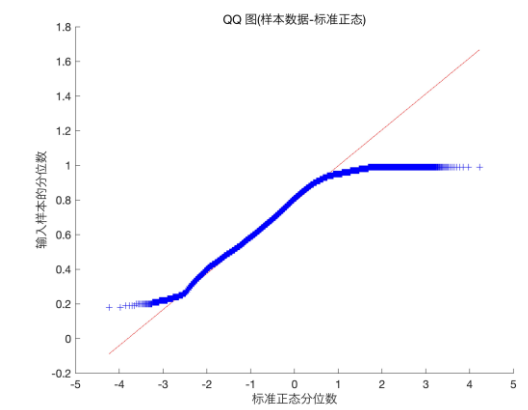


图 3 Line Current [A] DEV1 的 QQ 图

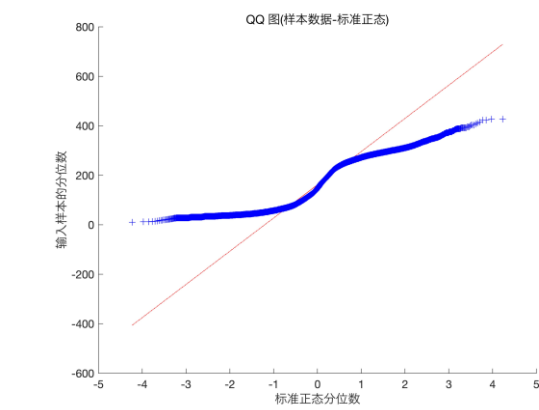


图 4 Line Current [A] DEV2 的 QQ 图

由 QQ 图可知，所有 QQ 图均不为近似地在一条直线附近，因此所鉴别的样本数据不近似于正态分布。

5.5 利用 Jarque-Bera test 检验正态分布

对于一个随机变量 $\{X_i\}$ ，假设其偏度为 S ，峰度为 K ，构造 JB 统计量：

$$JB = \frac{n}{6} \left[S^2 + \frac{(K - 3)^2}{4} \right] \tag{1}$$

可以证明，如果 $\{X_i\}$ 是正态分布。那么在大样本情况下 $JB \sim \chi^2(2)$ （自由度为 2 的卡方分布）

Step1) 提出假设

- 原假设 H_0 ：该随机变量服从正态分布
- 备择假设 H_1 ：该随机变量不服从正态分布
- 确定置信水平为 90%

Step2) 计算偏度和峰度

利用 MATLAB 计算样本数据的峰度系数和偏度系数，结果如表 1 所示：

表 1 样本数据偏度系数与峰度系数表

数据类型	DEV 1		DEV 2	
	偏度	峰度	偏度	峰度
Line Temperature [°C]	0.013	-0.767	-0.023	-0.744
Ambient Temperature OTLM [°C]	-0.018	-0.826	-0.067	-0.798
Line Current [A]	0.154	-1.513	0.186	-1.439
Humidity [%]	-0.694	-0.263	-0.884	0.340
Angle [°]	-0.852	0.146	0.268	-1.643

Step3) 计算 JB^* ，并计算 p 值

利用 MATLAB 计算样本数据的检验值 H 和 p 值，结果见表 2 所示：

表 2 样本数据检验值 H 和 p 值表

数据类型	DEV 1		DEV 2	
	H	p	H	p
Line Temperature [°C]	1.000	0.001	1.000	0.001
Ambient Temperature OTLM [°C]	1.000	0.001	1.000	0.001
Line Current [A]	1.000	0.001	1.000	0.001
Humidity [%]	1.000	0.001	1.000	0.001
Angle [°]	1.000	0.001	1.000	0.001

由表 2 可知, DEV 1 和 DEV 2 样本数据的检验值 H 均为 1, p 值均小于 0.05, 故拒绝原假设, 样本数据不服从正态分布, 不能利用 **Pearson 相关系数检验**变量之间是否具有线性关系。现利用 **Spearman 相关系数检验**变量之间是否具有相关性。

5.6 利用 Spearman 相关系数判断各变量之间的相关性

5.6.1 Spearman 相关系数的运用

Spearman 相关系数是一种非参数的相关性度量, 等级化变量之间的相关性, 用于分析两个连续变量之间的相关性。因样本数据为连续数据, 故本次建模采用 **Spearman 相关系数**用于分析样本数据的相关关系。

X 、 Y 为两组独立同分布的数据, 其样本个数为 N 。 X_i 、 Y_i 分别表示两组随机变量的第 i 个值, 其中 $i = 1, 2, \dots, N$ 。首先对 X 、 Y 集合同时降序或升序排列, 得到两个元素排序集合 x 、 y , 其中元素 x_i 、 y_i 分别为 X_i 、 Y_i 在各自集合中的排序。设定 d 集合为 X 与 Y 集合中相同位元素排序之差, d 集合各元素计算式如下:

$$d_i = x_i - y_i \quad (2)$$

Spearman 相关系数 r_s 计算公式如下:

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (3)$$

利用 MATLAB 求解样本数据的 **Spearman 相关系数**如下表 3 和表 4 所示(由于篇幅原因, 其余 Spearman 相关系数见附录):

表 3 DEV1 的 Spearman 相关系数表

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	Line Current [A]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9838	-0.0674	-0.3508	0.8982	0.2179	0.4331
Ambient Temperature OTLM [°C]	0.9838	1.0000	-0.1939	-0.3246	0.8906	0.2263	0.4487
Line Current [A]	-0.0674	-0.1939	1.0000	0.0588	-0.1324	-0.0411	-0.2288
Humidity [%]	-0.3508	-0.3246	0.0588	1.0000	-0.3499	-0.0732	-0.3627
Angle [°]	0.8982	0.8906	-0.1324	-0.3499	1.0000	0.2242	0.4482
Altitude [m]	0.2179	0.2263	-0.0411	-0.0732	0.2242	1.0000	0.0627
Vbat [V]	0.4331	0.4487	-0.2288	-0.3627	0.4482	0.0627	1.0000

表 4 DEV2 的 Spearman 相关系数表

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	Line Current [A]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9847	-0.0396	-0.3329	-0.7575	0.2146	0.4269
Ambient Temperature OTLM [°C]	0.9847	1.0000	-0.1547	-0.2966	-0.7714	0.2270	0.4292
Line Current [A]	-0.0396	-0.1547	1.0000	0.0074	0.0795	-0.0388	-0.2137
Humidity [%]	-0.3329	-0.2966	0.0074	1.0000	0.3208	-0.0693	-0.2970
Angle [°]	-0.7575	-0.7714	0.0795	0.3208	1.0000	-0.2343	-0.3742
Altitude [m]	0.2146	0.2270	-0.0388	-0.0693	-0.2343	1.0000	0.0627
Vbat [V]	0.4269	0.4292	-0.2137	-0.2970	-0.3742	0.0627	1.0000

表 5 DEV1 Ch1 的 Spearman 相关系数表

表 6 DEV1 Ch2 的 Spearman 相关系数表

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	应力 [MPa]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9838	0.2889	-0.3508	0.8982	0.2179	0.4331
Ambient Temperature OTLM [°C]	0.9838	1.0000	0.3190	-0.3246	0.8906	0.2263	0.4487
Line Current [A]	0.2889	0.3190	1.0000	0.0405	0.2358	0.1174	0.0344
Humidity [%]	-0.3508	-0.3246	0.0405	1.0000	-0.3499	-0.0732	-0.3627
Angle [°]	0.8982	0.8906	0.2358	-0.3499	1.0000	0.2242	0.4482
应力 [MPa]	0.2179	0.2263	0.1174	-0.0732	0.2242	1.0000	0.0627
Vbat [V]	0.4331	0.4487	0.0344	-0.3627	0.4482	0.0627	1.0000

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	应力 [MPa]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9838	-0.3137	-0.3508	0.8982	0.2179	0.4331
Ambient Temperature OTLM [°C]	0.9838	1.0000	-0.3305	-0.3246	0.8906	0.2263	0.4487
Line Current [A]	-0.3137	-0.3305	1.0000	0.1464	-0.2656	-0.0960	-0.1080
Humidity [%]	-0.3508	-0.3246	0.1464	1.0000	-0.3499	-0.0732	-0.3627
Angle [°]	0.8982	0.8906	-0.2656	-0.3499	1.0000	0.2242	0.4482
应力 [MPa]	0.2179	0.2263	-0.0960	-0.0732	0.2242	1.0000	0.0627
Vbat [V]	0.4331	0.4487	-0.1080	-0.3627	0.4482	0.0627	1.0000

由表 3 和表 4 可知：

- I. OTLM 采集的 Line Temperature [°C]与 Ambient Temperature OTLM [°C]之间相关系数较高，呈正相关。OTLM1 与 OTLM2 采集的 Line Temperature [°C]与 Ambient Temperature OTLM [°C]的 Spearman 相关系数分别为 0.9847 与 0.9838，表明它们之间存在较强的正相关关系。
- II. Humidity [%] 与 Line Temperature [°C]，Ambient Temperature OTLM [°C]的 Spearman 相关系数均小于 0，表明它们之间存在负相关关系。但 OTLM1 与 OTLM2 采集的 Humidity [%]与 Line Current [A]的相关系数分别为 0.0074 与 0.0588，相关性近似于 0。
- III. OTLM1 采集的 Angle [°]与 Line Current [A]，Ambient Temperature OTLM [°C]的 Spearman 相关系数分别为-0.7575 与-0.7714，表明它们之间存在一定的负相关关系。相反，OTLM2 采集的 Angle [°]与 Line Current [A]，Ambient Temperature OTLM [°C]的 Spearman 相关系数分别为 0.8982 与 0.8906，它们之间存在较强的正相关关系。

5. 6. 2 Spearman 相关系数的检验

现对样本数据采取 Spearman 相关系数检验，检验步骤如下：

Step1) 提出假设

原假设 H_0 ：Spearman 系数 $R \neq 0$

备择假设 H_1 ：Spearman 系数 $R = 0$

确定置信水平为 90%

Step2) 计算 p 值

利用 MATLAB 进行 Spearman 相关系数检验，结果如表所示(由于篇幅原因，其余 Spearman 相关系数检验 p 值表见附件)：

表 7 DEV1 的 Spearman 相关系数检验 p 值表

数据类型	LT [°C]	AT [°C]	LC [A]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
AT [°C]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag [°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0

表 8 DEV2 的 Spearman 相关系数检验 p 值表

数据类型	LT [°C]	AT [°C]	LC [A]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
AT [°C]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag [°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0

$v[V]$	0	0	0	0	0	0	1
--------	---	---	---	---	---	---	---

$v[V]$	0	0	0	0	0	0	1
--------	---	---	---	---	---	---	---

由表可知，样本数据任意两者之间的 p 值均远小于 0.05，故接受原假设，可认为各个样本数据之间存在显著相关性。

六、问题一模型的建立与求解

6.1 第一问：时间序列和神经网络预测动态载流量需求

首先建立时间序列预测模型，由题可知，导线温度直接决定弧垂的大小，现考虑关键影响因素 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]。利用 Python 处理 DEV 1 的 2019 年 7 月至 2020 年 7 月时间段内的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据，使用三次 Hermite 插值和三次样条插值，将样本数据转换为连续数据，以每小时数据为原始数据，预测 DEV 1 的 2020 年 8 月的 Line Temperature [°C]、Ambient Temperature OTLM [°C]和 Humidity [%]数据。然后，运用预测的数据，利用 BP 神经网络算法，预测 2020 年 8 月的动态载流量需求。

6.1.1 利用时间序列预测相关样本数据

利用 MATLAB 对 Line Temperature [°C]和 Ambient Temperature OTLM [°C]数据进行三次 Hermite 插值，对 Humidity [%]进行三次样条插值。然后利用 SPSS 软件构建时间序列模型，以专家建模器为基本算法，进行离群值检验，预测 2020 年 8 月的动态载流量需求。

表 9 各样本数据适用的时间序列模型

数据类型	模型
Line Temperature [°C]	SARIMA(1,1,6)(2,0,1) ₂₄
Ambient Temperature OTLM [°C]	SARIMA(2,1,3)(1,0,1) ₂₄
Humidity [%]	SARIMA(2,1,3)(1,0,1) ₂₄

Step1)模型解释

SARIMA(p, d, q)(P, D, Q)_m 模型公式

$$\left(1 - \sum_{i=1}^p \phi_i L^i\right) \left(1 - \sum_{i=1}^P \Phi_i L^{mi}\right) (1-L)^d (1-L^m)^D y_t = \alpha_0 + \left(1 + \sum_{i=1}^q \theta_i L^i\right) \left(1 - \sum_{i=1}^Q \Theta_i L^{mi}\right) \varepsilon_t \quad (4)$$

其中， L 为滞后算子， y_i 为第 i 期预测值

Step2)模型求解

模型拟合与预测结果如下(由于篇幅原因，现只展示 Ambient Temperature OTLM [°C]的拟合图，其余见附件)：

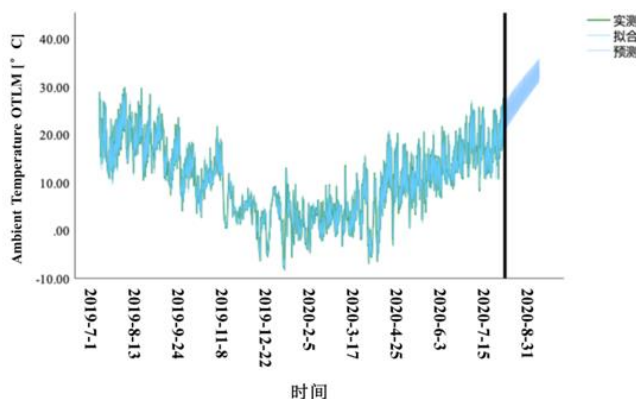


图 5 Ambient Temperature OTLM [°C]拟合与预测图

Step3)模型检验

表 10 模型拟合与预测效果检验

检测值	LT [°C]	AT [°C]	H [%]
平稳 R 方	0.444	0.476	0.283
R 方	0.988	0.994	0.980
Q 检验 p 值	0.012	0.123	0.019

由表可知，三次时间序列预测，平稳 R 方均较小， R 方均接近于 1，说明估计效果很好，其次， Q 检验在置信区间为 95%的情况下， p 值均小于 0.01，因此这三个模型能够较好地识别本题中的样本数据。

6.1.2 利用 BP 神经网络算法预测动态载流量需求

Step1) BP 神经网络层数的选取

网络层数的选取决定了训练的精度和速度，增加层数可以降低误差，提高精度，但是同时也会增加网络的复杂性，从而增加训练时间。Kolmogorov 理论指出：具有单个隐含层的 BP 神经网络可以映射所有连续函数，而具有双隐含层的 BP 神经网络可以映射任何函数（包括不连续函数）。在此，并不清楚所要建立的映射关系是否连续，所以先考虑有单个隐含层的 BP 神经网络，如果不能满足要求，就考虑再加一个隐含层。

Step2) 神经网络的设计

输入层起缓冲器的作用，把数据源加到网络上，其节点数目取决于数据源的维数。基于 BP 算法的神经元网络输出层神经元可以根据待求的问题和数据表示的方式而定。在设计输入层和输出层时，应该尽可能的减小系统规模,使系统的学习时间和复杂性减小。

神经网络结构如下：

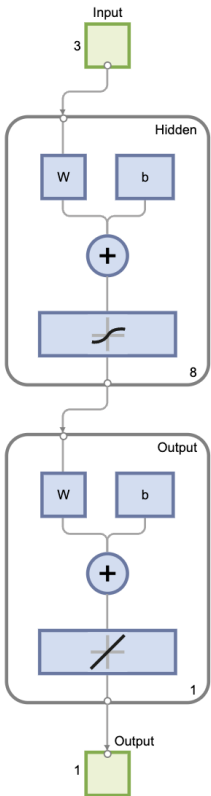


图 6 具有十层神经元的神经网络结构图

我们用于训练的指标为 Line Current [A] , Ambient Temperature OTLM [°C] 和 Humidity [%] 共 3 个指标,输出量为 Line Current [A] 共 1 个指标。选择训练层数为 10, 设计神经网络。

Step3)学习速率与训练方法的选择

学习速率决定每一次循环训练所产生的权值的变化量。过大的学习速率可能导致系统的不稳定,但是过小的学习速率将导致训练时间较长,收敛速度很慢,不过能保证网络的误差值趋于最小。一般情况下,学习速率的选取范围在 0.01~0.9 之间。通过保证稳定训练的前提下,达到了合理的高速率,可以减少训练时间。在本模型中,我们采用贝叶斯正则化训练方法。

神经网络的训练效果如下图:

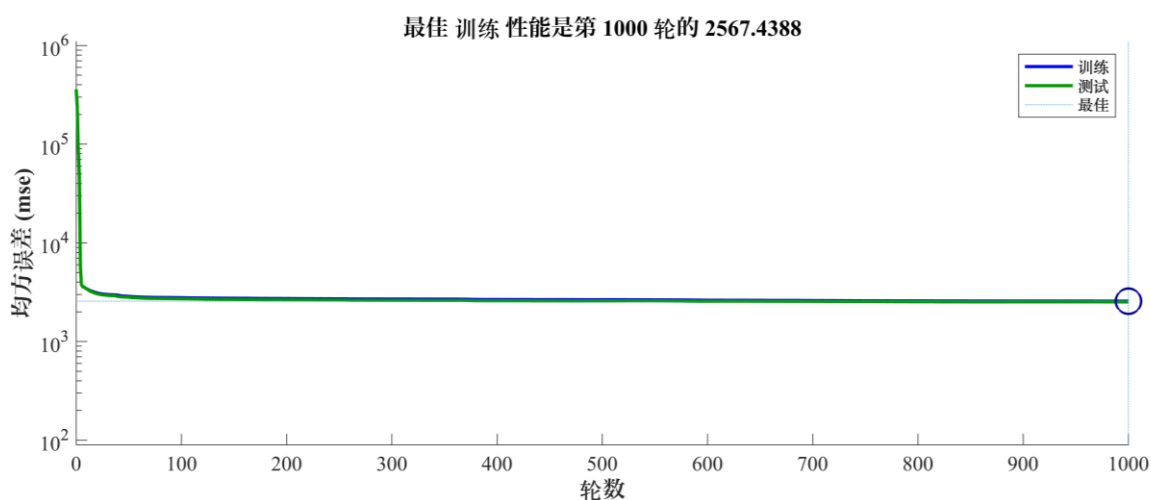


图 7 BP 神经网络训练效果图

神经网络的误差直方图如下:

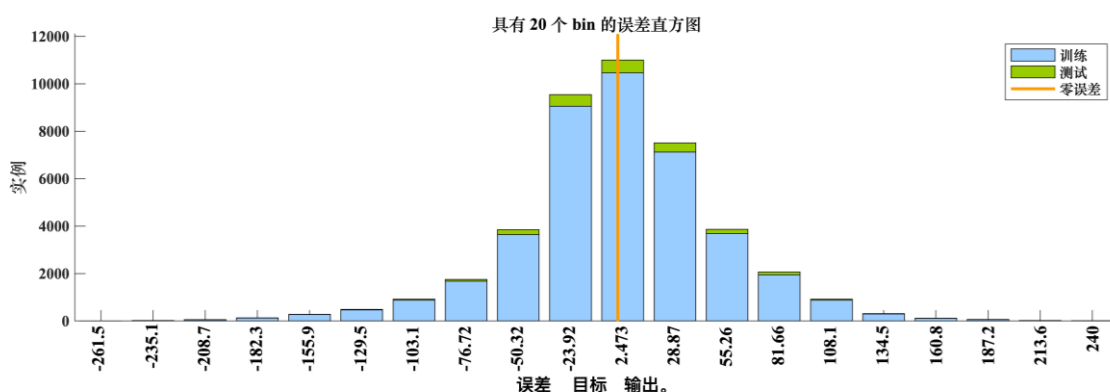


图 8 BP 神经网络误差直方图

使用 MATLAB 神经网络拟合工具训练神经网络,以 Line Current [A] , Ambient Temperature OTLM [°C] 和 Humidity [%]为指标,预测 Line Current [A],得到的结果如下图:

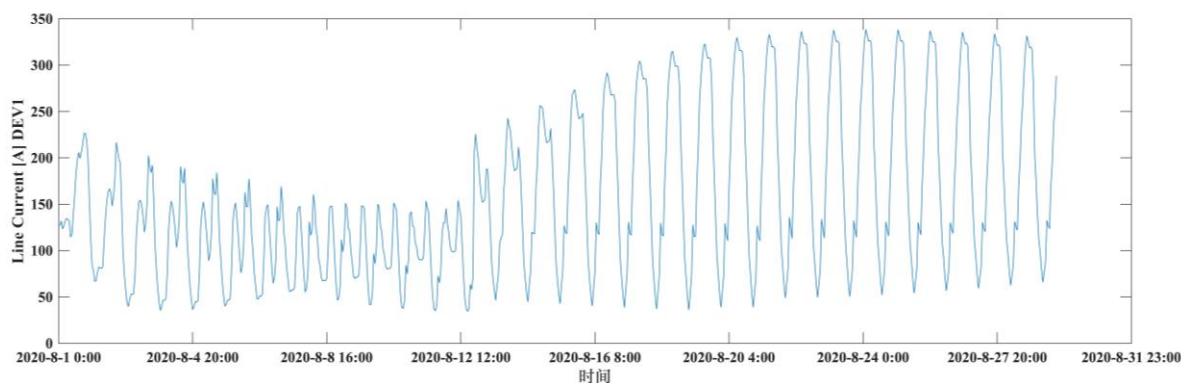


图 9 BP 神经网络预测结果图

由上图可知，Line Current[A]在 8 月随昼夜变化，呈现出较强的周期性。在 2020-8-12 12:00 之前，Line Current[A]振幅较小。在 2020-8-12 12:00 之后，振幅出现了明显的增大趋势，最高峰值 Line Current[A]达到了 340[A]左右，这与夏季用电高峰有关。

6.2 利用傅里叶变换分析载流量的周期性变化规律

傅里叶变换是一种用于对数据周期性分析的方法，其将时域中的数据转换为频域中的表示。如果一组数据包含有特定周期的成分，那么在频域表示中，与这个周期相对应的频率处将会出现一个峰值。这个峰值的大小代表了该频率成分的强度，即大部分数据都以该频率重复出现，再由频率和周期之间互为倒数的关系，便可分析出这组数据的周期性。

首先对附件中给出的载流量数据进行处理。对于缺失的数据，利用三次 Hermite 插值法进行插值。具体方法在前文已经给出，此处不再赘述。利用 Python 对插值后的数据进行傅立叶变换分析，得到结果如下图所示：

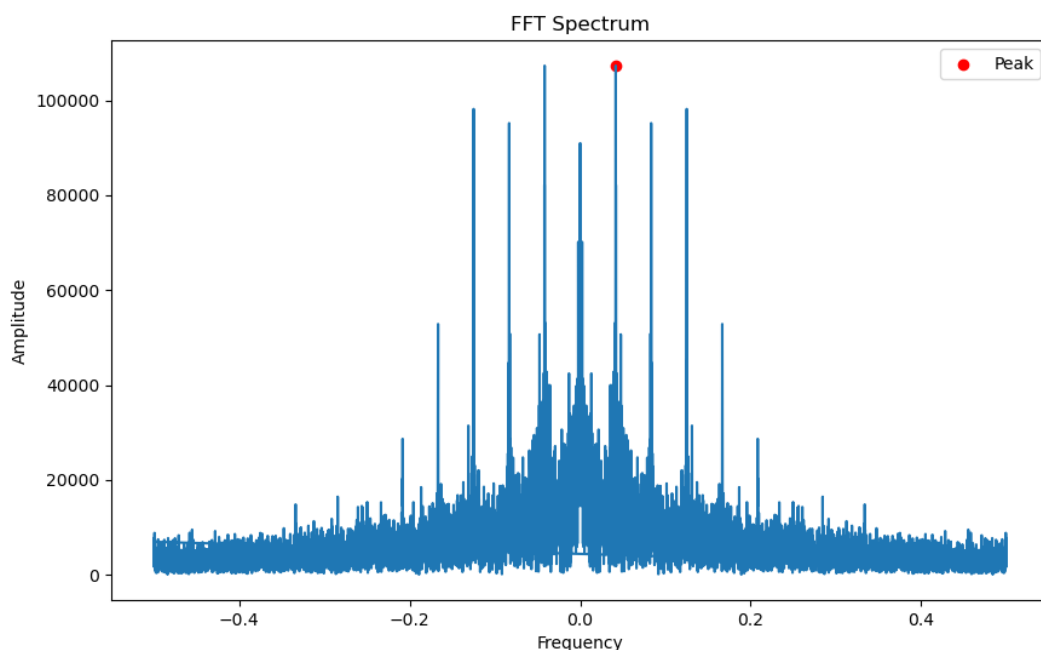


图 10 傅里叶变换频率分布图

分析图像可知，傅立叶变换图像的峰值对应的频率为 0.04 左右，而这一频率恰好对应了载流量的 24 小时周期性变化。另外，利用 Python 输出傅立叶变换后的频率数据，取倒数后所得的周期为 23.984 小时，与 24 小时非常接近，印证了这一结论（具体代码

见附件)。另外几个峰值对应的周期分别为 4 小时、8 小时、12 小时，这一结论与昼夜用电量变化的规律相符。考虑最显著的周期表现，取 24 小时为载流量变化的周期。

七、问题二模型的建立与求解

7.1 回归自变量的确定

本文的数据侧写部分计算了 DEV1 和 DEV2 的样本数据之间的 **Spearman** 相关系数,现只考虑 Line Current [A]与 Line Temperature [°C]、Ambient Temperature OTLM [°C]、Angle [°]之间的相关性大小,如下表所示:

表 11 部分样布数据 Spearman 相关系数表

	Line Current [A] DEV1		Line Current [A] DEV2	
	数值	相关性大小	数值	相关性大小
Line Temperature [°C]	-0.0674	小	-0.0396	小
Ambient Temperature OTLM [°C]	-0.1939	较大	-0.1547	较大
Angle [°]	-0.1324	较大	0.0795	较大

由表可知,与 Line Temperature [°C]相关性较强的有 Ambient Temperature OTLM [°C]和 Angle [°]两个变量,故分析温度、弧垂角度对载流量的定量影响,选择 Ambient Temperature OTLM [°C]和 Angle [°]作为回归自变量,分别记作 x_1 , x_2 。

7.2 回归模型的基本假定

对于回归模型的计算结果,需要一个随机误差项对该模型的误差程度进行评价,本文遵循 Gauss-Markov supposition: 随机误差项服从一个零均值、同方差的正态分布,即: $\mu \sim N(0, \sigma^2)$ 。

7.3 多元线性回归模型

对于因变量 Line Temperature 和两个自变量 Ambient Temperature OTLM [°C]、Angle [°]可以得到一个多线线性回归模型:

$$\begin{cases} y = \beta_0 + \sum_{j=1}^2 \beta_j x_j + \mu_i \\ \mu \sim N(0, \sigma^2) \end{cases} \quad (5)$$

其中, β_j 为回归系数

7.4 模型求解与结论

7.4.1 模型求解

利用 Stata 进行多元线性回归,结果如下:

表 12L Line Temperature [°C]DEV1 回归系数表

(1)	
	Line Temperature [°C]DEV1
Ambient Temperature OTLM [°C]DEV1	-3.293*** (-32.053)
Angle [°]DEV1	10.729*** (9.081)
_cons	63.258*** (4.490)

N	42879
*** p<0.01 ** p<0.05 * p<0.1	

表 13L Line Temperature [°C]DEV2 回归系数表

(1)	
	Line Temperature [°C]DEV1
Ambient Temperature OTLM [°C]DEV2	-3.196*** (-33.409)
Angle [°]DEV2	-10.862*** (-13.478)
_cons	324.720*** (31.222)
N	42879
*** p<0.01 ** p<0.05 * p<0.1	

由两表可知，所有回归系数 p 值均小于 0.05，代表在 95%置信水平下，回归系数显著地异于 0。

联合显著性检验结果如下：

表 14 联合显著性检验结果表

	Line Temperature [°C]DEV1	Line Temperature [°C]DEV2
F(2,42876)	983.19	750.93
Prob > F	0.0000	0.0000

由表可知，联合显著性检验 p 值均小于 0.05，故拒绝原假设 $\beta_1 = \beta_2 = 0$ ，认为回归模型可信

故得到 DEV1 和 DEV2 的 Line Temperature [°C]与 Ambient Temperature OTLM [°C]、Angle [°]变量之间的队员线性回归方程，方程如下：

$$LC1 = 63.258 - 3.293AT1 + 10.729AG1 \quad (6)$$

$$LC2 = 324.720 - 3.196AT1 - 10.862AG1 \quad (7)$$

7.4.2 求解结论

由方程可知，对于 DEV1，Ambient Temperature OTLM [°C]每增加一个单位，Line Temperature [°C]平均减少 3.293 个单位，Angle [°]每增加一个单位，Line Temperature [°C]平均增加 10.729 个单位；对于 DEV2，Ambient Temperature OTLM [°C]每增加一个单位，Line Temperature [°C]平均减少 3.196 个单位，Angle [°]每增加一个单位，Line Temperature [°C]平均减少 10.862 个单位。

7.5 White 检验

现利用 White 检验异方差：

原假设 H_0 ：不存在异方差

备择假设 H_1 ：存在异方差

利用 Stata 软件进行 White 检验，两次检验结果如下：

第一次检验

Chi2(5) = 1770.40
Prob > chi2 = 0.0000

第一次检验

Chi2(5) = 2492.36
Prob > chi2 = 0.0000

可知， p 值均小于 0.05，故接受原假设，回归模型不存在异方差

八、问题三模型的建立与求解

附件中数据给出了钢架四条腿部每侧应力的相关数据。首先使用因子分析，通过研究变量间的**相关系数矩阵**，将钢架每侧的应力变量反映成一个综合变量，这一过程又可称为**降维**。然后建立**典型相关分析**，从 Line Temperature [°C]、Ambient Temperature OTLM [°C]、Humidity [%]、Angle [°]、Altitude [m]和 Vbat [V]中选取若干典型变量，其目的是分析电流与应力之间存在的关系。最后建立**一元线性回归方程**，定量描述应力与弧垂角度之间的关系。

8.1 利用因子分析进行应力数据的处理分析模型的求解

8.1.1 模型解释

因子分析是主成分分析的推广，它也是一种把多个变量化为少数几个综合变量的多元分析方法，其目的是用有限个不可观测的隐变量来解释原始变量之间的相关关系。

假设 p 维随机向量 $x = (x_1, x_2, \dots, x_p)'$ 的均值 $u = (u_1, u_2, \dots, u_p)'$ ，协方差矩阵 $\Sigma_{p \times p} = (\sigma_{ij})$ ，因子分析的一般模型为：

$$\begin{cases} x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m + \varepsilon_1 \\ x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m + \varepsilon_2 \\ \vdots \\ x_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \dots + a_{pm}f_m + \varepsilon_p \end{cases} \quad (8)$$

其中 f_1, f_2, \dots, f_m 被称为公共因子， $\varepsilon_i (i = 1, 2, \dots, p)$ 为特殊因子，均为无法观测的随机变量。公共因子 f_1, f_2, \dots, f_m 出现在每一个原始变量 $x_i (i = 1, 2, \dots, p)$ 的表达式中，即原始变量共同拥有的某些特征；每个特殊因子 $\varepsilon_i (i = 1, 2, \dots, p)$ 仅出现在与之相应的第 i 个原始变量 x_i 的表达式中，其只对该原始变量起作用。

上式用矩阵形式可记为： $x = u + Af + \varepsilon$

其中 $f = (f_1, f_2, \dots, f_p)'$ ($m \leq p$)为公因子向量， $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)'$ 为特殊因子向量， $A_{p \times m} = (a_{ij})$ 称为因子载荷矩阵，并假设 A 的秩为 m 。

8.1.2 KMO 检验和巴特利特球形检验

I.KMO 检验

该检验是对原始变量之间的简单相关系数和偏相关系数的相对大小进行检验，主要应用于多元统计的因子分析。KMO 统计量是取值在 0 和 1 之间，当所有变量间的简单相关系数平方和远远大于偏相关系数平方和时，**KMO 值越接近于 1，意味着变量间的相关性越强，原有变量越适合作因子分析**；当所有变量间的简单相关系数平方和接近 0 时，**KMO 值越接近于 0，意味着变量间的相关性越弱，原有变量越不适合作因子分析**。

II.巴特利特球形检验

巴特利特球形检验是一种检验各个变量之间相关性程度的检验方法。巴特利特球形检验是以变量的相关系数矩阵为出发点的。原假设是**相关系数矩阵是一个单位阵(不适合做因子分析，指标之间的相关性很差，不适合降维)**，即相关系数矩阵对角线上的所有元素都是 1，所有非对角线上的元素都为 0。巴特利特球形检验的统计量是根据相关系数矩阵的行列式得到的。如果该值较大，且其对应的 p 值小于用户心中的显著性水平(一般为 0.05)，那么应该拒绝原假设，认为相关系数不可能是单位阵，即原始变量之间存在相关性，适合于作因子分析。相反不适合作因子分析。

III.假设检验

表 15 KMO 和巴特利特检验(Ch1)表

KMO 取样适切性量数	.537
-------------	------

巴特利特球形度检验	近似卡方	106914.762
	自由度	6
	显著性	<0.001

表 16 KMO 和巴特利特检验(Ch1)表

KMO 取样适切性量数		.549
巴特利特球形度检验	近似卡方	178172.216
	自由度	6
	显著性	<0.001

原假设 H_0 : 数据不适合进行因子分析

备择假设 H_1 : 数据适合进行因子分析

对因子进行 KMO 和巴特利特检验, 由 KMO 取样适切性量数和巴特利特检验的显著性水平可知, 在 99%的置信水平下, 我们拒绝原假设, 即认为原数据适合因子分析, 通过了 KMO 和巴特利特检验。

8.1.3 因子分析

通过对附件中的数据建立预处理、特征工程、特征选择, 变量从原来的8个变成了2个。而特征维度太高对主观法与客观法建立评价体系来说都很复杂, 因此我们利用因子分析降维, 将8个变量降成了2维。这里展示重要公共因子:

表 17 公共因子方差(Ch1)表

	初始	提取
S1 Ch1 Avg [MPa]	1.000	.590
S2 Ch1 Avg [MPa]	1.000	.924
S3 Ch1 Avg [MPa]	1.000	.799
S4 Ch1 Avg [MPa]	1.000	.081

提取方法: 主成分分析法

表 18 公共因子方差(Ch2)表

	初始	提取
S1 Ch2 Avg [MPa]	1.000	.669
S2 Ch2 Avg [MPa]	1.000	.674
S3 Ch2 Avg [MPa]	1.000	.619
S4 Ch2 Avg [MPa]	1.000	.797

提取方法: 主成分分析法

由表和表可看出, 通过主成分分析法, 对于 Ch1 侧, 提取的公共因子可以反映 S1 脚 59.0%的信息, S2 脚 92.4%的信息, S3 脚 79.9%的信息, S4 脚 8.1%的信息。对于 Ch2 侧, 提取的公共因子可以反映 S1 脚 66.9%的信息, S2 脚 67.4%的信息, S3 脚 61.9%的信息, S4 脚 79.7%的信息。

通过公共因子方差可以看出, 提取的公因子对各个变量的贡献度均较大, 这表明这两个公因子可以较好地反映各个变量的信息。

表 19 Ch1 成分矩阵表

Ch1 成分矩阵 ^a	
	成分
	1
S1 Ch1 Avg [MPa]	.768
S2 Ch1 Avg [MPa]	.961

S3 Ch1 Avg [MPa]	-. 894
S4 Ch1 Avg [MPa]	-. 285
提取方法：主成分分析法。	
a. 提取了 1 个成分。	

表 20 Ch2 成分矩阵表

Ch2 成分矩阵 ^a	
	成分
	1
S1 Ch2 Avg [MPa]	. 818
S2 Ch2 Avg [MPa]	-. 821
S3 Ch2 Avg [MPa]	. 787
S4 Ch2 Avg [MPa]	. 893
提取方法：主成分分析法。	
a. 提取了 1 个成分。	

因子载荷是变量与公共因子的相关系数。当某变量在某公共因子中的载荷绝对值越大，表明该变量与该公共因子更密切，即该公共因子更能代表该变量。由成分矩阵可得，合成的公共因子可以写成以下形式：

$$F_1 = 0.748 \times S1Ch1Avg + 0.961 \times S2Ch1Avg - 0.894 \times S3Ch1Avg - 0.285 \times S4Ch1Avg \quad (9)$$

$F_2 = 0.818 \times S1Ch2Avg - 0.821 \times S2Ch2Avg + 0.787 \times S3Ch2Avg + 0.893 \times S4Ch2Avg$	(10)
---	------

8.2 利用典型相关分析分析电流与应力变化的关系

为了研究电流与应力变化之间的关系，令电流为输入变量，应力变化为输出变量，采用典型相关分析法。典型相关分析基本思想是利用主成分思想，分别找出输入变量与输出变量的线性组合，然后讨论线性组合之间的相关关系。首先在每组变量中找出变量的线性组合，使得两组的线性组合之间具有最大的相关系数；然后选取和最初挑选的这对线性组合不相关的线性组合，使其配对，并选取相关系数最大的一对；如此继续下去，直到两组变量之间的相关性被提取完毕为止。被选出的线性组合配对称为典型变量，该组合相关系数称为典型相关系数，典型相关系数度量了这两组变量之间联系的强度。

表 21 典型相关性设置表

	值	符号
电流显著型 指标	LT[°C]DEV1	Z_1
	LT[°C]DEV2	Z_2
	AT[°C]DEV1	Z_3
	AT[°C]DEV2	Z_4
	AL[m]	Z_5
	AG[°]DEV1	Z_6
	AG[°]DEV2	Z_7
应力显著性 指标	H[%]DEV1	Z_8
	H[%]DEV2	Z_9
	V[V]	Z_{10}

若总体典型相关系数 $\lambda_k = 0$ ，则相应的典型变量 U_k ， V_k 之间无相关关系，因此对分析 $X^{(1)}$ 对 $X^{(2)}$ 的影响不起作用。这样的典型变量可以不予考虑，于是提出如何根据样本资料来判断总体典型相关系数是否为零，以便确定应该取几个典型变量的问题。巴特莱

特(Bartlett)提出了一个根据样本数据检验总体典型相关系数 $\lambda_1, \lambda_2, \dots, \lambda_r$ 是否等于零的方法。

假设电流与应力数据服从联合正态分布。构造似然比统计量，对两组变量的相关性进行检验。检验假设为

原假设 H_0 ：两组变量之间不存在相关关系

备择假设 H_1 ：两组变量之间存在相关关系

用于检验的似然比统计量为：

$$\Lambda_k = \prod_{i=k+1}^r (1 - \widehat{\lambda}_i^2) \quad (11)$$

利用 SPSS 软件进行典型相关分析，可以得到在 99%的置信水平下，我们拒绝原假设，即认为两组变量之间存在相关性，可以进行典型相关分析。

表 22 典型相关系数表

	相关系数	特征值	威尔克统计	F	分子自由度	分母自由度	显著性
典型相关变量	.131	.018	.983	150.353	5.000	42871.000	.000

H0 for Wilks 检验是指当前行和后续行中的相关性均为零

考虑到各个变量之间的量纲存在较大差异，为消除量纲对线性组合系数的影响，对线性组合系数进行**标准化**。标准化后**典型相关变量**对应的线性组合系数见下表。未标准化的典型相关变量对应的线性组合系数见附录。

表 23 集合 1 标准化典型相关变量对应的线性组合系数表

变量	典型相关变量
LT[°C]DEV1	10.606
LT[°C]DEV2	-10.650
AT[°C]DEV1	-1.090
AT[°C]DEV2	.882
AL[m]	-.087
AG[°]DEV1	-.625
AG[°]DEV2	-.446

表 24 集合 2 标准化典型相关变量对应的线性组合系数表

变量	典型相关变量
H[%]DEV1	-1.642
H[%]DEV2	1.466
V[V]	-.836

典型相关变量的典型载荷与已解释的方差比例见附录。可得到标准化后的典型变量：可得到标准化的第一对典型变量：

$$U_1^* = 10.606Z_1^{(1)} - 10.65Z_2^{(1)} - 1.09Z_3^{(1)} + 0.882Z_4^{(1)} - 0.087Z_5^{(1)} - 0.625Z_6^{(1)} - 0.446Z_7^{(1)} \quad (12)$$

$$V_1^* = -1.642Z_8^{(2)} + 1.466Z_9^{(2)} - 0.836Z_{10}^{(2)} \quad (13)$$

其中， $Z_i^{(1)}$ 和 $Z_j^{(2)}$ 分别为原始变量 X_i 和 Y_j 标准化后的结果。

通过分析标准化典型相关变量对应的线性组合系数可知，对于 LC[A], LT [°C] DEV1 和 AT [°C] DEV1 的线性组合系数的绝对值分别为 2.929 和 5.494，表明其对 L C [A]的

影响较大。而 $AL[m]$ 的线性组合系数仅为 0.029, 表明其对 $LC[A]$ 的影响几乎可以忽略。对于 $ST[MPa]$, 影响较为显著的是 $H[\%]DEV2$ 与 $V[V]$ 。

8.3 利用一元线性回归建立应力与弧垂角之间的线性关系

8.3.1 线性回归

利用一元线性回归, 结合题目已知条件, 利用 Stata 软件进行回归分析, 建立一元线性回归方程如下:

$$ST = -9.08805AG + 38.82115 \quad (14)$$

其中, ST 为应力, AG 为弧垂角。

利用 Stata 软件可以得到 p 值均小于 0.05, 代表在 95% 的置信水平下, 该回归系数显著的异于 0

8.3.2 White 检验

现利用 White 检验异方差:

原假设 H_0 : 不存在异方差

备择假设 H_1 : 存在异方差

利用 Stata 软件进行 White 检验, 结果如下:

$$\text{Chi2}(2) = 7897.76$$

$$\text{Prob} > \text{chi2} = 0.0000$$

可知, p 值小于 0.05, 故接受原假设, 不存在异方差

九、问题四模型的建立与求解

9.1 问题背景

输电线路是高效、快捷的能源输送通道, 是电网安全运行的关键环节, 截至 2017 年底, 中国 110kV 及以上的输电线路长度已达 $9.87 \times 10^5 \text{km}$ 。随着用电需求的急剧增加和可再生能源的大量接入, 负荷集中地区如长江三角洲、珠江三角洲以及新能源发电集中送出地区电网“卡脖子”问题突出, 提升输电系统的传输能力已成为亟待解决的重要课题。考虑到负荷曲线的波动性和可再生能源的间歇性, 输电线路增容往往不需要全天候运行, 通过新建输电线路来提升电网传输能力不仅周期长、投资大, 而且会带来环境污染问题。

目前解决电力供不应求问题的主要方法为动态增容技术(dynamic line rating, DLR), 即根据环境变化随时间动态调整输电线路容量的方法, 该方法的动态性体现在它是一个时变值, 并基于早期发展的热方程理论来确定。

DLR 方法中最重要的两类包括: 环境调节式 DLR(简记为 AA-DLR)和实时监测式 DLR(简记为 RTM-DLR)。输电线路动态增容技术可以在保证安全的前提下显著提高输电线路的输送能力, 整合可再生能源, 有效解决局部电网“卡脖子”问题, 具有很好的应用前景。

物理模型的构建主要基于输电线路稳态和暂态热平衡方程, 典型的有 IEEE 模型、CIGRE 模型、IEC 模型以及有限元模型等。物理模型原理简单, 可以根据输入参数对线路限额进行实时的计算或预测, 但是难以涵盖线路所有影响因素, 如降雨对线路冷却效果的影响。物理模型中的某些参数难以准确获取, 如导线表面辐射系数和吸热系数与导线新旧程度和表面污秽程度有关, 其数值往往基于经验判断。改进物理模型的一种方法是通过在模型中考虑被忽略的影响因素进行补充, 如考虑了降水冷却效果的扩展 CIGRE 模型^[2]。

输电线路动态增容技术的实现依托于输电线路关键状态信息的采集，包括环境信息和导线状态信息，如**环境温度、光照强度、风速风向、导线温度、导线弧垂**等。通过在输电线路路上安装各类状态监测装置来实时采集输电线路环境和导线状态信息，然后通过无线通信技术将采集的数据发送到数据中心进行处理和展示^[2]。

9.2 利用灰色预测模型对最大载流量进行预估

灰色预测是对既含有已知信息又含有不确定信息的系统进行预测，就是对在一定范围内变化的、与时间有关的灰色过程进行预测。对原始数据通常进行累加处理来寻找系统变动的规律，并生成有较强规律性的数据序列，然后建立相应的微分方程模型，从而预测事物未来发展趋势的状况。

在本题中，因为没有给出具体的用电量变化情况，最大载流量中含有较多的未知因素。为了对最大载流量进行估计，我们采用**灰色预测模型**。由于电流的变化具有周期性，选择前 13 个月的幅值电流作为灰色预测的序列，它们中包含了同样因为用电量增加而引起的载流量增大的部分信息。使用灰色预测模型对最大载流量进行预测，可以预知 8 月 15 日因用电量增加而需要增加的载流量，进而预知合理的动态增容量。

9.3 模型求解

首先进行准指数规律的检验，以验证数据可以进行**灰色预测**。根据前 13 个月的幅值载流量数据，光滑比小于 0.5 的数据占比是 83.33333333333334% (>50%)。除去前两个时期外，光滑比小于 0.5 的数据占比是 100.0 % (>90%)，满足灰色预测的条件，因此所选数据**可以进行灰色预测**。

进行灰色预测的结果如下图所示：

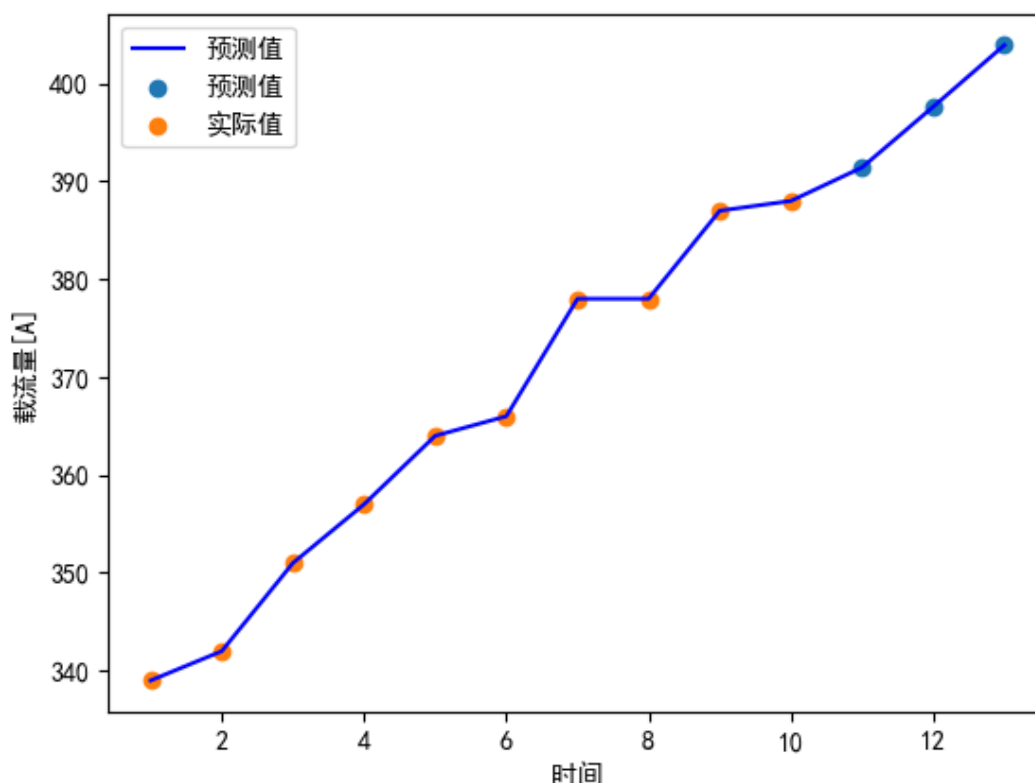


图 11 传统 GM(1,1) 灰色预测图

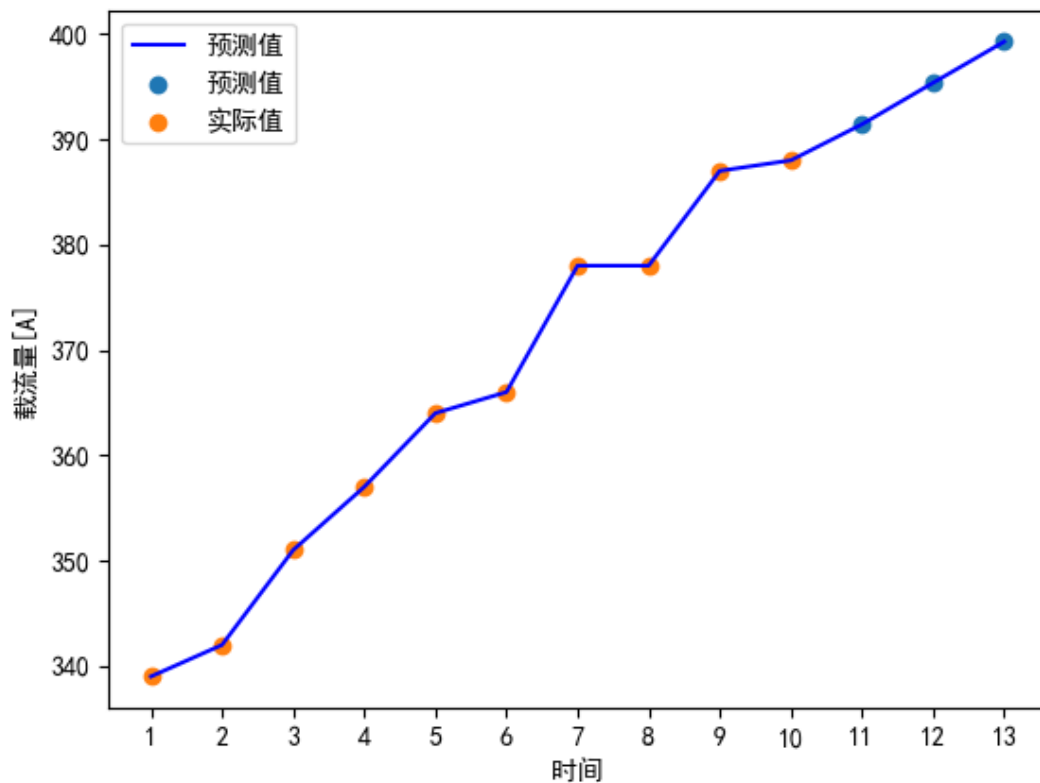


图 12 新 GM11 灰色预测图

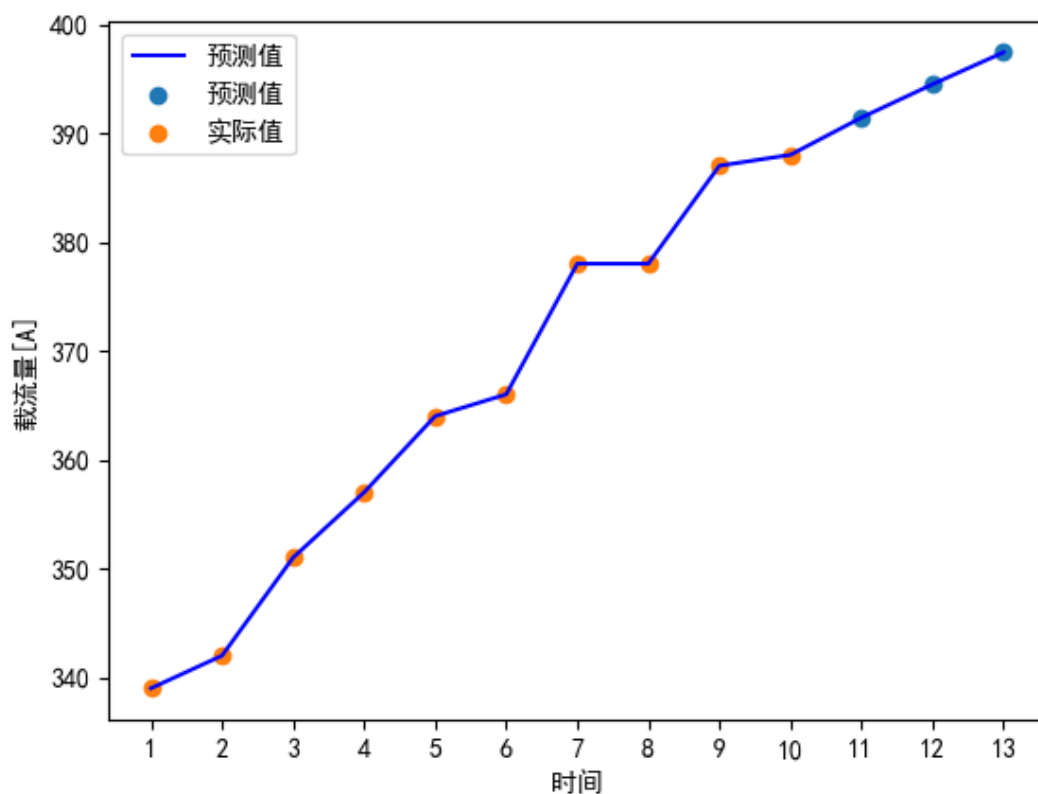


图 13 新陈代谢 GM11 灰色预测图

灰色预测结果如下表：

表 25 灰色预测结果表

灰色预测模型	传统 GM11	新 GM11	新陈代谢 GM11
--------	---------	--------	-----------

	灰色预测模型	灰色预测模型	灰色预测模型
预测的最大载流量[A]	391.42	397.62	403.91

9.4 动态扩容方案

根据最大载流量可以计算出需要的动态增容量。对于实际的动态扩容过程，主站根据现场检测到的环境温度、湿度、风速、日照和导线温度等参数计算出导线的最大载流量以及线路的隐性容量，并结合线路的实际载流量、额定载流量实时绘制出线路的载流量曲线的实时对比图。同时为了保证线路安全，现场采集导线弧垂参数保证线路安全^[3]。

十、模型的评价与改进

10.1 模型的优点

1. 利用多种工具处理数据，较好地分析附件所给出的大量数据，并且运用到模型的建立与求解中，实现了对数据的充分利用。
2. 运用了多种预测模型预测动态载流量，取长补短，避免了单一预测模型预测数据缺失的问题，能够较好地提取预测量的特征，且给出合理预测量。
3. 在因子分析中，将多个应力变量提取为两个公共因子，降低了模型求解的复杂度，且更利于得到合理的结果。
4. 充分利用了题目所给背景与条件，建立了变量之间的定量关系，更加全面地解决了题目所给问题。

10.2 模型的缺点

1. 部分模型的假设存在过于理想化的问题，导致模型出现局限性，无法进行更加全面的推广。
2. BP 神经网络预测自变量个数不足，导致预测结果存在一定误差。

10.3 模型的改进

1. 选择更加合理的模型，使模型能够更好地解决问题，提高模型的复用性，增强模型的说服力。
2. 运用更加合理的数据处理方法，使所补充数据更符合实际，以达到更好的模型拟合效果。

十一、参考文献

- [1] 柴永剑,张立军,严雨灵,等.基于 ALO-BP 神经网络的 SRAM 读时序预测[J].电子设计工程,2024,32(08):82-86+91.DOI:10.14022/j.issn1674-6236.2024.08.018.
- [2] 侯宇,王伟,韦微,等.输电线路动态扩容技术研究及应用[J].电力系统自动化,2021,45(17):189-198.
- [3] 周赞东,苟晓侃,曾胜强.输电线路动态扩容模块化计算技术研究[J].能源研究与管理,2017(02):55-57.DOI:10.16056/j.1005-7676.2017.02.013.
- [4] 王孔森. 输电线路动态容量系统应用分析及其风险评估方法[D].上海交通大学,2012.
- [5] 王孔森,盛戈皞,孙旭日,等.基于径向基神经网络的输电线路动态容量在线预测[J].电网技术,2013,37(06):1719-1725.DOI:10.13335/j.1000-3673.pst.2013.06.009.

附录

附录 1：支撑材料文件列表

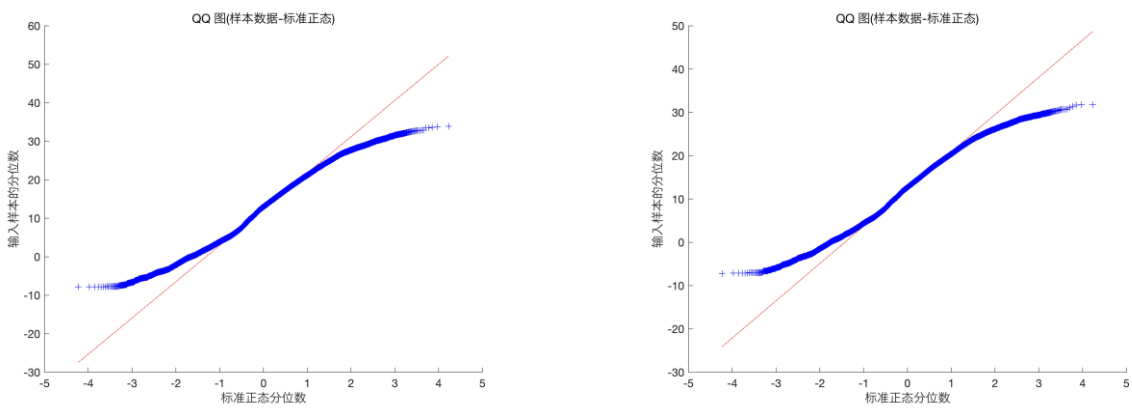
附录 2：补充表格、图片

附录 3：代码

附录 1 支撑材料文件列表	
	文件列表名
数据处理	整合处理数据.xlsx
	Spearman.m
问题一	Spearman.mat
问题四	傅里叶变换.py
	灰色预测.py

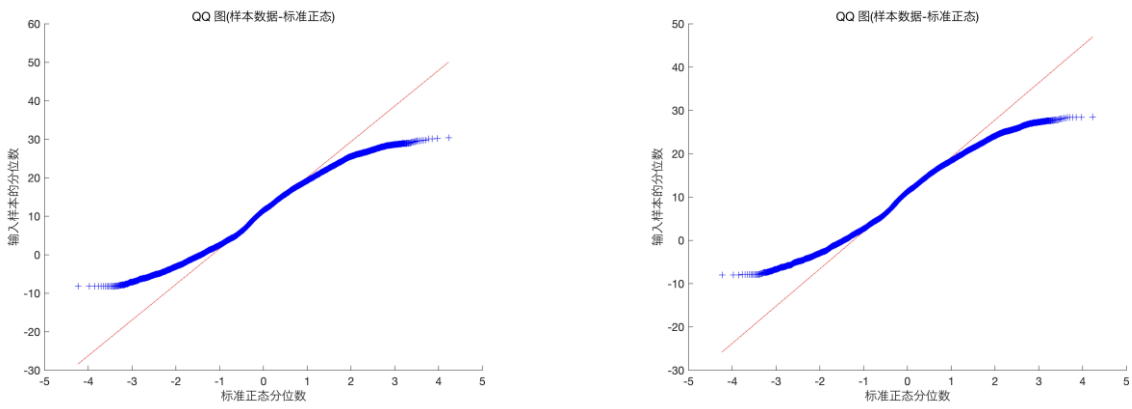
附录 2 补充表格、图片

样本数据的 QQ 图



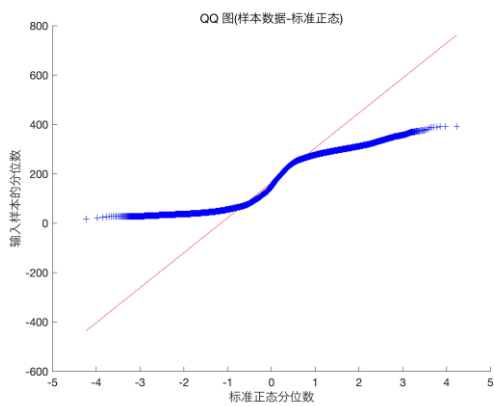
Line Temperature [°C] DEV1 的 QQ 图

Line Temperature [°C] DEV2 的 QQ 图

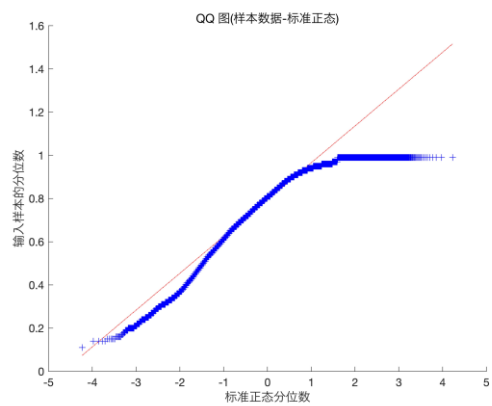


Ambient Temperature OTLM [°C] DEV1 的 QQ 图

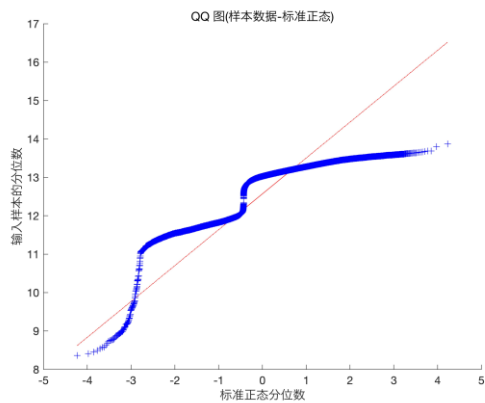
Ambient Temperature OTLM [°C] DEV2 的 QQ 图



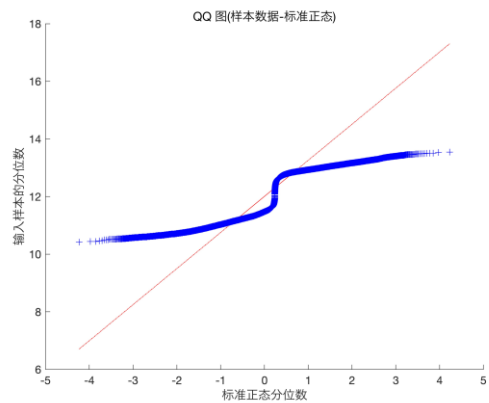
Humidity [%] DEV1 的 QQ 图



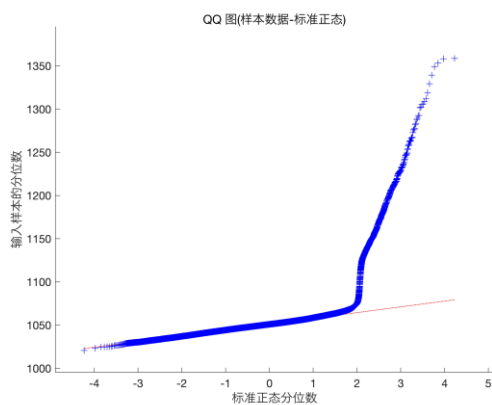
Humidity [%] DEV2 的 QQ 图



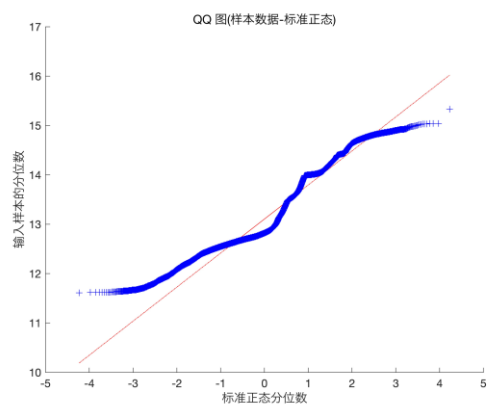
Angle [°] DEV1 的 QQ 图



Angle [°] DEV2 的 QQ 图



Altitude [m] 的 QQ 图



Vbat [V] 的 QQ 图

Spearman 相关系数表

表 26 DEV2 Ch1 的 Spearman 相关系数表

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	应力 [MPa]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9847	0.3068	-0.3329	-0.7575	0.2146	0.4269
Ambient Temperature OTLM [°C]	0.9847	1.0000	0.3278	-0.2966	-0.7714	0.2270	0.4292
Line Current [A]	0.3068	0.3278	1.0000	0.0896	-0.2366	0.1174	0.0344
Humidity [%]	-0.3329	-0.2966	0.0896	1.0000	0.3208	-0.0693	-0.2970
Angle [°]	-0.7575	-0.7714	-0.2366	0.3208	1.0000	-0.2343	-0.3742
应力 [MPa]	0.2146	0.2270	0.1174	-0.0693	-0.2343	1.0000	0.0627
Vbat [V]	0.4269	0.4292	0.0344	-0.2970	-0.3742	0.0627	1.0000

表 27 DEV2 Ch2 的 Spearman 相关系数表

Spearman相关系数							
	Line Temperature [°C] DEV1	Ambient Temperature OTLM [°C]	应力 [MPa]	Humidity [%]	Angle [°]	Altitude [m]	Vbat [V]
Line Temperature [°C]	1.0000	0.9847	-0.3327	-0.3329	-0.7575	0.2146	0.4269
Ambient Temperature OTLM [°C]	0.9847	1.0000	-0.3372	-0.2966	-0.7714	0.2270	0.4292
Line Current [A]	-0.3327	-0.3372	1.0000	0.1208	0.1835	-0.0960	-0.1080
Humidity [%]	-0.3329	-0.2966	0.1208	1.0000	0.3208	-0.0693	-0.2970
Angle [°]	-0.7575	-0.7714	0.1835	0.3208	1.0000	-0.2343	-0.3742
应力 [MPa]	0.2146	0.2270	-0.0960	-0.0693	-0.2343	1.0000	0.0627
Vbat [V]	0.4269	0.4292	-0.1080	-0.2970	-0.3742	0.0627	1.0000

Spearman 相关系数检验p值表:

表 28 DEV1Ch1 的 Spearman 相关系数检验p值表

数据类型	LT [°C]	应力 [MPa]	LC [A]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
应力 [MPa]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag[°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0
V [V]	0	0	0	0	0	0	1

表 29 DEV1Ch1 的 Spearman 相关系数检验p值表

数据类型	LT [°C]	AT [°C]	应力 [MPa]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
应力 [MPa]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag[°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0
V [V]	0	0	0	0	0	0	1

表 30 DEV2Ch1 的 Spearman 相关系数检验p值表

数据类型	LT [°C]	应力 [MPa]	LC [A]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
应力 [MPa]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag[°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0
V [V]	0	0	0	0	0	0	1

表 31 DEV2Ch2 的 Spearman 相关系数检验p值表

数据类型	LT [°C]	AT [°C]	应力 [MPa]	H [%]	A [°]	AL [m]	V [V]
LT [°C]	1	0	0	0	0	0	0
应力 [MPa]	0	1	0	0	0	0	0
LC [A]	0	0	1	0	0	0	0
H [%]	0	0	0	1	0	0	0
Ag[°]	0	0	0	0	1	0	0
AL [m]	0	0	0	0	0	1	0
V [V]	0	0	0	0	0	0	1

拟合与预测图

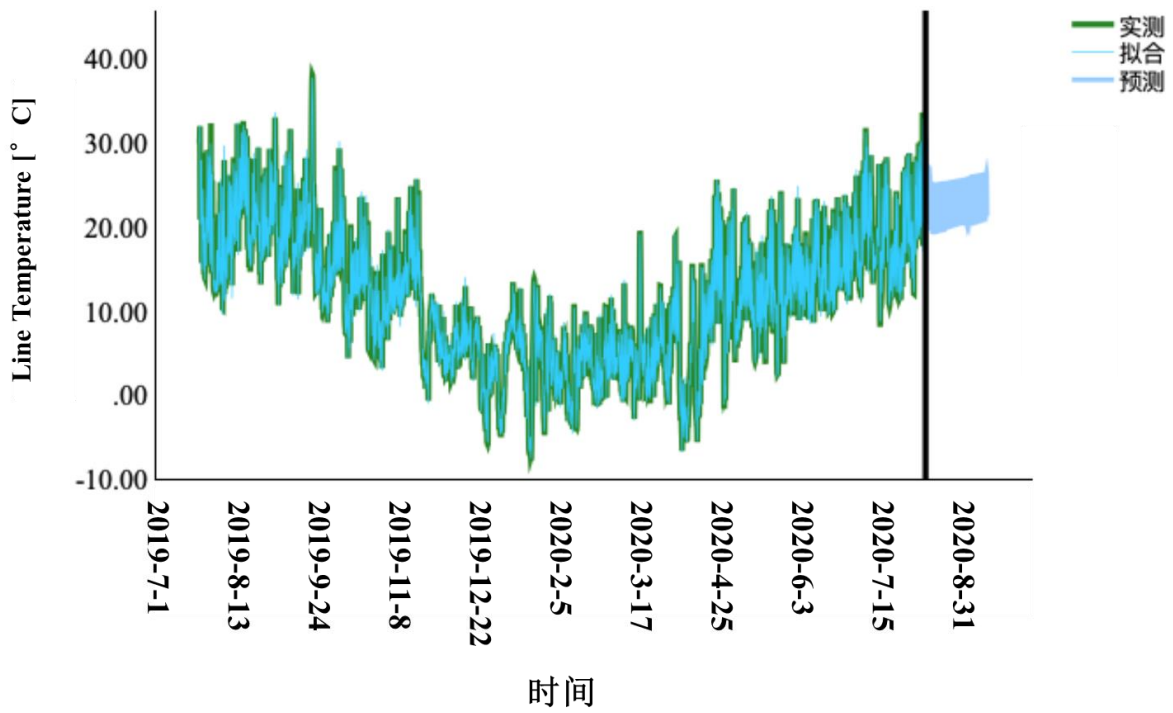


图 14 Line Temperature [°C]拟合与预测图

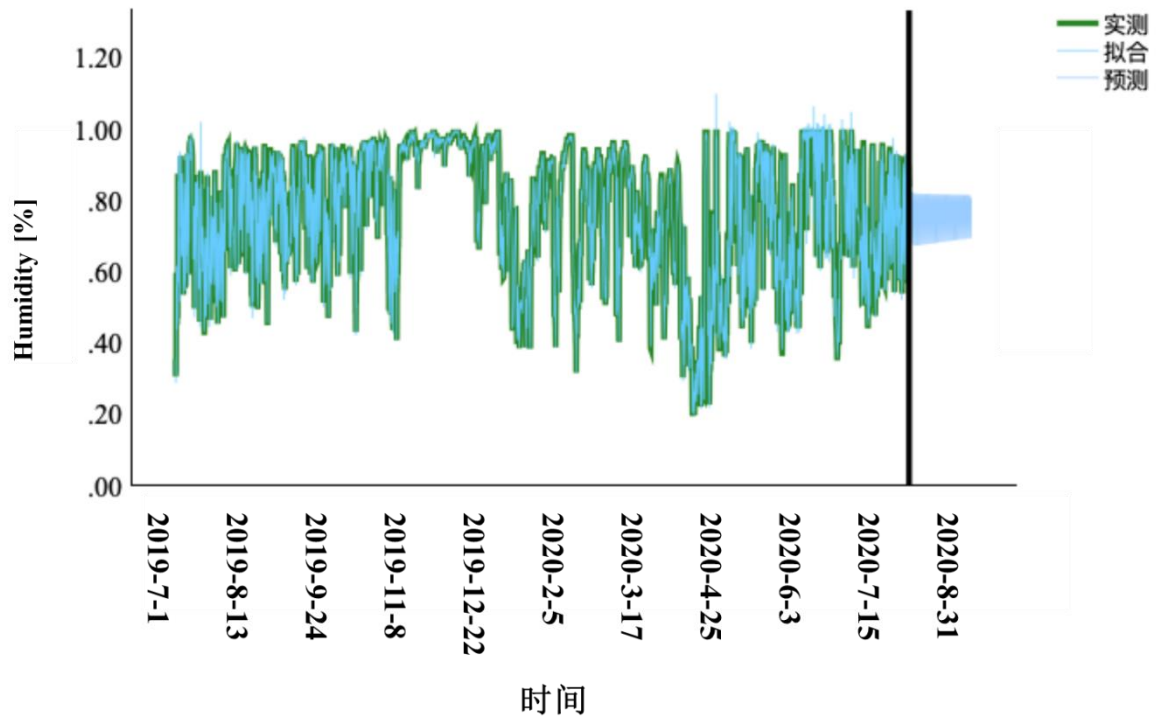


图 15 Humidity [%]拟合与预测图

非标准化典型相关变量对应的线性组合系数

表 32 集合 1 非标准化典型相关变量对应的线性组合系数表

变量	典型相关变量
LT[°C]DEV1	1.323
LT[°C]DEV2	-1.432
AT[°C]DEV1	-.141
AT[°C]DEV2	.121
AL[m]	-.005
AG[°]DEV1	-.933
AG[°]DEV2	-.518

表 33 集合 2 非标准化典型相关变量对应的线性组合系数表

变量	典型相关变量
H[%]DEV1	-9.623
H[%]DEV2	8.988
V[V]	-1.253

典型相关变量的典型载荷表与已解释的方差比例表

表 34 集合 1 典型相关变量 典型载荷表

变量	典型相关变量
LT[°C]DEV1	-.386
LT[°C]DEV2	-.459
AT[°C]DEV1	-.419
AT[°C]DEV2	-.446
AL[m]	-.150
AG[°]DEV1	-.223
AG[°]DEV2	.022

表 35 集合 2 典型相关变量 典型载荷表

变量	典型相关变量
H[%]DEV1	-.116
H[%]DEV2	.190
V[V]	-.634

表 36 已解释的方差比例

典型 变量	集合 1* 自身	集合 1* 集合 2	集合 2* 自身	集合 2* 集合 1
3	.115	.002	.151	.003

附录 3 代码

代码 1

介绍：该段代码是由 MATLAB 编写，绘制 qq 图，进行 JB 检验，计算 Spearman 相关系数

```
% qq 图进行粗检验
qqplot(x1(:,1))
% 依次处理 x1, x2 的每一列
% qqplot(x1(:,2))
% qqplot(x1(:,3))
% qqplot(x1(:,4))
```

```

% qqplot(x1(:,5))
% qqplot(x1(:,6))
% qqplot(x1(:,7))
% qqplot(x2(:,1))
% qqplot(x2(:,2))
% qqplot(x2(:,3))
% qqplot(x2(:,4))
% qqplot(x2(:,5))
% qqplot(x2(:,6))
% qqplot(x2(:,7))

%正态分布 J-B 检验，H 输出 1 代表不符合，输出 0 代表符合
%用循环检验所有列的数据
columns = size(x1,2); % 获取变量列数
P = zeros(1,5);
for i = 1:columns
[h,p] = jbtest(x1(:,i),0.05);
H(i)=h;
P(i)=p;
end
disp(H)
disp(P)

% columns = size(x2,2);
% P = zeros(1,5);
% for i = 1:columns
% [h,p] = jbtest(x1(:,i),0.05);
% H(i)=h;
% P(i)=p;
% end
% disp(H)
% disp(P)

%斯皮尔曼相关系数检验
[R,P]=corr(x1, 'type' , 'Spearman')

% [R,P]=corr(x2, 'type' , 'Spearman')

```

代码 2

介绍：该段代码是有 Python 编写，绘制傅里叶变换频率分布图，研究周期性

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_excel('整合处理数据.xlsx', sheet_name =
'Sheet5',header=None, skiprows=1)

```

```

data = df.iloc[:, 0].values

# 傅立叶变换
fft_result = np.fft.fft(data)
fft_freq = np.fft.fftfreq(len(data))

# 去除直流分量
nonzero_indices = np.where(fft_freq != 0)[0]
nonzero_fft_result = np.abs(fft_result[nonzero_indices])
nonzero_fft_freq = fft_freq[nonzero_indices]

main_freq_index = np.argmax(nonzero_fft_result)
main_freq = nonzero_fft_freq[main_freq_index]

# 计算周期
period = 1 / np.abs(main_freq)

# 作图
plt.figure(figsize=(10, 6))
plt.plot(nonzero_fft_freq, nonzero_fft_result)
plt.xlabel('Frequency')
plt.ylabel('Amplitude')
plt.title('FFT Spectrum')

# 添加峰值标记
plt.scatter(nonzero_fft_freq[main_freq_index],
            nonzero_fft_result[main_freq_index], color='red', label='Peak')

# 导出图片
plt.savefig('fft 结果图.png')
plt.legend()
plt.show()

print("数据的周期为:", period)

```

代码 3

介绍：该段代码由 Python 编写，实现灰色预测

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import gm

# 让图片可以显示中文
plt.rcParams['font.sans-serif'] = 'SimHei'
# 让图片可以显示负号
plt.rcParams['axes.unicode_minus'] = False

```

```

all_data = pd.read_excel(io='整合处理数
据.xlsx',sheet_name='Sheet4',engine='openpyxl')
data = all_data['MAX 排序'].tolist()
years = [1,2,3,4,5,6,7,8,9,10,11,12,13]

data_1 = np.cumsum(data)
print(data_1)

#判断是否符合准指数规律
rho = np.divide(data[1:],data_1[0:-1])# 两数组间点除
plt.plot(years[1:],rho,linestyle = '-',)
plt.axhline(y=0.5,color='r',linestyle='--')
plt.scatter(years[1:],rho,c = 'r')
plt.xlabel('时间')
plt.ylabel('原始数据的光滑度')
plt.show()

result_rho_1 = float(np.sum(rho<0.5)/len(rho))
print("光滑比小于 0.5 的数据占比是",result_rho_1 * 100,'%')
result_rho_2 = float(np.sum(rho[2:]<0.5)/len(rho[2:]))
print("除去前两个时期外, 光滑比小于 0.5 的数据占比是", result_rho_2
* 100,'%')

test_num = 0
if len(data) <= 7:
    test_num = 2
elif len(data) >7:
    test_num = 3

data_train = data[:-test_num]
print('实验数据为: \n',data_train)
data_test = data[-test_num:]
print('测试数据为: \n',data_test)

(result1,x0_hat1,_,_) = gm.gm(data_train,test_num)
print(result1)
print(x0_hat1)

data_result1 = np.concatenate((data_train,result1))

plt.xlabel('时间')
plt.ylabel('载流量[A]')
plt.plot(years,data_result1,c = 'b',label = '预测值')
plt.scatter(years[10:13],data_result1[10:13],label = '预测值')
plt.scatter(years[:10],data[:10],label = '实际值')

```



```

plt.legend()
plt.show()

result2 = gm.new_gm(data_train,test_num)
data_result2 = np.concatenate((data_train,result2))
print(data_result2)

plt.xlabel('时间')
plt.ylabel('载流量[A]')
plt.plot(years,data_result2,c = 'b',label = '预测值')
plt.legend()
plt.scatter(years[10:13],data_result2[10:13],label = '预测值')
plt.scatter(years[:10],data[:10],label = '实际值')
plt.legend()
plt.xticks(years)
plt.show()

result3 = gm.met_gm(data_train,test_num)
data_result3 = np.concatenate((data_train,result3))

plt.xlabel('时间')
plt.ylabel('载流量[A]')
plt.plot(years,data_result3,c = 'b',label = '预测值')
plt.legend()
plt.scatter(years[10:13],data_result3[10:13],label = '预测值')
plt.scatter(years[:10],data[:10],label = '实际值')
plt.legend()
plt.xticks(years)
plt.show()

print(data_result1)
print(data_result2)
print(data_result3)

```