

第一届长三角高校数学建模竞赛

题目 基于 BP 神经网络以及模糊综合评价的管道温度曲线分析

摘要:

为了探究在燃煤发电实际生产过程中影响管道温度的主要因素以及给出超温段的最优调节策略。本文基于 BP 神经网络, 主要通过 BP 神经网络和模糊综合评价的结合, 分析各个管道的温度数据以及对其进行评价, 并建立了数学模型对十个管道求出其温度变化规律以及判断出引发超温现象的主要操作变量并对超温段进行了优化。

针对问题一, 我们从简单的平均值, 方差, 最大值最小值出发, 并且给出十根管道的温度变化曲线, 发现其对称性较强, 从而考虑到偏度和峰度。同时, 再从题目所提到的时间序列的角度出发, 发现十个管道的序列性不强, 没有趋势性, 表现出很强的随机性, 猜测其为白噪声序列, 于是紧接着从协方差和自相关系数去进行分析。最终发现曲线存在很强的内生性问题, 对于后面的回归分析有很大的影响。

针对问题二, 我们需要对十个管道的曲线进行评价并确定最优和最劣的管道温度曲线, 我们利用模糊综合评价的思想, 不难确定几个判断好坏的目标: 1、从安全和使用周期的长短来讲, 平均温度越低越好; 2、从温度的波动情况来看, 波动越小越好; 3、同时我们注意到, 445℃为临界的温度线, 如果高于它必然扣分。运用三个隶属度函数的构建和求解, 得出每根管道的分数后不难得出最优管道为管道六, 最劣管道为管道十。

针对问题三, 我们先通过岭回归, 发现出线性回归越到后面效果越差, 故而用基于 BP 神经网络对总体情况进行分析来解决内生性强的问题, 并且在此过程中优化了自适应率以及隐藏层的个数, 从而使结果更准确。由于参数过于庞大, 我们借助了 STATA 进行全变量的回归拟合进行了验证。最后用模糊矩阵的传递闭包得出 153 个参数可以合并为六个, 并且得出合并后的参数的回归系数。本题处理的亮点是: BP 神经网络和最后的回归系数都很精确, 前者的拟合优度为 10 的-4 次方级别, 最后的拟合误差控制在百分之 0.5 左右。

针对问题四, 我们通过双类多元回归以及向后逐步回归对超温影响因素进行判断和分析。对于双类多元回归的建立, 我们主要考虑单参数中值突变的情况。在这种状况下, 我们进行优化线性回归, 采用问题数据和最少的数据量对比, 使得错误参数体现的更加明显, 将最大错误参数作为结果; 同时, 我们采用向后逐步回归的求解模式进行进一步验证, 得出结论, 导致超温现象的主要操作变量为#2 角 B 层二次风调节挡板位置反馈。

针对问题五, 我们将题目的模型简化为已知函数最佳值, 调整位置求解最优解的过程。故而采用粒子群算法, 并在传统的求解过程中两次改变惯性权重, 使得最后求解的过程更加准确。同时和第四题的结论进行比对, 我们得出将#2 角 B 层二次风调节挡板位置反馈的权重增大一倍, 同时对于炉膛压力和炉膛与风箱差压进行平稳性处理。此外, 由于第十根管道的平均温度偏高, 所以对于其他的权重分别下降 0.1 的随即惯性权重比的调整。经过 BP 神经网络和聚类分析发现, 不仅调整后的模型误差小, 而且与曲线六完美的契合。

关键词: 内生性的研究, 模糊综合评价, BP 神经网络, 模糊矩阵的传递闭包, 向后逐步回归, 粒子群算法

目录

一、 问题重述.....	1
1.1 问题背景.....	1
1.2 问题提出.....	1
二、 问题分析.....	1
2.1 问题一的分析.....	1
2.2 问题二的分析.....	1
2.3 问题三的分析.....	2
2.4 问题四的分析.....	2
2.5 问题五的分析.....	2
三、 模型假设.....	3
3.1 第二问模型假设.....	3
3.2 第三问模型假设.....	3
3.3 第四问模型假设.....	3
3.4 第五问模型假设.....	3
四、 符号说明.....	3
五、 模型的建立和求解.....	4
5.1 问题一的模型建立和求解.....	4
5.1.1 多角度确定指标.....	4
5.1.2 指标的求解.....	5
5.2 问题二的模型建立和求解.....	6
5.2.1 建立模糊综合评价的模型.....	6
5.2.2 通过隶属度对其进行评分处理获得最优最劣温度曲线.....	7
5.3 问题三的模型建立和求解.....	9
5.3.1 岭回归尝试解决线性回归.....	9
5.3.2 BP 神经网络的建立及其优化.....	9
5.3.3 BP 神经网络求解，并用 STATA 进行辅助性验证.....	11
5.3.4 传递闭包聚类求主要参数并确定误差.....	11
5.4 问题四的模型建立和求解.....	13
5.4.1 双类多元线性回归的建立.....	13
5.4.2 对各个变量的向后逐步回归求解.....	13
5.5 问题五的模型建立和求解.....	14
5.5.1 粒子群求解模型建立.....	14
5.5.2 运用优化粒子群算法求解并做出解释.....	15
六、 模型的分析与检验.....	16
七、 模型的推广与评价.....	17
7.1 模型的优点.....	17
7.2 模型的缺点.....	18
7.3 模型的推广.....	18
八、 参考文献.....	18
九、 附录.....	19
9.1 三个隶属度函数的模板（分段函数）.....	19
9.2 BP 神经网络基础代码.....	19
9.3 模糊矩阵的传递闭包聚类.....	20
9.4 管道温度的系统聚类.....	22
9.5 粒子群代码.....	22
9.6 灰色预测代码.....	24

一、问题重述

1.1 问题背景

在中国当代的社会发展中，火力发电仍然是主要发电形式。主要是以燃烧煤炭来作为总动力的来源。煤炭从磨煤机出来后，经煤粉管道直接送入燃烧器，并由燃烧器将其输入炉膛进行燃烧，使得煤粉在炉膛内燃烧释放出大量热量，锅炉四周的水冷壁再对其进行降温。此时，我们要保证炉膛内的温度不能太高，且要保证其稳定。显然，再不同的情况下，水冷的效果必然不一样。那么，在已有的温度曲线以及各种影响指标的情况下，如何通过现有的测量数据去判断主要的影响因素并且对其进行改进，成为了一个重要的问题。同时要对于现实中的超温现象做出合理的解释，并且在改变变量数少，操作变量的调控量小的前提下使得工作曲线（温度）达到一个比较合理的状态。

1.2 问题提出

第一问通过对于十根水冷壁管道的温度数据，进行曲线刻画，并且通过平均温度，温度波动情况，最高最低温度或者是变化趋势等，对其温度的变化进行一个描述。

第二问则是确定自己的评价标准体系，对于管道的温度数据线进行评价，同时要判断出最优和最差的曲线。

第三问通过对于附件中数据，建立起各个影响因素与温度之间的关系，构建出函数的同时对于函数的各项指标进行评价。

第四问要在前面几题的基础上对于样本点的超温现象做出解释，同时根据其数据，找出引发超温现象形成的主要操作变量。

第五问主要是对于曲线的超温段的优化，在改变操控变量少，调控量小，与最优解尽可能吻合的前提下，建立起优化模型

二、问题分析

2.1 问题一的分析

第一题要求对于管道的温度做出一些描述，开放程度较大。对于这样类型曲线，我们不光可以从简单的平均值，方差，最大值最小值去分析。仔细观察的话不难发现，十根曲线呈比较明显的对称性，故而可以从偏度和峰度的角度去考虑。同时，题目中说从时间序列的角度去进行分析。对于曲线的初步判断中，我们发现，十个管道的序列性不强，没有趋势性，表现出很强的随机性，故而为白噪声序列。那么接下来就可以从协方差和自相关系数去进行分析并做出评价。最后可以适当分析以下各个温度曲线的趋势，并将其关联，为后几题做出铺垫。

2.2 问题二的分析

第二问要求我们确定最优和最劣的管道温度曲线，不难发现要自行确定几个判断好坏的目标。1、从安全和使用周期的长短来讲，平均温度越低越好。2、从温度的波动情

况来看，波动越小越好。3、同时我们注意到， 445°C 为临界的温度线，如果高于它必然扣分。同时为了增加判断的精度，我们设立 420°C 的临界线作为次参考标准。因为评价并不是准确的，所以我们采用模糊综合评价的模型，通过柯西分布的偏小型函数， T 分布的偏小型函数和 K 次抛物形分布的偏大型扣分系统对其进行评分，自然的，评分越高，说明温度曲线越好，从而得出最后的结论。并且，在后面对于参数的分析中，我们也可以通过聚类对于本小题的解题结果做出一个相应的印证来判断是否符合逻辑。

2.3 问题三的分析

第三问要求根据附录中的数据来建立模型。此题因为数据量大，用单一的模型不一定保证准确，所以我们采用先通过 BP 神经网络对总体的情况进行分析，并且再次过程中，我们优化了自适应率以及隐藏层的个数，使得最后求解的结果更加准确。因为这种方法的变量参数依然很大，所以用 STATA 进行全变量的回归拟合，来验证刚才的 BP 神经网络是否合理。为了得到最为有效的参数，我们使用模糊矩阵传递闭包的方法分别对曲线和参数进行聚类，将其分类后再用多元线性函数拟合的思想或者相关性系数去做，既使得变量少，表达式简单，避免龙格现象的产生；又使得拟合的效果收到内生性的干扰减小。

2.4 问题四的分析

第四问要求根据第 10 个管道的温度曲线以及附件 2 中的数据分析并定位引发超温现象的主要操作变量。此题数据量大，无法从数据中很快的找出哪些数据出了问题，但可以从图像以及题目所给的第 3172 个样本点出发，先在第 3172 个样本点附近找出出现了超温现象的数据，再用这些问题数据与其附近的正常数据量对比，从而使造成超温现象的变量产生错误更明显，进而定位出引发超温现象的主要操作变量。

2.5 问题五的分析

第五问要求针对第 10 个管道的温度变化曲线建立模型并给出超温段的最优调节策略。我们可以采用多目标优化，将最少的调控的变量数、最少的调控的操控变量数、调整后最优曲线与温度曲线的差值最小看作目标函数，再将三个目标函数合成一个目标函数，最后采用粒子群算法，改变了它们的自适应惯性权重及随机惯性权重，最后得出最优的调节策略。

三、模型假设

3.1 第二问模型假设

- 1、对于分数的判定符合主观的模糊函数的隶属度
- 2、函数具体的参数可以自行确定

3.2 第三问模型假设

- 1、不考虑题中给的数据有比较强烈的内生性情况出现
- 2、优化后的 BP 神经网络能够较好的进行收敛
- 3、模糊矩阵的传递闭包处理的 F 统计量只采用欧几里得距离法，从而得到最佳的 λ 值。

3.3 第四问模型假设

对于内生性的问题，在使用向后逐步回归后基本能排除其影响。

3.4 第五问模型假设

- 1、粒子群算法得到结果可以当成最优解，即在循环过程中必能找到解
- 2、惯性权重的设定对于参数的适应度符合预期

四、符号说明

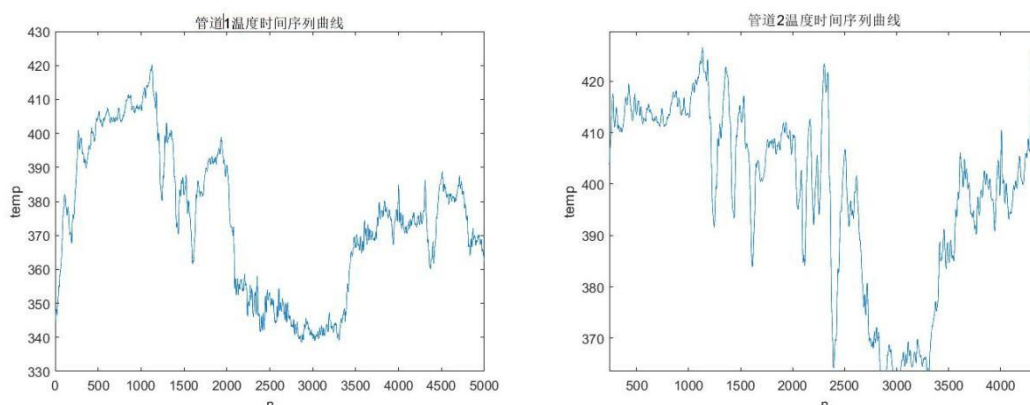
符号	说明	单位
k_2	二阶累积量的无偏估计	\
x_i	管道温度	\
r_{nn}	自相关系数	\
$\delta(\tau)$	自相关函数	\

五、模型的建立和求解

5.1 问题一的模型建立和求解

5.1.1 多角度确定指标

首先，我们用 excel 画出十个管道的温度曲线变化图，这边以管道一和管道二作为例子进行描述（具体见附录）。



5-1-1 管道温度曲线

不难发现，这些图形基本以中间为轴，向两边呈对称的趋势，故而我们对于附件一中的十个管道的温度赋予统计特征：样本容量（Obs）、平均值（Mean）、标准差（Std. Dev.）、最小值（Smallest）、最大值（Largest）、偏度（Skewness）、峰度（Kurtosis）进行初步的分析，具体公式如下图所示：

$$\text{标准差: } S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

$$\text{极差: } Range = Highest - Lowest$$

$$\text{峰度: } \frac{(n+1)n}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{k_2^2} - 3 \frac{(n-1)^2}{(n-2)(n-3)}$$

$$\text{偏度: } \frac{\sqrt{n(n-2)}}{n-2} \left[\frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{3}{2}}} \right]$$

5-1-2 公式

根据十张图的大致形状可猜测十个管道的温度变化满足：1、序列不相关 2、无趋势性 3、随机性强，可以得出该十个温度变化曲线可能属于白噪声序列。对于白噪声序列，我们一般从自相关系数和协方差两方面进行判断。

$$r_{nn} = E\{n(t)n(t-\tau)\} = \delta(\tau)$$

在一般的情况下，对于不想管的随机变量序列构成的数中，即对于所有的 S 不等于 T，故而我们可以得到随机变量 X_t 和 X_s 的协方差为 0。接下来就从实际运行结果中去进行判断。

$$COVMAT(i,j) = ((A(i,:)-E(i)) * (A(j,:)-E(j)))' ./ (n-1)$$

这会为后面求解时判断函数之间的拟合关系是否存在内生性提供重要的指标线索。

5.1.2 指标的求解

根据公式，我们可以得到：

	管道 1	管道 2	管道 3	管道 4	管道 5
标准差	21.4519	16.85238	10.67595	16.37582	12.97768
极差	81.6	69.8	56	70.3	59.9
偏度	0.0420382	-0.7799583	-0.0637034	-0.1732019	-0.0743229
峰度	1.992701	2.650248	2.506752	2.013424	2.112343
	管道 6	管道 7	管道 8	管道 9	管道 10
标准差	16.65687	16.0218	17.21068	17.70641	18.69408
极差	71.5	65.9	70.1	85.6	86.6
偏度	-0.1662122	-0.2756255	0.4428037	0.1411771	-0.2498083
峰度	1.970328	2.118567	2.193412	2.599508	2.137842

5-1-3 四个指标的数据展示

管道1温度					
	Percentiles	Smallest			
1%	339.7	338.6			
5%	341.4	338.6			
10%	343.6	338.6	Obs	5,000	
25%	354.5	338.6	Sum of Wgt.	5,000	
50%	374.6		Mean	374.3109	
		Largest	Std. Dev.	21.4519	
75%	391	419.9			
90%	404.7	420	Variance	460.1841	
95%	408.1	420.1	Skewness	.0420382	
99%	414.8	420.2	Kurtosis	1.992701	
管道2温度					
	Percentiles	Smallest			
1%	359.7	356.8			
5%	364.6	356.8			
10%	367.5	356.8	Obs	5,000	
25%	389.45	356.8	Sum of Wgt.	5,000	
50%	401.3		Mean	397.9088	
		Largest	Std. Dev.	16.85238	
75%	411.2	426.3			
90%	415.8	426.5	Variance	284.0029	
95%	419.25	426.5	Skewness	-.7799583	
99%	423.6	426.6	Kurtosis	2.650248	

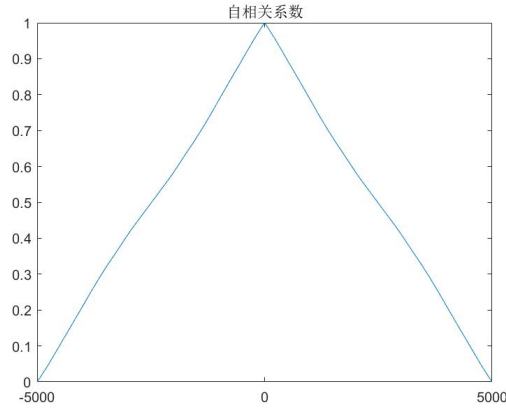
5-1-4 统计指标的峰度与偏度

首先偏度为零，则随机变量为对称分布；峰度大于 3，则其密度函数的最高处比正态分布更尖，而两侧尾部更厚（正态分布的峰度为 3）。

而十个管道的偏度都不超过 0 ± 0.8 ，峰度都小于 3，可以将十个管道的温度近似看成对称分布，而其密度函数的最高处与两侧尾部的差距却不如正态分布明显。

其中，可以发现管道 3 的温度变化曲线更趋近于正态分布。

从管道的自相关系数和协方差进行讨论，得出以下的关系图和数据：



5-1-5 自相关系数

$$COVMAT = 460.1841$$

不难发现，曲线虽然说没有呈现出较强的周期性，但是很明显存在强烈的内生性问题 and 外界的干扰，所以，对于后面参数的把握就不能简单通过多元线性回归去解决。

从整体的趋势上来看，每根曲线大体的趋势都是先下降后上升，过程中会有几次比较大的波动。

5.2 问题二的模型建立和求解

5.2.1 建立模糊综合评价的模型

对于水冷壁管道的温度进行评价，我们主要从以下几个方面入手：

- 1、总体的温度越小越好
- 2、若温度超过了 445℃，则对其进行扣分处理（超过警戒值的占空比）
- 3、温度的平稳性，即方差不要太大

那么根据实践经验，我们分别对于这三个指标指派一个隶属度函数，用模糊综合评价的方式对于曲线进行评分。以下就是三个隶属度函数：

$$A_1(x) = \begin{cases} 100 & x \leq a \\ 100 \exp[-k(x-a)] & x \geq a (k > 0) \end{cases}$$

5-2-1 \mathcal{T} 分布的偏小型函数

$$A_2(x) = \begin{cases} 0, & x \leq a, \\ \left(\frac{x-a}{b-a}\right)^k, & a < x < b, \\ 1, & x \geq b, \end{cases}$$

5-2-2 \mathcal{K} 次抛物形分布的偏大型扣分系统

$$A_3(x) = \begin{cases} 1, & x \leq a \\ \frac{1}{1 + \alpha(x-a)^\beta}, & x > a \end{cases}$$

5-2-3 柯西分布的偏小型函数

然后，我们需要得到是个水冷壁管道温度的平均值，方差以及其温度大于 445℃（420℃）的占空比，按照以上三个函数的模型进行打分以及扣分的处理，选出最大值和最小值。

令最后的得分为 P，三个隶属度函数分别对应的值为 X，Y，Z，则最后得分的计算公式为：

$$P = \frac{X - Y + Z}{3}$$

5.2.2 通过隶属度对其进行评分处理获得最优最劣温度曲线

下表是对于管道温度的一个初步数据分析：

Variable	Obs	Mean	Std. Dev.	Min	Max
管道1温度	5,000	374.3109	21.4519	338.6	420.2
管道2温度	5,000	397.9088	16.85238	356.8	426.6
管道3温度	5,000	396.7092	10.67595	364.2	420.2
管道4温度	5,000	387.6063	16.37582	350.6	420.9
管道5温度	5,000	391.3618	12.97768	360	419.9
管道6温度	5,000	376.8434	16.65687	338	409.5
管道7温度	5,000	386.1856	16.0218	350.6	416.5
管道8温度	5,000	378.6845	17.21068	351.2	421.3
管道9温度	5,000	408.0336	17.70641	368.7	454.3
管道10温度	5,000	424.9091	18.69408	378.8	465.4

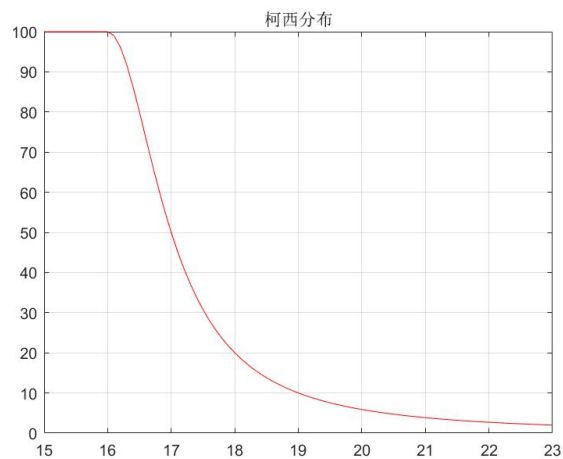
5-2-4 管道温度的初步数据分析

其次，我们统计十个管道中温度分别大于 445℃ 和 420℃ 的数量，得到的结果如下图所示：

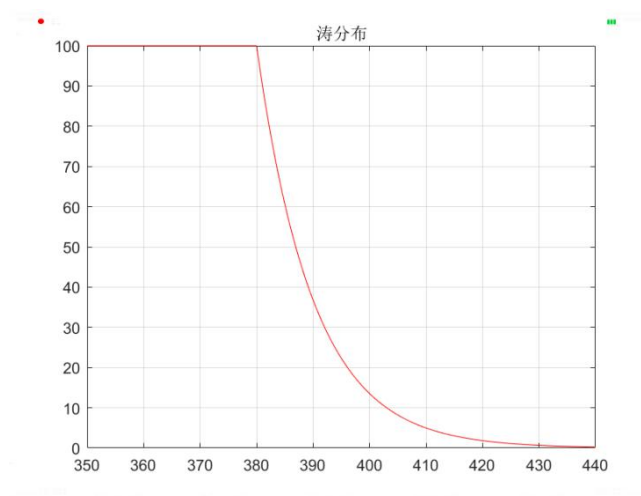
	管道一	管道二	管道三	管道四	管道五	管道六	管道七	管道八	管道九	管道十
>445℃	0	0	0	0	0	0	0	0	146	803
>420℃	2	209	2	25	0	0	0	0	1098	2704
占空比 1	0	0	0	0	0	0	0	0	2.92%	16.06%
占空比 2	0	4.18%	0	0.5%	0	0	0	0	21.96%	54.08%

5-2-5 超温现象的占空比数据统计

接下来就是确定三个隶属度函数各个变量的值，首先，我们先画出三个隶属度函数的大致形状，确定其加分和扣分的梯度标准点。

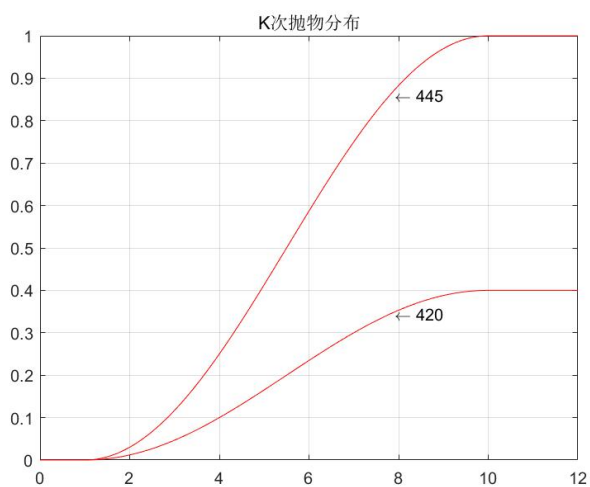


5-2-6 柯西分布偏小型



5-2-7 T 分布偏小型

K 次抛物型分布由于分为 445℃ 和 420℃ 的两个评价标准, 所以选用同一函数的两个不同参数设定的形式解决, 得到以下图示:



5-2-8 K 次抛物分布偏大型

然后对是个管道温度曲线进行打分, 可以得到得分表:

	管道一	管道二	管道三	管道四	管道五	管道六	管道七	管道八	管道九	管道十
温差	3	95	100	96	100	97	100	48	46	15
平均值	100	30	30	32	31	100	75	100	8	2
445℃	-0	-0	-0	-0	-0	-0	-0	-0	-5	-50
430℃	-0	-5	-0	-0	-0	-0	-0	-0	-50	-50
得分	34	40	40	43	43	66	58	49	0	-28

所以我们得到最优曲线为管道六，最劣管道为管道十。

5.3 问题三的模型建立和求解

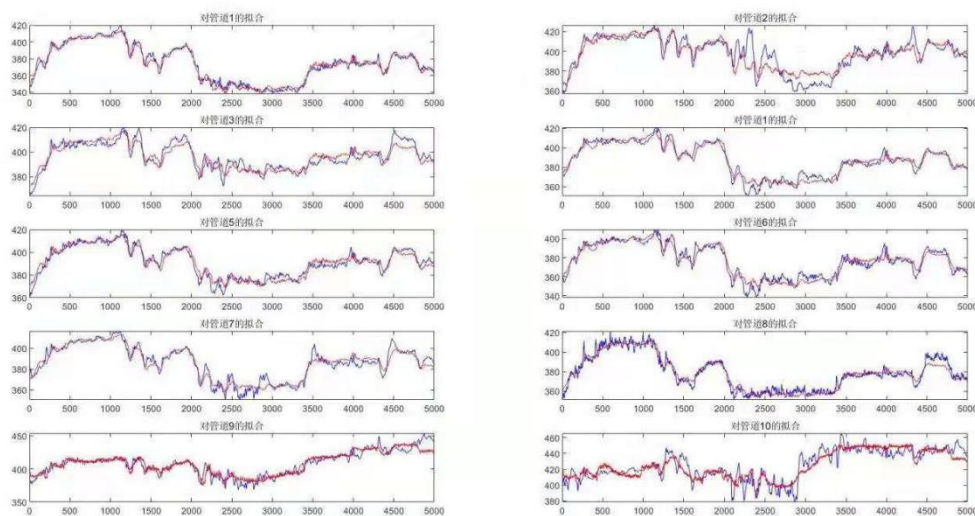
5.3.1 岭回归尝试解决线性回归

第一题中，我们就发现参数的内生性情况比较严重，所以在对于数据的初步尝试中，首先排除用岭回归去解决多元线性函数的回归问题。考虑到此题中的参数数量过于庞大，会产生内生性的影响而导致部分数据拟合的结果并不适用于其他管道的情况，我们对此进行验证，以便于后期使用更加精密的算法得到最优解。

在加入了惩罚函数的情况下：

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \left\{ \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^P x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^P \beta_j^2 \right\}$$

我们用软件对其进行初步的拟合得到下图：

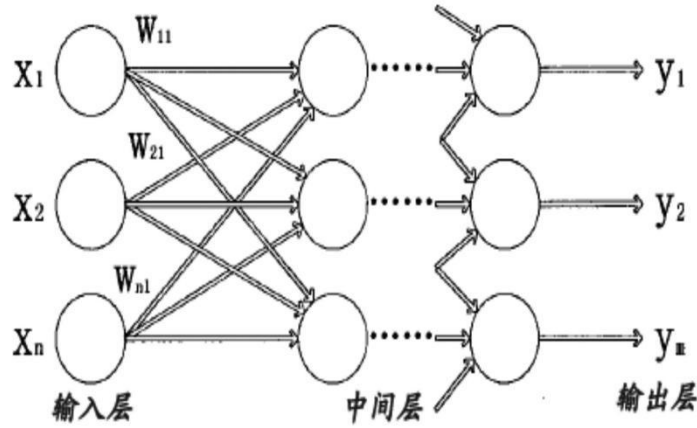


5-3-1 岭回归初步判断

不难发现，与我们预判的结果一样，越到后面，预测的结果越差，所以我们不得不采用其他的办法对问题进行求解。

5.3.2 BP 神经网络的建立及其优化

于是我们采用 BP 神经网络的思路去解决水冷壁的温度变化规律。首先，我们构造一个神经 BP 网络，如下图所示：



5-3-2 神经网络示意图

BP 神经网络的过程主要分为两个阶段，第一阶段是信号的前向传播，从输入层经过隐含层，最后到达输出层；第二阶段是误差的反向传播，从输出层到隐含层，最后到输入层，依次调节隐含层到输出层的权重和偏置，输入层到隐含层的权重和偏置。首先，我们打乱所有的样本数据，然后抽取 50 组数据作为样本，对其进行训练。然后抽取 10 组来判断训练得到的结果是否符合预期。同时我们可以得到隐层权值，隐层阈值，隐层到输出层权值和输出层阈值

因为本题中的参数众多，所以我们在传统的 BP 神经网络模型中做出如下的改进：

- 1、大大增加隐藏层的数量，使得函数的对应关系更加的紧密
- 2、由于不同参数之间可能存在内生性的关系，且不同参数因为数值不同会导致权重不同，故而在过程中要进行归一化处理，并且要用蒙特卡罗模拟验证内生性的问题
- 3、固定的自适应率容易产生不稳定性的震荡或者是收敛速度慢，导致拟合的效果差。所以，我们要实施改变自适应率的数值，使其震荡减小，拟合优度高。

这边，我们提出自适应学习率的表达式：

$$\beta(t) = \frac{2}{1 + e^{-|t| \times 10^n}} - 1$$

其中，t 是 BP 神经网络的权值变化：

$$t = \frac{\partial err}{\partial W_j^L} = \frac{\partial err}{\partial f(S_j^L)} \times \frac{\partial f(S_j^L)}{\partial S_j^L} \times \frac{\partial S_j^L}{\partial W_j^L}$$

我们观察以上两个式子，不难发现， β 的函数中 t 为权值的变化，即误差曲面函数的梯度。所以说再训练的过程中，学习率会不断的减小，达到一个最小值，从而达到了自适应的目的。同时各个参数的权值也在此过程中不断的进行变化。

误差的偏导为：

$$\frac{\partial err}{\partial W_j^L} = f(S_j^L) - (1 - f(S_j^L)) \times (d - Y) \times X_n$$

其中 d 为期望输出，Y 为实际输出。在同时结合梯度上升、正态分布模型和自适应学习率的权值更新方式后，我们可以得到：

$$W_{nj}^L = W_j^L + \left(\frac{2}{1 + e^{-|t| \times 10^n}} - 1 \right) \times \frac{1}{\delta^2} \times \frac{-1}{\delta \times \sqrt{2\pi}} \times e^{-\left(\frac{err}{\delta}\right)^2} \times \text{sgn}(err) \times e^{|err|} \times f(S_j^L) \times (d - y) \times (1 - f(S_j^L)) \times X_n$$

其中：

$$\begin{cases} err = \text{sgn}(err) \times e^{|err|} \\ \text{sgn}(err) = \begin{cases} 1, & err \geq 0 \\ -1, & err < 0 \end{cases} \end{cases}$$

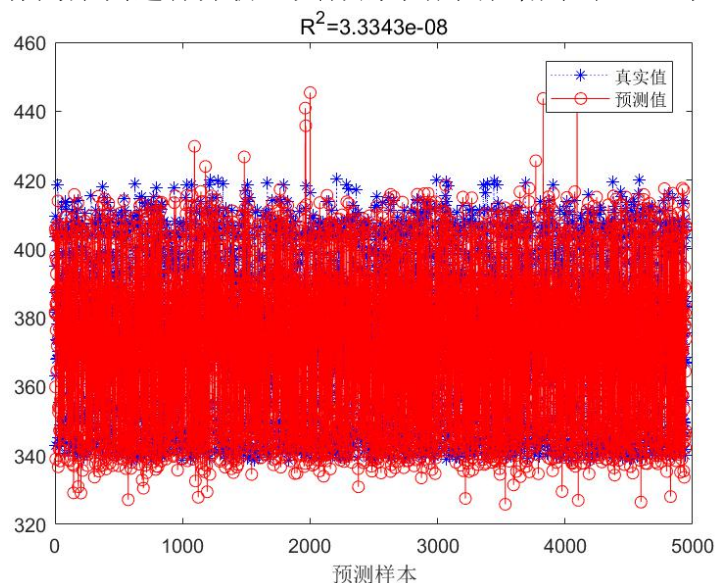
这样的设计可以保证在网络训练的后期，收敛的速度依然很快。相比于传统的梯度下降法或者是固定值的自适应率，这样的设计能让搜索的结果更加精细。

其次，我们在 BP 的神经网络搜索中，放入了 15 个隐藏层，使得函数的对应关系更加的多样，有利于函数的拟合（本题中参数数量大）。在计算完毕后，用蒙特卡罗模拟计算参数之间的内生性，若有内生性，考虑合并，并重复以上的步骤。

5.3.3 BP 神经网络求解，并用 STATA 进行辅助性验证

通过代码的编写，我们求出了 4950 组样本的测试值（有 50 组数据是样本，不算入其中），计算出误差，隐层权值，隐层阈值，隐层到输出层权值和输出层阈值等，并通过 `genFunction (net)` 调用出每个变量的权值。

我们以管道一作为案例进行分析（具体的十张图在附录中），不难发现：



5-3-3 管道一的 BP 神经网络预测样本

在如此大样本的情况下，模拟的情况非常的不错，拟合优度达标。具体的数据，误差等详见附录，权重也在附录中体现，误差极小。

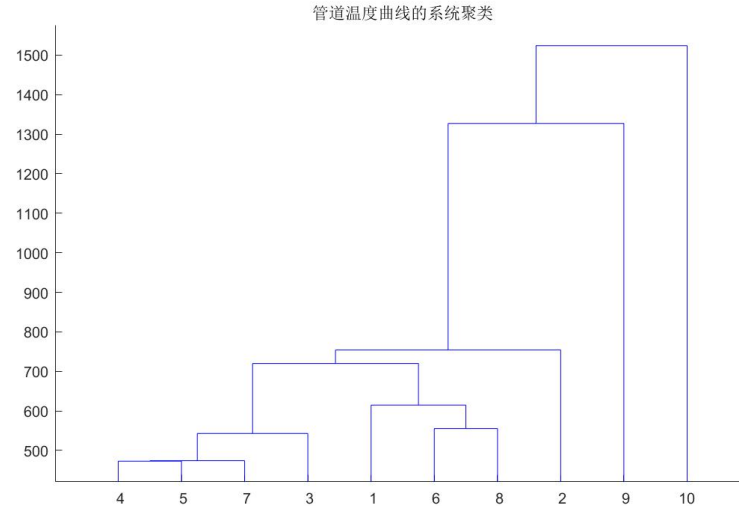
同时，我们用 `regress` 函数再 STATA 中对 153 个数据进行简单的多元线性回归。得到的结果详见附录中的 excel 表格。在得到的 95%置信区间中，我们寻找出 $P > |z|$ 值小于或等于 0.05 表示结果显著的参数。

5.3.4 传递闭包聚类求主要参数并确定误差

为了使得参数的聚集效果更加的明显，我们采用模糊矩阵的传递闭包，分别对温度曲线和参数进行聚类，将重要参数合并，简化拟合的函数，防止过于复杂。

运行代码（详见附录）：

将每个管道的 5000 组数据做成一个行矩阵，得到的聚类结果如下图所示：



5-3-4 管道温度的系统聚类

同时，这张图也能够变相的验证我们第二小题的结论（管道 10 最劣，管道 6 或者 7 最优，只不过在我们第二题的综合评价的体系下，第六根管道为最佳）

然后我们对于 153 个变量进行 5 层的变量进行合并，得到的结果因为太多，放到附录中进行描述。

我们发现，基本所有的指标都为二级指标，但是第 123 和 124 个指标为一级指标，第 91 个参数为三级指标，第 88 个参数为四级指标，第三个数据为 5 级指标，第 113 个数据为六级指标。同时对照 STATA 上的数据可得，P 全部为 0.000，说明显著，两者相互印证。

故而，我们将 153 个参数总共合并成 6 个指标进行分析。将六组数据分类处理后用 STATA 进行拟合得到下图：

A	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
B	.0242249	.0024679	9.82	0.000	.0193867	.029063
C	-1.208658	.1229805	-9.83	0.000	-1.449754	-.9675626
D	1.088626	.0150873	72.16	0.000	1.059048	1.118204
E	-.8542441	.0765839	-11.15	0.000	-1.004382	-.704106
F	1.095264	.0725686	15.09	0.000	.9529976	1.23753
G	.266319	.0329755	8.08	0.000	.2016725	.3309654
_cons	-68.67193	5.405972	-12.70	0.000	-79.27001	-58.07385

5-3-5 合并参数后的拟合

从上图中我们发现，P 的值基本等于零，说明拟合的效果非常好。

将其系数带入不同的十组数据中进行检验，得到误差表图如下所示：

	管道一	管道二	管道三	管道四	管道五	管道六	管道七	管道八	管道九	管道十
误差	0.00%	0.17%	0.11%	0.02%	0.14	0.02	0.06%	0.11%	0.49%	1.39%

误差在大数据的情况下小于百分之一，在可接受的范围内。

5.4 问题四的模型建立和求解

5.4.1 双类多元线性回归的建立

引发超温现象必然我们从下面两种情况着手进行分析。第一种情况是参数中的某值突变，导致总体的函数值突变。第二种情况是对于权重较大的几个参数受到了较大的波动，也会对函数值的呈现产生较大的影响。

所以，对于第一种情况，我们画出再 3172 号数据前后的各个参数之间的变化曲线，通过对比哪里呈现出上三角或下三角的情况，来分析那个参数的值产生了突变的情况，从而定位分析求解。

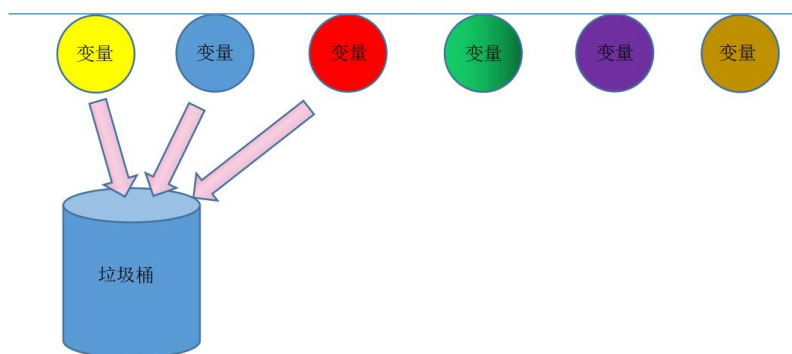
对于第二种情况，根据我们第三题的聚类分析，我们将指标分成了六级，级数越高表明其改变对于最终的影响更大。定位高权重参数后，通过分析此参数的数据变化率与其他参数之间的关系，从而得出结论。

用下面的函数进行描述：

$$Y = K_1 X_1 + \cdots + K_n X_n + A_1 Z_1 + \cdots + A_2 Z_n$$

其中， X 为第一类变量， Z 为第二类变量。

在此过程中，为了确保参数的数量少，但是仍然能够对于模型进行解释，我们采用向后逐步回归，先将所有变量均放入模型，之后尝试将其中一个自变量从模型中剔除，看整个模型解释因变量的变异是否有显著变化，之后将最没有解释力的那个自变量剔除；此过程不断迭代，直到没有自变量符合剔除的条件。（缺点：一开始把全部变量都引入回归方程，这样计算量比较大。若对一些不重要的变量，一开始就不引入，这样就可以减少一些计算。但是这个缺点由于我们采用代码进行计算，所以不需要考虑在内），形象的分析如下图所示：



5-4-1 向后逐步回归

就这样，不断的将无关变量扔进垃圾桶中，最后便能够得到比较简洁的表达式以及超温影响因素。

5.4.2 对各个变量的向后逐步回归求解

通过 STATA 中数据的筛选以及变量的处理（具体数据详见附录）

我们可以将主要的参数变量定位到以下的两张截图中，他们分别是：

- 1、#2 角 B 层二次风调节挡板位置反馈
- 2、炉膛压力和炉膛与风箱差压

	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BU	BV
3160	14.297	17.882	17.722	39.846	17.989	15.616	14.922	14.67	29.959	5.432	21.125	5.012	4.661	5.088	5.088
3161	14.304	17.852	17.661	39.854	17.951	15.616	14.922	14.67	29.829	5.409	16.669	4.799	4.715	5.088	5.088
3162	14.312	17.882	18.988	39.846	17.928	15.616	15.09	14.663	30.05	5.424	19.866	4.799	4.707	5.088	5.088
3163	14.304	17.882	18.073	39.838	17.951	15.548	15.174	14.663	30.043	5.424	16.265	4.654	4.707	5.088	5.073
3164	14.319	17.882	17.829	39.854	17.951	15.548	14.922	14.625	30.043	5.424	19.24	4.936	4.623	5.195	5.088
3165	14.312	17.859	17.867	39.701	18.012	15.449	14.922	14.312	30.012	5.401	18.76	4.799	4.715	5.226	5.088
3166	14.312	17.875	17.882	39.808	17.928	15.571	15.082	14.312	30.012	5.424	18.523	4.799	4.6	5.226	5.088
3167	14.297	17.875	17.821	39.792	17.928	15.456	14.907	14.67	30.043	5.424	17.18	4.799	4.623	5.218	5.096
3168	14.297	17.89	17.806	39.709	17.943	15.525	14.922	14.67	29.905	5.424	11.108	4.646	4.616	5.203	5.088
3169	14.266	17.89	17.806	39.846	17.928	15.456	14.93	14.487	29.898	5.409	19.85	4.57	4.692	5.233	5.088
3170	14.304	17.875	17.836	39.854	17.936	15.449	15.09	14.663	30.05	5.432	17.28	4.562	4.844	5.233	5.088
3171	14.297	17.89	17.867	39.854	17.852	15.624	14.915	14.739	29.944	5.417	11.588	4.44	4.616	5.233	5.088
3172	14.289	17.852	17.966	39.854	17.852	15.616	14.922	14.792	30.005	5.417	6.614	4.585	4.593	5.233	5.088
3173	14.289	17.875	17.814	39.785	17.829	15.449	14.922	14.495	29.837	5.439	12.397	4.585	4.623	5.233	5.088
3174	14.289	17.882	17.814	39.663	17.859	15.487	14.922	14.67	30.561	5.424	18.287	4.608	4.623	5.088	5.088
3175	14.319	17.867	17.806	39.701	17.844	15.449	14.922	14.8	29.921	5.424	13.015	4.6	4.661	5.088	5.088
3176	14.281	17.897	17.974	39.777	17.859	15.899	14.922	14.319	30.043	5.417	18.058	4.516	4.677	5.073	5.088
3177	14.304	17.882	17.981	39.701	17.745	15.891	14.922	14.67	29.883	5.417	14.464	4.402	4.722	5.088	5.073
3178	14.289	17.882	17.981	39.93	17.745	15.899	14.93	14.701	29.982	5.417	18.142	4.371	4.814	5.088	5.088
3179	14.335	17.844	17.981	39.937	17.707	15.899	15.113	14.67	29.913	5.417	17.585	5.264	4.738	5.088	5.073
3180	14.281	17.875	17.981	39.93	17.714	15.899	14.922	14.67	30.043	5.417	12.595	5.264	4.6	5.088	5.088
3181	14.281	17.89	18.142	39.968	17.691	15.891	14.922	14.663	29.928	5.424	12.58	5.264	4.623	5.081	5.088
3182	14.304	17.867	18.035	39.922	17.676	15.899	14.922	14.663	29.989	5.424	18.34	5.264	4.677	5.081	5.088

5-4-2 单一参数的大幅突变

	DH	DI	DJ	DK	DL	DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW
3155	99.672	257.876	178.071	12.201	1.773	569.101	553.59	26.341	359.323	4.272	537.661	-45.4	238.609	0.001	8.305	0.001
3156	99.672	257.876	178.071	12.212	1.773	569.19	553.635	26.142	359.118	4.287	536.847	-43.16	237.372	0.001	8.45	0.001
3157	99.672	257.876	178.071	12.258	1.773	569.19	553.679	25.871	358.843	4.282	535.22	-39.346	241.145	0.001	8.273	0.001
3158	99.672	257.876	178.071	12.27	1.773	569.369	553.724	25.542	358.569	4.281	535.13	-27.342	249.436	0.001	8.305	0.001
3159	99.672	257.924	178.071	12.281	1.773	569.458	553.813	25.273	358.432	4.264	537.48	-51.848	233.793	0.001	8.419	0.001
3160	99.672	257.876	178.071	12.27	1.773	569.414	553.859	25.079	358.273	4.236	540.102	-47.584	232.832	0.001	8.298	0.001
3161	99.672	257.97	178.071	12.27	1.773	569.548	553.992	24.913	358.158	4.243	540.283	-49.093	231.299	7.362	8.475	0.001
3162	99.672	257.924	178.071	12.26	1.781	569.504	554.038	24.822	357.975	4.228	539.74	-56.084	227.324	0.001	0.001	0.001
3163	99.672	257.924	178.071	12.249	1.781	569.504	554.127	24.738	357.884	4.218	539.379	-47.987	229.813	0.001	8.456	0.001
3164	99.672	258.064	178.071	12.238	1.781	569.548	554.216	24.665	357.86	4.205	539.108	-59.318	225.304	0.001	8.45	0.001
3165	99.672	258.158	178.071	12.238	1.781	569.458	554.26	24.606	357.861	4.203	539.56	-50.728	235.672	0.001	8.411	0.001
3166	99.672	258.158	178.071	12.238	1.781	569.458	554.26	24.584	357.883	4.214	538.656	-59.489	231.764	0.001	8.374	0.001
3167	99.672	258.252	178.071	12.27	1.781	569.369	554.305	24.526	357.929	4.199	534.949	-50.727	234.269	0.001	8.418	0.001
3168	99.672	258.252	178.071	12.27	1.771	569.369	554.35	24.524	357.998	4.23	535.853	-61.227	226.91	0.001	8.43	0.001
3169	99.672	258.252	178.071	12.254	1.771	569.369	554.35	24.65	358.089	4.198	533.955	-54.415	229.572	0.001	8.555	0.001
3170	99.672	258.346	178.071	12.322	1.761	569.458	554.395	24.533	358.25	4.193	528.35	-67.795	226.666	0.001	8.342	0.001
3171	99.672	258.44	178.071	12.32	1.75	569.548	554.395	24.624	358.478	4.213	527.084	-42.463	239.086	0.001	8.498	0.001
3172	99.672	258.346	178.071	12.331	1.75	569.681	554.483	24.634	358.638	4.271	528.982	-49.883	232.245	0.001	8.473	0.001
3173	99.672	258.44	178.071	12.343	1.75	569.727	554.528	24.749	358.684	4.271	529.706	-55.917	234.663	0.001	8.373	0.001
3174	99.672	258.534	178.071	12.354	1.75	569.816	554.618	24.763	358.776	4.251	530.158	-46.333	237.927	0.001	8.328	0.001
3175	99.672	258.534	178.071	12.366	1.75	569.906	554.707	24.812	358.867	4.253	529.254	-45.802	236.807	0.001	8.207	0.001
3176	99.672	258.628	178.071	12.389	1.75	569.995	554.797	24.827	358.913	4.263	529.163	-35.047	232.478	0.001	8.238	0.001
3177	99.672	258.534	178.071	12.4	1.75	570.084	554.841	24.849	359.027	4.282	528.078	-48.97	219.125	0.001	8.315	0.001

5-4-3 第一级参数的剧烈波动

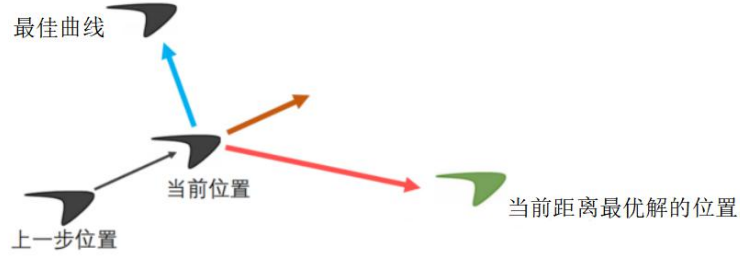
所以导致超温现象发生的主要原因就是：#2 角 B 层二次风调节挡板位置反馈以及炉膛压力和炉膛与风箱差压。

其中，第一种情况的分析为操作变量，而第二种情况二点改变为状态变量。所以对于本问题的分析中，主要操作变量为#2 角 B 层二次风调节挡板位置反馈。

5.5 问题五的模型建立和求解

5.5.1 粒子群求解模型建立

最后一题的本质就是对于参数的不断修改以达到最优解的这么一个过程，简单的理解为已知具体函数和函数的最优解，如何改变变量达到最小值。于是，我们考虑用粒子群算法，在已知最优解的情况下，和鸟类群体行为的特征类似的算法去进行不断的靠近。核心思想是利用群体中的个体对信息的共享，使整个群体的运动在问题求解空间中产生从无序到有序的演化过程，从而获得问题的可行解。



5-5-1 粒子群算法的路径求截图

不难得到以下的两个关系：

- 1、改变参数的值=前一步改变的值+前一步改变的速度*改变的时间

$$x(d) = x(d-1) + v(d-1) * t \quad (\text{每一步运动的时间 } t \text{ 一般取 } 1)$$

- 2、值改变的速度=上一个值改变的速度惯性+自我认知部分+社会认知部分

$$v(d) = w * v(d-1) + c1 * r1 * (pbest(d) - x(d)) + c2 * r2 * (gbest(d) - x(d))$$

对于粒子群算法找到近似的最优解，我们要改变自适应惯性权重，使得在搜索前期和后期都能拥有较好的能效，既能使得收敛的速度快，精度高，还不容易陷入局部极值点的陷阱，集体的表达式如下：

$$v_i^d = \omega v_i^{d-1} + c_1 r_1 (pbest_i^d - x_i^d) + c_2 r_2 (gbest^d - x_i^d)$$

假设现在求最大值问题，那么：

$$\omega_i^d = \begin{cases} \omega_{\min} + (\omega_{\max} - \omega_{\min}) \frac{f(x_i^d) - f_{\min}^d}{f_{\text{average}}^d - f_m^d}, & f(x_i^d) \leq f_{\text{average}}^d \\ \omega_{\max}, & f(x_i^d) > f_{\text{average}}^d \end{cases}$$

其中：

- (1) ω_{\min} 和 ω_{\max} 是我们预定给定的最小惯性系数和最大惯性系数，一般取 0.4 和 0.9
 - (2) $f_{\text{average}}^d = \sum_{i=1}^n f(x_i^d) / n$ ，即第 d 次迭代时所有粒子的平均适应度
 - (3) $f_{\min}^d = \min\{f(x_1^d), f(x_2^d), \dots, f(x_n^d)\}$ ，即第 d 次迭代时所有粒子的最小适应度
- 随即惯性权重：

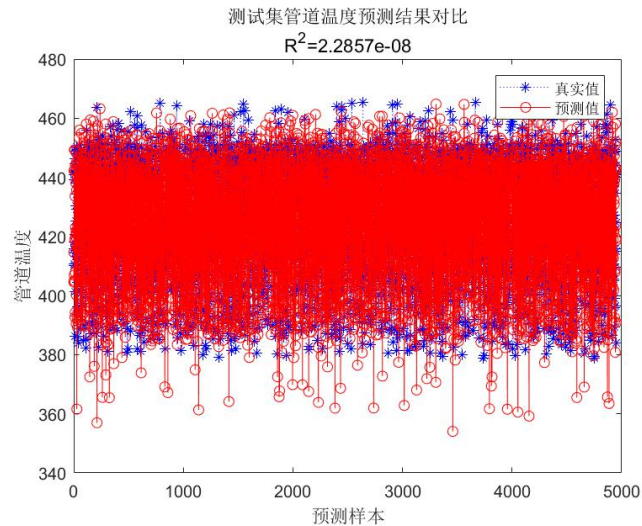
$$\omega = \mu_{\min} + (\mu_{\max} - \mu_{\min}) \times rand() + \sigma \times randn()$$

通过对于惯性权重的修改，能够使得对于参数进行修改的同时能够更好的去寻找到最优解，可以避免在迭代的前期局部搜索能力的不足，也可以避免全局搜索能力的不足。

5.5.2 运用优化粒子群算法求解并做出解释

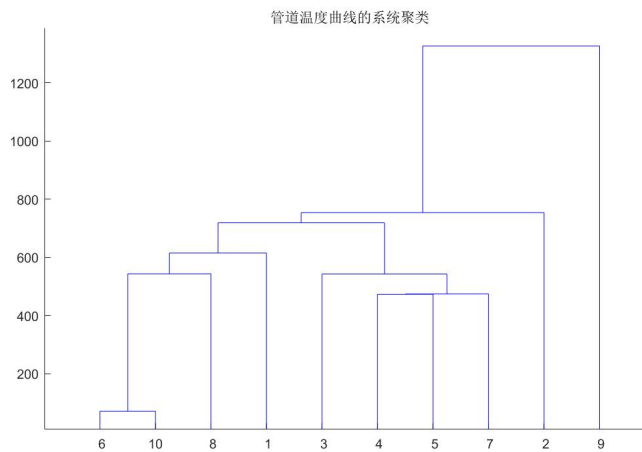
通过粒子群算法的代码运行，得到的结果（详见附录）与第四问进行结合看待，最终决定将#2 角 B 层二次风调节挡板位置反馈的权重增大一倍，同时对于炉膛压力和炉膛与风箱差压进行平稳性处理。此外，由于第十根管道的平均温度偏高，所以对于其他的权重分别下降 0.1 的随机惯性权重比。这样，总共只改变了三个参数以及调节了一次权重，操作简单，而且拟合效果极佳，具体情况如下图所示：

- 1、修改后用 BP 神经网络再次进行拟合，误差已经能控制到很小的范围内



5-5-2 BP 神经网络修改

2、将修改后的数据用模糊矩阵的传递闭包进行聚类



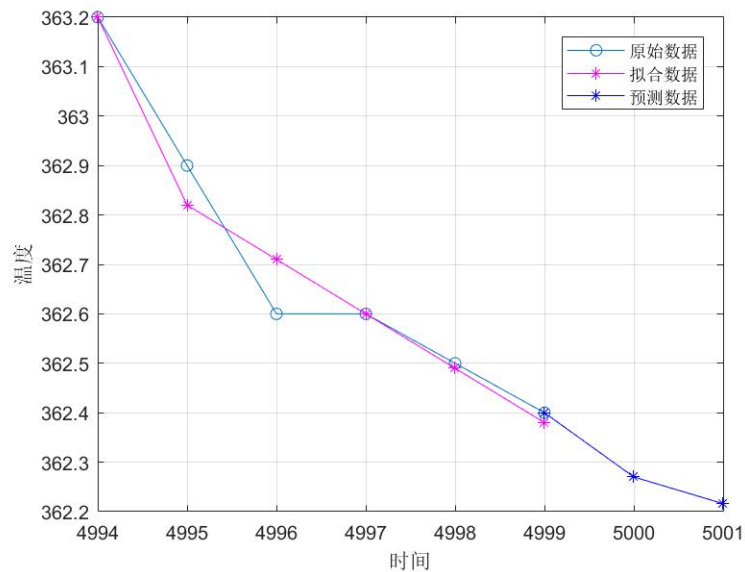
5-5-3 聚类

从图中发现，第6跟管道和第10跟管道属于同一级，且处于最高的分类级别，证明参数的修改是正确的。

六、模型的分析与检验

对于模型灵敏度的分析中，我们的思路是：先使用灰色预测对于短期内的数据进行预测。通过我们前面得出的参数指标进行检验。如果对于新产生的模拟值依然有很好的拟合程度，证明模型是稳定的，那么在此之前，我们要对灰色预测的正确性做出判断。

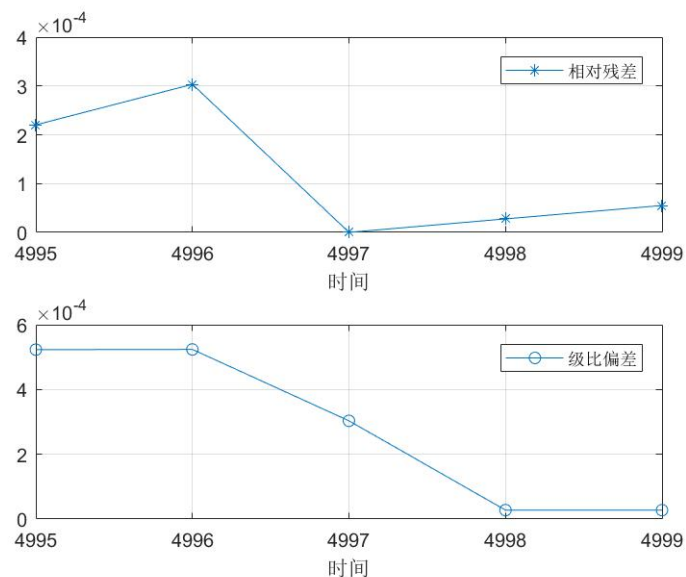
通过代码的运行我们得到以下的图示以及结论：



6-1-1 预测数据及其对比

与附录中的实际值进行对比，不难发现拟合的数据和元数据之间及其的接近。不光是在接近尾端的数据中可以体现，在刚性曲线段的转折点依然可以做到完美的预测。

接下来观察预测的级比偏差和相对残差，发现其误差都在 10^{-4} 的级别。这样的数据证明我们可以通过灰色预测来生成数据，从而对于我们的灵敏度进行分析。



6-1-2 相对残差和级比偏差

以第 5001 组数据为例，我们预测的值为 361.6°C ，用刚才得到的 net 网络导入得到预计会产生的温度为 363.2°C ，误差在可控的范围内，说明模型的稳定度可行。

七、模型的推广与评价

7.1 模型的优点

1. 我们的基于 BP 神经网络的管道温度变化评价模型使用多种方法进行优化，结果非常的合理，可以通过多组数据测试；并且我们使用了模糊矩阵的传递闭包，使得参数

的聚集效果更加明显。

2. 我们对于粒子群算法惯性权重进行了优化,使得时间复杂度和空间复杂度在大数据量的计算情况下,大幅度下降,增加了模型的效率。
3. 我们对于单一模型采用多种方法进行验证,对结果进行横向对比,从而得出更为精确的结果;并且,我们的模型能较好的解决难以量化的问题,适合各种非确定性问题的解决
4. 我们的敏感度分析模型使用灰色预测进行数据拓展,进行模型的验证,保证了我们的模型更为可靠。

7.2 模型的缺点

1. 温度异常分析模型的各个参数的权重难以确定,很难给出一个具体的函数进行表达;同时,由于数据的局限性,我们所验证的权重不是非常精准。
2. 我们的模型在变量过于庞大情况下,容易出现拟合度不高的情况。

7.3 模型的推广

我们的模型可以在多个工程项目中拥有运用场景;可以在包括但不限于工业生产中仪器损伤定位、损伤检测,变压器故障诊断等多个领域都得到应用。

八、参考文献

- [1]朱振国,田松禄.基于权值变化的BP神经网络自适应学习率改进研究.计算机系统应用,2018,27(7):205~210.
- [2]赵志刚,黄树运,王伟倩.基于随机惯性权重的简化粒子群优化算法.计算机系统应用,2014.02.010
- [3]富宇,李倩,刘澎.“改进的BP神经网络算法的研究与应用.”计算机与数字工程.(2019):28-32.Print.
- [4]张浩.(2011).PSO粒子群优化算法的混沌时间序列优化.科技创新导报,29.doi:CNKI:SUN:ZXDB.0.2011-26-012.

九、附录

9.1 三个隶属度函数的模板（分段函数）

```
x=0:0.1:12;
y1=0.*(x<1)+(0.5+0.5*sin((pi/9)*(x-5.5))).*(x>=1&x<=10)+1.*(x>10);
plot(x,y1,'r');
text(7.8,0.8597,'\leftarrow 445');
hold on
y2=0.*(x<1)+0.4*(0.5+0.5*sin((pi/9)*(x-5.5))).*(x>=1&x<=10)+0.4.*(x>10);
plot(x,y2,'r');
text(7.8,0.3439,'\leftarrow 420');
title('K 次抛物分布')
grid on
```

9.2 BP 神经网络基础代码

```
%%I. 清空
clear;clc
%%
% II. 随机产生训练集和测试集
temp = randperm(size(NIR,1)); %打乱样本排序
% 训练集——50 个样本
P_train = NIR(temp(1:50),:);
T_train = octane(temp(1:50),:);
% 测试集——10 个样本
P_test = NIR(temp(51:end),:);
T_test = octane(temp(51:end),:);
N = size(P_test,2);
%% III. 数据归一化
[p_train, ps_input] = mapminmax(P_train,0,1);
p_test = mapminmax('apply',P_test,ps_input);
[t_train, ps_output] = mapminmax(T_train,0,1);
%% IV. BP 神经网络创建、训练及仿真测试
%%
% 1. 创建网络
net = newff(p_train,t_train,15);
%%
% 2. 设置训练参数
net.trainParam.epochs = 1000; %迭代次数
net.trainParam.goal = 1e-3; %mse 均方根误差小于这个值训练结束
net.trainParam.lr = 0.01; %学习率
%%
% 3. 训练网络
net = train(net,p_train,t_train);
%%
% 4. 仿真测试
t_sim = sim(net,p_test); %返回 10 个样本的预测值
%%
% 5. 数据反归一化
T_sim = mapminmax('reverse',t_sim,ps_output); %反归一化结果
%% V. 性能评价
%%
```

```

% 1. 相对误差 error
error = abs(T_sim - T_test)./T_test;
%%
% 2. 决定系数 R^2
R2 = ((N * sum(T_sim .* T_test) - sum(T_sim) * sum(T_test))^2 / ((N * sum((T_sim).^2) -
(sum(T_sim))^2) * (N * sum((T_test).^2) - (sum(T_test))^2)))/(N*(N-1));
%%
% 3. 结果对比
result = [T_test' T_sim' error']      %输出真实值，预测值，误差
%% VI. 绘图
figure
plot(1:N,T_test,'b:*',1:N,T_sim,'r-o')
legend('真实值','预测值')
xlabel('预测样本')
ylabel('管道温度')
string = {'测试集管道温度预测结果对比';['R^2=' num2str(R2)]};
title(string)

```

9.3 模糊矩阵的传递闭包聚类

```

function[Imd,xhsz,flsz,flrlsz]=F_DtJIt(n,R)%定义函数
%函数功能：模糊聚类分析--模糊聚类图
%n, 模糊相似矩阵的阶数；R, 模糊相似矩阵
%返回值：Imd, 分类水平数组；xhsz,最终序号数组
%返回值：flsz, 分类数数组；flrlsz, 子类容量数组
for(i=1:n)R(i,i)=1;%修正错误
    for(j=i+1:n)
        if(R(i,j)<0)R(i,j)=0;elseif(R(i,j)>1)R(i,j)=1;end
        R(i,j)=round(10000*R(i,j))/10000;
        R(j,i)=R(i,j);
    end
end
js0=0;
while(1)%求传递闭包
    R1=Max_Min(R,R);js0=js0+1;
    if(R1==R)break;else R=R1;end
end
Imd(1)=1;k=1;
for(i=1:k-1)for(j=i+1:n)pd=1;%找出所有不相同的元素
    for(x=1:k)if(R(i,j)==Imd(x))pd=0;break;end;end
    if(pd)k=k+1;Imd(k)=R(i,j);end
end;end
for(i=1:k-1)for(j=i+1:k)if(Imd(i)<Imd(j))%从大到小排序
    x=Imd(j);Imd(j)=Imd(i);Imd(i)=x;
    end;end;end
for(x=1:k)%按 Imd(x)分类，分类数为 flsz(x),临时用 SZ 记录元素序号
    js=0;flsz(x)=0;
    for(i=1:n)pd=1;
        for(y=1:js)if(Sz(y)==i)pd=0;break;end;end
        if(pd)
            for(j=1:n)if(R(i,j)>=Imd(x))js=js+1;Sz(js)=j;end;end
            flsz(x)=flsz(x)+1;
        end
    end
end
end
end

```

```

for(i=1:k-1)for(j=i+1:k)if(flsz(j)==flsz(j)=0;end;end;end
fl=0;
for(i=1:k)if(flsz(i))fl=fl+1;Imd(fl)=Imd(i);end;end
for(i=1:n)xhsz(i)=i;end
for(x=1:fl)%获得分类情况，对分类元素进行排序
    js=0;flsz(x)=0;
    for(i=1:n)pd=1;
        for(y=1:js)if(Sz(y)==i)pd=0;break;end;end
        if(pd)id(js==0)y=0;end
        for(j=1:n)if(R(i,j)>=Imd(x))js=js+1;Sz(js)=j;end
        flsz(x)=flsz(x)+1;
        Sz0(flsz(x))=js-y;
    end
end
js0=0;
for(i=1:flsz(x))
    for(j=1:Sz0(i))Sz1(j)=Sz(js0+j);end
    for(j=1:n)for(y=1:Sz0(i))
        if(xhsz(j)==Sz1(y))js0=js0+1;Sz(js0)=xhsz(j);end
    end;end
end
for(i=1:n)xhsz(i)=Sz(i);end
end
for(x=1:fl)%获得分类中每一子类的元素个数
    js=0;flsz(x)=0;
    for(i=1:n)pd=1;
        for(y=1:js)if(Sz(y)==i)pd=0;break;end;end
        if(pd)if(js==0)y=0;end
        for(j=1:n)if(R(i,j)>=Imd(x))js=js+1;Sz(js)=j;end;end
        flsz(x)=flsz(x)+1;Sz0(flsz(x))=js-y;
    end
end
js0=1;
for(i=1:flsz(x))y=1;
    for(j=1:flsz(x))
        if(Sz(y)==xhsz(js0))flrlsz(x,i)=Sz0(j);js0=js0+Sz0(j);break;end
        y=y+Sz0(j);
    end
end
end
xhsz,flsz,Imd,flrlsz
F_dtltx=figure('name','动态聚类图','color','w');
axis('off');
Kd=30;Gd=40;
if(n<20)Ix=60;else Ix=80;end
if(flsz(1)==n)y=fl*Kd+Gd;text(24,y+Gd/2+8,'\lambda');
    for(i=1:n)
        text(Ix-5+i*Kd-0.4*Kd*(xhsz(i)>9),y+Gd/2+8,int2str(xhsz(i)));
        line([Ix+i*Kd,Ix+i*Kd],[y,y-Gd]);hxsz(i)=Ix+i*Kd;
    end
text(Ix*1.5+n*Kd,y+Gd/1+8,'分类数');
js0=1;js1=0;
text(16,y-Gd/2,'1.0000');
for(i=1:flsz(1))js1=flrlsz(1,i)-1;
    if(js1)line([hxsz(js0),1+hxsz(js0+js1)],[y,y]);end
    line([(hxsz(js0+js1)+hxsz(js0))/2,(hxsz(js0+js1)+hxsz(js0))/2],[y,y-Gd]);
    hxsz(i)=(hxsz(js0+js1)+hxsz(js0))/2;js0=js0+js1+1;

```

```

end
text(Ix*1.5+n*Kd,y-Gd/2,int2str(flsz(1)));
else y=fl*Kd;text(24,y,'λ');
for(i=1:n)
    text(Ix-5+i*Kd-0.4*Kd*(xhsz(i)>9),y,int2str(xhsz(i)));
    line([Ix+i*Kd,Ix+i*Kd],[y-Gd/4,y-Gd/2]);hxs(i)=Ix+i*Kd;
end
text(Ix*1.5+n*Kd,y,'分类数');
js0=1;js1=0;y=y-Gd/2;
text(16,y-Gd/2,'1.0000');
for(i=1:flsz(1))js1=flrlsz(1,i);
    for(j=1:n)js2=js1+1;
        if(js2==js1)break;end
    end
    if(j~=js0)line([hxs(js0),1+hxs(j)],[y,y]);end
    line([(hxs(js0)+hxs(j))/2,(hxs(js0)+hxs(j))/2],[y,y-Gd]);
    hxs(i)=(hxs(js0)+hxs(j))/2;js0=j+1;
end
text(Ix*1.5+n*Kd,y-Gd/2,int2str(flsz(1)));
end
for(x=2:fl)js0=1;js1=0;y=y-Gd;
    text(16,y-Gd/2,num2str(Imd(x)));
    for(i=1:flsz(x))js2=0;js1=js1+flrlsz(x,i);
        for(j=1:flsz(x-1))js2=0;js1=js1+flrlsz(x-1,j);
            if(js2==js1)break;end
        end
        if(j~=js0)line([hxs(js0),1+hxs(j)],[y,y]);end
        line([(hxs(js0)+hxs(j))/2,(hxs(js0)+hxs(j))/2],[y,y-Gd]);
        hxs(i)=(hxs(js0)+hxs(j))/2;js0=j+1;
    end
    text(Ix*1.5+n*Kd,y-Gd/2,int2str(flsz(x)));
end
end

```

9.4 管道温度的系统聚类

```

X=[A;B;C;D;E;F;G;H;I;J];
Y=pdist(X);
SF=squareform(Y);
Z=linkage(Y,'single');
dendrogram(Z);%显示系统聚类树
T=cluster(Z,'maxclust',3);
title('管道温度曲线的系统聚类')

```

9.5 粒子群代码

```

x0 = 0;
A=[]; b=[];
Aeq=[];beq=[];
x_lb = d1
x_ub = d2
[x,fval] = fmincon(@Obj_fun1,x0,A,b,Aeq,beq,x_lb,x_ub)
fval = -fval
function y = Obj_fun1(x)

for i=1:n

```



```

        for j=1:n
            y+=a[i]*x[j];
        end

    if x[i]-x[i-1]>10%
        {X=temp;
        x=x1;
        temp=x1;}}
    end

n = 10; % 粒子数量
narvs = 111; % 变量个数(函数中有几个自变量)
c1 = 2; % 每个粒子的个体学习因子
c2 = 2; % 每个粒子的社会学习因子
K = 50; % 迭代的次数
vmax = 1.2; % 粒子的最大速度
x_lb = d1; % x 的下界
x_ub = d2; % x 的上界

x = zeros(n,narvs);
for i = 1: narvs
    % 随机初始化粒子所在的位置在定义域内
    x(:,i) = x_lb(i) + (x_ub(i)-x_lb(i))*rand(n,1);
end
% 随机初始化粒子的速度, 设置为[-vmax,vmax]
v = -vmax + 2*vmax.*rand(n,narvs);
fit = zeros(n,1); % 初始化这 n 个粒子的适应度全为 0
for i = 1:n % 循环整个粒子群, 计算每一个粒子的适应度
    fit(i) = Obj_fun1(x(i,:)); % 调用 Obj_fun1 函数来计算适应度
    (这里写成 x(i,:)主要是为了和以后遇到的多元函数互通)
end
pbest = x; % 初始化这 n 个粒子迄今为止找到的最佳位置 (是一个 n*narvs 的向量)
ind = find(fit == max(fit), 1); % 找到适应度最大的那个粒子的下标
gbest = x(ind,:); % 定义所有粒子迄今为止找到的最佳位置 (是一个 1*narvs 的向量)
function y = Obj_fun1(x)
    y += a[i]*x[j];
end
fit = zeros(n,1); % 初始化这 n 个粒子的适应度全为 0
for i = 1:n % 循环整个粒子群, 计算每一个粒子的适应度
    fit(i) = Obj_fun1(x(i,:)); % 调用 Obj_fun1 函数来计算适应度
    (这里写成 x(i,:)主要是为了和以后遇到的多元函数互通)
end
pbest = x; % 初始化这 n 个粒子迄今为止找到的最佳位置 (是一个 n*narvs 的向量)
ind = find(fit == max(fit), 1); % 找到适应度最大的那个粒子的下标
gbest = x(ind,:); % 定义所有粒子迄今为止找到的最佳位置 (是一个 1*narvs 的向量)
function y = Obj_fun1(x)
    y =;
end
fit(i) = Obj_fun1(x(i,:)); % 重新计算第 i 个粒子的适应度
% 如果第 i 个粒子的适应度大于这个粒子迄今为止找到的最佳位置对应的适应度
if fit(i) > Obj_fun1(pbest(i,:))
    pbest(i,:) = x(i,:); % 那就更新第 i 个粒子迄今为止找到的最佳位置
end
% 如果第 i 个粒子的适应度大于所有的粒子迄今为止找到的最佳位置对应的适应度

```

```

if Obj_fun1(pbest(i,:)) > Obj_fun1(gbest)
gbest = pbest(i,:); % 那就更新所有粒子迄今为止找到的最佳位置
end
end
end

```

9.6 灰色预测代码

```

clear;clc
year =[4995:1:5000]';
x0 = [362.9,362.6,362.6,362.5,362.4,362.2]';
figure(1);
plot(year,x0,'o-'); grid on;
set(gca,'xtick',year(1:1:end))
xlabel('时间'); ylabel('温度');
ERROR = 0;
if sum(x0<0) > 0
    disp('灰色预测的时间序列中不能有负数')
    ERROR = 1;
end
n = length(x0);
disp(strcat('原始数据的长度为',num2str(n)))
if n<=3
    disp('数据量太小')
    ERROR = 1;
end
if n>10
    disp('多数据量')
end
if size(x0,1) == 1
    x0 = x0';
end
if size(year,1) == 1
    year = year';
end
if ERROR == 0
    disp('-----')
    disp('准指数规律检验')
    x1 = cumsum(x0);
    rho = x0(2:end) ./ x1(1:end-1) ;

    figure(2)
    plot(year(2:end),rho,'o-',[year(2),year(end)],[0.5,0.5],'-'); grid on;
    text(year(end-1)+0.2,0.55,'临界线')
    set(gca,'xtick',year(2:1:end))
    xlabel('年份'); ylabel('原始数据的光滑度');
    disp(strcat('指标 1: 光滑比小于 0.5 的数据占比为',num2str(100*sum(rho<0.5)/(n-1)),'%'))
    disp(strcat('指标 2: 除去前两个时期外, 光滑比小于 0.5 的数据占比为',num2str(100*sum(rho(3:end)<0.5)/(n-3)),'%'))
    disp('参考标准: 指标 1 一般要大于 60%, 指标 2 要大于 90%')
    Judge = input('你认为可以通过准指数规律的检验吗? 可以通过请输入 1, 不能请输入 0: ');
    if Judge == 0
        disp('灰色预测模型不适合')
        ERROR = 1;
    end
end

```

```

end
disp('-----')
end
if ERROR == 0
    if n > 4
        disp('因为原数据的期数大于 4， 所以将数据组分为训练组和试验组')
        if n > 7
            test_num = 3;
        else
            test_num = 2;
        end
        train_x0 = x0(1:end-test_num);
        disp('训练数据是: ')
        disp(mat2str(train_x0))
        test_x0 = x0(end-test_num+1:end);
        disp('试验数据是: ')
        disp(mat2str(test_x0))
        disp('-----')
        disp(' ')
        disp('***下面是传统的 GM(1,1)模型预测的详细过程***')
        result1 = gml1(train_x0, test_num);
        disp(' ')
        disp('***下面是进行新信息的 GM(1,1)模型预测的详细过程***')
        result2 = new_gml1(train_x0, test_num);
        disp(' ')
        disp('***下面是进行新陈代谢的 GM(1,1)模型预测的详细过程***')
        result3 = metabolism_gml1(train_x0, test_num);
        disp(' ')
        disp('-----')
        test_year = year(end-test_num+1:end);
        figure(3)
        plot(test_year, test_x0, 'o-', test_year, result1, '*-', test_year, result2, '+-', test_year, result3, 'x-');
        grid on;
        set(gca, 'xtick', year(end-test_num+1):1:year(end))
        legend('试验组的真实数据', '传统 GM(1,1)预测结果', '新信息 GM(1,1)预测结果', '新陈代谢 GM(1,1)预测结果')
        xlabel('年份'); ylabel('排污总量'); % 给坐标轴加上标签
        SSE1 = sum((test_x0 - result1).^2);
        SSE2 = sum((test_x0 - result2).^2);
        SSE3 = sum((test_x0 - result3).^2);
        disp(strcat('传统 GM(1,1)对于试验组预测的误差平方和为', num2str(SSE1)))
        disp(strcat('新信息 GM(1,1)对于试验组预测的误差平方和为', num2str(SSE2)))
        disp(strcat('新陈代谢 GM(1,1)对于试验组预测的误差平方和为', num2str(SSE3)))
        if SSE1 < SSE2
            if SSE1 < SSE3
                choose = 1;
            else
                choose = 3;
            end
        elseif SSE2 < SSE3
            choose = 2;
        else
            choose = 3;
        end
        Model = {'传统 GM(1,1)模型', '新信息 GM(1,1)模型', '新陈代谢 GM(1,1)模型'};
        disp(strcat('因为', Model(choose), '的误差平方和最小， 所以我们应该选择其进行预测'))
    end
end

```

```

disp('-----')
predict_num = input('请输入你要往后面预测的期数: ');
[result, x0_hat, relative_residuals, eta] = gm11(x0, predict_num); % 先利用 gm11 函数得
到对原数据拟合的详细结果
if choose == 2
    result = new_gm11(x0, predict_num);
end
if choose == 3
    result = metabolism_gm11(x0, predict_num);
end
disp('-----')
disp('对原始数据的拟合结果: ')
for i = 1:n
    disp(strcat(num2str(year(i)), ' : ', num2str(x0_hat(i))))
end
disp(strcat('往后预测', num2str(predict_num), '期的得到的结果: '))
for i = 1:predict_num
    disp(strcat(num2str(year(end)+i), ' : ', num2str(result(i))))
end
else
disp('因为数据只有 4 期, 因此我们直接将三种方法的结果求平均即可~')
predict_num = input('请输入你要往后面预测的期数: ');
disp(' ')
disp('***下面是传统的 GM(1,1)模型预测的详细过程***')
[result1, x0_hat, relative_residuals, eta] = gm11(x0, predict_num);
disp(' ')
disp('***下面是进行新信息的 GM(1,1)模型预测的详细过程***')
result2 = new_gm11(x0, predict_num);
disp(' ')
disp('***下面是进行新陈代谢的 GM(1,1)模型预测的详细过程***')
result3 = metabolism_gm11(x0, predict_num);
result = (result1+result2+result3)/3;
disp('对原始数据的拟合结果: ')
for i = 1:n
    disp(strcat(num2str(year(i)), ' : ', num2str(x0_hat(i))))
end
disp(strcat('传统 GM(1,1)往后预测', num2str(predict_num), '期的得到的结果: '))
for i = 1:predict_num
    disp(strcat(num2str(year(end)+i), ' : ', num2str(result1(i))))
end
disp(strcat('新信息 GM(1,1)往后预测', num2str(predict_num), '期的得到的结果: '))
for i = 1:predict_num
    disp(strcat(num2str(year(end)+i), ' : ', num2str(result2(i))))
end
disp(strcat('新陈代谢 GM(1,1)往后预测', num2str(predict_num), '期的得到的结果: '))
for i = 1:predict_num
    disp(strcat(num2str(year(end)+i), ' : ', num2str(result3(i))))
end
disp(strcat('三种方法求平均得到的往后预测', num2str(predict_num), '期的得到的结果: '))
for i = 1:predict_num
    disp(strcat(num2str(year(end)+i), ' : ', num2str(result(i))))
end
end
figure(4)
subplot(2,1,1)

```

```

plot(year(2:end), relative_residuals,'*-'); grid on;
legend('相对残差'); xlabel('时间');
set(gca,'xtick',year(2:1:end))
subplot(2,1,2)
plot(year(2:end), eta,'o-'); grid on;
legend('级比偏差'); xlabel('时间');
set(gca,'xtick',year(2:1:end))
disp(' ')
disp('****下面将输出对原数据拟合的评价结果****')
average_relative_residuals = mean(relative_residuals);
disp(strcat('平均相对残差为',num2str(average_relative_residuals)))
if average_relative_residuals<0.1
    disp('残差检验的结果表明：该模型对原数据的拟合程度非常不错')
elseif average_relative_residuals<0.2
    disp('残差检验的结果表明：该模型对原数据的拟合程度达到一般要求')
else
    disp('残差检验的结果表明：该模型对原数据的拟合程度不太好')
end
average_eta = mean(eta);
disp(strcat('平均级比偏差为',num2str(average_eta)))
if average_eta<0.1
    disp('级比偏差检验的结果表明：该模型对原数据的拟合程度非常不错')
elseif average_eta<0.2
    disp('级比偏差检验的结果表明：该模型对原数据的拟合程度达到一般要求')
else
    disp('级比偏差检验的结果表明：该模型对原数据的拟合程度不太好，建议使用其他模
型预测')
end
disp(' ')
disp('-----')
figure(5)
plot(year,x0,'-o', year,x0_hat,'-m', year(end)+1:year(end)+predict_num,result,'-b'); grid
on;

hold on;
plot([year(end),year(end)+1],[x0(end),result(1)],'-b')
legend('原始数据','拟合数据','预测数据')
set(gca,'xtick',[year(1):1:year(end)+predict_num])
xlabel('时间'); ylabel('温度');
end

function [result, x0_hat, relative_residuals, eta] = gml1(x0, predict_num)
    n = length(x0); % 数据的长度
    x1=cumsum(x0); % 计算一次累加值
    z1 = (x1(1:end-1) + x1(2:end)) / 2; % 计算紧邻均值生成数列（长度为 n-1）
    y = x0(2:end); x = z1;
    k = ((n-1)*sum(x.*y)-sum(x)*sum(y))/((n-1)*sum(x.*x)-sum(x)*sum(x));
    b = (sum(x.*x)*sum(y)-sum(x)*sum(x.*y))/((n-1)*sum(x.*x)-sum(x)*sum(x));
    a = -k; %注意：k = -a
    disp('现在进行 GM(1,1)预测的原始数据是:')
    disp(mat2str(x0'))
    disp(strcat('最小二乘法拟合得到的发展系数为',num2str(-a),', 灰作用量是',num2str(b)))
    disp('*****分割线*****')
    x0_hat=zeros(n,1); x0_hat(1)=x0(1);
    for m = 1: n-1
        x0_hat(m+1) = (1-exp(a))*(x0(1)-b/a)*exp(-a*m);
    end
end

```

```

end
result = zeros(predict_num,1);
for i = 1: predict_num
    result(i) = (1-exp(a))*(x0(1)-b/a)*exp(-a*(n+i-1));
end
absolute_residuals = x0(2:end) - x0_hat(2:end);
relative_residuals = abs(absolute_residuals) ./ x0(2:end);
class_ratio = x0(2:end) ./ x0(1:end-1) ;
eta = abs(1-(1-0.5*a)/(1+0.5*a)*(1./class_ratio)); end

```