Team Number :	APMCM0642

Problem Chosen :	A

2016 APMCM summary sheet

The control of temperature and the content of key elements have important significance for acquiring the best performance of objective metal in the process of metal smelting. The core objective of this paper is to establish a mathematical model or algorithm model to desctibe the relationship between the temperature , the content of key elements and the light intensity under the condition of specific cumulative consumption of combustion gas and the cumulative consumption ratio of combustion gas.

In the first question, we adopted the principal component analysis method to find the characteristic value from the 2048 sets of light intensity data recorded every 0.5 seconds. According to the calculation results of the original light intensity data calculated by using SPSS software,13 principal components could be used to represent the 2048 sets of data of the optical information. With the contribution rate of each principal component as the weight, the characteristic values obtained by the weighted sum ,which could represent the vast majority of information on the original light intensity data.

In the second question, first we explore the rules for optical information characteristic data $\boldsymbol{\lambda}$, and find it can be fitted well in a linear function. Especially, we adopt segmentation function for $\boldsymbol{\lambda}$ in process 1 because of its particular scatter diagram. Then through our observation and exploration, we decide to use mutiple linear regression to predict Kelvin temperature $\boldsymbol{T}$and key element content $\boldsymbol{C}$. At last we make correlation test to check if the result is significant, and the answer is true. So we make the conclusion that mutiple linear regression can predict $\boldsymbol{T}$ and $\boldsymbol{C}$ and give the predict expressions of each process.

In the third question, we first conduct a crossover experiment among three prediction model and three process data and caculate mean square error in each cell of crossover experiment.And we make a generally analysis of the universality of the model.Then,to improve the universality of the model and control the error,we analyze the sources of errors comprehensively and detailedly in the whole process.We distinguish the part of error which can be artificially controlled.According to different kinds of error,we put forward an error control scheme to improve model accuracy and universality from four aspect at error in the given data,error in PCA algorithm,error in regression forecast model and random error.

Key words: characteristic value; principal component analysis; linear regression model; error characteristic analysis

Contents

(At the beginning of this text)

1.Introduction

1. 1 Problem Description

Background:

The control of temperature and the content of key elements can reflect the performance of objective metal in the process of metal smelting. According to the theory of pinhole imaging, the photo detector recorded the light intensity data of the flame at every 0.5s. The data in the three process of metal smelting are presented,which include the time ***t***、the cumulative consumption of combustion gas ***Q***, the cumulative consumption ratio of combustion gas ***p***, optical information data（***f_1-f_2048***，light intensity at different frequencies）, Kelvin temperature ***T*** *and c*ontent of key element ***C***.

The problem to be solved:

Problem 1: Find the characteristic λ of the optical information data from the data given in the annex recorded in the three process of metal smelting.

Problem 2: Establish the mathematical model or algorithm model to predict the Kelvin temperature ***T*** and key element content ***C*** by using the optical information characteristic data extracted in problem one and the data given in the data table, such as the time ***t*** and the accumulated consumption ***Q*** of the combustion- supporting gas, and to explore the relationship between the Kelvin temperature ***T*** and key element content ***C***.

Problem 3: Design the crossover experiment scheme, cross-validate the error will be generated by the prediction target; provide the error control scheme on the basis of the error analysis.

1. 2Terminology and Definitions

- Characteristic values: The greater the variance of the characteristic values, the greater the amount of information contained in the feature vector direction. The characteristic value is often used to reduce the dimensionality of the data,which will remove the directional data of smaller characteristic value variance corresponding[1].
- Optical information: The intensity, phase, color and polarization state of light[2].
- Mathematical model：Mathematical model is a scientific or engineering model that uses mathematical logic and mathematical language.
- Kelvin temperature: With absolute zero as the starting point of the calculation of the temperature, the temperature of the water triple point is exactly defined as the temperature after 273.15K [3].
- Error analysis: When the system function is complete, the causes, consequences and the stages of the deviation from the target are analyzed, and the error is reduced to a minimum.

1. 3 Our work

THe paper tries to establish mathematical model with strong universality by information characteristic data λ、time t and the accumulated consumption Q of the combustion- supporting gas to predict the Kelvin temperature T and key element content C.In section 1, we find out the characteristic value of the 2048 groups of light intensity data detected in each 0.5s using the scientific and reasonable mathematical method to.In section 2, we establish mathematical model predict the kelvin temperature T and key element content C and explore explore the relationship. In section 3,we analysis the error characteristics of the results in the cross experiment,and provide the error control scheme.

2.Problem Analysis

2. 1 Analysis of problem 1

,Three sets of data extracted from metal smelting process are presented, including the time t、the cumulative consumption of combustion gas Q, the cumulative consumption ratio of combustion gas p, optical information data（f_1-f_2048，light intensity at different frequencies）, Kelvin temperature T *and c*ontent of key element C. If considering 2048 light intensity data generated at a time as the input and the flame temperature and the Key element content as the output, constructing a mathematical model is bound to be complex, so our core task is to find one or more characteristic values of light intensity data to represent most of the information in the original data. The method of principal component analysis is selected to reduce the dimension of the 2048 groups of data, and the original data is replaced by principal components which has less dimension and not related to each other. We use the SPSS software to calculate the characteristic value of the light intensity data.

2. 2 Analysis of problem 2

The problem 2 asks us to use the optical information characteristic data λ, the time t and the accumulated consumption Q, to predict the Kelvin temperature T and key element content C.To solve this problem we need to choose an appropriate prediction model. Through our observations, we find that varibles t, Q, T and C increase linearly, and varible λ's scatter diagram indicates that it can be fitted to a line. Therefore, we decide to choose multiple linear regression to predict the Kelvin temperature T and key element content C.

2.3 Analysis of problem 3

The problem 3 asks us to design a crossover experiment firstly,and cross-validate the error generated by the prediction target based on the result of crossover experiment. Further we can design error control scheme according to the analysis of the conclusion.In our research, The composition of error may include but not limited to the model and algorithm of problem solving and the random error.An error control method should be proposed based on the analysis of the error sources.So,We will test the error

can be divided into two types of can be controlled and not by artificially power.,and puts forward different error control methods.So as to achieve the goal of improving the model's universality.

3.Models

3.1The model of problem 1

3.1.1Assumptions

- Assuming that the selected principal components are highly representative.
- Assuming that the characteristic value which represent more than 85% of information can be used to replace all the original light intensity data.
- Assuming that the original light intensity data have strong correlation with each other.

3.1.2 Terms, Definitions and Symbols

In this section, we will give some basic symbols and definitions used in the following pages for convenience.

Table1： Terms, Definitions and Symbols of problem 1

Symbols	Definitions
F_i	i-th principal component
α_i	Variance of principal component i
X_i	Raw data of light intensity information
Cov	covariance
S	covariance matrix
$G(m)$	Cumulative contribution rate of the m-th principal component
a_i	Principal component score coefficient
l_{ij}	Principal component load
λ	characteristic value

3.1.3Establishment of model

Principal component analysis is used to reduce the dimension of light intensity data.The value of the principal component can replace the original amount of data, and then we will seek the characteristic values of light intensity data calculated by SPSS.

F_1present the principal component index of the first linear combination of 2048 groups light intensity data recorded during every 0.5 seconds:

$$F_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \quad (1)$$

The amount of information extracted from each principal component can be measured by its variance, the greater the variance α_1, the more information that the principal component F_1 contains.

If the first principal component F_1 is not sufficient to represent the original data of the five groups of information, and then consider the selection of second principal component F_2. In order to effectively reflect all the information of the original data, the information contained in the F_1 does not need to appear in F_2,and the information contained in F_1 and F_2 remain independent, covariance as follows:

$$Cov(F_1, F_2) = 0 \tag{2}$$

F_2 is known as the second principal component when F_1 and F_2 are not related and F_2 has the maximum variance in all linear combinations. And so on,constructing principal components $F_1, F_2, \ldots, F_m$ were used as the first、second、...、the m-th component of the original data：

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \\ F_2 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \\ \quad \ldots\ldots \\ F_m = a_{m1}X_1 + a_{m2}X_2 + \ldots + a_{mp}X_p \end{cases} \tag{3}$$

According to the above analysis:

(1) F_i and F_j are not related, and the covariance $Cov(F_i, F_j) = 0$.

（2）F_1 has the greatest variance in in all linear combinations. F_1 and $F_2, F_3, \ldots, F_m$ are not related.

Therefore,there are two main tasks to find the characteristic value of the light intensity information:

(1) Determine mathematical expression between principal components $F_i (i = 1, 2, \ldots, m)$ and original variables $X_j (j = 1, 2, \ldots, p)$, that is coefficient $a_{ij} (i = 1, 2, \ldots, m; j = 1, 2, \ldots, \mathrm{p})$.

（2）Calculate principal component load l_{ij} , which can reflect the relationship between the principal components F_i and the original variables X_j.

$$l_{ij} = \sqrt{F_k} a_{ki} \ (i = 1, 2, \ldots, \mathrm{p}; \mathrm{k} = 1, 2, \ldots, \mathrm{m}) \tag{4}$$

3.1.4Model solving process

The specific steps to calculate the characteristic value of light intensity information using principal component analysis are as follows：

（1）Computing covariance matrix

Calculate the covariance matrix of the original data which can reflect the light intensity information:

$$S = \left(s_{ij}\right)_{p \times p} \qquad i, j = 1, 2, \ldots, p \tag{5}$$

（2）Select principal components

Determine the principal component $F_1, F_2, \ldots, F_m$ according to the cumulative contribution rate of variance $G(m)$:

$$G(m) = \sum_{i=1}^{m} F_i \Big/ \sum_{k=1}^{p} F_k \tag{6}$$

Principal components were selected when their cumulative contribution rate was more than 85%, and it was considered that the information of the original variables could be enough to reflect by the selected principal components.

（3）Find the principal component score F_i and its variance contribution rate α_i

The score coefficient matrix is the coefficient between principal components and original data, the score of the principal component F_i as follows:

$$F_i = a_i X \tag{7}$$

The variance contribution rate α_i of the principal component can reflect the amount of information:

$$\alpha_i = F_i \Big/ \sum_{i=1}^{m} F_i \tag{8}$$

（4）Calculate principal component load

The principal component load can reflect the correlation degree between the principal component F_i and the original variable X_j, the load $l_{ij}\left(i = 1, 2, \ldots, m; j = 1, 2, \ldots, p\right)$ on the principal component $F_i\left(i = 1, 2, \ldots, m\right)$ of the original data $X_j\left(j = 1, 2, \ldots, p\right)$:

$$l(Z_i, X_j) = \sqrt{\lambda_i} a_{ij} (i = 1, 2, \cdots, m; j = 1, 2, \cdots, p) \tag{9}$$

（5）Caculate principal component score

Calculate the score of the original data in a principal component:

$$F_i = a_{1i}X_1 + a_{2i}X_2 + ... + a_{pi}X_p \qquad i = 1,2,\ldots,m \tag{10}$$

（6）Calculate the characteristic value of the light intensity data

For the score of a principal components, the contribution rate as its weight, the characteristic value of the 2048 light intensity information detected by the photo detector in each 0.5s as follows:

$$\lambda_j = \sum_{i=1}^{m} F_i \alpha_i \tag{11}$$

3.1.5 Solution of characteristic value

Taking the first set of metal smelting data as an example(the same below), the calculation results of the other two groups are shown in the appendix. the data is calculated by using SPSS, the results obtained as following:

（1）The calculated covariance matrix S is shown in the following table：

Table 2:Component score covariance matrix

成份	1	2	3	4	5	6	7	8
1	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
7	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
8	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
9	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
11	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
12	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
13	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

（2）According to the analysis results, selecting 13 principal components which cumulative contribution rate is greater than 85%,and the cumulative contribution rate is shown in the following table:

Table 3: Cumulative Contribution Rate of Principal Components

F_i	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}
α_i	81.741%	1.139%	0.728%	0.201%	0.171%	0.163%	0.162%	0.159%	0.156%	0.155%	0.154%	0.152%	0.150%

(3) The component score coefficient matrix is shown in appendix.

(4) In the analysis results after principal component analysis by SPSS, "component matrix" is the principal component load matrix l_{ij} ,which is shown in the appendix.

(5) A total of thirteen principal components were obtained from the analysis results, the 2048 sets of data detected in each 0.5s can be represented by 13 principal components. The principal component score matrix of 405*13 can be obtained,which is shown in the appendix.

(6) The information available in the 2048 set of data in each 0.5s is represented by the characteristic value λ_j, so a total of 405 characteristic values are obtained, which is shown in the feature_output_1.xlsx.

3.1.6Strength and Weakness

Strength:

（1）We used principal component analysis to select 13 principal components of which the contribution rate of was above 85% to replace the original 2048 groups of light intensity information.The 13 principal components represent at least 85% of the original amount of data. The selected principal components are highly representative.

（2）SPSS software is used to extract the principal components of the original data, which makes the calculation more simple and easy to implement.

（3）With the variance contribution rate of the principal component as the weight, the characteristic values of the 2048 sets of data are obtained by weighted sum. The results showed that the decision making process using principal component analysis is scientific, which can lay a good foundation for the establishment of the model in second problems.

Weakness:

（1）Some information was lost in the process of reducing the dimension of original light intensity data by sing principal component analysis method.

（2）Because the selected principal component could not represent all the information of the data, the information contained in the characteristic value is fuzzy to a certain extent.

3.2The model of problem 2

3.2.1Assumptions

- The data given in the Appendix 1-3 are true and effective.
- The optical information characteristic data $\boldsymbol{\lambda}$get from Problem 1 are right.
- The optical information characteristic data $\boldsymbol{\lambda}$canwave in an acceptable range, some few outlier data could be discarded..
- No more elements will influence the prediction of Kelvin temperature $\boldsymbol{T}$ and key element content $\boldsymbol{C}$.

3.2.2 Terms, Definitions and Symbols

Table4：Terms, Definitions and Symbols of problem 1

Symbols	Definitions

λ_i	Optical information characteristic data in process i
t_i	Time data in process i
Q_i	Accumulated consumption of the gas data in process i
T_i	Kelvin temperature data in process i
C_i	Key element content data in process i
α_i	The coefficient of linear fitting for $\boldsymbol{\lambda}$
β_i	The coefficient of multiple linear regression for $\boldsymbol{T}$
γ_i	The coefficient of multiple linear regression for $\boldsymbol{C}$
$\hat{\lambda}_i$	Fitted characteristic data in process i
$\hat{T}_i$	Predicted Kelvin temperature data in process i
$\hat{C}_i$	Predicted Key element content data in process i

3.2.3 The Foudation of Model

From the analysis we sketch that varibles $\boldsymbol{t}$, $\boldsymbol{Q}$, $\boldsymbol{T}$ and $\boldsymbol{C}$ increase linearly, thus in the next step, exploring the fitting situation of variable $\boldsymbol{\lambda}$will be significant. Through the scatter diagram of $\boldsymbol{\lambda}$ in process 1,2 and 3, we find that $\boldsymbol{\lambda}$'s numerical value fluctuates roughly around a line, so we decide to perform linear fitting aimed at $\boldsymbol{\lambda}$as the formula below.

$$\hat{\lambda}_i = \alpha_0 + \alpha_1 t_i \tag{12}$$

Then, we find that $\boldsymbol{\lambda}$ in process 1, its scatter diagram(***Figure 3-X***) has an obvious mutation. So we choose the Segmentation function to fit $\boldsymbol{\lambda}$.

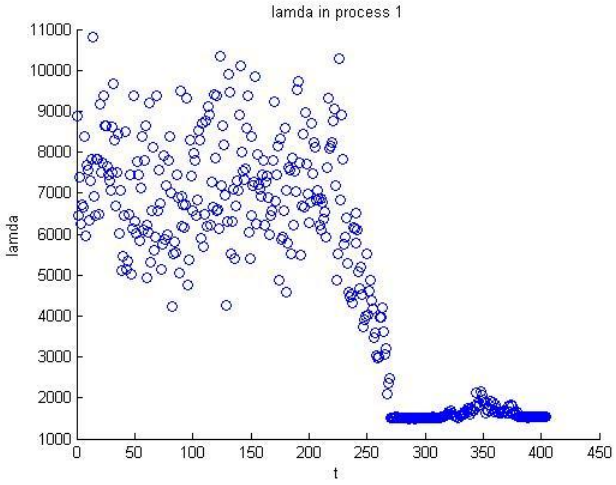


Figure 1 λ distribution in process 1

After detailed examination of the data, we choose the data 269 as the segment point. Make the linear fitting process, we can get the expression of ***λ***and the fitting diagram(***Figure 3-x***) below.

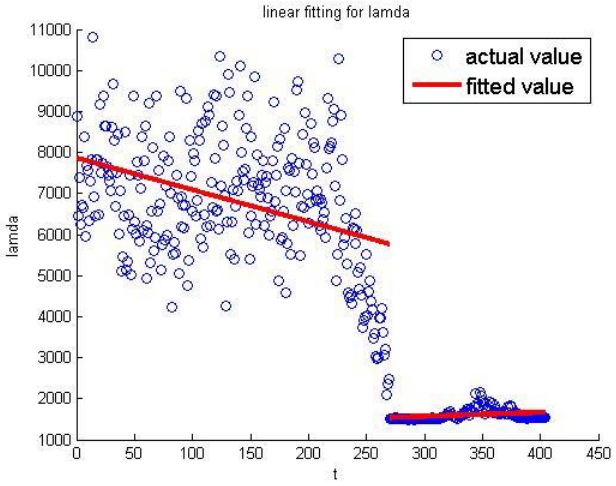


Figure 2 Linear fitting for λ in process 1

$$\lambda_1 = \begin{cases} -15.5385t_1 + 7847.9, 0 < t_1 \le 134 \\ 1.9426t_1 + 1276.9, 134 < t_1 \le 201.5 \end{cases} \tag{13}$$

Then, using SPSS to make correlation test for ***λ***. The result(***Figure 3-x***) reveals that the actual value and the fitted value of ***λ***have significant correlation. It means ***λ***'s fitted value,in a manner, can represent its actual value.

相关系数

			actual	fitted
Kendall 的 tau_b	actual	相关系数	1.000	.324**
		Sig.（双侧）	.	.000
		N	135	135
	fitted	相关系数	.324**	1.000
		Sig.（双侧）	.000	.
		N	135	135
Spearman 的 rho	actual	相关系数	1.000	.521**
		Sig.（双侧）	.	.000
		N	135	135
	fitted	相关系数	.521**	1.000
		Sig.（双侧）	.000	.
		N	135	135

**. 在置信度（双测）为 0.01 时，相关性是显著的。

1) λin segment 1

相关系数

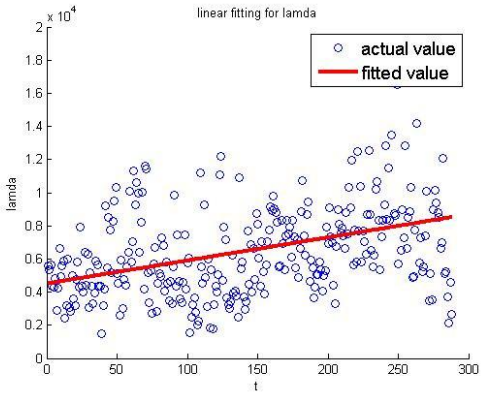
			actual	fitted
Kendall 的 tau_b	actual	相关系数	1.000	.220**
		Sig.（双侧）	.	.000
		N	269	269
	fitted	相关系数	.220**	1.000
		Sig.（双侧）	.000	.
		N	269	269
Spearman 的 rho	actual	相关系数	1.000	.319**
		Sig.（双侧）	.	.000
		N	269	269
	fitted	相关系数	.319**	1.000
		Sig.（双侧）	.000	.
		N	269	269

**. 在置信度（双测）为 0.01 时，相关性是显著的。

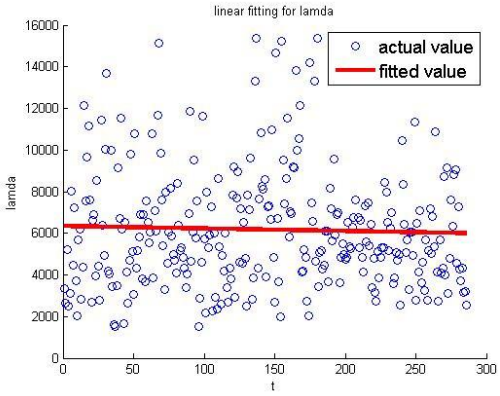
2) λ in segment 2

Figure 3 The result of λ correlation test in process 1

Using this fitting method to fit λin process 2 and 3, we can get another 2 fitting expressions and the figures(***Figure 3-x***). In regrad of the figures, though there are few outlier data, the fitting result is still significant in general.



1) Linear fitting for λ in process 2



2) Linear fitting for λ in process 3

Fiugre 4 Linear fitting for λ in process 2 and 3

$$\lambda_2 = 27.8384t_2 + 4533.6, 0 < t_2 \le 143.5 \tag{14}$$

$$\lambda_3 = -2.4104t_3 + 6344.8, 0 < t_3 \le 142.5 \tag{15}$$

Now we get the linear fitted λ for every process. Because the 3 independent varibles are all linear, so we choose multiple linear regression to predict Kelvin temperature T and Key element content C. The flow chart of this problem(***Figure 3-x***) and the formula are shown below.

$$\hat{T}_i = \beta_0 + \beta_1 t_i + \beta_2 Q_i + \beta_3 \hat{\lambda}_i \tag{16}$$

$$\hat{C}_i = \gamma_0 + \gamma_1 t_i + \gamma_2 Q_i + \gamma_3 \hat{\lambda}_i \tag{17}$$

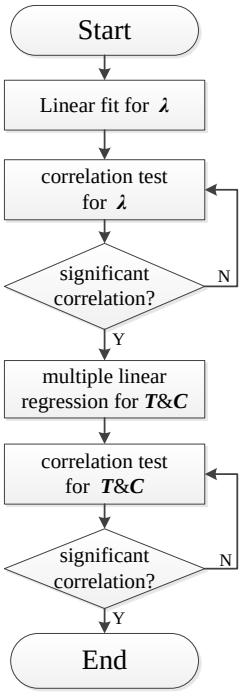
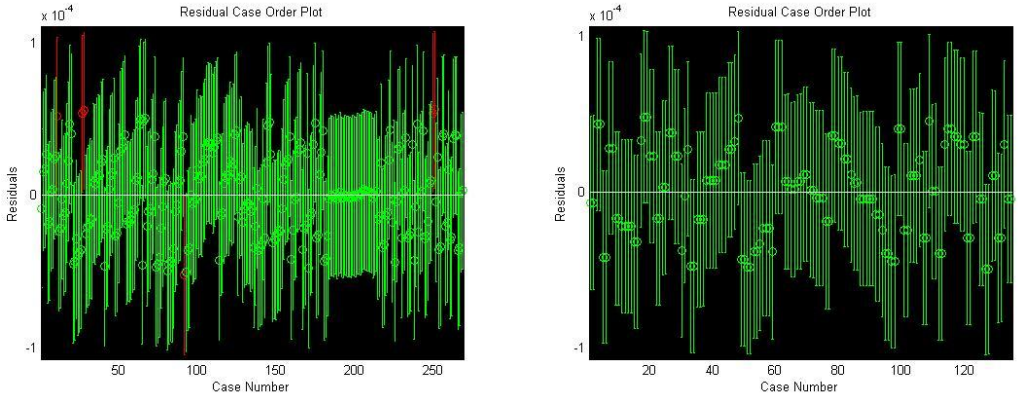


Figure 5 The flow chart of Problem 2

3.2.4Solution and Result

For this part, we use MATLAB to multiple linear regression(***code can be seen in Appendix***). Using ***T*** prediction in process 1 as an example, first we should examine the residual of the regression process, the result(***Figure 3-x***) is shown below. According to the result we can find that most of the residuals are acceptable in its confidence interval, thus we can get the conclusion that the regression process is successful.



1) ***T*** prediction process in segment 1 2) ***T*** prediction process in segment 2

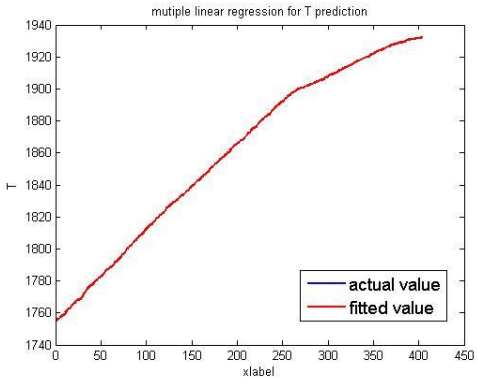
Figure 6 residuals generated during T prediction regression process

Then, we can calculate the coefficient and get the expression for T prediction in process 1.

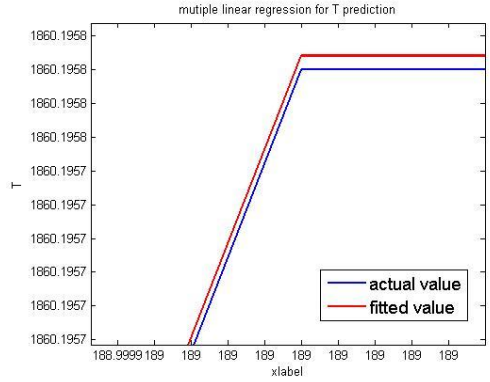
$$\hat{T}_1 = \begin{cases} 2.6554t_1 + 0.1605Q_1 + 0.1709\hat{\lambda}_1, segment1 \\ -2.0404t_1 + 0.1605Q_1 + 1.0503\hat{\lambda}_1, segment2 \end{cases} \tag{18}$$

So we can figure this segmentation expression comparing to the actual value ***T***. The result***(Figure 3-x)*** shows that the fitted value and the actual value of ***T*** have a very

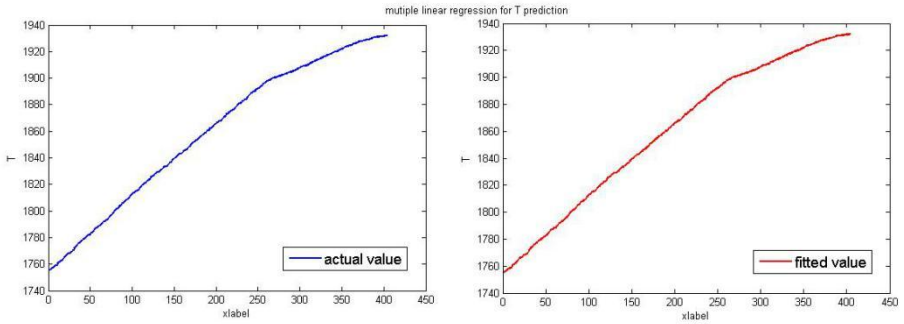
high fitted degree.



1) Both value in a diagram



2) Fractionated gain diagram



3) Comparison between actual value and fitted value

Figure 7 The result of T regression prediction in process 1

For the last step, we need to make correlation test for ***T***. The result(***Figure 3-x***) reveals that the actual value and the fitted value of ***T*** have significant correlation. So we can consider that our model for ***T*** prediction is successful.

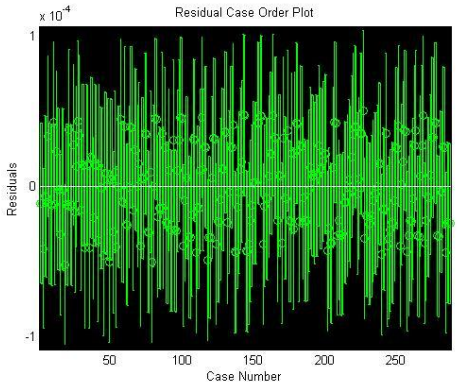
相关系数

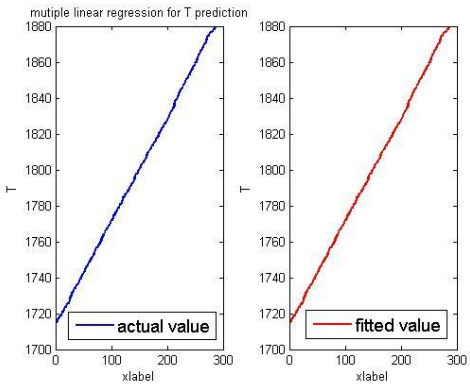
			actual	fitted
Kendall 的 tau_b	actual	相关系数	1.000	.999**
		Sig.（双侧）	.	.000
		N	404	404
	fitted	相关系数	.999**	1.000
		Sig.（双侧）	.000	.
		N	404	404
Spearman 的 rho	actual	相关系数	1.000	1.000**
		Sig.（双侧）	.	.000
		N	404	404
	fitted	相关系数	1.000**	1.000
		Sig.（双侧）	.000	.
		N	404	404

**. 在置信度（双测）为 0.01 时，相关性是显著的。

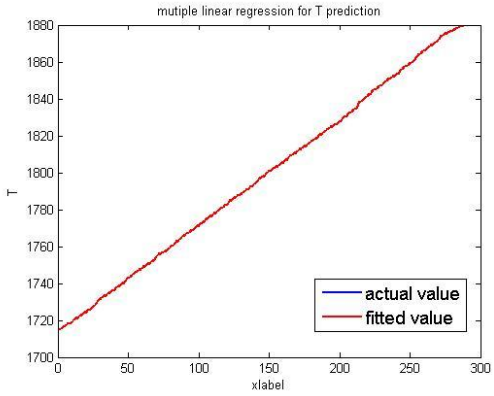
Figure 8 The result of T correlation test

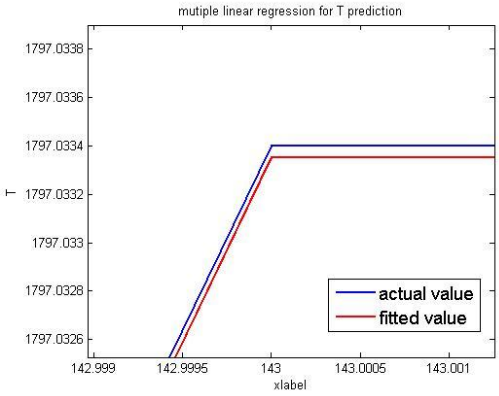
Then by this mean we can get other 2 ***T*** prediction results and 3 ***C*** prediction results as the figures (***Figure 3-x to 3-x***) and expressions below.





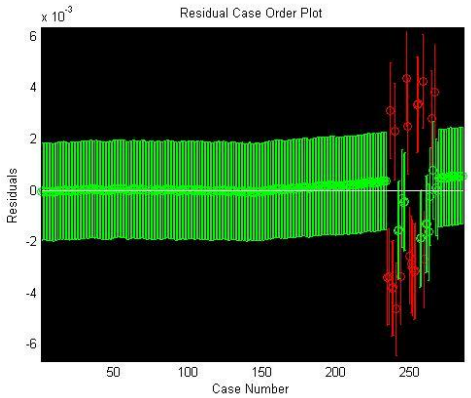
1) Residual diagram 2) Comparison between actual value and fitted value

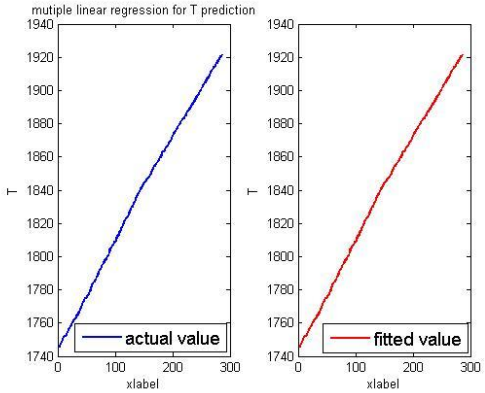




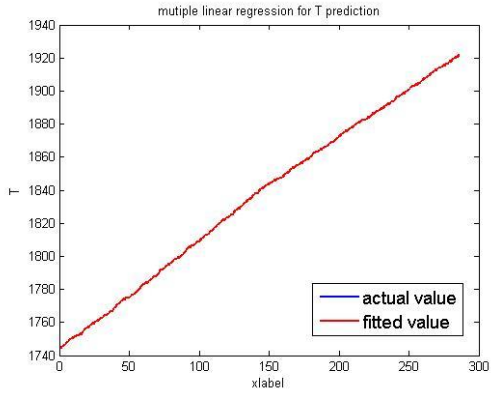
3) Both value in a diagram 4) Fractionated gain diagram

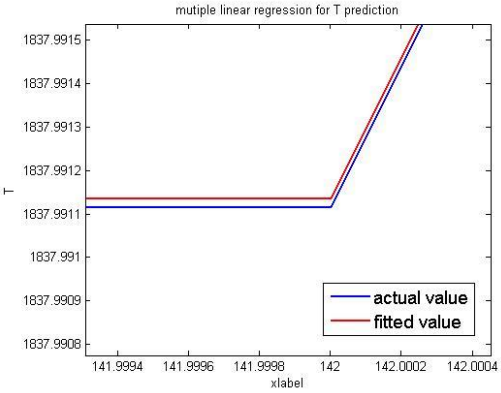
Figure 9 The result of T regression prediction in process 2





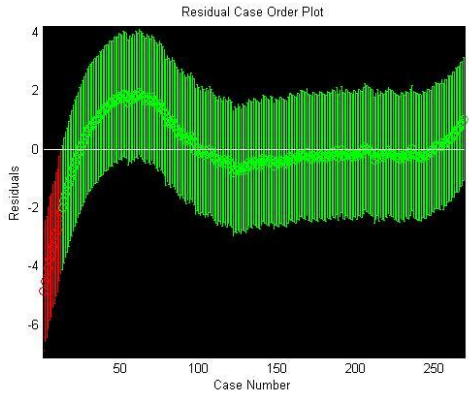
1) Residual diagram 2) Comparison between actual value and fitted value

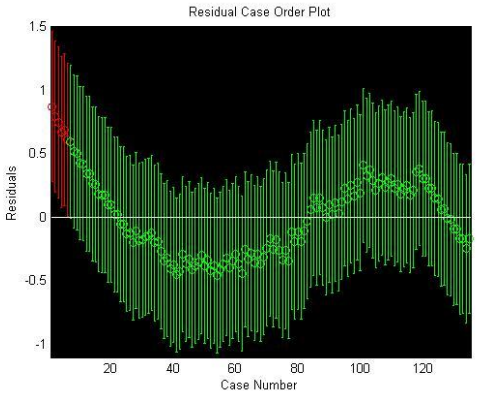




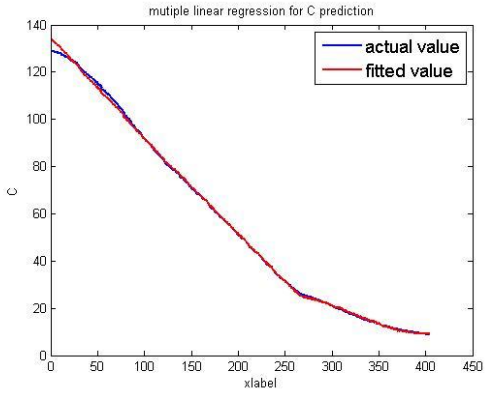
3) Both value in a diagram 4) Fractionated gain diagram

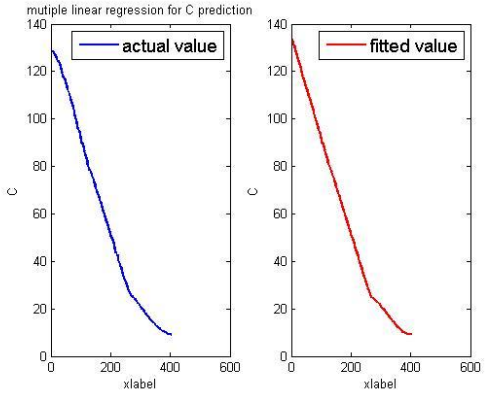
Figure 10 The result of T regression prediction in process 3





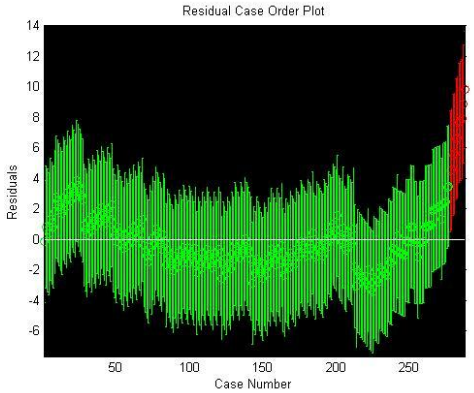
1) Residual diagram in segment 1 2) Residual diagram in segment 2

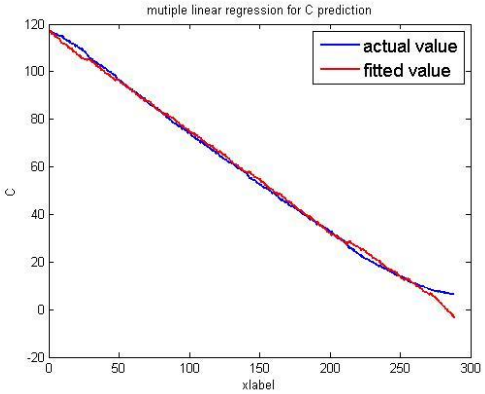




3) Both value in a diagram 4) Comparison between actual value and fitted value

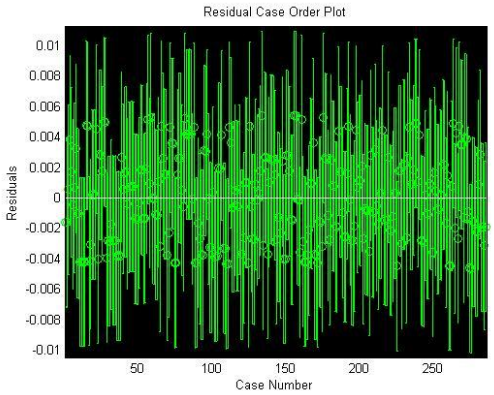
Figure 11 The result of C regression prediction in process 1



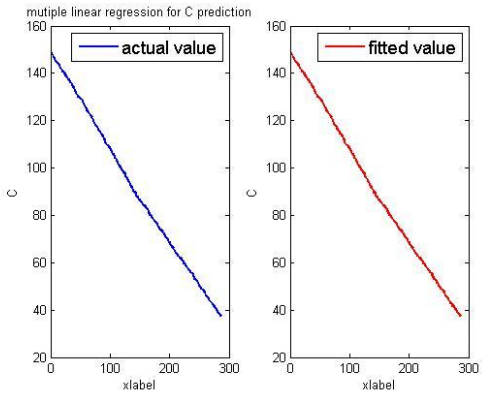


1) Residual diagram 2) both value in a diagram

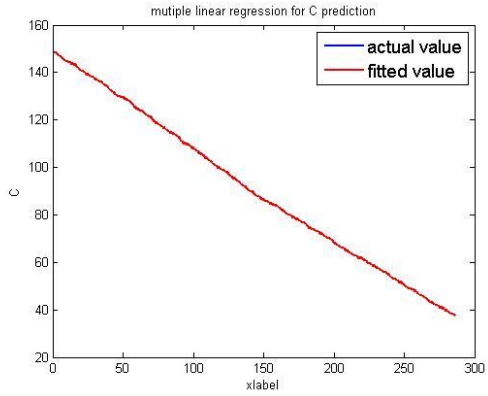
Figure 12 The result of C regression prediction in process 2



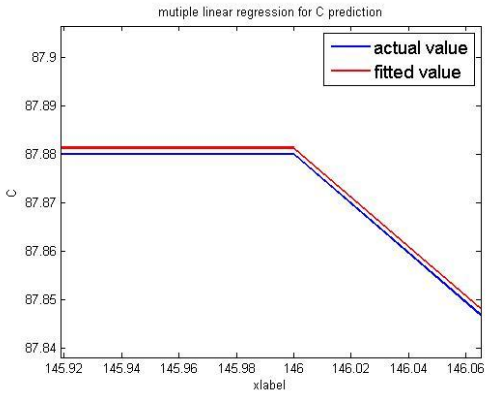
1) Residual diagram



2) Comparison between actual value and fitted value



3) Both value in a diagram



4) Fractionated gain diagram

Figure 13 The result of C regression prediction in process 3

Underneath are expressions for ***T*&*C*** prediction in each of process.

$$\hat{T}_2 = -8.1739t_2 + 0.1605Q_2 + 0.2936\hat{\lambda}_2 \tag{19}$$

$$\hat{T}_3 = 0.5056t_3 + 0.1698Q_3 + 0.2098\hat{\lambda}_3 \tag{20}$$

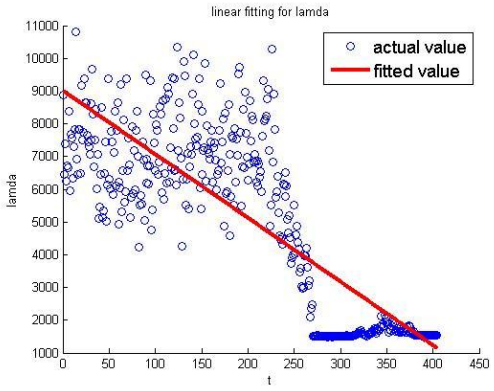
$$\hat{C}_1 = \begin{cases} 0.4141t_1 - 0.0813Q_1 + 0.0438\hat{\lambda}_1, segment1 \\ -0.5440t_1 - 0.1312Q_1 + 0.3608\hat{\lambda}_1, segment2 \end{cases} \tag{21}$$

$$\hat{C}_2 = -0.4149t_2 + 0.1511Q_2 - 0.0540\hat{\lambda}_2 \tag{22}$$

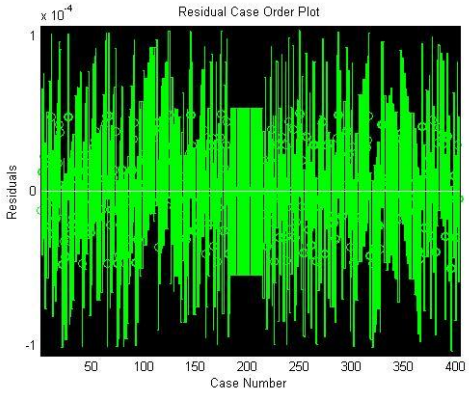
$$\hat{C}_3 = 0.1552t_3 - 0.1067Q_3 + 0.0644\hat{\lambda}_3 \tag{23}$$

When do the correlation test for every of ***T*** or ***C***, all of the results show that the actual value and the fitted value of ***T*** have significant correlation. Thus we can conclude that using mutiple linear regression can better predict Kelvin temperature ***T*** and key element content ***C***.

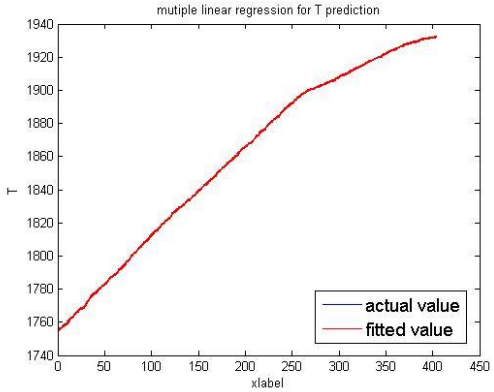
Fianlly, considering Problem 3 and cross-validate expriment, we cancel the segmentation function in process 1. It may cause that the error generated by the model becomes a little bigger, but after our exploration, the error is still acceptable. Therefore, for Problem 3, we will adopt another expression for process 1. Its regression process(***Figure 3-x***) and expression are listed underneath.



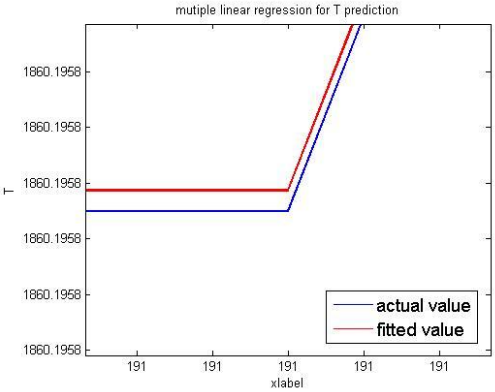
1) λ fitting diagram



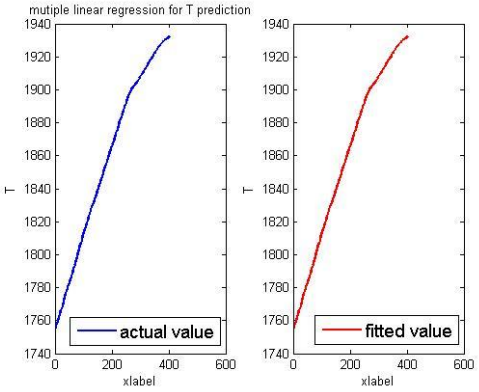
2) Residual diagram for T



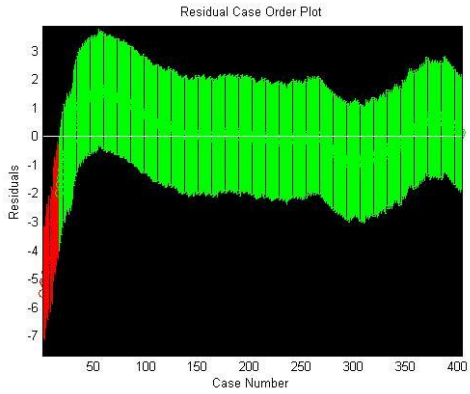
3) T regression result diagram



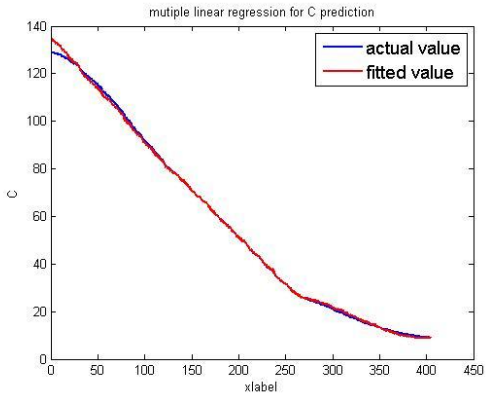
4) Fractionated gain diagram



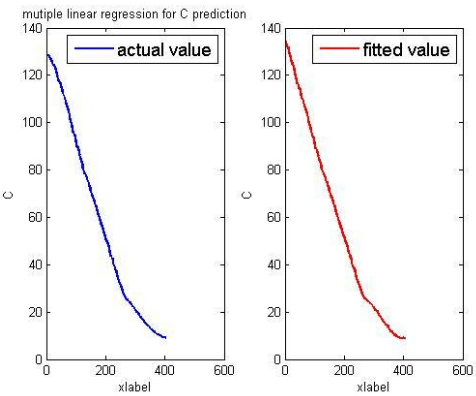
5) Comparison mode for T regression



6) Residual diagram for C



7) C regression result diagram



8) Comparison mode for T regression

Figure 14 The result of T andC regression prediction in process 1

Underneath are expressions for ***T*&*C*** prediction.

$$\hat{T}_1 = 5.8054t_1 + 0.1605Q_1 + 0.1492\hat{\lambda}_1 \tag{24}$$

$$\hat{C}_1 = 2.4239t_1 - 0.1481Q_1 + 0.0575\hat{\lambda}_1 \tag{25}$$

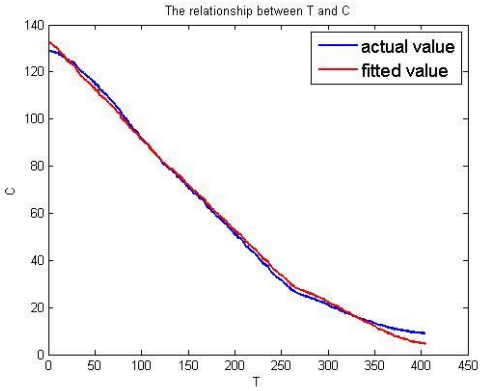
Now we summarize the predict expressions as ***Table 3-x***.

Table 5 Predict expressions for each process

	T prediction	***C*** prediction
Process 1	$\hat{T}_1 = 5.8054t_1 + 0.1605Q_1 + 0.1492\hat{\lambda}_1$	$\hat{C}_1 = 2.4239t_1 - 0.1481Q_1 + 0.0575\hat{\lambda}_1$
Process 2	$\hat{T}_2 = -8.1739t_2 + 0.1605Q_2 + 0.2936\hat{\lambda}_2$	$\hat{C}_2 = -0.4149t_2 + 0.1511Q_2 - 0.0540\hat{\lambda}_2$
Process 3	$\hat{T}_3 = 0.5056t_3 + 0.1698Q_3 + 0.2098\hat{\lambda}_3$	$\hat{C}_3 = 0.1552t_3 - 0.1067Q_3 + 0.0644\hat{\lambda}_3$

From the Table we can find that the variable ***t***'s coefficients are much larger than others' and they have significant difference, and the variable ***Q*** and variable ***λ***'s coefficient in each process are similar. So we can infer that the difference of heat and key element generated by different metal smelting process depend mostly on its lasting time.

When it comes to the relationship between ***T*** and ***C***, through our observation, obviously when ***T*** increases ***C*** decreases. Owing to each variable is linear, so we do linear fitting for them and get the result(***Figure 3-x, Table 3-x***) below.



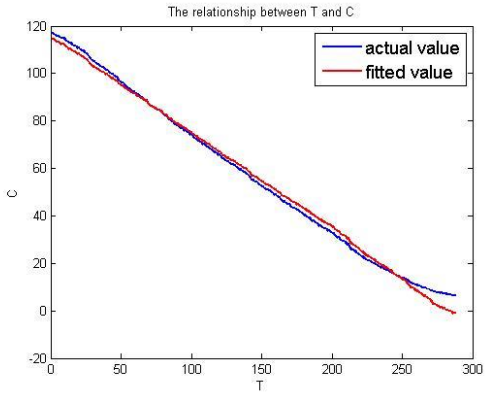
1) Process 1 result

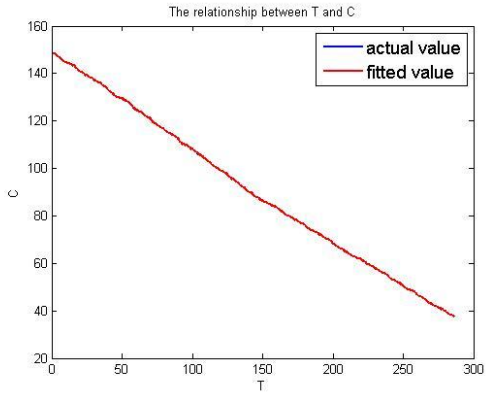
相关系数

			actual	fitted
Kendall 的 tau_b	actual	相关系数	1.000	1.000**
		Sig.（双侧）	.	.000
		N	404	404
	fitted	相关系数	1.000**	1.000
		Sig.（双侧）	.000	.
		N	404	404
Spearman 的 rho	actual	相关系数	1.000	1.000**
		Sig.（双侧）	.	.000
		N	404	404
	fitted	相关系数	1.000**	1.000
		Sig.（双侧）	.000	.
		N	404	404

**. 在置信度（双测）为 0.01 时，相关性是显著的。

2) Correlation test result in process 1(example)





3) Process 2 result 2) Process 3 result

Figure 15 The result of the relationship between T and C

Table 6 The relationship between T and C

	Expressions
Process 1	$\hat{C}_1 = -0.7235T_1 + 1402.5$
Process 2	$\hat{C}_2 = -0.7029T_2 + 1320.0$
Process 3	$\hat{C}_3 = -0.6284T_3 + 1244.9$

From the table we can know that each of process has the similar original temperature, and heat changes by different metal smelting process have almost the same influence to its key element content.

3.2.5Strength and Weakness

Strength:

- We used multiple linear regression models to predict predict the Kelvin temperature ***T*** and key element content ***C*** by the data given in the data table and characteristic values of light intensity data calculated in the problem 1, which could accurately measure the degree of correlation between each factor and the degree of regression fitting.
- We have carried out the error test to the forecast result，and it turns out to be the case that the error of the predicted temperature value with the given value is within the acceptable range, which had a good forecast effect and could improve the scientific and accuracy of prediction.
- The problem can be solved by using SPSS and MATLAB software trough using multiple linear regression model, which makes the solving process more simple and more feasible.

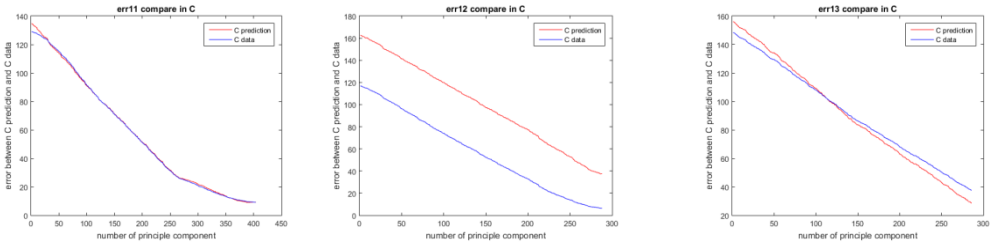
Weakness:

- The process of determination the expression of each factor is just a conjecture of decision maker's**,** which has subjective to a certain extent.

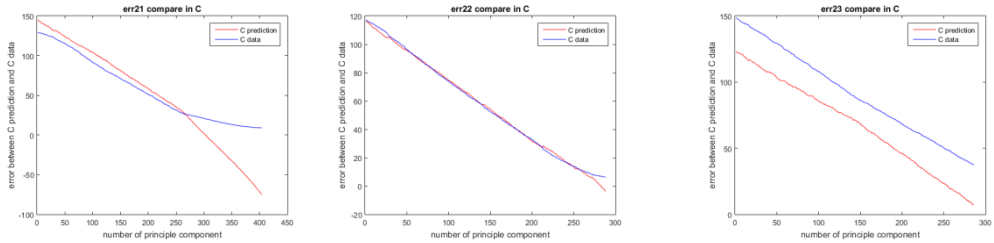
3.3The model of problem 3

3.3.1 Error analysis of crossover experiment

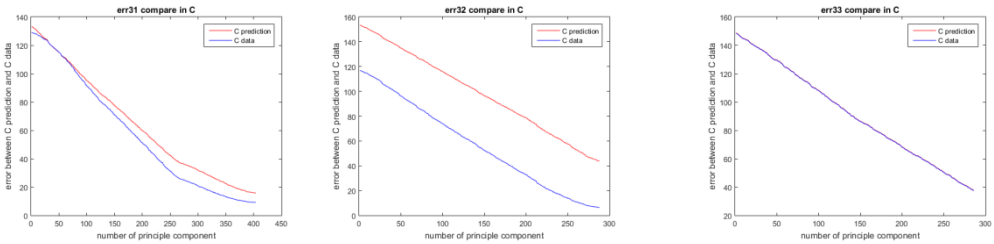
According to crossover experiment designed in the problem, we obtained the experimental results by using MATLAB. Figure err11 compare in C showed the predicted and actual values of the key element content based on the prediction model of 1 process and the data of 1 process,figure err21 compare in C showed the predicted and actual values of the key element content based on the prediction model of 1 process and the data of 2 process, and so on.



err11 compare in C err12 compare in C err13 compare in C



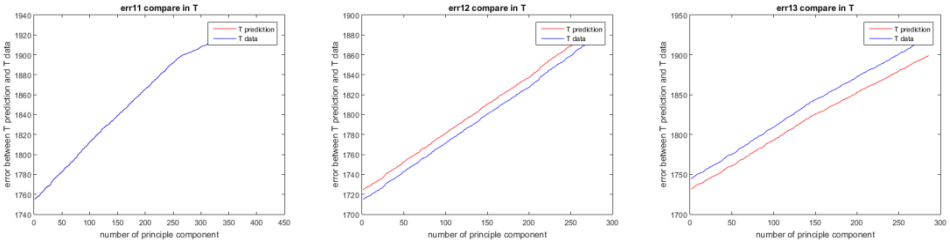
err21 compare in C err22 compare in C err23 compare in C



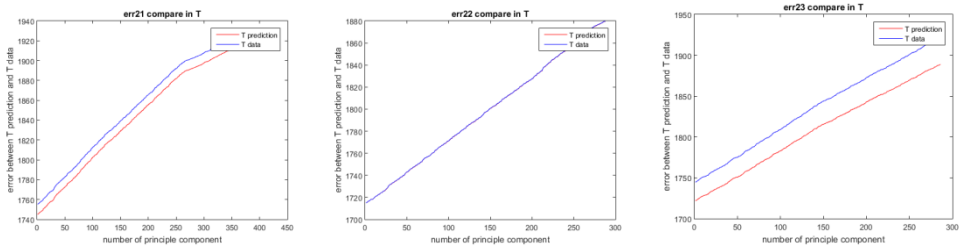
err31 compare in C err32 compare in C err33 compare in C

Figure 16 err compare in C of crossover experiment

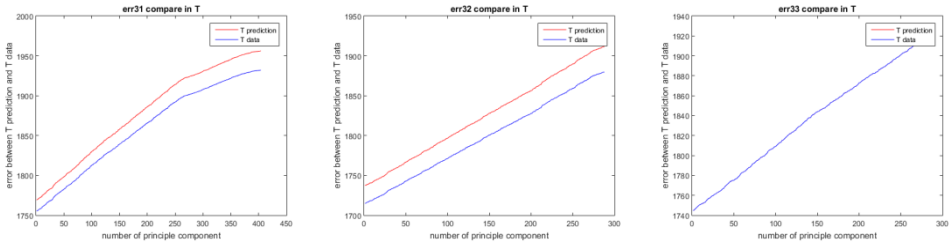
Figure err11 compare in T showed the predicted and actual values of the temperature based on the prediction model of 1 process and the data of 1 process,figure err21 compare in C showed the predicted and actual values of the temperature based on the prediction model of 1 process and the data of 2 process, and so on.



err11 compare in T err12 compare in T err13 compare in T



err21 compare in T err22 compare in T err23 compare in T



err31 compare in T err32 compare in T err33 compare in T

Figure 17 err compare in T of crossover experiment

From the results of the crossover experiment, the cross simulation curves of the two groups between T and C produced deviation to a certain extent,whats's more, the tail of the curve appears bifurcation in err21 compare in C. Through the assistance of MATLAB software, we calculated the mean square error of each unit in the cross test to have a rough description of the difference between the predicted data and the given data. The error results analysed are shown in the following table.

Table7：Mean square deviation T

Data Model	Prediction model based on 1 process	Prediction model based on 2 process	Prediction model based on 3 process
1 process data	0.124	10.153	20.271
2 process data	9.879	0.128	27.186
3 process data	18.033	27.963	0.0189

Table8：Mean square deviationC

Data Model	Prediction model based on 1 process	Prediction model based on 2 process	Prediction model based on 3 process
1 process data	1.062	27.788	7.809
2 process data	43.542	2.068	42.134
3 process data	5.112	23.836	0.204

According to the above table, predictive value performed well based on the prediction model of 1 process and the data of 3 processes for T,while forecast result performed poor based on the prediction model of process 3. The performance of the three prediction models in the cross prediction isn't very good for C, which can be seen in the figure.

The error of the model is derived from many aspects, the analysis of the error will help to improve the general applicability of the model, and this section will be described in detail in the next section.

3.3.4 Error control scheme

1. Error source analysis

The error of the model can be divided into several aspects:error generated by modeling method、error in a given data、random error. Because of the random error is not controlled and the impact is not as high as the other two, so here we only analyze the error of the modeling method and the error of the data itself.

In the data processing method we used, the error of principal component analysis

is mainly derived from the dimension of data reduction and the selection of cumulative variance contribution rate. The error caused by the data reduction is caused by the method itself, which belongs to uncontrollable factors,while cumulative variance contribution rate is determined by the selection of people.

(1) Error analysis of principal component algorithm

1) Error generated by data dimension reduction

Three m-dimensional samples were given. Assuming that the goal is to reduce the n samples from the m dimension to the k dimension,this process should be as much as possible to ensure that this dimension reduction operation does not make the important information loss. In other words, we need to project the n samples from the m dimensional space to the k dimension. For each sample point, we can use the following formula to represent the projection process：

$$Z = A^T X \tag{26}$$

X is m dimensional sample point；Z is the K dimension sample point obtained after projection； A is a m*k dimensional matrix. In the principal component analysis, we first need to find out the mean value of the sample:

$$u = \frac{1}{n}\sum_{i=1}^{n} X^{(i)} \tag{27}$$

Then seek out the dispersion matrix:

$$S = \sum_{i=1}^{n} (X^{(i)} - u)(X^{(i)} - u)^T \tag{28}$$

Then seek the characteristic vector $S_1, S_2, \ldots\ldots, S_k$ that the K large eigenvalues of the scatter matrix corresponds to. After unit characteristic vector $S_1, S_2, \ldots\ldots, S_k$, we can get the matrix A:

$$A = \left[S_1, S_2, \ldots S_k\right]^T \tag{29}$$

Taking a set of 10 sample point data as an example, the loss of this set of sample points can be expressed by the following formula when we reduce it to 1 dimension using PCA：

$$L = \sum_{i=1}^{n} \left\| X^{(i)} - AA^T X^{(i)} \right\|^2 \tag{30}$$

The meaning of this formula is the sum of the distances projecet each sample point from the high dimensional space to the low dimensional space. In order to achieve the purpose of dimension reduction, this error is not controlled. This part of the error cannot be eliminated or reduced unless it is found to be a more reliable dimension reduction method.

2）Parameter error selection

The parameter here refers to the cumulative variance contribution rate. In the principal component algorithm.The variance contribution rate of the first k principal

component Y_k is defined by the following formula:

$$\alpha_k = \frac{\lambda_k}{\sum_{i=1}^{p} \lambda_i} (i = 1,2,\ldots, p) \tag{31}$$

Cumulative variance contribution rate of the first m principal components $Y_1, Y_2, \ldots, Y_m$ is as follows:

$$\sum_{i=1}^{m} \alpha_i = \frac{\sum_{i=1}^{m} \lambda_i}{\sum_{i=1}^{p} \lambda_i}, m < p \tag{32}$$

In actual analysis, the value of m is determined according to the cumulative variance contribution rate. The greater the cumulative variance contribution rate is, the stronger the variance of the random vector X is explained by the few selected principal components. However, there is a contradiction here. If we select a larger cumulative variance contribution rate $\sum_{i=1}^{m} \alpha_i$, although it can reduce the loss of data information, but will lead to an increase in the selection of the main components, and ultimately lead to deviation from the dimension of the target. If the cumulative variance contribution rate is too small, although it will greatly reduce the number of principal components, but will lead to the data representation is not enough, and the significance of modeling will lost.

In the model of principal component analysis , we generate the interpretation total variance table and gravel map by SPSS software. After analysis and comparison, we have selected the main components of which the cumulative variance contribution rate is more than 85%. As we can see from the rubble and form, not only the selection of the principal component is suitable for quantity, but also the loss of the data information controled in the acceptable range. We take the gravel map and the total variance table generated by the principal component analysis of the first set of characteristic values as an example：

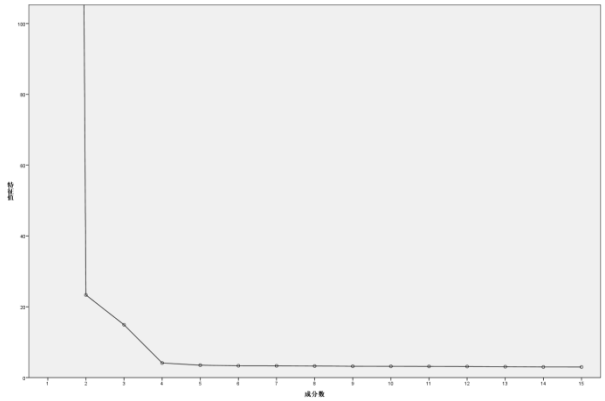


Figure 18 gravel map

解释的总方差

成份	初始特征值			提取平方和载入			旋转平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	1674.063	81.741	81.741	1674.063	81.741	81.741	1670.254	81.555	81.555
2	23.336	1.139	82.881	23.336	1.139	82.881	24.623	1.202	82.758
3	14.912	.728	83.609	14.912	.728	83.609	16.405	.801	83.559
4	4.121	.201	83.810	4.121	.201	83.810	4.072	.199	83.758
5	3.499	.171	83.981	3.499	.171	83.981	3.683	.180	83.937
6	3.343	.163	84.144	3.343	.163	84.144	3.502	.171	84.108
7	3.323	.162	84.307	3.323	.162	84.307	3.396	.166	84.274
8	3.264	.159	84.466	3.264	.159	84.466	3.380	.165	84.439
9	3.204	.156	84.622	3.204	.156	84.622	3.309	.162	84.601
10	3.181	.155	84.778	3.181	.155	84.778	3.269	.160	84.760
11	3.156	.154	84.932	3.156	.154	84.932	3.237	.158	84.919
12	3.122	.152	85.084	3.122	.152	85.084	3.235	.158	85.076
13	3.069	.150	85.234	3.069	.150	85.234	3.226	.158	85.234
14	2.998	.146	85.380						
15	2.988	.146	85.526						
16	2.940	.144	85.670						
17	2.906	.142	85.812						
18	2.878	.141	85.952						
19	2.840	.139	86.091						
20	2.818	.138	86.228						

Figure 19 Total variance table（Partial screenshots）

（2）Error analysis of regression model

1）Least square error

The error of the regression model is mainly derived from the least squares estimation method by SPSS. For each sample unit, the difference between the observed value y_i and the mean value $\beta_0 + \beta_1 x_i$ of the dependent variable is considered, the smaller of the difference the better. Comprehensive consideration of n differences, define the sum of squared deviations:

$$Q(\beta_0, \beta_1) = \sum_{i=1}^{n} [y_i - E(Y_i \mid x_i)]^2 = \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x_i)^2 \tag{33}$$

The least square method is to find the estimated value $\hat{\beta}_0, \hat{\beta}_0$ of the temperature β_0, β_1. The least square estimation of regression parameters is brought into the regression function, the deviation between the predicted variables $\hat{y}_i$ and the observed values y_i of the regression function is called residual error. In the modeling process, tables generated automatically by SPSS in the fitting process can be intuitive to express the residuals of the regression model. The following table is first set of table of variance analysis table and residual error statistics for fitting the optical characteristic value λ and time t.

***Table9*：Table of variance analysis**

Anova[b]

模型		平方和	df	均方	F	Sig.
1	回归	2.079E9	1	2.079E9	783.593	.000[a]
	残差	1.067E9	402	2653422.842		
	总计	3.146E9	403			

a. Predictive variable:(constant),t

b.. dependent variable: lamda

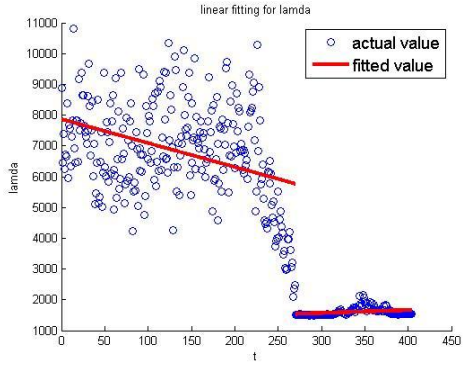
Table10*：*Residual statistic table

	极小值	极大值	均值	标准偏差	N
预测值	1148.3977	8987.6309	5068.0141	2271.41294	404
标准预测值	-1.726	1.726	.000	1.000	404
预测值的标准误差	81.043	161.784	111.853	25.026	404
调整的预测值	1144.5404	8993.1162	5069.5245	2272.53831	404
残差	-3176.24365	5658.90137	.00000	1626.91077	404
标准残差	-1.950	3.474	.000	.999	404
Student 化残差	-1.955	3.478	.000	1.001	404
已删除的残差	-3192.58228	5673.51514	-1.51037	1633.51280	404
Student 化已删除的残差	-1.962	3.528	.000	1.003	404
Mahal。距离	.000	2.978	.998	.893	404
Cook 的距离	.000	.016	.002	.003	404
居中杠杆值	.000	.007	.002	.002	404

a. dependent variable：lamda

(3) Source data error analysis

According to the analysis of the source data, the fluctuation of the optical characteristic value λ is very violent, and there is no law to follow, while C, K and Q, P and t showed a significant linear correlation. Therefore, in regression analysis of the factors P, Q, C, λ toC、K，it should be considered linear characteristics of P, Q, K, C, and linear regression is the best choice. However, we find that the fitting effect of the linear regression model is not good when combined with the principal component analysis results of optical characteristic values,the results shown in the fit diagram graph and goodness of fit table：



模型	R	R 方	调整 R 方	标准估计的误差
1	.813[a]	.661	.660	1628.93304

a. 预测变量: (常量), t。

b. 因变量: lamda

Figure 20 the fit diagram grapha and goodness of fit table

As shown in the above table，R square is only 0.661, which showed that there is an error in the process of collecting the source data.

2. Error control method

The error can be divided into controllable and uncontrollable errors by analyzing the source of error. From the source, process and results, reducing the negative effects of controllable errors can improve the model's universal., we put forward the method of controlling the error ,according to the above analysis:

(1) Improve the technology of metal smelting process, increase the accuracy of measurement and recording for optical information data. Reducing the volatility of data and the error has a great help for accurate prediction of T and C.

(2) Aim at the volatility of the initial data, smoothing should be used. Through the design of reasonable algorithm, the abnormal fluctuation value is filtered, so as to reduce the negative impact on the accuracy of the model.

(3) In the principal component algorithm, the larger cumulative variance contribution rate is chosen to improve the data loss in the dimension reduction process, so as to improve the control error and improve the model's universality.

(4) In the regression model, the relationship between T and Q is needed to be further explored.For example, through the data processing before regression or using different regression model to carry on the thorough analysis, is helpful to obtain the more accurate regression coefficient.

(5) The influence of random error can be reduced by means of the method of multiple measurements of the same kind of process.

4.Conclusions

4.1 Methods used in our models

Firstly,we used principal component analysis method to extract the characteristic value of light intensity data.Then We established multiple linear regression models to predict the Kelvin temperature ***T*** and key element content ***C*** by three sets of data given in the appendix, and the results have a small error between the predicted results and the actual values given.Lastly, we designed the crossover experiment,analyzed the error characteristic scientifically,and we put forward a scheme of control error.

4.2 Advantages of the model

- The principal component analysis method can reduce the dimensionality of the original data，getting a few comprehensive value to replace the original large light intensity data. At the same time, the weight is determined according to the variance contribution rate, which reduced subjective factors and uncertain factors in the process of finding the characteristic value. The characteristic values of light intensity data could be calculated by SPSS,which is convenient in the process of caculation and can reduce the workload.
- The process of multiple linear regression model building is relatively simple, which only need to bring influencing factors into multiple linear regression model. We can get the general effect of various independent variables on the dependent variable by using multiple linear regression model.The result of the prediction has a small error, which proved that this method is accurate and avaiable well.

4.3 Disadvantages of the models

- The principal component analysis can lead to some loss of some information of the original data in the process of reducing the dimension of the original data,and the characteristic values has some fuzziness, which could not represent all the information of the original data.
- Due to the possible presence of multicollinearity between the independent variables multiple linear regression model, which added the time of the model calculation and will take a lot of time to test the model.

5.Future Work

5.1Improved method of the model

- In the process of applying principal component analysis method，decision makers should be as far as possible to select more elements in order to minimize the loss of information loss. In the application of multiple linear regression model, more data should be used to design more cross experiments, which makes the model more universal.

5.2Proposal

- More scientific methods and significance should be used when light intensity data are collected, which can reduce the error of calculation.
- More group of data should be recorded in order to design more cross experiments to make the model more universal.

5.3 Model Application

- Principal component analysis is used to extract the characteristic value of light intensity data，which request the characteristics of sample data should be fully analyzed In the practical application.
- In the application of multiple linear regression model, a large amount of data should

be used to test the model, and the model can be modified constantly ,which will make the model has more universal.

6.References

[1] Mao Shisong，Cheng Yiming，Pu Xiaolong.Probability theory and mathematical statistics [M].Bei—jing：Advanced Education Press，2004.

[2] Cheng Naiping, Jiang Xiufu, Shao Dingrong. Acoustic optic signal processing and its application[M].Beijing: National Defence Industry Press,2004.

[3] Shen Naicheng. The history, current situation and development trend of measurement and basic physical constant temperature unit Kelvin[J]. China Metrology,2013,01:57-59.

[4] Lin Haiming. Analysis of ten problems in the application of principal component analysis[J].Statistics & Decision,2007,16:16-18.

[5] Shang Liqun,Wang Shoupeng. Effect of ultra-high pressure treatment on aroma compounds in cucumber water analyzed by principal component[J]. Power System Technology,2014,07:1928-1933.

[6] YIN Renkun．Data Structure[M]．Beijing：Tsinghua University Press，2007.

[7] Brian P．O'connor，SPSS and SAS programs for determining the number of components using parallel analysis and Velicer'S MAP test[J]． Behavior Research Method ， Instruments&Computers，2000，32(3):396—402.

[8] Bi Jianwu, Jia Jinzhang, Liu Dan. Prediction of gas emission in coal mining face based on SPSS multiple regression analysis[J]. Journal of Safety and Environment,2013,05:183-186.

[9] Wang Shengtao, Liang Xiaoyong, Zhou Yitao. Discussion on data regression analysis of tunnel monitoring measurement[J]. Tunnel Construction,2009,06:629-632+663.

[10] Qin Xiaobo,Li Yue,Shi Shengwei,Wan Yunfan,Ji Xionghui,Liao Yulin,Liu Yuntong,Li Yong. Multivariate regression analysis of greenhouse gas emissions associated activities and populations of soil microbes in a double-rice paddy soil[J]. Acta Ecologica Sinica, 2012,06:1811-1819.

[11] Rao C R. Some comments on the minimum mean square error as criterion of estimation statistics and related topics.Ameserdam:North Holland Press,1981:123-143.

[12] Jolliffe I. Principal component analysis[M]. New York:Springer-Verlag, 1986: 10-28.

7.Appendix

Statement:In this paper,we mainly used matlab software and spss software.

1. Code of Problem two

```
%T 总程序
%T1 总程序
clear;clc
load APMCM1
y1=polyfit(time1,lamda1,1); %对散点图进行线性拟合
lamdatest1=y1(1)*time1+y1(2);
figure (1)
x1=1:269;
plot(x1,lamda1,'b-');
hold on
plot(x1,lamdatest1,'r-');
title('lamda 拟合');
xlabel('t');
ylabel('lamda');
legend('实验值','预测值');
%对 T1 进行多元回归
e1=ones(269,1);
X1=[e1,time1,Q1,lamdatest1];
[b1,bint1,r1,rint1,stats1]=regress(T1,X1,0.05);
figure (2)
rcoplot(r1,rint1);
%绘制 T1 与 Ttest1 曲线
Ttest1=b1(1)+b1(2)*time1+b1(3)*Q1+b1(4)*lamdatest1;
figure (3)
x1=1:269;
subplot(1,2,1)
plot(x1,T1,'b-');
title('温度 T 的多元回归');
xlabel('横坐标');
ylabel('T');
legend('实验值');
subplot(1,2,2)
plot(x1,Ttest1,'r-');
xlabel('横坐标');
ylabel('T');
legend('预测值');
figure (4)
plot(x1,T1,'b-');
hold on
plot(x1,Ttest1,'r-');
```

```
title('温度 T 的多元回归');
xlabel('横坐标');
ylabel('T');
legend('实验值','预测值');
%T2 总程序
y2=polyfit(time2,lamda2,1);% 对散点图进行线性拟合
lamdatest2=y2(1)*time2+y2(2);
%lamdatest2=[];
%for  i=1:135
%      lamdatest2(i,1)=1.9.*time2(i)+1276.9;
%end
figure (5)
x2=1:135;
plot(x2,lamda2,'b-');
hold  on
plot(x2,lamdatest2,'r-');
title('lamda 拟合');
xlabel('t');
ylabel('lamda');
legend('实验值','预测值');
%用 SPSS 对 lamda 和 lamdatest 进行相关性检测
%对 T2 进行多元回归
e2=ones(135,1);
X2=[e2,time2,Q2,lamdatest2];
[b2,bint2,r2,rint2,stats2]=regress(T2,X2,0.05);
figure (6)
rcoplot(r2,rint2)
%绘制 T2 与 Ttest2 曲线
Ttest2=b2(1)+b2(2)*time2+b2(3)*Q2+b2(4)*lamdatest2;
%Ttest2=[];
%for  i=1:135
%      Ttest2(i)=-1.9956.*time2(i)+0.1605.*Q2(i)+1.0503.*lamdatest2(i);
%end
figure (7)
x2=1:135;
subplot(1,2,1)
plot(x2,T2,'b-');
title('温度 T 的多元回归');
xlabel('横坐标');
ylabel('T');
legend('实验值');
subplot(1,2,2)
plot(x2,Ttest2,'r-');
xlabel('横坐标');
```

```
ylabel('T');
legend('预测值');

figure (8)
plot(x2,T2,'b-');
hold on
plot(x2,Ttest2,'r-');
title('温度 T 的多元回归');
xlabel('横坐标');
ylabel('T');
legend('实验值','预测值');
%绘制 T 与 Ttest 曲线
Ttest=[];
for i=1:404
if i<=269
        Ttest(i)=Ttest1(i);
else
        Ttest(i)=Ttest2(i-269);
end
end
figure (9)
x=1:404;
plot(x,T,'b-','linewidth',2)
hold on
plot(x,Ttest,'r-','linewidth',2)
title('mutiple linear regression for T prediction');
xlabel('xlabel');
ylabel('T');
h1=legend('actual value','fitted value');
set(h1,'Fontsize',14)
figure (10)
subplot(1,2,1)
plot(x,T,'b-','linewidth',2)
title('mutiple linear regression for T prediction');
xlabel('xlabel');
ylabel('T');
h2=legend('actual value');
subplot(1,2,2)
plot(x,Ttest,'r-','linewidth',2)
xlabel('xlabel');
ylabel('T');
h3=legend('fitted value');
set(h2,'Fontsize',14);
set(h3,'Fontsize',14);
```

代码 2

```
%探索 T、C 之间的关系
clear;
clc;
load APMCM1;
%C 和 T 的线性回归
e=ones(404,1);
X=[e,T];
[b,bint,r,rint,stats]=regress(C,X,0.05);
figure (1)
rcoplot(r,rint)
Ctest=b(1)+b(2)*T;
%y3=polyfit(T,C,1);
%Ctest=y3(1)*T+y3(2);

ttt=1:404;
figure (2)
plot(ttt,C,'b-','linewidth',2)
hold on
plot(ttt,Ctest,'r-','linewidth',2)
title('The relationship between T and C')
xlabel('T');
ylabel('C');
h=legend('actual value','fitted value');
set(h,'Fontsize',14)
b
%y3 %线性拟合参数 常数项 α 1
```