Composition Analysis and Identification of Ancient Glass Products

Yuchen Wang*
Jiangsu University of Science and Technology, Suzhou, Jiangsu, 215699 China
*Corresponding author: 18851357465@163.com

Abstract: In this paper, aiming at the problem of composition analysis and identification of ancient glass products, a correlation analysis model based on chi square test is established to solve the relationship between the surface weathering of cultural relics and glass types, patterns and colors; ARMA model is established to solve the statistical law of chemical composition content with or without weathering and the problem of predicting chemical composition content before weathering; A classification model based on K-means++clustering is established to solve the problem of classification rules; An evaluation model based on principal component analysis is established to solve the problem of type identification of glass relics.

Keywords: Chi-square test, ARMA model, K-means++clustering algorithm.

1. Introduction

The Silk Road, which originated in the Western Han Dynasty, was an important exploration and discovery in the history of China and even the world since ancient times, as a channel for communication, trade and cultural exchanges between China and foreign countries. Between them, glass, as an important part of the early trade exchanges, provided a lot of evidence for the long-term trade relationship between China and the West and the resulting interaction of different cultures. The glass of the Silk Road period was made of fine beads from Western Asia and Egypt and then introduced into China. The research shows that in the 3rd century BC, China absorbed the glass making technology from Western Asia and Egypt and began to use local materials to make glass. The inlaid glass beads were the first glass products that China copied from Western Asia.

According to chemical composition or other basis, existing archaeologists divide a batch of ancient glass products into high potassium glass and lead barium glass. Glassware plays an important role in long-term trade relations. Therefore, it is of great significance to analyze the internal relationship and chemical composition of glass cultural relics through mathematical modeling for studying the cross-cultural communication interaction caused by different art and style elements incorporated in glassware. The main problems to be solved in this paper are as follows:

(1) The relationship between surface weathering of glass relics and glass type, decoration and color is analyzed. Then, according to the glass type, analyze the statistical rule of whether there is weathering chemical composition content on the surface of the cultural relic sample. Finally, combine the weathering point detection data, establish a mathematical model to predict the chemical composition content before weathering.

(2) Analyze the classification rules of high potassium glass and lead barium glass according to the chemical composition proportion of the classified cultural relics, and use specific classification methods to select appropriate chemical composition for each category to classify, give the classification results, and finally give the rationality and sensitivity analysis of the classification results.

2. Analysis of the factors affecting the weathering of glass relics and their chemical composition

2.1 Factors affecting the weathering of glass relics

2.1.1 Establishment of association analysis model based on chi-square test

The basic concept of chi square test is mainly to compare the difference analysis between categorical variables and categorical variables. According to the deviation degree between the actual observation value and the theoretical inference value of the statistical sample, the deviation degree between the actual observation value and the theoretical inference value determines the chi square value[1]. If the chi square value is larger, the deviation between the actual observation value and the theoretical inference value is larger; On the contrary, the deviation between the actual observation value and the theoretical inference value is smaller; If the two values are completely equal, the chi square value will be 0, indicating that the actual value is completely consistent with the theoretical value[2].

2.1.2 Relationship between surface weathering and ornamentation

The pattern A/B/C is represented by the number 1/2/3, the weathering is represented as 1, and the non weathering is represented as 2, and the corresponding total number of cultural relics is filtered out in excel, the table data is imported, weighted, and then chi square test is performed, and the chi square test results are shown in Table 1.

Table.1. Chi-square test result 1

Pearson chi-square	5.747ᵃ	2	0.056
Likelihood ratio	7.993	2	0.018
Linear correlation	0.205	1	0.650
Valid cases	54		

2.1.3 Relationship between surface weathering and classification

The glass type high potassium and lead barium are respectively represented by the number 1/2, the weathering is represented as 1, and the non weathering is represented as 2. The corresponding total number of cultural relics is screened out in excel, the table data is imported, weighted, and then chi square test is conducted. The chi square test results are shown in Table 2.

Table.2. Chi-square test result 2

Pearson chi-square	5.400ᵃ	1	0.020		
Continuity correctionᵇ	4.134	1	0.042		
Likelihood ratio	5.448	1	0.020		
Fisher's exact test				0.040	0.021
Linear correlation	5.300	1	0.021		
Valid cases	54				

2.1.4 Relationship between surface weathering and color

The colors of cultural relics are blue-green, light blue, purple, dark green, dark blue, light green, black, and green, represented by the number 1/2/3/4/5/6/7/8 respectively. Weathering is represented as 1, and non weathering is represented as 2. The corresponding total number of cultural relics is filtered out in excel, the table data is imported, weighted, and chi square test is performed. The chi square test results are shown in Table 3.

Table.3. Chi-square test result 3

Pearson chi-square	6.287[a]	7	0.507
Likelihood ratio	8.156	7	0.319
Linear correlation	0.707	1	0.401
Valid cases	54		

It can be seen from the above chi square test that the Pearson chi square values of whether the surface of the cultural relics is weathered and its corresponding patterns, types and colors are 5.747/5.400/6.287 respectively. In the chi square test critical value table, their corresponding P values are less than 0.05, so it is concluded that there are significant differences in the results, that is, whether the surface of the cultural relics is weathered has a great relationship with its glass type, patterns and colors.

2.2 Chemical composition analysis of glass relics before and after weathering

2.2.1 ARMA model

Autoregressive moving average model (ARMA) is a method used to predict stationary time series[3]. It is often used to track long-term data, estimate irregular stationary fluctuations or time series with seasonal changes. The model is composed of autoregressive model (AR) and moving average model (MA). The basic principle of the model is to regard the data sequence generated by the change of the indicators required for prediction with time as a random sequence, and show the continuity of this group of random variables in time through mathematical operation[4].

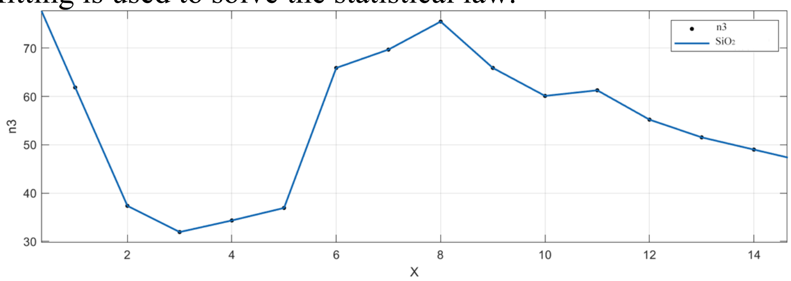
The autoregressive moving average process has the characteristics of randomness, which includes two different parts, namely autoregressive and moving average. If *p* of AR represents the upper limit of some order values and *q* represents the upper limit of the later order values, then the autoregressive moving average process can be expressed as ARMA (*p*, *q*). The specific expression is as follows.

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_p \varepsilon_{t-p} \tag{1}$$

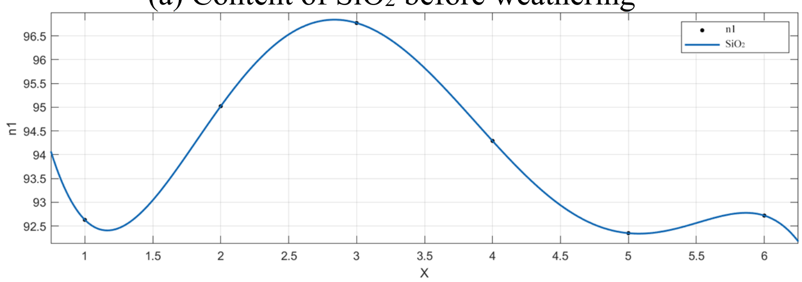
Where: ε_t is white noise, φ_1, $\varphi_2, \cdots \varphi_p$ are parameters of autoregressive model, θ_1, $\theta_2, \cdots \theta_p$ are parameters of the moving average model.

2.2.2 Chemical composition content of glass relics before and after weathering

Scatter plot curve fitting is used to solve the statistical law.



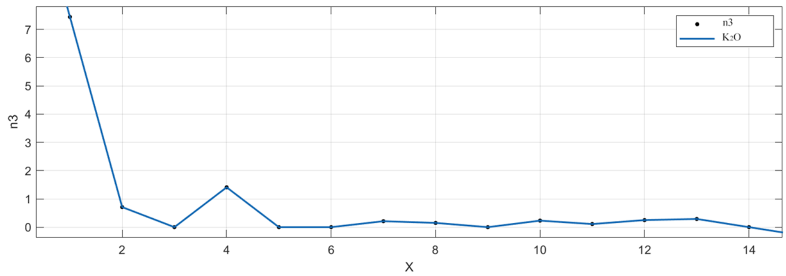
(a) Content of SiO_2 before weathering



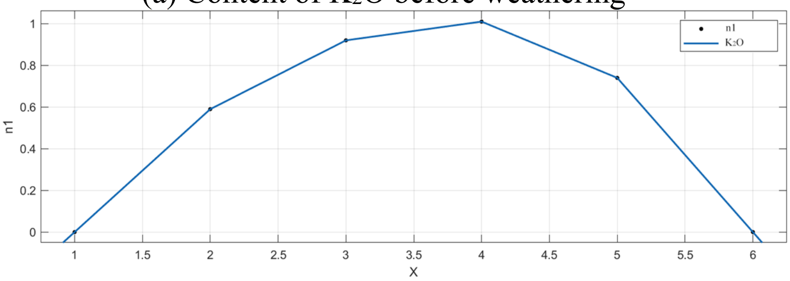
(b) Content of SiO_2 after weathering

Figure 1. SiO_2 content before and after glass weathering

It can be seen from Figure 1 that the content of SiO_2 in potassium glass increases significantly after weathering.



(a) Content of K_2O before weathering



(b) Content of K_2O after weathering

Figure 2. K_2O content before and after glass weathering

It can be seen from Figure 2 that the potassium oxide content decreases slightly after weathering of potassium glass. According to the analysis by the same method, the silicon dioxide content of the lead barium glass decreases significantly after weathering, the lead oxide content increases significantly, and the barium oxide content increases significantly.

2.2.3 Prediction of chemical composition content

Combined with AIC criterion, the model is identified and ordered. Select *p* and *q* to make the following formula hold.

$$\min \mathrm{AIC}=\mathrm{n}\ln\hat{\sigma}_{\varepsilon}^{2}+2(\mathrm{p}+\mathrm{q}+1) \tag{2}$$

Where, n is the sample size, $\hat{\sigma}_{\varepsilon}^{2}$ is the estimate of σ_{ε}^{2}, which is related to *p* and *q*. If the above formula reaches the minimum value when $p=\hat{p}$, $q=\hat{q}$, the sequential formula $\mathrm{ARMA}(\hat{p}, \hat{q})$ is considered。 When ARMA (*p*, *q*) sequence contains unknown mean parameter μ, the model is:

$$\varphi(\mathrm{B})(X_{\mathrm{t}}-\mu)=\theta(\mathrm{B})\varepsilon_{\mathrm{t}} \tag{3}$$

The m-step prediction of time series is to estimate the random variable X_{k+m} (m>0) at the future $k+m$ time according to the value of $\{X_k, X_{k-1},\cdots\}$. The estimator is recorded as $\hat{X}_k(m)$, which is a linear combination of $X_k, X_{k-1},\cdots$.

For ARMA (*p*, *q*), there are:

$$\hat{X}_k(m)=\varphi_1\hat{X}_k(m-1)+\varphi_2\hat{X}_k(m-2)+\cdots+\varphi_p\hat{X}_k(m-p), m>p \tag{4}$$

Therefore, you only need to know $\hat{X}_k(1),\hat{X}_k(2),\cdots,\hat{X}_k(p)$ to get $\hat{X}_k(m)$, $m>p$ recursively. The prediction vector is defined as follows.

$$\hat{X}_k(m)=\left[\hat{X}_k(1),\hat{X}_k(2),\cdots,\hat{X}_k(p)\right] \tag{5}$$

$$\varphi_j^{*}=\begin{cases}\varphi_j, j=1,2,\cdots,p,\\ 0, j>p\end{cases} \tag{6}$$

The following recursive prediction formula can be obtained:

$$\hat{X}_{k+1}^{(q)}=\begin{bmatrix} -G_1 & 1 & 0 & \cdots & 0 \\ -G_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -G_{q-1} & 0 & 0 & \cdots & 1 \\ -G_q+\varphi_q^* & \varphi_{q-1}^* & \varphi_{q-2}^* & \cdots & \varphi_1^* \end{bmatrix}\hat{X}_k^{(q)}+\begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{q-1} \\ G_q \end{bmatrix}X_{k+1}+\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \sum_{j=q+1}^{p}\varphi_j^* X_{k+q+1-j} \end{bmatrix} \tag{7}$$

Where: G_j meets $X_t=\sum_{j=0}^{\infty}G_j\varepsilon_{t-j}$. The third item in the formula is 0 when $p \le q$. Guaranteed by the reversibility condition, when k_0 is small, the initial value $\hat{X}_{k0}^{(q)}=0$ can be set.

In practice, the model parameters are unknown. If the time series model has been established. The unknown parameters of are replaced by their estimates, and then the method described above is used for prediction. Five layers of data were predicted to obtain the chemical composition content of the surface of five groups of cultural relics.

In order to simplify the calculation, only the main chemical components of the glass relics are analyzed below. The selected weathering point detection data are sorted in the order of severe weathering, weathering and non weathering, and are substituted into the ARMA sequence model for solution. The results are shown in Table 4.

Table.4. Prediction table of chemical composition content of unweathered surfaces of 5 groups of cultural relics

SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
61.74	2.9	0.2	2.04	1.5	9.20	0	0.6	19.21	11	0	0.27	0	0
65	5.47	0.3	1.2	0	9.77	1.21	0.21	23.02	8.54	3.54	0.25	0	0
57	0	0.5	3	0.2	8.43	0.43	1.47	15.64	5.31	0	0.31	0	0
58.02	3.22	0.4	0.95	1.2	5.62	0.37	2.03	27.51	10.64	4.61	0.26	0	0
53	0.59	0	2	0.77	3.36	0.76	0.92	17.01	4.65	1.02	0.1	0	0

The content of silica, the main component of glass, in potassium glass increases obviously, and in lead barium glass decreases obviously; For different glass types, the content of potassium oxide in potassium glass decreased after weathering, while the content of lead oxide and barium oxide in lead barium glass increased significantly; However, the content of calcium oxide, aluminum oxide and copper oxide, which are common to the two kinds of glasses and have low content, will decrease after weathering.

For intuitive and accurate analysis, calculate the average value of each content before and after weathering, compare it with the conclusion obtained above, and then check it. The results are shown in Table 5.

Table.5. Average content of each component before and after weathering

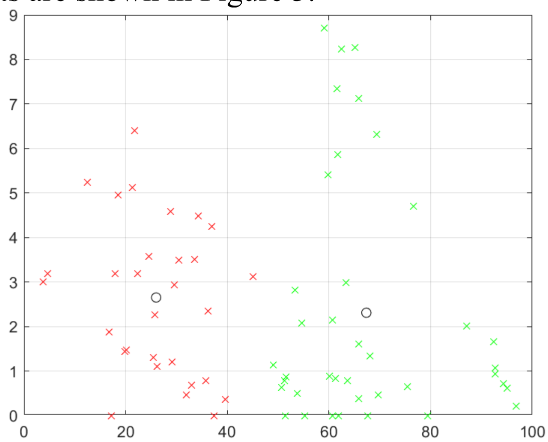
	Potassium is not weathered	After potassium weathering	Lead barium not weathered	After lead barium weathering
SiO_2	54	94	67	34
K_2O	0.77	0.54		
CaO	1.14	0.87	4.57	2.40
Al_2O_3	3.19	1.93	6.35	3.95
CuO	1.79	1.56	2.45	1.08
PbO			0.59	35.30
BaO			1.02	11.48

It can be seen from the above table that after weathering of potassium glass, the content of silicon dioxide increases significantly and the content of other components decreases; After the lead barium glass is weathered, the content of silicon dioxide, calcium oxide, aluminum oxide and copper oxide decreases, while the content of lead oxide and barium oxide increases significantly. Therefore, the rationality of the above conclusions is verified.

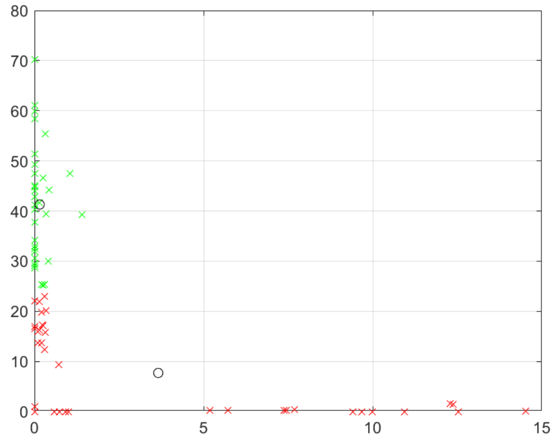
3. Classification of glass

3.1 Classification Model Based on K-means++Clustering Algorithm

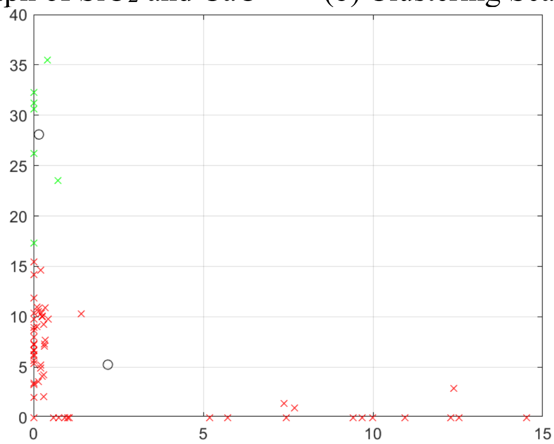
The cultural relics samples are divided into two categories, so K is taken as 2. The content data of silicon dioxide is selected as the first cluster center. Calculate the shortest distance between each sample and the cluster center., To give the classification rules of glass types, the preliminary judgment is based on its chemical composition[5]. Therefore, the clustering is divided into three categories: silica and calcium oxide, potassium oxide and lead oxide, and potassium oxide and barium oxide. Among them, silicon dioxide and calcium oxide are the components with large content in both types of glass, potassium oxide is the unique component of potassium glass, and lead oxide and barium oxide are the unique components of lead barium glass. Clustering and drawing. Clustering scatter points are divided into two categories, which are represented by red and green respectively[6]. The different results are shown in Figure 3.



(a) Clustering Scatter Graph of SiO_2 and CaO



(b) Clustering Scatter Graph of K_2O and PbO



(C) Clustering Scatter Graph of K_2O and BaO

Figure 3. Cluster Scatter Graph

According to the above clustering analysis and inspection, it can be concluded that high potassium glass and lead barium glass are classified according to their main chemical composition. The specific classification rules are: high potassium glass with high potassium oxide content and low lead oxide and barium oxide content; Lead barium glass has low potassium oxide content and high lead oxide and barium oxide content.

The evaluation of a clustering effect can be expressed by the contour value. Profile value, i.e. profile coefficient. The value of profile coefficient is between - 1 and 1. The closer the value is to 1, the more compact the cluster is, and the better the clustering is. When the profile coefficient is close to 1, the cluster is compact and far away from other clusters. The contour value analysis of potassium oxide and barium oxide clustering is shown in Figure 4.

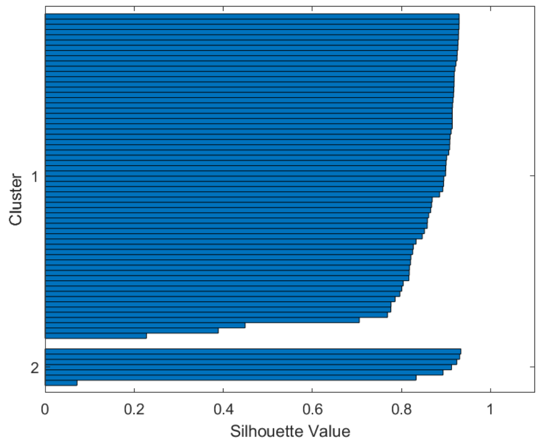


Figure 4. Profile coefficient histogram

3.2 Glass classification method

From the point of view of chemical composition, the high potassium glass and lead barium glass are classified into subcategories. The data of chemical composition content of high potassium glass are analyzed by principal component analysis, which is arranged from large to small. The results are shown in Table 6.

Table.6. Chemical composition content of high potassium glass Principal component analysis results

Number	Characteristic value	Contribution rate
1	4.2004	30.0026
2	2.6509	18.9347
3	1.8471	13.1638
4	1.6310	11.6497
5	0.9419	6.7280
6	0.9008	6.4344
7	0.6993	4.9953
8	0.5187	3.7052
9	0.3878	2.7700
10	0.1058	0.7554
11	0.0851	0.6076
12	0.0160	0.1141
13	0.0145	0.1039
14	0	0.0053

It can be seen from the above table that the first four characteristic values are greater than 1, and the principal component analysis is very effective. The corresponding chemical components of No. 1-4 are: MgO, PbO, BaO, SO_2.

Similarly, the chemical composition of lead barium glass was analyzed by principal component analysis, and the results are shown in Table 7.

Table.7. Chemical composition content of high potassium glass Principal component analysis results

Number	Characteristic value	Contribution rate
1	3.8681	27.6295
2	2.8947	20.6763
3	1.7766	12.6898
4	1.1154	7.9674
5	1.0329	7.2777
6	0.7957	6.6834
7	0.6648	4.7486
8	0.5277	3.7693
9	0.4416	3.1539
10	0.3681	2.6294
11	0.2250	1.6070
12	0.1682	1.2016
13	0.1163	0.8311
14	0.0049	0.0349

It can be seen from the above table that the first five characteristic values are greater than 1, and the principal component analysis is very effective.

3.3 Rationality and sensitivity analysis

The factor contribution after rotation is the contribution rate. Taking lead barium glass as an example, the chemical composition results of the top 5 weights are shown in Table 8.

Table.8. Factor contribution rate

chemical composition	initial	extract
K_2O	1.000	0.783
Na_2O	1.000	0.765
MgO	1.000	0.901
Fe_2O_3	1.000	0.926
P_2O_5	1.000	0.914

It can be seen from the above table that the common factor variances of the two variables, iron oxide and phosphorus pentoxide, exceed 0.91, which means that the variance contribution rate of the two common factors we extracted to the two variables exceeds 91%, that is, the two common factors can reflect 95% of the information of the two variables.

Table.9. Analysis of total variance

	Initial characteristic value			Extract the sum of squares of the load			Sum of squares of rotating loads		
Component	Total	Percent Variance	Accumulate	Total	Percent Variance	Accumulate	Total	Percent Variance	Accumulate
1	3.401	73.120	71.114	3.796	71.340	71.293	2.744	54.885	54.770
2	0.727	14.425	85.464	0.727	14.270	14.270	1.529	30.579	85.582
3	0.450	8.685	94.139						
4	0.230	4.374	98.811						
5	0.047	1.182	100.00						

Table 9 is the explanation table of total variance, giving the variance and cumulative sum explained by each common factor. It can be seen from the column of "initial eigenvalue" that the cumulative variance explained by the first two public factors is 85.64%, while the eigenvalue of the latter public factors is smaller and smaller, making less and less contribution to the interpretation of the original variables. Therefore, it is appropriate to extract two public factors.

The column "Sum of squares of extracted loads" refers to the variance contribution information of the two common factors extracted without rotation, which is the same as the values in the first two lines of the column "Initial eigenvalue". The "sum of squares of rotating loads" refers to the variance contribution information of the new common factor obtained after rotation. Compared with the contribution information of non-rotating, the variance contribution rate of each common factor changes, but the final cumulative variance contribution rate remains unchanged.

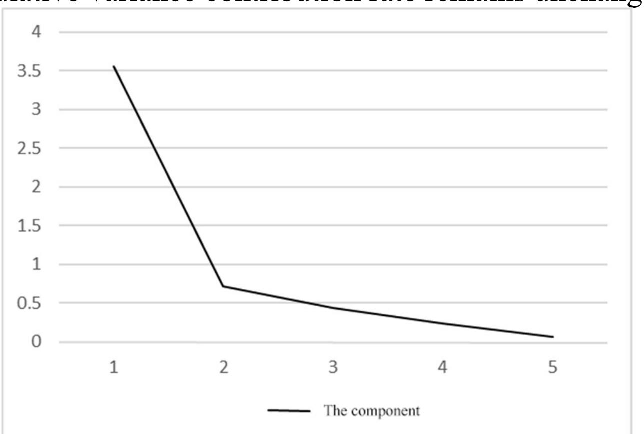


Figure 5. Profile coefficient histogram

It can be seen from the gravel map that the change of the eigenvalue corresponding to the first two factors is steep, and the change of the eigenvalue is relatively flat from the third factor, so we should select two factors for analysis. Based on the above figures, the factor analysis method is used to check, and it is found that the principal components obtained by the principal component analysis method are consistent with the common factors obtained by the factor analysis method, and the corresponding contribution rate error of the two methods is very small. Therefore, the principal component analysis model is reasonable. Therefore, the division result is reasonable.

According to the above analysis, the classification results of high potassium glass and lead barium glass are shown in Table 10.

Table.10. Classification of high potassium glass and lead barium glass

High potassium glass			Lead barium glass		
MgO	Y	N	K_2O	Y	N
PbO	Y	N	Na_2O	Y	N
BaO	Y	N	MgO	Y	N
SO_2	Y	N	Fe_2O_3	Y	N
			P_2O_5	Y	N

On the basis of the above model, control the content of other components to remain unchanged, adjust the content of calcium oxide in samples 1-10 for high potassium glass and lead barium glass, calculate the contribution rate to magnesium oxide again, and perform sensitivity analysis to verify the sensitivity of the results. The results are shown in Figure 11 and Figure 12.

Table.11. Numerical table of magnesium oxide contribution rate changing with calcium oxide in high potassium glass

Content of CaO	+1	+5	+10	+15	+20
MgO's contribution rate	30.0026	30.0127	31.0208	31.1025	31.2102

Table.12. Numerical table of variation of magnesium oxide contribution rate with calcium oxide in lead barium glass

Content of CaO	+1	+5	+10	+15	+20
MgO's contribution rate	12.6898	12.7085	12.7121	12.7202	12.7294

The results obtained from the table show that the contribution rate of magnesium oxide is almost unchanged when the content of calcium oxide changes, indicating that the classification results are not affected by the composition, so the sensitivity of the classification results is low. The results are stable.

4. Conclusions

(1) In this paper, the method of scatter plot curve fitting is used to analyze the statistical law of chemical composition on the surface of samples with or without weathering, and ARMA model is established to solve the prediction problem of chemical composition before weathering. In the inspection and analysis stage, the method of calculating the average value of each content before and after weathering is adopted. Through comparison with the above results, the results are consistent, which verifies the rationality of the model and the solution method.

(2) The principal component analysis method is used to establish the evaluation model, and the principal component analysis is carried out on the chemical composition of high potassium glass and lead barium glass respectively. The final result is that there are 4 sub types of high potassium glass and 5 sub types of lead barium glass. Then the factor analysis method is used to analyze the rationality of the classification results, and the principal components obtained by the principal component analysis method are consistent with the common factors obtained by the factor analysis method, which verifies the rationality of the classification results. Sensitivity analysis is used for inspection and analysis.

Under the premise of controlling the content of other components, change the parameter of calcium oxide content in samples 1-10 of high potassium glass and lead barium glass, and calculate the contribution rate of magnesium oxide again. When the calcium oxide content changes, the contribution rate of magnesium oxide is almost unchanged, so the classification results are less sensitive to the changes of system parameters, which proves that the model is stable.

References

[1] Li Chunming. Computing simulation and dynamics modeling of elastic system [J]. Journal of System Simulation, 2008 (01): 62-64+168.

[2] Xie Yu, Ren Jinbo, Zhang Xiang. Grid independence analysis of computational fluid dynamics based on chi-square [J]. Science Technology and Engineering, 2020, 20(01): 123-127.

[3] Qian Jiangu, Wu Haian, Ji Jun. Prediction for nonlinear time series of geotechnical engineering based on wavelet-optimized LSTM-ARMA model [J]. Journal of Tongji University(Natural Science), 2021, 49(08): 1107-1115.

[4] You J, Yu C, Sun J, et al. Generalized maximum entropy based identification of graphical ARMA models [J]. Automatica, 2022(141-).

[5] Yu Haofang, Sun Lifan, Fu Zhumu. Multi-extended target tracking algorithm based on improved K-means++ clustering [J]. Journal of Computer Applications, 2020, 40(01): 271-277.

[6] Bian Yongming, Gao Fei, Li Mengru, et al. Fire detection method using K-means++ clustering and features of mixed color and geometry [J]. Chinese Journal of Construction Machinery, 2020, 18(01): 1-6.