Research on component analysis and identification of ancient glass products based on cluster analysis

Hengyi Zhang #, Hongchen Guo #, Hao Zhang *, #

School of electrical and electronic engineering, North China Electric Power University, Baoding, China

* Corresponding Author Email: 3439887078@qq.com

#These authors contributed equally.

Abstract. In this paper, we study the ancient glass component analysis and category to identify problems, through the establishment of principal component analysis, cluster analysis and multiple regression model mathematical model, using SPSS software, realized by analyzing the chemical composition content on the glass before and after weathering type classification problem and for different categories of the relationship between the chemical composition of glass sample problem solving, According to the change rate, the chemical composition content before weathering is predicted, and the predicted value before weathering is obtained.

Keywords: Ancient glass products, chemical composition content, correlation analysis, cluster analysis, principal component analysis

1. Introduction

The Silk Road was an important trade link between China and the West in ancient times, and glass products are one of the important physical evidence of the early Silk Road.Glass products came to China along the Silk Road and were then made locally in China according to its techniques.Therefore, they are similar in appearance to the early imported glass products, but their chemical composition and structure are not the same.Burial environment is very easy to affect the weathering of ancient glass. Due to different burial environment, the degree and type of weathering of glass in the burial process are also different.Some cultural relics are marked as no weathering, but some superficial weathering is not excluded.Some cultural relics are marked as weathered, but do not exclude that some surfaces are not weathered.Due to the different types of glass and surface ornamentation, the degree of weathering is also different, which makes the glass show different colors[1].

Existing data on a batch of Chinese ancient glass products, by detecting the chemical composition of samples of these cultural relics and other testing means, this batch of glass has been divided into two types, that is, high potassium and lead barium glass, based on the analysis of the different types of chemical composition and content before and after weathering of cultural relics to find out high potassium and lead barium glass classification rule;Based on the analysis of chemical composition content, the specific subcategory classification method and classification results of each category of glass are given, and the rationality and sensitivity of the classification results are analyzed[2].

2. Establishment of composition analysis model

Considering that the type of glass, ornamentations and colors belong to the categorical variables, which belong to the categorical correlation analysis, the chi-square test should be used for analysis.

2.1. Grouping of data

Based on the known data, the data were first divided into two groups according to the type of glass: high-potassium glass and lead-barium glass, and then each group of data was processed separately to find the pattern.Each chemical composition of each group of data was treated as the dependent variable one by one to obtain the predicted value of each chemical content before weathering.Then, the known data were quantified and statistically analyzed according to the weathering condition, and the rate of change was calculated and tables and charts were drawn.Observe the statistical law of chemical composition content of cultural relics samples.Then, according to the change law of chemical composition between weathered samples and unweathered samples, a model was established to predict the chemical composition content before weathering[3].

2.2. Color-related data processing

Since there were missing values in the color part, the whole row of the missing values was deleted first, and the correlation analysis of the data was performed using SPSS software, and then the chi-square test was performed on the data.The significance test value of the correlation analysis between surface weathering and glass type was 0.020, the significance test value of the correlation analysis between surface weathering and glass ornamentation was 0.056, and the significance test value of the correlation analysis between surface weathering and glass color was 0.507.According to the chi-square test, when the significance test value is greater than 0.05, the correlation is not strong, and when it is less than 0.05, the correlation is strong. In conclusion, the surface weathering is strongly correlated with the type of glass, while the surface weathering is not strongly correlated with the ornamentation and color of glass[4].

2.3. Analysis of statistical regularities

Firstly, the data were divided into two categories of high-potassium glass and lead-barium glass according to the type of glass.Secondly, the glass was classified according to non-weathering and weathering, and the mean value of each chemical component content in non-weathering and weathering periods was counted as the reference amount.The change before and after weathering of each chemical composition was divided by the change without weathering as the change rate before and after weathering, and the statistical rule of chemical composition content with or without weathering was obtained by plotting statistics.Due to the limited space, this paper mainly presents the results of correlation analysis of high-potassium glass.The statistical rule of high-potassium glass is shown in Figure 1[5].

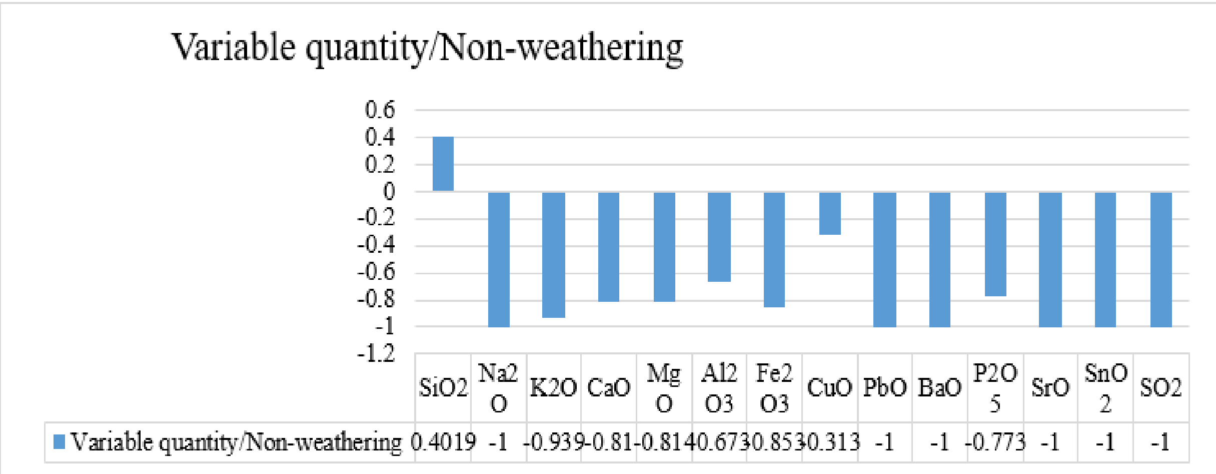

Figure 1. Statistical law of high potassium glass.

Since it is change/no weathering, the minimum rate of change is -1, which means that all of the component is consumed during weathering.As can be seen from the chart, in the weathering process of high-potassium glass, the proportion of silica increases before and after weathering, while the proportion of other chemical components decreases during weathering, especially sodium oxide, lead oxide, barium oxide, strontium oxide, tin oxide and sulfur dioxide, which are all consumed.Pre-weathering prediction: The contents of high-potassium glass and lead-barium glass before and after weathering were predicted according to the proportion of changes before and after weathering.Figure 2 shows the high potassium content before weathering. [6]

	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
1	92.63	0	0	1.07	0	1.98	0.17	3.24	0	0	0.61	0	0	0
1	95.02	0	0.59	0.62	0	1.32	0.32	1.55	0	0	0.35	0	0	0
1	96.77	0	0.92	0.21	0	0.81	0.26	0.84	0	0	0	0	0	0
1	94.29	0	1.01	0.72	0	1.46	0.29	1.65	0	0	0.15	0	0	0
1	92.35	0	0.74	1.66	0.64	3.5	0.35	0.55	0	0	0.21	0	0	0
1	92.72	0	0	0.94	0.54	2.51	0.2	1.54	0	0	0.36	0	0	0
1	55.40617597	0	0	1.936333802	0	3.312226662	0.315031671	4.253115373	0	0	1.081540243	0	0	0
1	56.83574264	0	1.144130968	1.121987811	0	2.208151108	0.593000792	2.034669391	0	0	0.620555877	0	0	0
1	57.88249647	0	1.784068627	0.380028129	0	1.355001816	0.481813143	1.102659541	0	0	0	0	0	0
1	56.39909675	0	1.95859708	1.302953586	0	2.442348953	0.537406968	2.165938384	0	0	0.265952519	0	0	0
1	55.23869535	0	1.435011722	3.00403188	1.161097616	5.854946119	0.648594616	0.721979461	0	0	0.372333526	0	0	0
1	55.46000902	0	0	1.701078293	0.979676113	4.198832788	0.370625495	2.021542492	0	0	0.638286045	0	0	0

Figure 2. Basic information before high potassium weathering.

3. Glass partition model

After data preprocessing, it was found that part of the data was missing, so the blank space was replaced by the value 0 to represent that the chemical composition contained in the sample was 0. Moreover, it was found that the cumulative proportion of the components of the samples numbered 15 and 17 were 79.47% and 71.89%, respectively, which were deleted according to the meaning of the question.According to the question stem, the sum of all chemical components is approximately equal to 100%, which can be regarded as the data normalization of the chemical components of each sample point. [7]

Firstly, the samples were divided into high-potassium glass and lead-barium glass, and then they were divided into weathered glass and unweathered glass, respectively. The chemical components contained in the samples were used as variables for cluster analysis. The classification rules of high-potassium glass and lead-barium glass could be obtained by comparing the changes of the chemical components in the weathered and unweathered glasses.

Firstly, the sample types were divided into high-potassium glass and lead-barium glass, and 14 chemical components were used as indicators for high-potassium glass and lead-barium glass, respectively. Sixty-seven monitoring sites were comprehensively evaluated according to principal component analysis (PCA). [8]

3.1. Results of cluster analysis based on SPSS

SPSS software was used to conduct k-means cluster analysis on the contents of chemical components of the two types of glass weathered and unweathered [9], and the results are shown in Table 1 and Table 2.

Table 1. Results of cluster analysis of each component after weathering.

	Cluster category (mean ± standard deviation)		F	P
	Lead barium glass (n=25)	High potassium glass (n=7)		
Silica (SiO2)	23.776 + / - 9.065	88.159 + / - 15.439	199.883	0.000 * * *
Sodium oxide (Na2O)	0.193 + / - 0.555	0.114 + / - 0.302	0.127	0.724
Potassium hydroxide (K2O)	0.126 + / - 0.242	0.511 + / - 0.415	10.001	0.004 * * *
(CaO CaO)	2.69 + / - 1.694	1.149 + / - 0.861	5.32	0.028 * *
Magnesium oxide (MgO style)	0.614 + / - 0.697	0.389 + / - 0.58	0.612	0.440

Alumina (Al2O3)	2.543 + / - 1.512	3.604 + / - 4.516	1.043	0.315
Iron oxide (Fe2O3)	0.567 + / - 0.746	0.374 + / - 0.296	0.438	0.513
Copper oxide (CuO)	2.367 + / - 2.839	1.339 + / - 1.038	0.867	0.359
Lead oxide (PbO)	44.418 + / - 11.081	2.244 + / - 5.938	92.386	0.000 * * *
Baryta (BaO)	11.987 + / - 10.141	1.044 + / - 2.763	7.815	0.009 * * *
Phosphorus pentoxide (P2O5)	5.444 + / - 4.194	0.397 + / - 0.364	9.882	0.004 * * *
Strontium oxide ((SrO) SrO)	0.425 + / - 0.268	0.036 + / - 0.094	14.001	0.001 * * *
Tin oxide (SnO2)	0.019 + / - 0.094	0.187 + / - 0.495	2.763	0.107
Sulfur dioxide (SO2)	1.421 + / - 4.283	0.0 + / - 0.0	0.752	0.393

Note: ***, **, * represent 1%, 5%, and 10% significance levels, respectively.

Table 2. Results of cluster analysis of each component without weathering.

	Cluster category (mean ± standard deviation)		F	P
	Lead barium glass (n=13)	High potassium glass (n=24)		
Silica (SiO2)	46.381 + / - 8.298	66.359 + / - 7.035	59.96	0.000 * * *
Sodium oxide (Na2O)	2.232 + / - 2.851	0.973 + / - 1.361	3.339	0.076 *
Potassium hydroxide (K2O)	0.266 + / - 0.403	5.279 + / - 5.248	11.672	0.002 * * *
(CaO CaO)	1.409 + / - 1.579	3.168 + / - 3.136	3.566	0.067 *
Magnesium oxide (MgO style)	0.652 + / - 0.55	0.878 + / - 0.629	1.191	0.283
Alumina (Al2O3)	3.671 + / - 1.707	5.723 + / - 3.604	3.727	0.062 *
Iron oxide (Fe2O3)	0.707 + / - 0.992	1.375 + / - 1.555	1.957	0.171
Copper oxide (CuO)	2.013 + / - 2.396	1.607 + / - 1.536	0.395	0.534
Lead oxide (PbO)	26.422 + / - 8.395	7.075 + / - 8.305	45.42	0.000 * * *
Baryta (BaO)	11.379 + / - 6.397	2.762 + / - 3.412	28.884	0.000 * * *
Phosphorus pentoxide (P2O5)	1.452 + / - 2.358	0.938 + / - 1.169	0.794	0.379
Strontium oxide ((SrO) SrO)	0.341 + / - 0.288	0.093 + / - 0.113	14.04	0.001 * * *
Tin oxide (SnO2)	0.065 + / - 0.158	0.108 + / - 0.482	0.098	0.756
Sulfur dioxide (SO2)	0.0 + / - 0.0	0.203 + / - 0.749	0.945	0.338

Note: ***, ** and * represent the significance levels of 1%, 5% and 10%, respectively.

The classification rule of high-potassium glass and lead-barium glass can be obtained by comparing and analyzing the proportion of main chemical components in the charts after weathering and before weathering: 1. The distribution law of main components: silica occupies the largest proportion in both high-potassium glass and lead-barium glass.In the high potassium glass, the proportion of sodium oxide and potassium oxide is relatively large before weathering.In the lead-barium glass, the proportion of lead oxide and barium oxide is relatively large.2. The change law of main components: potassium oxide (K) is the component of high potassium glass before and after weathering$_2$The proportion of O decreased significantly, and the proportion of lead oxide (PBO) increased significantly before and after weathering, while the proportion of barium oxide (BaO) did not change much [9].

3.2. Results of principal component analysis

On the premise that the characteristic root of principal components >1 and the explanatory contribution rate of variables >85%, five components were selected as principal components. The variance interpretation table of principal component analysis of high potassium glass by SPSS software is shown in Table 3.

Table 3. Total variance explanation table.

composition	Characteristics of the root		
	Characteristics of the root	Variance explanation rate (%)	Cumulative variance explained rate (%)
1	5.400	38.573	38.573
2	2.430	17.359	55.932
3	1.715	12.253	68.185
4	1.649	11.777	79.962
5	0.949	6.775	86.737
6	0.644	4.601	91.339
7	0.513	3.666	95.004
8	0.302	2.16	97.164
9	0.195	1.392	98.557
10	0.111	0.795	99.352
11	0.061	0.435	99.787
12	0.018	0.129	99.916
13	0.007	0.051	99.967
14	0.005	0.033	100

Five components were selected as principal components on the premise that the characteristic root of principal components >1 and the explanatory contribution rate of variables >85% as far as possible. Through factor analysis, SPSS software was used to calculate the factor loading coefficient table as shown in Table 4.The composition matrix table of high potassium glass is shown in Table 5.

Table 4. Factor loading coefficient table of high potassium glass.

	Principal component 1	Principal component 2	Principal component 3	The principal components of 4	The principal components of 5	
Silica (SiO2)	0.820	0.415	0.124	0.176	0.073	0.897
Sodium oxide (Na2O)	0.075	0.776	0.082	0.462	0.318	0.930
Potassium hydroxide (K2O)	0.680	0.507	0.285	0.260	0.164	0.895
(CaO CaO)	0.621	0.607	0.036	0.227	0.141	0.827
Magnesium oxide (MgO style)	0.741	0.330	0.332	0.147	0.122	0.805
Alumina (Al2O3)	0.860	0.005	0.004	0.093	0.180	0.780
Iron oxide (Fe2O3)	0.833	0.163	0.006	0.234	0.319	0.877
Copper oxide (CuO)	0.530	0.069	0.277	0.603	0.167	0.754
Lead oxide (PbO)	0.423	0.227	0.663	0.304	0.323	0.867
Baryta (BaO)	0.583	0.356	0.632	0.014	0.298	● 0.954
Phosphorus pentoxide (P2O5)	0.680	0.560	0.095	0.037	0.392	0.940
Strontium oxide ((SrO) SrO)	0.730	0.506	0.043	0.247	0.035	0.853
Tin oxide (SnO2)	0.069	0.322	0.453	0.586	0.433	0.845
Sulfur dioxide (SO2)	0.301	0.196	0.608	0.582	0.289	0.921

Table 5. Composition matrix table of high potassium glass.

The name of the	Table of Composition Matrix composition				
	Component 1	Component 2	Component 3	Composition of 4	Composition of 5
Silica (SiO_2)	0.152	0.171	0.072	0.107	0.077
Sodium oxide (Na_2O)	0.014	0.32	0.048	0.28	0.335
Potassium hydroxide (K_2O)	0.126	0.209	0.166	0.158	0.173
(CaO CaO)	0.115	0.25	0.021	0.137	0.149
Magnesium oxide (MgO style)	0.137	0.136	0.193	0.089	0.129
Alumina (Al_2O_3)	0.159	0.002	0.002	0.056	0.189
Iron oxide (Fe_2O_3)	0.154	0.067	0.004	0.142	0.336
Copper oxide (CuO)	0.098	0.029	0.162	0.366	0.176
Lead oxide (PbO)	0.078	0.093	0.386	0.185	0.341
Baryta (BaO)	0.108	0.147	0.368	0.009	0.314
Phosphorus pentoxide (P_2O_5)	0.126	0.23	0.055	0.023	0.413
Strontium oxide ((SrO) SrO)	0.135	0.208	0.025	0.15	0.037
Tin oxide (SnO_2)	0.013	0.132	0.264	0.355	0.457
Sulfur dioxide (SO_2)	0.056	0.081	0.355	0.353	0.305

(1) Principal component 1: the representative level is basically consistent, and only negatively correlated with silicon dioxide and tin oxide

(2) Principal component 2: represents phosphorus pentoxide, tin oxide, and is positively correlated represents sodium oxide, potassium oxide, and negatively correlated

(3) Principal component 3: represents tin oxide, sulfur dioxide, and is positively correlated represents lead oxide, barium oxide, and negatively correlated

(4) Principal component 4: represents sodium oxide, tin oxide, and is positively correlated represents copper oxide, sulfur dioxide, and negatively correlated

(5) Principal component 5: represents lead oxide, barium oxide, tin oxide, sulfur dioxide, and is positively correlated represents sodium oxide, iron oxide, phosphorus pentoxide, and negatively correlated

The weight table of each factor of high potassium glass is shown in Table 6. [10]

Table 6. Weight table of each factor of high potassium glass.

The name of the	Variance explanation rate (%)	Cumulative variance explained rate (%)	Weight (%)
Principal component 1	38.573	38.573	44.471
Principal component 2	17.359	55.932	20.014
Principal component 3	12.253	68.185	14.126
The principal components of 4	11.777	79.962	13.578
The principal components of 5	6.775	86.737	7.811

3.3. Reasonableness Analysis

Next, KMO test and Barlett test were used to test the rationality of the subclass classification results of glass by SPSS software

Note: ***, ** and * respectively represent the significance level of 1%, 5% and 10%, where P<0.05 is significant, and Bartlett test can be used for principal component analysis.

The KMO test and Bartlett test of high potassium glass are shown in Table 7.

Table 7. KMO test and Bartlett's test for high potassium glass.

KMO value	0.367	
Bartlett's test for sphericity	The approximate chi-square	265.919
	df	91.000
	P	0.000 * * *

Note: ***, ** and * represent the significance levels of 1%, 5% and 10%, respectively

P<0.05 is significant, and Bartlett test can be used for principal component analysis.

In summary, both lead-barium glass and high-potassium glass have passed Bartlett test, and only one KMO test and Bartlett test can be passed. Therefore, principal component analysis can be conducted for them, which means that the classification results are reasonable.

3.4. Sensitivity analysis of the model

For the analysis of high-potassium glasses, according to the composition matrix table, the following results can be obtained:

F = x + F1 (0.386/0.867) (0.174/0.867) * F3 + F2 + (0.123/0.867) (0.118/0.867) * F5 F4 + (0.068/0.867)

Clean F = 0.445 + 0.201 * * F1 F2 + 0.078 + 0.142 + 0.136 * * F3 F4 * F5

The control variable method was used, and the other variables were regarded as constants. One F was analyzed separately, and the slope was used to show the sensitivity.The coefficient before F1 is very large, so F1 is very sensitive to F;The coefficient before F5 is close to zero, which has little effect on F value, so the sensitivity is very weak.The coefficients before F2,F3 and F4 belong to the mean value, and their sensitivity to F is general.

4. Conclusion

In this paper, we study the ancient glass component analysis and category to identify problems, through the establishment of principal component analysis, cluster analysis and multiple regression model mathematical model, using SPSS software, realized by analyzing the chemical composition content on the glass before and after weathering type classification problem and for different categories of the relationship between the chemical composition of glass sample problem solving, According to the change rate, the chemical composition content before weathering is predicted, and the predicted value before weathering is obtained.The principle of the model applied in this paper is simple, the implementation is easy, the convergence speed is fast, and the interpretation degree of the algorithm is strong. When dealing with large data sets, the algorithm is relatively scalable and efficient

References

[1] Votto Martina, Fasola Salvatore, Cilluffo Giovanna, Ferrante Giuliana, La Grutta Stefania, Marseglia Gian Luigi, Licari Amelia. Cluster analysis of clinical data reveals three pediatric eosinophilic gastrointestinal disorder phenotypes[J]. Pediatric Allergy and Immunology,2022,33(2).

[2] Hatano Yutaka, Morita Tatsuya, Mori Masanori, Maeda Isseki, Oyamada Shunsuke, Naito Akemi Shirado, Oya Kiyofumi, Sakashita Akihiro, Ito Satoko, Hiratsuka Yusuke, Tsuneto Satoru. Complexity of desire for hastened death in terminally ill cancer patients: A cluster analysis[J]. Palliative and Supportive Care,2021,19(6).

[3] Zhipeng WANG, Bo LIN. Research on Credit Decision Issues of the Small and Medium-Sized Enterprises Based on TOPSIS and Hierarchical Cluster Analysis[C]//.Proceedings of the 2021 International

Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS 2021).,2021:1134-1144.

[4] Dan Wu, Chenhui Ji. Study on the Correlation Analysis of Energy Consumption in Economic Industries and Evaluation of Carbon Emission Decoupling in Beijing-Tianjin-Hebei Region[C]//. Proceedings of the first International Symposium on innovation management and Economics (ISIME 2021).,2021: 175-189.DOI: 10.26914/c.cnkihy.2021.040684.

[5] Xiyuan Su. Comprehensive evaluation and classification of urban economic benefits under typical correlation analysis[C]//. Proceedings of 2021 International Conference on Intelligent Transportation,Big Data & Smart City(ICITBS 2021)PartIII.,2021:208-211.

[6] ZHOU Jingkun, TIAN Juan, PAN Pengfei. The Interactive Relationship between the Atmospheric Input and Output of Major Cities in China Based on Canonical Correlation Analysis[C]//. Proceedings of The 8th Academic Conference of Geology Resource Management and SustainableDevelopment.,2020:484-491.

[7] Blazowski Lukasz, Jerzynska Joanna, Kurzawa Ryszard, Kuna Piotr, Majak Pawel. Cluster analysis of 505 real-life food-induced anaphylaxis in children reveals two stable clinical phenotypes. [J]. Allergy,2020,76(5).

[8] Keenan Tiarnán D.L. The Hitchhiker’s Guide to Cluster Analysis: Multi Pertransibunt et Augebitur Scientia[J]. Ophthalmology Retina,2020,4(12).

[9] Calo Cristina Marilin, Rizzutto Marcia A, CarmelloGuerreiro Sandra M,Dias Carlos S B,Watling Jennifer,Shock Myrtle P,Zimpel Carlos A,Furquim Laura P,Pugliese Francisco,Neves Eduardo G. A correlation analysis of Light Microscopy and X-ray MicroCT imaging methods applied to archaeological plant remains' morphological attributes visualization. [J]. Scientific reports,2020,10(1).

[10] Engineering; Reports from Science University of Malaysia Describe Recent Advances in Engineering (Canonical Correlation Analysis Feature Fusion with Patch of Interest: A Dynamic Local Feature Matching for Face Sketch Image Retrieval) [J]. Journal of Engineering,2020.