Analysis and identification of ancient glass artifacts based on their chemical composition

Haonan Yu*, Hao Cheng, Wanting Zhan

Fujian Normal University, Fuzhou, China

*Corresponding author: 1269878150@qq.com

Abstract: The Silk Road led to the development of ancient glass products, and common ancient glass products include high potassium glass and lead-barium glass. Ancient glass is susceptible to weathering by elemental exchange with the surrounding environment, resulting in changes in internal chemical composition and appearance. In this paper, the correlation and influence of relevant factors on the degree of weathering on the surface of glass artifacts are fully analyzed, and subclasses and unknown types are identified based on their chemical composition, while the necessary sensitivity analysis is conducted for each of these results, and finally, correlations between chemical compositions and differences in correlations are presented. This study can guide for glass heritage workers to help them better carry out their conservation work.

Keywords: Chemical composition; correlation; glass artifacts; degree of weathering.

1. Introduction

Because of the Silk Road West Asia and Egypt region bead-shaped jewelry spread, and our ancient glass absorbed its shape technology for production. And the chemical composition of our ancient glass is different from it. The main chemical composition of ancient glass products is silica[1]. The chemical composition of different glass products, and the forging process require different auxiliary additives. Due to the need to reduce the melting point when calcining quartz sand, commonly used grass ash, natural alkali, and other co-solvents as well as the addition of limestone as a stabilizer to assist in refining; lead barium glass chooses lead oxide, barium oxide with a high content of lead ore as a flux; potassium glass is a high content of potassium grass ash, etc. as a flux [2-3].

As ancient glass is very susceptible to weathering by the surrounding environment. The exchange of elements between the environment and the interior, leads to changes in the ratio of internal chemical composition, which eventually makes it difficult to judge the category. The specific external change is reflected in the color change, while the substantial change lies in the change of its chemical composition content [4-6].

In this paper, we analyze the relationship between weathering and the type, decoration, and color of the surface of the given glass artifacts based on the classification information and the ratio of the main components of the two glass types, high potassium, and lead-barium, and analyze the statistical pattern of the weathering of the surface of the glass artifacts, i.e., their chemical content, based on the glass type, and then predict the chemical content before weathering based on the given data. The classification pattern of high potassium and lead-barium glass is analyzed and each category is divided into subcategories based on the appropriate chemical composition, and specific classification methods and results are proposed and analyzed for reasonableness and sensitivity of the results. Analyze the chemical composition of unknown categories of glass products, identify their types, and then analyze the sensitivity of the classification results [7-9]. The correlations between the chemical compositions of each glass sample are analyzed according to the categories, and the differences in the correlations between the compositions of different categories are compared [10].

2. Methodology

2.1 Spearman's correlation coefficient and chi-square test

In this paper, we establish a prediction model based on correlation analysis and chi-square test to analyze the correlation between surface weathering and its type, decoration, and color, and to give

the statistical pattern of its chemical content, to predict their chemical content before weathering, based on data related to glass artifacts. The correlation is solved using Spearman's correlation coefficient and chi-square test; furthermore, the chemical content data are visualized by histograms, and the statistical patterns are observed; finally, the prediction results are obtained by multiplying the mean values of the chemical content with and without weathering back to the original data after making a ratio and normalizing them.

Correlation analyticity is a statistical analysis method to study the correlation between two or more random variables in equal status, the main body of which is the analysis of the overall with the sign of causality, it is the process of describing the closeness of the relationship between objective things and expressing it with appropriate statistical indicators. Based on the problem analysis, it was determined that the relatively more inclusive Spearman's correlation coefficient was used for correlation analysis. The Spearman's correlation coefficient is a rank statistical parameter of non-parametric nature (independent of the distribution), an indicator describing the existence of the same or opposite convergence between two sets of variables, which is of a better nature since the test does not require the assumption of obedience to a normal distribution and is obtained by determining only the rank of the variables at each point.

It, like the correlation coefficient r, takes values between -1 and +1, the difference being that it is calculated based on rank. Thus, the definition is given: X and Y are two sets of data and their Spearman (rank) correlation coefficients.

$$r_s = 1 - 6\sum \frac{d_i}{n(n^2-1)} \tag{1}$$

where n denotes the number of data and di denotes the difference between the two data orders, i.e.

$$d_i = rg(X_i) - rg(Y_i) \tag{2}$$

The chi-square test is the degree of deviation between the actual observed value of a statistical sample and the theoretical inferred value. The degree of deviation between the actual observed value and the theoretical inferred value determines the size of the chi-square value; the larger the chi-square value, the more incompatible it is, and the smaller the deviation, the smaller the chi-square value, the more it tends to comply, and if the quantities are exactly equal, the chi-square value is 0, indicating that the theoretical value is fully compatible. The chi-square test is often used to test the degree of association between two variables.

2.2 Clustering evaluation

Using visual analysis, the data were converted into bar charts and then compared to analyze the classification patterns of high potassium and lead-barium glass, and further subclassification of the glass using k-clustering analysis, and subsequent evaluation of the validity of the clusters by field variability analysis and interclass distances within the sample classes to analyze their reasonableness and sensitivity.

The k-cluster analysis comes from the vector quantization method in signal processing, and the goal is to assign n points (which can be a single observation or an instance of a sample) to k clusters, so that each point belongs to the cluster corresponding to the mean (that is, the cluster center) closest to it, and use it as the principle of clustering.

The principle of k-means cluster analysis: First, k points are randomly selected as the initial cluster centers, then the distance from each data object to each cluster center is calculated, and the data object is assigned to the class in which the nearest one is located; the new class after adjustment calculates the new cluster center, and if there is no change in the cluster centers of the two adjacent times, it means that the data object adjustment is finished and the clustering criterion function f has converged. In each iteration, the correct or incorrect classification of each sample is considered, and if there is an error, it is adjusted. After all the data are adjusted, the clustering center is then modified and the next iteration is entered. If all the data objects are correctly classified in one iteration of the algorithm, there will be no adjustment and the clustering centers will not change again, which means that f has converged and the algorithm ends at this point. This is essentially evaluating the error and modifying the model until the required error range is reached.

2.3 Multiple linear regression

The sensitivity analysis model based on multiple linear regression was established by identifying the types of unknown types of glass artifacts and analyzing their sensitivity by calculating the sensitivity factors.

A multiple linear regression model is used to study the linear relationship between a dependent variable and multiple independent variables, when the variables and the dependent variable have a certain linear relationship, then a multiple linear models can be established to analyze, multiple linear regression, which is the diversification of independent variables based on linear regression, is used to describe the stochastic linear relationship between multiple variables and y.

2.4 Correlation analysis

To analyze the correlations related to the chemical composition of glass artifacts, the subcategory data were integrated with the numbering, etc. The 14 chemical components were sorted in order of concentration to find the basis of classification, and the classification was labeled with the chemical composition relationships. Then the mean values were taken for the subcategories, and the variability of the data of nine different categories was explored by the chi-square test.

Gray correlation reflects the degree of association between data and standard data through gray correlation degree, so gray correlation analysis is the basis of gray system analysis, prediction, and decision making. Its specific operation steps are as follows.

(1) Determine the reference series and the series to be compared.

(2) dimensionless processing of the data.

(3) Find the gray correlation coefficient of the reference series and the comparison series.

(4) The degree of correlation is the degree of difference in geometry between the curves.

The correlation coefficient can also be simplified as the following equation.

$$\xi_{oi} = \frac{\Delta(min) + \rho\Delta(max)}{\Delta_{oi}(k) + \rho\Delta(max)} \tag{3}$$

3. Results and discussion

3.1 Results

Using significance tests (i.e., hypothesis tests), the data of each indicator were tested for correlation coefficients, and the final correlation coefficient results are shown in Table 1.

Table 1 Correlation coefficient results

	Color	Degree of weathering
color	1.000(0.000***)	-0.115(0.406)
Degree of weathering	0.115(0.406)	1.000(0.000***)

There is no significant difference in color, and since its correlation coefficient is negative, there is some negative correlation between weathering and type. Since the selected indicator variables were unordered, the chi-square test was used again. The results of the chi-square test analysis showed that, based on weathering and color, the significant P value was 0.507, which did not show significance at the level, and the original hypothesis was accepted, so there was no significant difference between weathering and color data; based on weathering and type, the significant P was 0.009***, which showed significance at the level, and the original hypothesis was rejected, so there was a significant difference for weathering and type data For the weathering and ornamentation data, the significant P value was 0.084*, which was not significant at the level of the original hypothesis.

The results of the above correlation analysis and chi-square test present that there is a significant difference between the degree of weathering of the surface of glass artifact samples and their types, while there is no significant difference between their ornamentation and color. According to the correlation coefficient analysis, there is a negative correlation between the degree of surface

weathering and the type of glass artifacts. This suggests that the main factor affecting the degree of surface weathering of glass artifacts is the chemical composition of the glass. From the background of the problem, it is clear that weathering of the surface of glass artifacts causes changes in their internal chemical composition, and weathering is related to the chemical exchange between the glass artifacts and their surrounding buried environment, and different production processes result in different chemical compositions, so the resulting chemical exchange is equally different, thus making the degree of weathering different. This further confirms the accuracy of our data and the reasonableness of our conclusions.

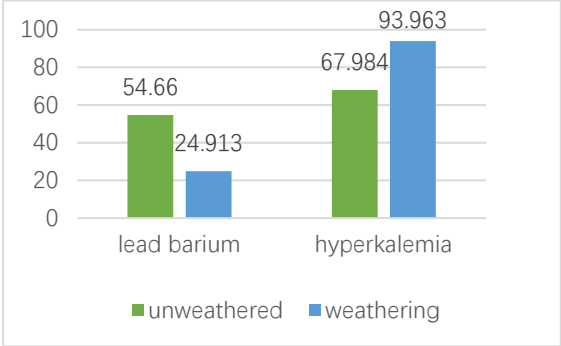


Figure 1 Silicon dioxide content

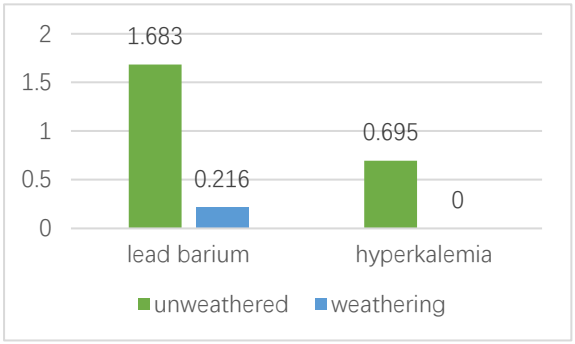


Figure 2 Sodium oxide content

Figures 1 and 2 show the silica content and sodium oxide content, respectively. The high potassium glass produces an enrichment of silica and a significant decrease of sodium oxide, potassium oxide, calcium oxide, magnesium oxide, aluminum oxide, iron oxide, and copper oxide before and after weathering, while the lead barium produces a significant decrease of silica and sodium oxide and a significant increase of sulfur dioxide, phosphorus pentoxide, lead oxide, calcium oxide, and copper oxide before and after weathering. It is also possible to compare the differences in weathering changes between high potassium and lead barium, especially reflected in silica, copper oxide, lead oxide, and calcium oxide elements.

Table 2 Analysis of variance

Chemical composition content	P	Differences
SiO_2	0.000***	Significant differences exist
Na_2O	0.242	There is no significant difference
K_2O	0.657	There is no significant difference
CaO	0.001***	Significant differences exist
MgO	0.102	There is no significant difference
(Al_2O_3	0.247	There is no significant difference
Fe_2O_3	0.243	There is no significant difference
CuO	0.000***	Significant differences exist
PbO	0.000***	Significant differences exist
BaO	0.000***	Significant differences exist
P_2O_5	0.329	There is no significant difference
SrO	0.007***	Significant differences exist
SnO_2	0.025**	Significant differences exist
SO_2	0.055**	There is no significant difference

Using the glass type as a classification criterion, the chemical composition content before weathering was predicted. First, the four mean values of high potassium glass and lead-barium glass with and without differentiation were obtained separately from the available data, and the three severely differentiated lead-barium data were removed. Then, the ratio was calculated, i.e., a set of mean values before differentiation and a set of mean values after differentiation. Then, the ratio is multiplied back by the original data. It is equivalent to weighting the weathered data, at this time, the

chemical content should be normalized and then multiplied by the form of a percent so that the chemical composition is still 100% as a benchmark, and thus, the prediction results are obtained.

The ANOVA results are shown in Table 2.

Regressions were performed separately for weathered and unweathered data sets, and the regressions for unweathered will be presented below; the regression steps are similar, so they are not shown. By analyzing the F-value, it is analyzed whether it can significantly reject the original hypothesis that the overall regression coefficient is 0 ($P<0.05$), and if it is significant, it indicates that there is a linear relationship between them. The model fit is analyzed by the R^2 value, and the VIF value is also analyzed, and if the model presents covariance then the covariance element is removed, and by analyzing it, silica is removed and analyzed again.

The analysis of the results of the F-test can be obtained that the significance P-value is 0.000***, which presents significance at the level and rejects the original hypothesis that the regression coefficient is 0. Therefore, the model meets the requirements. For the performance of variable co-linearity, VIF is all less than 10, so the model has no problem with multiple co-linearity and the model is well constructed. The equation of the model is as follows:

$$y = 1.531 + 0.014 \times Na_2O - 0.032 \times K_2O - 0.02 \times CaO - 0.021 \times MgO + 0.013 \times Al_2O_3 - 0.008 \times Fe_2O_3 - 0.047 \times CuO + 0.014 \times PbO + 0.02 \times BaO - 0.008 \times P_2O_5 - 0.073 \times SrO - 0.103 \times SnO_2 + 0.052 \times SO_2 \quad (3)$$

Similarly, we can obtain the multiple linear regression equation after weathering as

$$y = 0.86 + 0.08 \times Na_2O + 0.073 \times K_2O - 0.013 \times CaO - 0.005 \times MgO + 0.042 \times Al_2O_3 + 0.014 \times Fe_2O_3 + 0.019 \times CuO + 0.016 \times PbO + 0.022 \times BaO + 0.026 \times P2O5 - 0.17 \times SrO + 0.07SnO_2 - 0.11 \times SO_2 \quad (4)$$

The data divided into subcategories in question 2 were integrated with the numbering, etc., and the data were divided into high potassium and lead-barium groups, and by sorting, the groups were obtained as the following categories by chemical concentration: ancient high potassium glasses were further divided into 1 - no potassium (n-K), 2 - high potassium (m-K-Al) and 3 - high aluminophosphates (m-Al-P) based on the concentrations of K2O, SrO, CaO, and SiO2 in potassium glasses m indicates that the substance exists in the form of oxide, and n indicates that the substance exists in the form of ion, in which the content is given as the standard of oxide. Further classification of ancient lead-barium glass into 1-high silicon (m-Si), 2-low silicon and high lead (m-Pb), 3-medium silicon (m-Si), 4-low silicon and high barium (m-Ba-Cu) according to the concentration of CuO, BaO, PbO and SiO2 in the lead-barium glass.

The subcategories of the second question were taken as the mean values and subjected to gray correlation analysis with the mean values of the overall, using the mean-measure treatment, and

taking correlation coefficients for determining the difference values.

Among them, the correlation coefficients of the lead-barium treatment are shown in Table 3.

From the table 3 data, we can see that there is a big difference between the first category and potassium oxide, the second category and potassium oxide, the third category and aluminum oxide, phosphorus pentoxide and silicon dioxide, the fourth category and silicon dioxide, potassium oxide, calcium oxide, and the fifth category and silicon dioxide.

Table 3 Correlation coefficient results

Clustering type	1	2	3	4	5
SiO_2	0.55593628	0.84166776	0.57366041	0.50632217	0.33336187
Na_2O	0.93929609	0.93929609	0.93929609	0.8854732	0.93929609
K_2O	0.43361678	0.66989873	0.79451445	0.70288082	0.77983568
CaO	0.91090027	0.91147939	0.72814863	0.73508924	0.62045521
MgO	0.94623328	0.92465933	0.90307419	0.98250771	0.95928282
Al_2O_3	0.87494296	0.77942007	0.64742013	0.88655379	0.69660333
Fe_2O_3	0.97653781	0.9818907	0.76062914	0.98276729	0.78616683
CuO	0.84304172	0.99671543	0.97727543	0.98966799	0.77160222
PbO	0.99178226	0.96328408	1	0.96662016	0.97951906
BaO	0.96517746	0.9474205	0.90261685	0.98942569	0.9474205
P_2O_5	0.90940569	0.93305797	0.71198024	0.90457874	0.945361
SrO	0.99621114	0.99621114	0.98810834	0.99929609	0.99902149
SnO_2	0.98186294	0.98186294	0.98186294	0.98186294	0.87079489
SO_2	0.99053065	0.98854699	0.99053065	0.99856072	0.99053065

3.2 Discussion

A chi-square test was used to analyze the variance of the data, which only counts whether the word is in the document or not, regardless of the number of occurrences. This would make him unreasonably exaggerate the role of low-frequency words, and at the same time, when using the mean ratio for weighting, it did not take into account that the weathered glass elements might disappear due to weathering, and as weighting could not assign a value to the 0 value.

In the process of promoting data visualization, although the bar chart can achieve the effect of intuitively reflecting the data pattern, because the classification is based on chemical composition, and the chemical composition of glass artifacts is 14 kinds, so if a bar chart is made for each chemical composition, even if the data is visualized, because there are multiple independent bar charts finally obtained, it will be complicated to observe and compare, and it is not easy to understand at a glance.

Although the data showed good linearity, there were weak nonlinearities in the data, and the inability to participate in the linear regression due to the covariance of silica may also affect the judgment of the type of discrimination.

When using gray correlation analysis, there is no exact parent series, and only mean substitution can be used, which may lead to some errors.

4. Conclusions

By correlation analysis, data dimensionality reduction can be achieved, and the speed and accuracy of prediction can be improved at the same time. Based on this theory, to make the accuracy of the relationship analysis among the variable indicators, this question uses the Spearman method for correlation analysis. At the same time, the cardinality test is used to determine whether there is a significant difference between two grouped variables to improve the accuracy of prediction. Next, a bar chart was chosen to visualize the data and present the statistical patterns clearly and visually. Finally, the mean-ratio calculation was used for prediction before differentiation, taking into account the totality of the data and increasing the degree of data availability.

By using k-type cluster analysis as a means of empirical research, the principles and indicators of classification were improved, the practice of quantitative classification of subclasses was carried out, and the chemical composition of glass artifacts was divided into subclasses based on the original soil subclasses, to avoid the irrationality caused by "insufficient basis" or originality in solving the problem.

Multiple linear regression was used for prediction, and sensitivity analysis was performed to enhance the accuracy of the results.

By first determining the classification basis through concentration ranking, the correlation coefficients between the subcategories of high potassium and lead-barium and the overall mean values were calculated by using gray correlation analysis in the variance analysis, reflecting the variance of the classification and corroborating the classification basis by the correlation coefficients together.

References

[1] REN Yubo,WEN Rui,XIAN Yiheng,ZHANG Cuimin. Chemical compositions and manufacturing techniques of glass ear pendants unearthed from Yingchengzi Cemetery of the Han Dynasties[J]. Sciences of Conservation and Archaeology,2021,34(3):45-54.

[2] LU Xuan, LIU Zixin, HE Daxin, LIU Xiang. Hejia Village Tang Dynasty gem and glass bowl detection report[J]. Archaeology and Cultural Relics, 2017(6):114-120.

[3] ZHOU jing. A general view of the obsession of Chinese lead and barium glass art[J]. Art Panorama, 2021(3):184-186.

[4] CUI Tianxing,ZHANG Jihua,HAN Guohe. Song glass ornaments unearthed from Nanwa Site in Dengfeng[J]. Sciences of Conservation and Archaeology,2022,32(2):51-56

[5] HU Zhi-zhong,LI Pei,JIANG Lu-man,WANG Tong-yang,DU Gu,YANG Bo. Application of LA-ICP-MS in the Analysis of Archaeological Glass and Source Discrimination[J]. Rock and Mineral Analysis, 2020, 39(4):505-514.

[6] GU Liangliang,ZHOU Jing. Some Problems in the Study of Ancient Chinese Glass Art[J]. Art and Design,2021,2(4):112-114.

[7] CHENG Xue-lei,LI Wen-dong,HAI Ran,LIU Yan,HE Xian-feng,XIA Jin-hong. Parametric sensitivity simulation analyses of subway station structure surrounded by saturated soft soil foundation under earthquake [J]. Journal of Guangxi University (Natural Science Edition), 2022,47(1):92-102.

[8] OUYANG Zilu,WANG Hongdong,WANG Jianyao,YI Hong. Sensitivity analysis method of ship hydrodynamic derivatives based on intelligent control[J]. Chinese Journal of Ship Research,2021,16(4):116-124.

[9] FU Qiang,KUANG Guirong,L Liangbo,MO Huixuan,LI Qinghui,GAN Fuxi. Nondestructive Analysis of the Glass Artifacts of the Han Dynasties Excavated from Guangzhou[J]. Journal of Ceramics,2013(7):994-1003

[10] FANG Xiangzhong. Application conditions of chi-square test[J]. China Statistics,2022(5):29-31