A study on the composition classification of ancient glass products based on factor analysis

Xin Luo *, #, Ao Hu #, Zezheng Wang #

School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei, China, 230601

* Corresponding author: greenparty_1234567@163.com

#These authors contributed equally.

Abstract. In the process of ancient Chinese and foreign technology and cultural exchanges, the Silk Road has an important value. Glass artifacts, as goods on the Silk Road, are widely distributed and have important significance for people to reveal the law of social development. In order to study the classification characteristics of high potassium glass and lead-barium glass from the perspective of chemical composition content, and to classify the two kinds of glass into subclasses on this basis, this paper establishes a factor analysis model and applies a factor analysis algorithm to classify the variables by dimensionality reduction; furthermore, the factor rotation is performed on the reduced variable family to further explain what kind of original variables the new variables represent, so as to explain the classification method. Finally, the sensitivity test with the addition of dynamic Gaussian white noise is used to analyze the sensitivity of the model.

Keywords: Factor Analysis, Dynamic Gaussian White Noise, Glassware.

1. Introduction

Glass, over thousands of years, has gone from rare to common to playing an important role in various industries today. Among these, silicate glass is again numerous and widely used. By studying the weathering and subclassification of ancient silicate glass, this work hopes to make some pioneering studies for contemporary research on the prevention of weathering, application and digitization of silicate glass.

Firstly, weathering, ancient glass is highly susceptible to weathering by the burial environment. Liangzhi Zhou [1] applied modern research methods to study the main factors affecting the weathering of silicate glass. The in-depth study on the mechanism of weathering conducted by Wang Chengxuan et al [2] has some implications for the work on composition prediction. Next is the classification. Although Zhao Fengyan et al [3] performed nondestructive analysis of the chemical composition of glassware by pXRF for classification, the existing chemical research tools cannot accurately classify the glass according to its composition in a more accurate and reasonable way. Therefore, the introduction of machine learning is considered to abstract the actual chemical problems into mathematical and theoretical models by using classification prediction models and intelligent algorithms to solve them without the constraints of physical samples and high-precision analytical instruments during analysis.

Although intelligent algorithms have been widely used in scientific research in the field of materials in recent years, such as the construction and application of database of ancient glass beads excavated in China studied by Bailing Feng [4], the main theoretical basis, simulation process and application status of each simulation method summarized by Liyan Zhang et al [5] using seven simulation methods of glass compositional properties and the glass defect detection done by Jiangang Li [6] based on deep learning, but there are many studies at home and abroad using machine learning to study However, there is still a gap in using machine learning to study the weathering and subclassification of ancient silicate glass at home and abroad.

In this paper, we use "factor analysis" to classify high potassium glass and lead-barium glass: the main factors affecting the classification are selected by factor dimensionality reduction, and then the classification basis is determined by factor rotation, which avoids the problem that traditional machine learning methods, such as k-means [7], cannot explain the newly classified factors. Finally, subspecies are classified according to the analysis results, and the validity of the model is tested by performing sensitivity tests on the model.

2. High potassium-lead barium glass classification

In this paper, we aim to analyze the classification basis of high potassium glass and lead-barium glass according to the chemical composition of the surface of the artifacts, and to subcategorize each category. First, if the variables are to be classified, they need to be downscaled. Since the meaning of the new variable group obtained after dimensionality reduction of variables is often not easily interpreted. Therefore, in order to clarify the basis of classification and elucidate the meaning of new variable groups, the article applies factor rotation in factor analysis algorithm to solve the problem and selects factor analysis algorithm for modeling. Finally, it is enough to perform sensitivity analysis on the results of subclass classification to prove the accuracy of the classification results. The flow chart is shown in Figure 1.

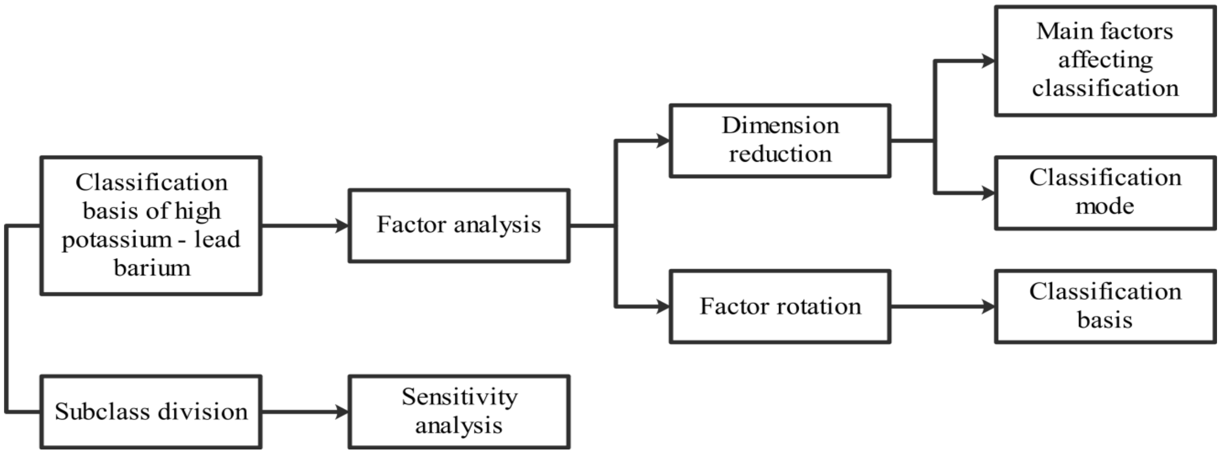

Figure 1. Classification flow chart

2.1. Subspecies division

2.1.1. Bartlett's spherical test

Bartlett's sphericity test [8] is a method to test the degree of correlation between individual variables. Before doing factor analysis, the correlation between variables needs to be judged, and if the correlation is too poor, then factor analysis is not suitable. Let the original hypothesis H_0 indicate that the correlation coefficient matrix is a unit array, which is not suitable for factor analysis, and the alternative hypothesis H_1 indicate that the correlation coefficient matrix is not a unit array, which is suitable for factor analysis. The results of Bartleet test for high potassium glass and lead-barium glass were obtained by SPSS calculation, as shown in Table 1.

Table 1. Bartleet test results

	High potassium glass	Lead barium glass
chi-square	265.919	416.381
freedom	91	91
P-value	0.003	0.005

According to the Bartleet test results in Table 1, the P-value of both high potassium glass and lead-barium glass is less than or equal to 0.01. This indicates that the original hypothesis is rejected at 99% confidence level, i.e., both high potassium glass and lead-barium glass can satisfy the prerequisites of factor analysis.

2.1.2. Factor reduction

The mean value of the vector of chemical components $x=(x_1,x_2,\ldots,x_{14})'$ is expressed as $u=(u_1,u_2,\ldots,u_{14})'$, and the common factor $f=(f_1,f_2,\ldots,f_m)$, m is the number of common factors obtained by dimensionality reduction. The relationship between the vector of chemical components and the common factor is modeled by the equation (1):

$$\begin{cases} x_1 = u_1 + a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 \\ x_2 = u_2 + a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2 \\ \quad \vdots \\ x_{14} = u_{14} + a_{14,1}f_1 + a_{14,2}f_2 + \cdots + a_{14,m}f_m + \varepsilon_{14} \end{cases} \tag{1}$$

$\varepsilon_i=(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_{14})$ is called the special factor and $A_{14\times m}=a_{i,j}, i=1,2,\ldots,14, j=1,2,\ldots m$ is called the factor loading matrix. Since the common factor appears f in each of the original variables, the common factor can be interpreted as some characteristics shared by the original fourteen chemical substances; since the special factor appears only in the corresponding original variables, it can be considered as acting only on the corresponding original variables.

The number of common factors after dimensionality reduction is determined in this model by using a fragmentation diagram [9], and the fragmentation diagrams of high potassium glass and lead-barium glass are shown in Figure 2 and Figure 3.

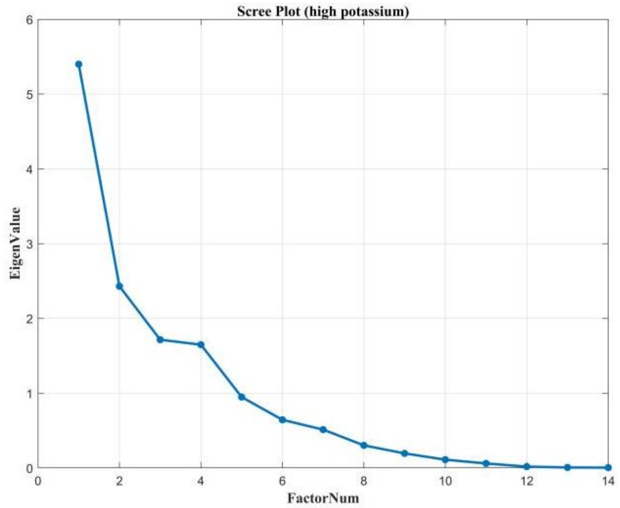

Figure 2. High potassium glass lithotripsy

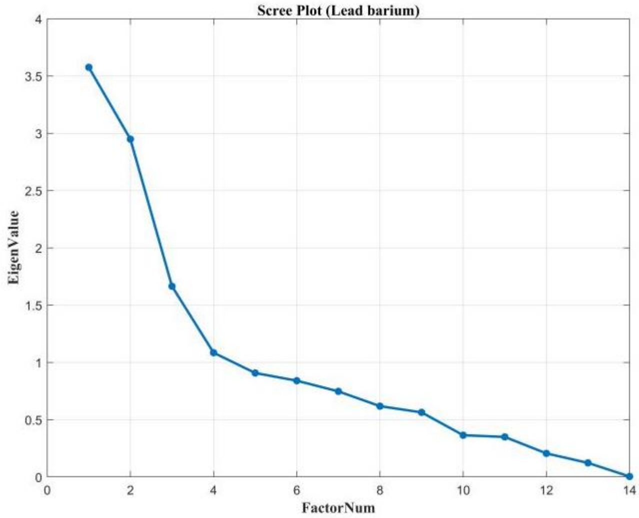

Figure 3. Lead and barium glass lithotripsy

Observing the above gravel diagram, both high potassium glass and lead-barium glass work best when the number of main factors is 4. After calculation, when m is taken as 4, the gravel diagram has an obvious downward slope and resides at the inflection point. Using Matlab, the load matrices of high potassium glass and lead-barium glass were calculated as follows in Tables 2 and 3.

Table 2. Factor loading matrix of high potassium glass

The original variables x / The main factor f	f_1	f_2	f_3	f_4
x_1	-0.419	0.833	0.015	0.147
x_2	-0.386	-0.762	-0.279	0.147
x_3	0.282	-0.883	0.092	0.021
x_4	0.104	-0.735	0.137	-0.486
x_5	0.834	-0.248	0.150	0.105
x_6	0.675	-0.491	-0.101	-0.202
x_7	0.730	-0.248	0.059	-0.421
x_8	0.288	-0.103	0.046	-0.794
x_9	0.130	-0.373	-0.738	-0.250
x_{10}	0.587	0.081	-0.582	-0.420
x_{11}	0.886	0.038	-0.002	-0.028
x_{12}	0.902	-0.084	-0.155	0.089
x_{13}	0.246	0.024	0.075	0.768
x_{14}	0.151	-0.244	0.810	-0.314

Table 3. Factor loading matrix of lead-barium glass

The original variables x / The main factor f	f_1	f_2	f_3	f_4
x_1	-0.406	-0.808	0.230	0.151
x_2	-0.116	-0.331	0.086	0.733
x_3	0.043	-0.134	0.511	-0.289
x_4	-0.057	0.658	0.336	-0.418
x_5	-0.439	0.343	0.637	0.075
x_6	-0.235	-0.164	0.810	0.073
x_7	-0.273	0.035	0.350	-0.622
x_8	0.734	0.087	-0.153	0.144
x_9	-0.199	0.727	-0.459	-0.150
x_{10}	0.952	-0.006	-0.115	0.059
x_{11}	0.049	0.712	0.047	-0.257
x_{12}	0.160	0.702	-0.024	0.369
x_{13}	-0.015	0.008	0.653	-0.093
x_{14}	0.753	0.133	0.019	-0.085

2.1.3. Factor rotation

The model for factor analysis was developed not only to identify the common factors and group the variables, but also to classify the subspecies of high potassium glass as well as lead-barium glass while knowing the meaning of each common factor. The model was rotated by using the maximum variance method. The loading matrix A is orthogonally rotated to maximize the variance of the resulting factor-rotated matrix $A' = AT$. T is the orthogonal matrix, represented as shown in Equation (2).

$$T = \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \tag{2}$$

The rotated loading matrix (see annex) is calculated. The absolute values of each row and column of the loading matrix are taken, with a higher value indicating a stronger common characteristic of the chemical. Values are divided by 0.5, and only values greater than 0.5 are involved in the analysis process of common factors. The chemical composition of the corresponding element classification, can be divided into six types: base class element - "sodium" "k", alkaline earth elements - "calcium magnesium" "barium strontium, transition elements - "copper" "iron", the main group metals - "aluminum" "tin" "lead", nonmetal - "s", "p" metals - "silicon".

In high potassium glasses, "magnesium oxide", "barium oxide" and "strontium oxide" have a strong common factor f_1, so f_1 can be set as oxides of alkaline earth elements, and similarly. f_2 is set as the oxide of alkaline elements, f_3 is set as the oxide of non-metallic elements, f_4 is set as the oxide of excessive elements; in lead-barium glass, "aluminum oxide" and "tin oxide" have a stronger common factor f_3, so c can be set as the oxide of the main group In the lead-barium glass, "aluminum oxide" and "tin oxide" have a stronger common factor f_3, so we can set c as the oxide of the main group, and similarly, a is set as the oxide of alkaline earth-like elements, f_2 is set as the oxide of metal-like elements, and f_4 is set as the oxide of alkaline elements.

The results of subspecies classification are shown in Figure 4.

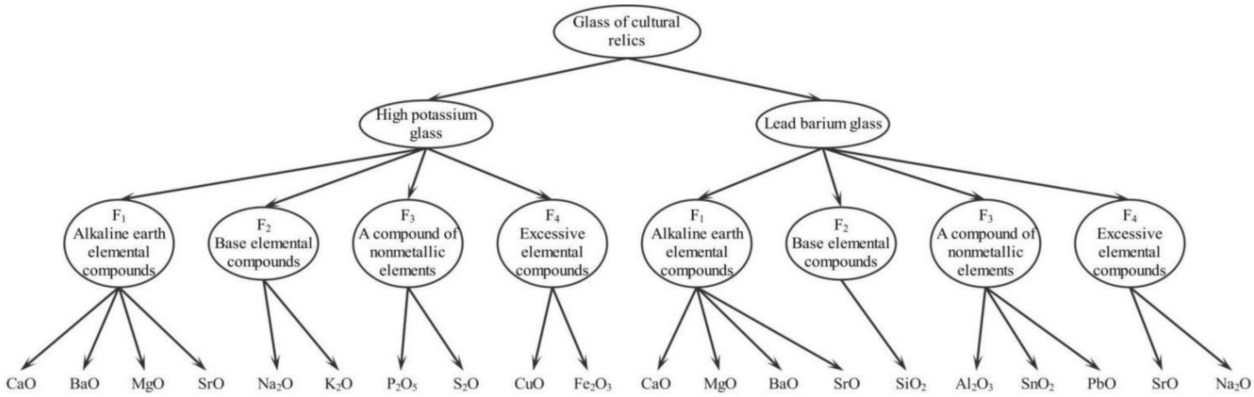

Figure 4. Results of subspecies classification of glass relics

2.1.4 Multiple linear regression

Immediately after equation (1), considering $x_i - u_i$ as the dependent variable and the factor loading matrix A as the prediction of the independent variables, we obtain equation (3).

$$\begin{cases} x_1 - \mu_1 = \alpha_{1,1} f_1 + \alpha_{1,2} f_2 + \cdots + \alpha_{1m} f_m + \varepsilon_1 \\ x_2 - \mu_2 = \alpha_{2,1} f_1 + \alpha_{2,2} f_2 + \cdots + \alpha_{2m} f_m + \varepsilon_2 \\ \vdots \\ x_{14} - \mu_{14} = \alpha_{14,1} f_1 + \alpha_{14,2} f_2 + \cdots + \alpha_{14,m} f_m + \varepsilon_{14} \end{cases} \tag{3}$$

Recorded in matrix form as $X - \mu = \Lambda F + \boldsymbol{\varepsilon}$. Since the variances of the special factors are different, the scores are obtained by weighted least squares so that equation (4) satisfies.

$$(X - \mu - \Lambda F)^{\mathrm{T}} D^{-1} (X - \mu - \Lambda F) \tag{4}$$

When the value of equation (4) is minimized. where $D = \operatorname{diag}\{\sigma_1^2, \ldots, \sigma_{14}^2\}$ is the variance contribution of the special factor σ_i^2 to x_i calculating equation (5) yields.

$$F=(A^{\mathrm{T}}D^{-1}A)^{-1}A^{\mathrm{T}}D^{-1}(X-\mu) \tag{5}$$

Matlab was used to calculate the scores of high potassium glass and lead-barium glass, and the new independent variables were f_1 score, f_2 score, f_3 score, and f_4 score, and the dependent variable was the degree of weathering of the glass. The box line diagram [10] is shown in Figures 5 and 6.

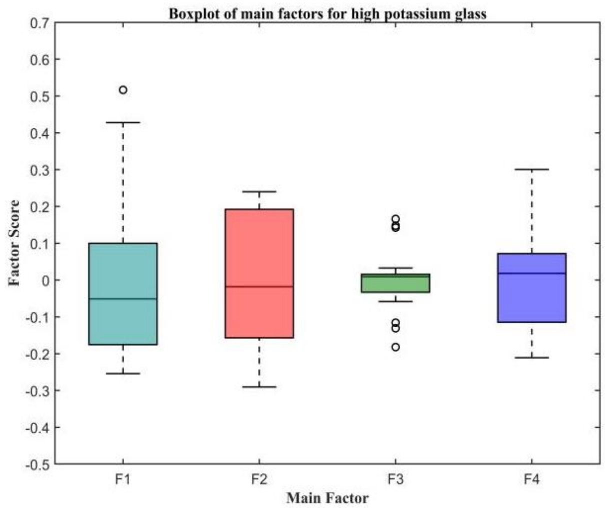

Figure 5. Boxplot of main factors for high potassium glass

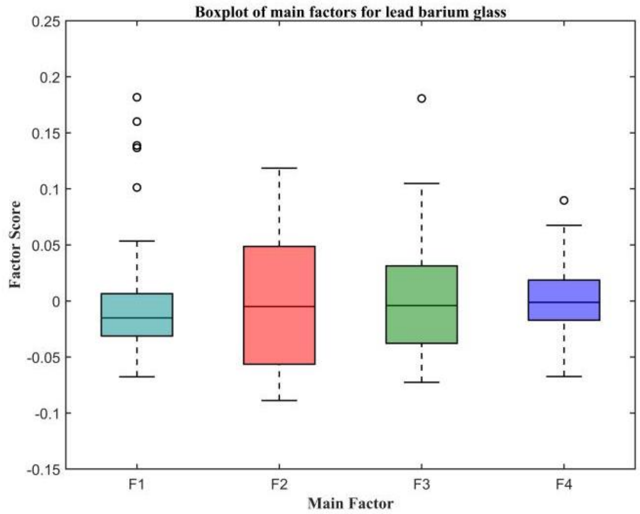

Figure 6. Main factor box diagram of lead and barium glass

According to the box plot of the main factors of high potassium glass, the median factor scores of the four main factors are similar, the volatility of factor f_1 and factor f_2 are larger, and the outliers of factor f_3 are the largest; according to the box plot of the main factors of lead barium glass, the median factor scores of the four main factors are similar, the volatility of factor f_2 is larger, and the outliers of factor f_1 are the largest.

The linear regression equations were obtained for high potassium glass and lead barium glass by multiple linear regression of "main factor score weathering degree", as in equation (6) for high potassium glass and equation (7) for lead barium glass.

$$y=0.3-0.6035f_1+2.1735f_2+0.0015f_3-1.0551f_4 \tag{6}$$

$$y=0.551+0.9884f_1+5.6427f_2-1.3406f_3-0.456f_4 \tag{7}$$

To verify the accuracy of the multiple linear regression, the fitted residuals of high potassium glass and lead-barium glass were plotted, as shown in Figures 7 and 8.

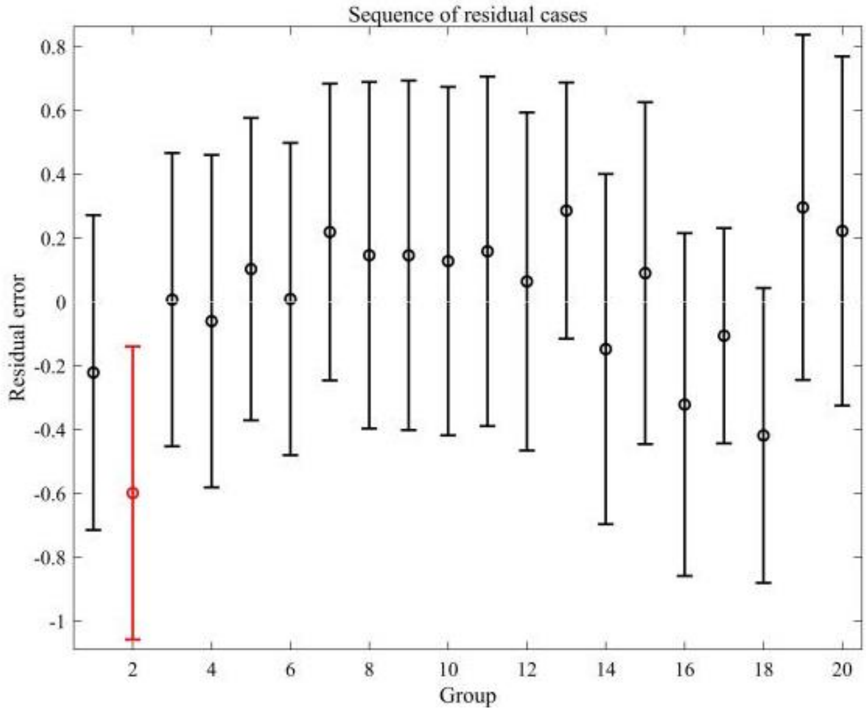

Figure 7. Residuals and confidence intervals after regression for high-potassium glass

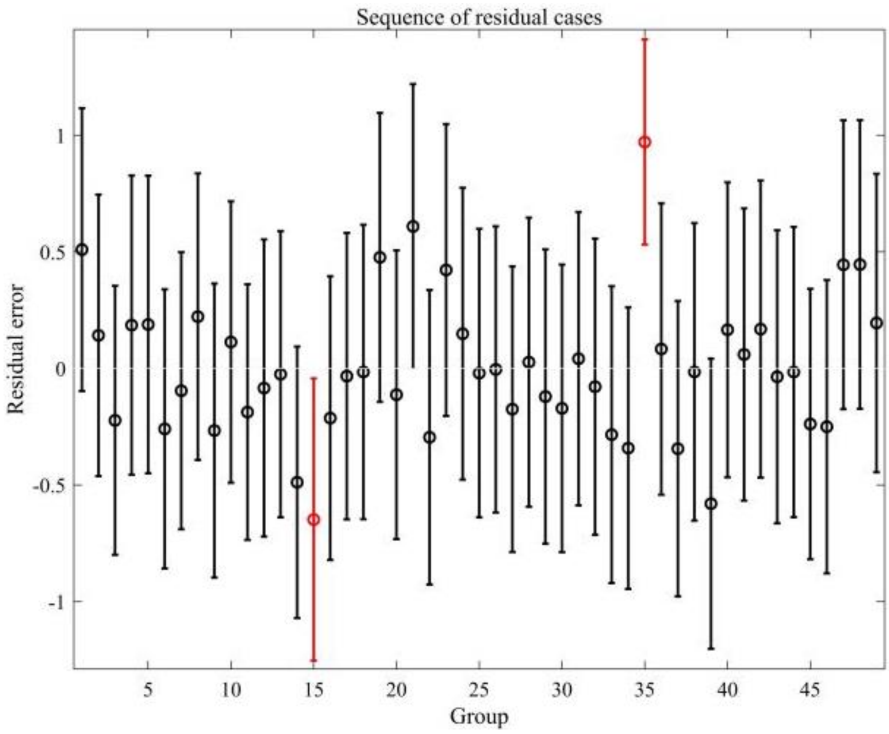

Figure 8. Residual and confidence interval plot of lead and barium glass after regression

Analysis of the above graph shows that there is an anomaly in the second group of data for high potassium glass, and there is an anomaly in the 15th and 35th groups of data for lead-acid glass. The above three groups of data are proposed, and the multiple covariance regression is re-run, and the regression equations are equation (8) and equation (9).

High potassium glass.

$$y = 0.3328 - 0.6861f_1 + 2.2443f_2 - 0.0859f_3 - 1.0331f_4 \tag{8}$$

Lead-barium glass.

$$y = 0.5342 + 0.109 f_1 + 5.2992 f_2 - 4.2587 f_3 - 2.0249 f_4 \tag{9}$$

2.2. Sensitivity analysis

Based on the residual plots of the two groups of glasses, it can be seen that the model is effective in classifying the two types of glass with only a few outliers, yielding results with a high degree of confidence. However, the chemical composition of the ancient glass products excavated in other environments and the data used to build the model may have deviations, so the four principal components obtained from the factor analysis of dimensionality reduction were changed to a little extent respectively. By adding dynamic Gaussian white noise [11] to the four principal components, we can make the model more flexible and reliable, and the results can be more reasonable.

When the value of the weathering term was above the splitting line y = 0.5, the glass products were classified as weathered glass products, and vice versa as undifferentiated glass products.

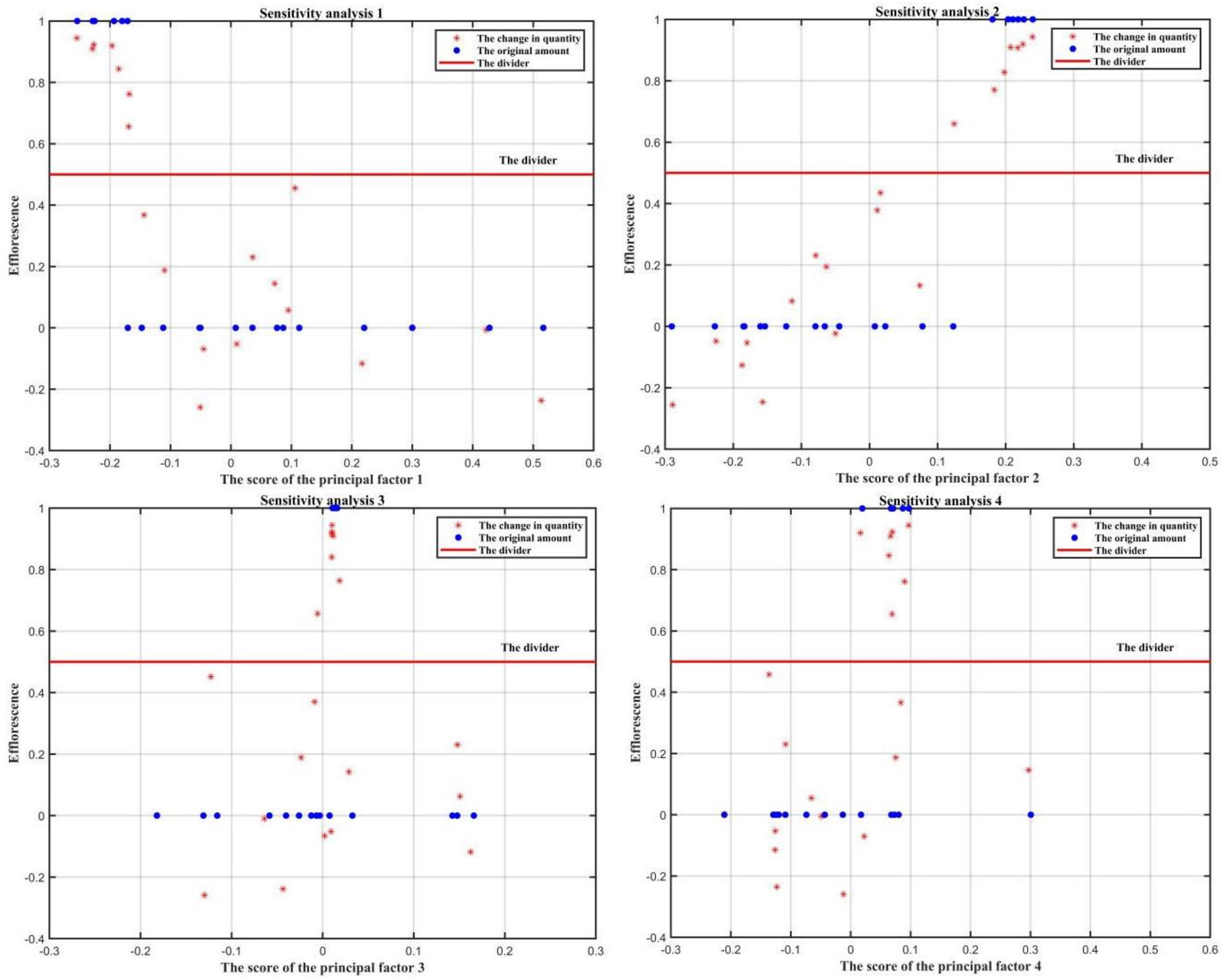

Figure 9. Sensitivity test chart of high potassium glass

According to Figure 9, the data processed with dynamic Gaussian white noise showed that the model was still very accurate for glass type classification when the model was tested. Therefore, it can be concluded that the classification of high-potassium glass passed the sensitivity test.

Next, a sensitivity analysis was performed on the lead-barium glass products to check the reasonableness of the model, and the results obtained are shown in Figure 10. When the value of the weathering term was above the division line y = 0.5, the glass products with the changed index were classified as weathering type glass products, and vice versa as undifferentiated type glass products.

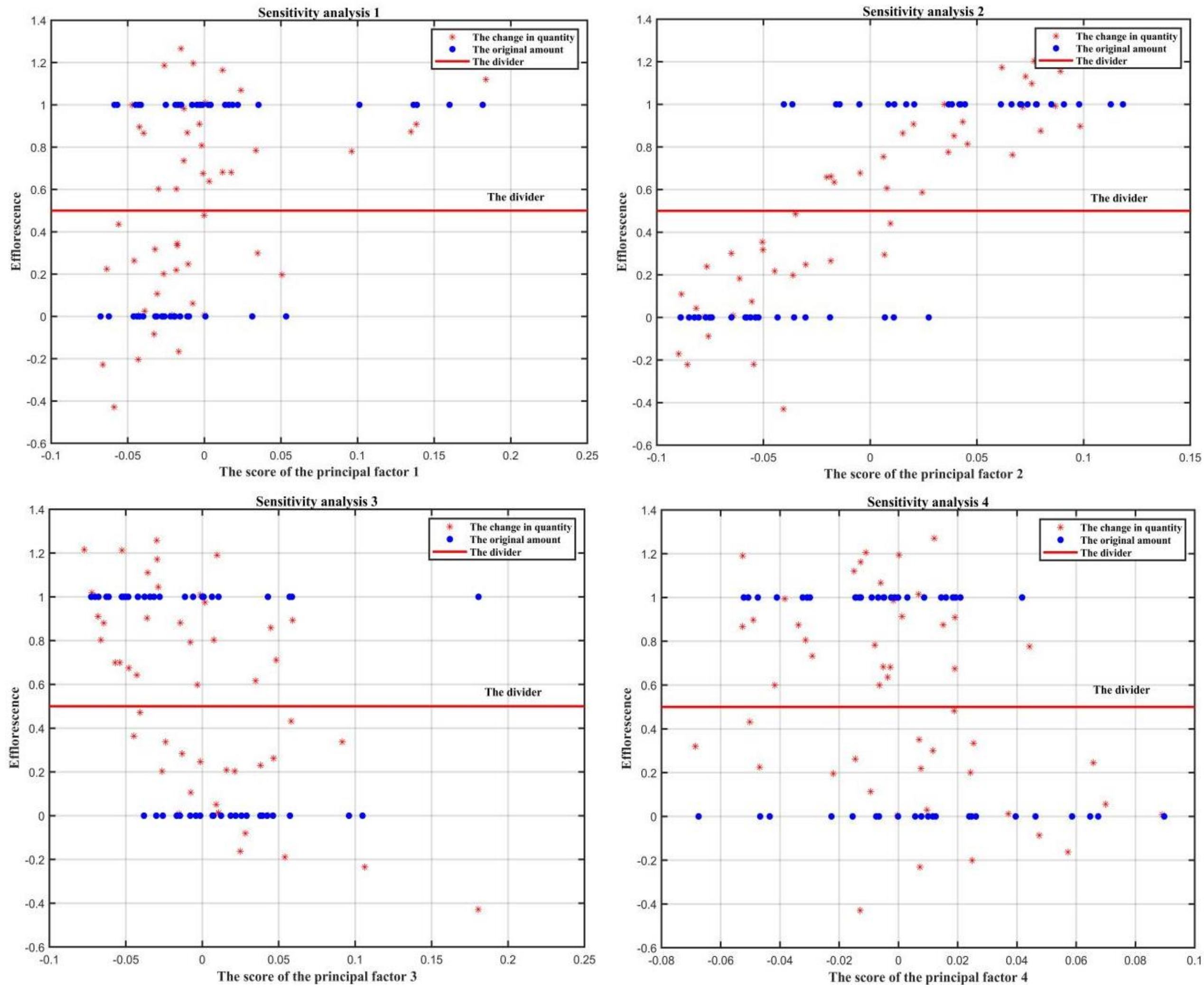

Figure 10. Sensitivity test chart of lead-barium glass

According to Figure 10, it can be seen that it is difficult to visualize the classification effect when the model is tested for the dynamic Gaussian white noise processed part of the data. Thus, when comparing each original data with the dynamic Gaussian white noise processed data with the shortest lateral distance from itself, we found that the model still had a high accuracy in classifying glass types. Therefore, it can be assumed that the classification of high-potassium glass passed the sensitivity test.

3. Conclusion

Factor analysis is used to reduce the dimensionality of the data, not to take away the original variables, but to recombine them according to the information of the original variables, to find out the common factors affecting the variables, and to simplify the data, so that the data will not have a large amount of missing data as in conventional dimensionality reduction algorithms, which leads to the reduction of the validity of the results. At the same time, factor analysis modeling makes the factor variables more interpretable by rotating the factor loading matrix, and the naming clarity is high. Finally, the sensitivity of the model is checked by adding appropriate Gaussian white noise. However, the calculation method used for factor analysis in calculating the factor scores is least squares, and this method may so metimes cause some errors in the results, after further enhancement can be carried out subsequently

References

[1] ZHOU Liangzhi. J Dalian Institute Light Ind (in Chinese), 1984 (1): 34 – 44.

[2] WANG Chengyu, TAO Ying, CHEN Min, et al. Bull Chin Ceram Soc (in Chinese), 1989 (6): 1 – 9.

[3] ZHAO Fengyan, CHEN Bin, CHAI Yi, et al. J Archaeal Cultural Relics (in Chinese), 2015(4): 111 – 119.

[4] FENG Bailing. Construction and application of ancient glass bead database unearthed in China (in Chinese, dissertation). Xi'an: Northwestern University, 2021.

[5] ZHANG Liyan, LI Hong, CHEN Shubin, et al. J Chin Ceram Soc, 2022, 50 (8): 2338 – 2350.

[6] LI Jiangang. Glass defect detection based on depth learning (inChinese, dissertation). Fuzhou: Fujian Institute of Engineering, 2021.

[7] HE Xuansen, HE Fan, XU Li, et al. J Electric Sci Technol Univ (in Chinese), 2022, 51 (6): 904 – 912.

[8] Sun X, Lei Y. Research on financial early warning of mining listed companies based on BP neural network model[J]. Resources Policy, 2021, 73: 102223.

[9] Mulenga, Sunday. "Evaluation of factors influencing rock fragmentation by blasting using interrelations diagram method." Journal of Physical Sciences 2.1 (2020): 1 - 15.

[10] Pius R, Sen A. Unitarity of the box diagram[J]. Journal of High Energy Physics, 2018, 2018 (11): 1 - 20.

[11] Tripura T, Chakraborty S. A sparse Bayesian framework for discovering interpretable nonlinear stochastic dynamical systems with Gaussian white noise [J]. Mechanical Systems and Signal Processing, 2023, 187: 109939.