Study on Composition Analysis and Identification of Ancient Glass Products

ShiXing Han [1,*], Yue Lv [1], Bin Hong [1], Shixuan Han [2], Yahui Li [3]

[1]College of Engineering, Tibet University, Lhasa, Tibet Autonomous Region, 850000 China

[2]Leicester International College of Mathematics and Basic Science, Dalian University of Technology, Panjin, Liaoning, 124221 China

[3]Faculty of Finance and Economics, Tibet University, Lhasa, Tibet Autonomous Region, 850000 China

*Corresponding author: 863580336@qq.com

Abstract: The introduction of ancient glass products, to a large extent, promoted the economic development at that time, but due to the influence of burial environment, a glass product was weathered, which brought great challenges to archaeologists. Therefore, it is particularly important to study the composition of ancient glass products. In this paper, multiple models are established for analysis, providing a certain reference for archaeologists.

Key words: time series; XGBoost classification; PSO-BP neural network; Glass relics

1. Introduction

Glass has a long history and was one of the earliest man-made materials invented by man, as well as one of the most expensive materials. In the early days, foreign glass was introduced to China, but after the continuous exploration and improvement of glass technology by our working people, there were significant differences between our ancient glass and foreign glass[1]. The glass of Western Asian civilization was made of soda calcium silicate as the main ingredient, while potassium oxide extracted from wood ash was used as the melting agent, which led to the significant difference between the glass of China and foreign countries in terms of material and other aspects.

However, the weathering process of ancient glass is complex, and there is a cross influence between the weathering of the surface of artifact samples and their glass type, decoration and color; there is also a certain pattern between the weathering of the surface of artifact samples and the content of their chemical composition; as a large number of elements will be exchanged between the internal and the environment during the weathering process of glass, the prediction of Since a large number of elements are exchanged between the environment and the interior in the process of glass weathering, it is particularly important to predict the chemical composition content before weathering to correctly determine its type; at the same time, it is essential to select the appropriate chemical composition for each category to classify, identify the type of artifact samples and analyze the correlation between its chemical composition.

Based on this, (1) the article firstly preprocessed the data, including filling in the missing data and eliminating the invalid data. Conditional logistic regression and Pearson correlation comparison were used to analyze the relationship between surface weathering and glass type, decoration and color, and the Ominbus test of model coefficients was verified to be reasonable. In addition, the data set was divided into two groups of four categories, and after visualization by histogram, the required statistical laws were analyzed. Finally, an ARIMA time series-based chemical composition prediction model was established to predict the chemical composition content before weathering. (2) A classification model based on XGBoost and principal component-cluster analysis was established, and the classification laws of the two glasses were summarized by using radar plot visualization. Then the number of principal components was determined with reference to the feature importance value of XGBoost model, and the subclasses were divided by using debris map and factor load quadrant 3D distribution map, and the color was used as the basis for the division, and the high potassium glass was divided into 5 subclasses and the lead-barium glass was divided into 9 subclasses. Finally, the intra-class similarity and inter-class difference evaluation indexes of chemical components were constructed, and the model was obtained to be reasonable and effective. Multi-parameter sensitivity analysis was performed, and the sensitivity of the model was obtained to be high.

2. Glass chemical composition analysis and prediction model

2.1 Type relationship analysis

After data processing and cleaning, to ensure the accuracy of the analysis results, Pearson correlation and conditional logistic regression are used here for comparative analysis to summarize the relationship between surface weathering and its type, ornamentation, and color to determine the exact final relationship[2]. Before that, this paper was visualized by Sanger diagram (as in Figure 1).

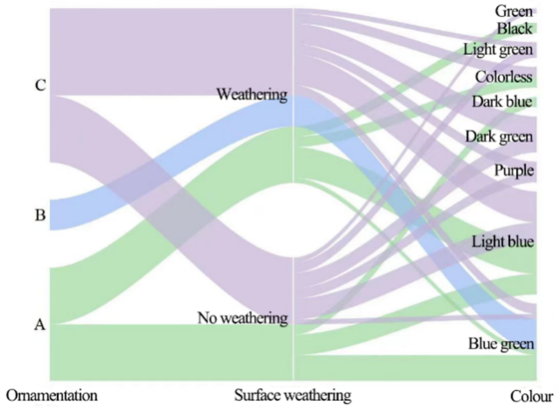


Figure 1 Sanjitu

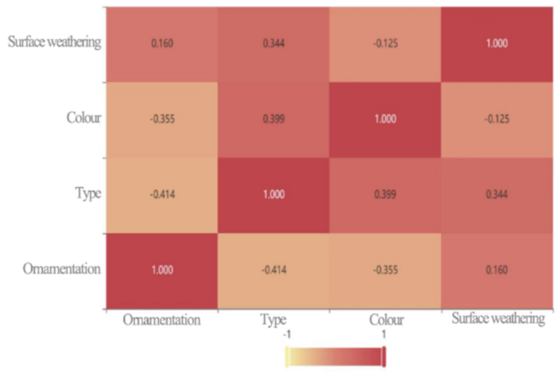


Figure 2 Heat map of correlation coefficient

From Figure 1, it can be seen that weathering is associated with motifs A, B, and C, and with black, blue-green, light blue, light green, dark green, and purple; no weathering is associated with motifs A and C, and with blue-green, green, light blue, light green, dark green, and purple.

(1) Pearson correlation

The correlation coefficients were analyzed for the positive and negative directions as well as the degree of correlation, and the heat map of correlation coefficients, as shown in Figure 2; it can be seen from the heat map that the correlation coefficient between surface weathering and type is 0.334, which becomes a strong positive correlation, and the correlation coefficient with ornamentation is 0.160, and the correlation coefficient with color is -0.125.

(2) Conditional logistic regression

1) Ominbus test of model coefficients was performed to check whether the conditional logistic regression model was valid.

2) The conditional logistic regression coefficients were summarized and displayed to observe whether the model was valid and to focus on the RR (relative risk) value to reveal the role of this variable on the occurrence of positive events. The results of the specific analysis are shown in Table 1.

Table 1 Conditional logistic regression results

item	Regression coefficients	Standard Error	Wald	df	P	RR value	Lower limit of RR value	Upper 95% confidence interval
Color	-0.27	0.141	3.633	1	0.057*	0.764	0.579	1.008
Type	1.278	0.495	6.655	1	0.010***	3.588	1.359	9.471
Ornamentation	0.159	0.21	0.573	1	0.449	1.172	0.777	1.768

From the above table, based on the variable (color), the significance P-value is 0.057*, which does not show significance at the level, therefore color will not have a significant effect on surface weathering, as well as an RR value of 0.764. Based on the variable - type, the significance P-value is 0.010***, which shows significance at the level, therefore type will have a significant effect on surface weathering, as well as an RR value of Based on the variable - ornamentation, the significance p-value is 0.449, which is not significant at the level, therefore ornamentation does not have a significant effect on surface weathering, and the RR value is 1.172.

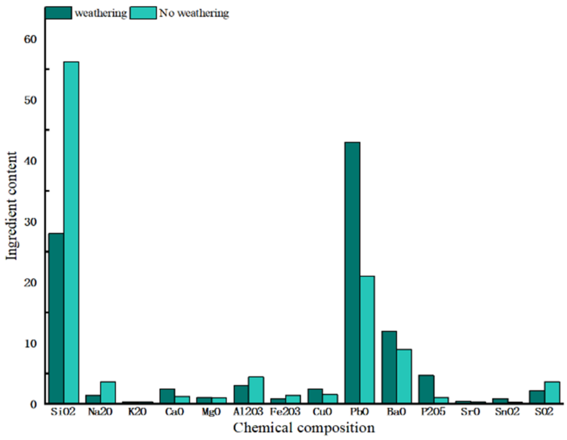


Figure 3 high potassium type glass

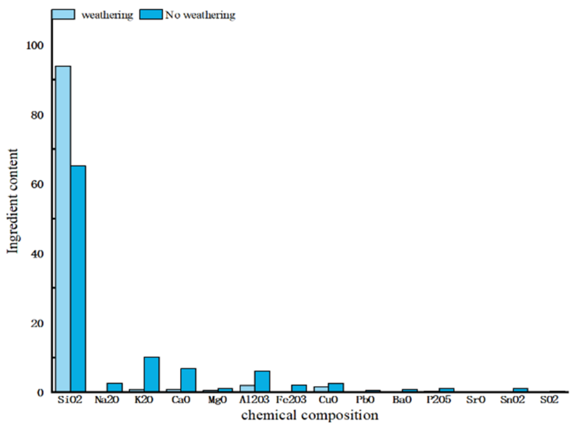


Figure 4 Lead and barium type glass

2.2 Statistical pattern analysis

To facilitate the analysis, the non-valid two sets of data were first excluded, and then the data were divided into two categories of four types (i.e., high potassium weathering, high potassium without weathering, lead-barium weathering, and lead-barium without weathering) and visualized using bar graphs, as shown in Figure 3 and Figure 4.

It can be seen from the graphs that the content of silica (SiO_2) increases, sodium oxide (Na_2O) decreases, potassium oxide (K_2O) decreases, calcium oxide (CaO) decreases, magnesium oxide (MgO) decreases, and aluminum oxide (Al_2O_3) decreases before and after weathering in the high potassium category, where the content of silica (SiO_2) remains the highest before and after weathering, while most of the other contents basically The content of silica (SiO_2) remains the highest before and after weathering, while most of the other contents are basically reduced. Before and after the weathering of lead-barium glass, the content of SiO2 decreases, the content of Na_2O decreases, the content of CaO increases, the content of MgO does not change much, the content of Al_2O_3 decreases, the content of PbO increases sharply, the content of BaO increases, and the content of P2O3 increases. The content of phosphorus pentoxide (P_2O_5) increased. In addition, the top three components with the highest content are silicon dioxide (SiO_2), lead oxide (PbO), and barium oxide (BaO), and they vary greatly before and after weathering.

2.3 Ingredient content prediction

(1) Prediction Process

Using SPSSPro software, the prediction of 14 chemical components based on a time series model, shuffling the data, ensuring a training percentage of 87%, and setting cross-validation to triple fold, the specific prediction effect before weathering can be obtained, which can be reflected by Table 2[3].

Table 2 Results of time series prediction model evaluation

	MSE	RMSE	MAE	MAPE	R^2
Training set	94.685	9.731	7.9	29.625	0.649
Cross-validation	8485.157	77.946	51.535	-27.461	87.667
Test set	74.617	8.638	7.261	135.565	0.762

The above table shows the prediction evaluation metrics for the cross-validation set, training set and test set, which measure the prediction effectiveness of the time series model through quantitative metrics. Among them, the MAE (mean absolute error) is 7.9 and 7.261, indicating a high accuracy of the model. the results of R^2 are 0.649 and 0.762 and converge to 1, indicating a high accuracy of the time series model.

(2) Forecast analysis

(1) Surface weathering has a strong positive correlation coefficient of 0.334 with type, a correlation coefficient of 0.160 with ornamentation, and a correlation coefficient of -0.125 with color. based on the variable (color), the significance p-value is 0.057*, which does not present significance at the level, so color does not have a significant effect on surface weathering, as well as an RR value of 0.764. based on the variable (type) Based on the variable (ornamentation), the significant P-value is 0.010***, which presents significance at the level, therefore type has a significant effect on surface weathering, as well as an RR value of 3.588. Based on the variable (ornamentation), the significant P-value is 0.449, which does not present significance at the level, therefore ornamentation does not have a significant effect on surface weathering, as well as an RR value of 1.172. In a comprehensive analysis, glass type has a significant effect on surface weathering is significantly influenced by and strongly correlated with color and ornamentation, and color and ornamentation do not have a significant effect on surface weathering.

(2) The content of silica (SiO_2) remains the highest before and after weathering of high potassium glass, while the content of most of the others basically decreases. Before and after weathering of lead-barium type glass, the top three components with the highest content are silica (SiO_2), lead oxide (PbO), and barium oxide (BaO), and each component fluctuates greatly.

(3) The predicted results of the chemical composition content of weathering points before weathering (only a part is listed) are.

Table 3 Predicted results of chemical composition content

Artifact Number	**SiO_2**	**K_2O**	**……**	**CaO**	**PbO**	**BaO**	**……**
7	67.65	7.37	……	0	0.2	1.38	……
22	29.64	0	……	2.93	42.82	5.35	……
…	…	…	…	…	…	…	…
44	28.79	0	……	4.58	34.18	6.1	……
58	26.25	0	……	1.11	61.03	7.22	……

3. Potassium glass and lead-barium glass classification model

3.1 Glass classification

By organizing the data and grouping the weathered data into one category and the unweathered data into one category, the article uses XGBoost model for analysis and Origin software for visualization, and uses radar plots to show each chemical composition of high potassium and lead-barium glass, respectively, as shown in Figure 5 and 6[4].

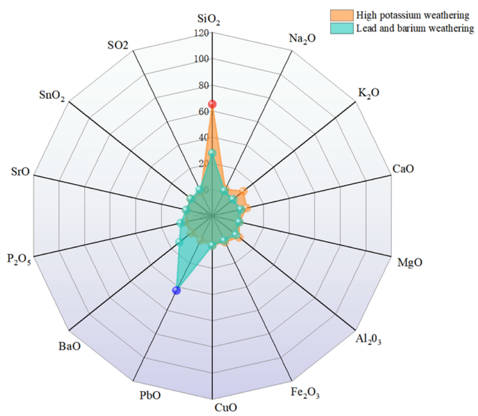


Figure 5 Comparison of chemical composition of high potassium and lead-barium glass (After weathering)

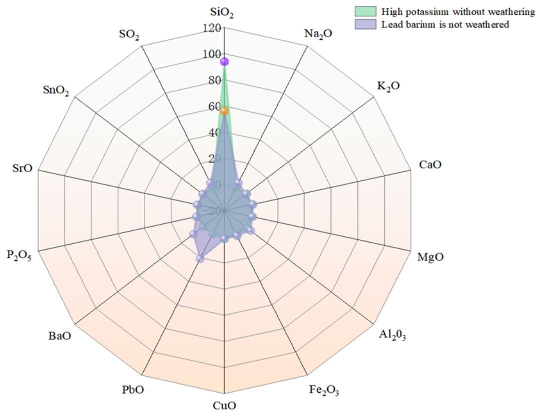


Figure 6 Comparison of chemical composition of high potassium and lead-barium glass (No weathering)

As can be seen from the radar diagram, under weathering, the content of silica (SiO_2) of high potassium glass is above 61.4% (the highest content), the content of potassium oxide (K_2O) is around 14.3%, the content of calcium oxide (CaO) is around 13.7%, and the rest of the content is low; while the content of lead barium glass oxide (PbO) is close to 20% ranking third, the content of silica (SiO_2) The content of lead oxide (PbO) is close to 20%, silicon dioxide (SiO_2) is close to 40%, and barium oxide (BaO) is more than 60%, ranking first (the highest content). It can be seen that the classification law of high potassium glass under weathering conditions can be approximated by the content of silica (SiO_2), potassium oxide (K_2O), and calcium oxide (CaO) The classification law of lead-barium glass can be approximated by the content of lead oxide (PbO), silica (SiO_2), and barium oxide (BaO).

In addition, it can be seen from the radar diagram that the classification law of high potassium glass under the condition of no weathering can be approximated by the highest content of silicon dioxide (SiO_2) and the lower content of all the rest 'The classification law of lead-barium glass can be approximated by the distribution of the content of lead oxide (PbO), silicon dioxide (SiO_2), barium oxide (BaO) and the lower content of the rest .

3.2 Glass subcategories

In this paper, firstly, XGBoost was used to analyze for 10 indicators (4 with very low chemical content were not included), and the characteristic importance value of each indicator was determined and obtained by using SPSSPro software, as shown in

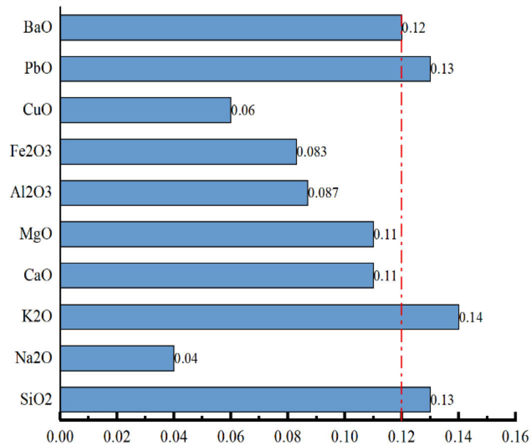


Figure 7 Characteristic importance values of chemical components(left)

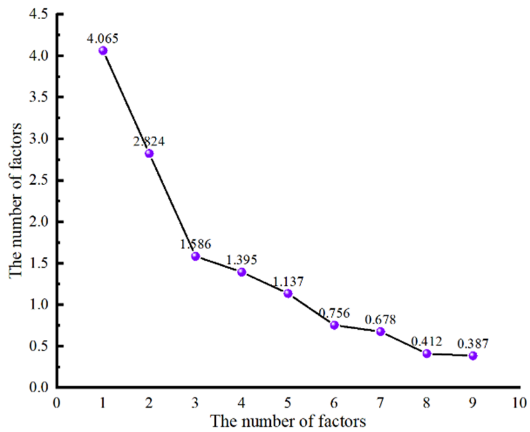


Figure 8 14 chemical compositions of the gravel map situation(right)

From the histogram, it can be seen that there are 4 characteristic importance values greater than 0.12, which are 0.13 for silicon dioxide (SiO_2), 0.14 for potassium oxide (K_2O), 0.13 for lead oxide (PbO), and 0.12 for barium oxide (BaO), so in order to facilitate subclass division, this paper uses the principal component-cluster analysis model to identify 4 principal components for analysis[5]. From Figure 7 and 8, it can be clearly found that through principal component and cluster analysis and importance analysis of eigenvalues, the eigenvalues of the top 3 ranked indicators are significantly larger than the eigenvalues of other indicators, which are 4.065, 2.824 and 1.586, respectively.

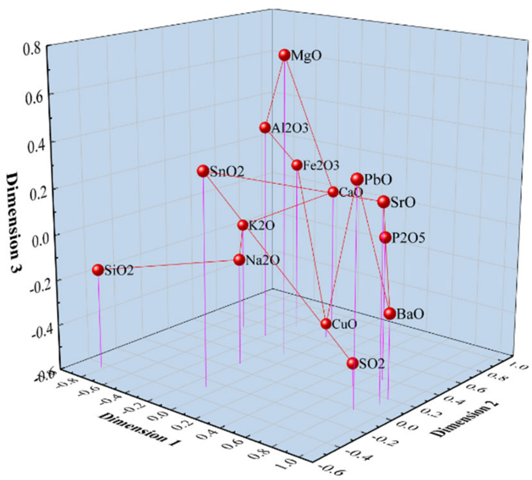
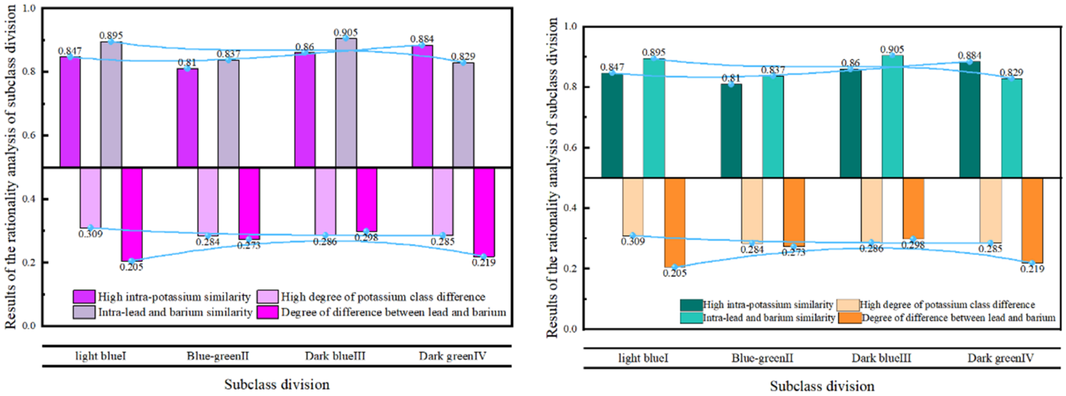


Figure 9 Three-dimensional distribution of the factor load quadrant

According to Figure 9, the spatial distribution of principal components is presented by means of quadrant plots. Since the extracted principal components in this paper are four, the three-dimensional loading principal component scatter plot can be directly presented, thus finding that silica (SiO_2), sodium oxide (Na_2O), potassium oxide (K_2O), lead oxide (PbO), copper oxide (CuO), lead oxide (PbO), and barium oxide (BaO) which chemical components deviate from other components, so they can be used as the basis for subclass classification.

3.3 Subcategory Reasonableness Evaluation

The larger value of ICS results indicates the higher similarity of chemical components within each subclass classification group; the larger value of ICD results indicates the higher difference of chemical components between each subclass classification group. In this question, the rationality of XGBoost classification model for chemical composition subclass classification was quantitatively analyzed by constructing the intraclass similarity and interclass difference evaluation indexes of chemical composition, combined with the subclass classification results[6].



Weathering-free conditions (a) Weathering conditions (b)

Figure 10 Results of the analysis of the reasonableness of the subclass division of high potassium glass and lead-barium glass

The intra-class similarity and inter-class variability of the subclasses of glass artifacts with and without weathering conditions are shown in Fig. It can be seen from the above figure that the intraclass similarity and interclass variability of each subclass classification under different weathering and type conditions are larger and the values of interclass variability are smaller, therefore, the classification model based on XGBoost is reasonable and the results are clearer and more intuitive.

3.4 Multi-parameter sensitivity analysis

In order to verify the effectiveness of the subclass classification results and the XGBoost classification model, a sensitivity analysis was performed on the intervals of subclass classification, assuming that the subclass classification results obeyed the normal distribution function, and a multi-parameter sensitivity analysis was performed on the solution results of the XGBoost classification model[7].

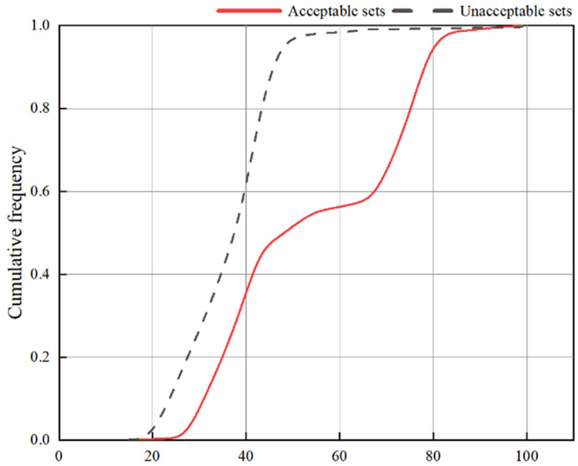


Figure 11 KS test for the parameter SiO_2

From the above figure, it can be seen that the difference between the cumulative frequencies of the acceptable and unacceptable sets of subclass classification results is large, indicating that SiO_2 is more sensitive to the subclass classification results, and by similar analysis, it shows that the sensitivity of other chemical components is higher than that of SiO_2 (see the annex of question 2 for details), indicating that the results of subclass classification are reliable.

3.5 Glass classification division

(1) Classification process

1) Under weathering conditions, the classification law of high potassium glass can be approximated by the content of silicon dioxide (SiO_2), potassium oxide (K_2O), calcium oxide (CaO), and the classification law of lead-barium glass can be approximated by the content of lead oxide (PbO), silicon dioxide (SiO_2), barium oxide (BaO). In the case of no weathering, the classification law of high potassium glass can be approximated by the highest content of silica (SiO_2) and the lower content of the rest. The classification law of lead-barium glass can be approximated by the distribution of the content of lead oxide (PbO), silica (SiO_2) and barium oxide (BaO) and the lower content of the rest.

2) Using the principal component-cluster analysis model based on XGBoost, it can be found that silica (SiO_2), sodium oxide (Na_2O), potassium oxide (K_2O), lead oxide (PbO), copper oxide (CuO), lead oxide (PbO), and barium oxide (BaO) where the chemical composition deviates from the other components, so it can be used as the basis for subclass classification.

(2) Classification analysis

Since the content of chemical composition affects the color of glass to a certain extent, in order to classify subclasses more precisely, this paper classifies the chemical composition based on the color change based on the classification law of high potassium glass and lead-barium glass, as shown in Tables 4 and 5 .

Table 4 Color and chemical content classification of high potassium glass and lead-barium glass under weathering-free conditions

Type	Color	Silica content	Lead oxide	Barium oxide
High Potassium	Pale BlueI	91.75		
	Blue-GreenII	63.47		
	Deep BlueIII	58.23		
	Dark GreenIV	35.95		
Lead barium	Pale BlueI	89.27	＜0.41	＜0.70
	Blue-GreenII	75.4	＜1.52	＜2.94
	Deep BlueIII	60.21	＜1.88	＜5.41
	Dark GreenIV	34.76	24.13＜	13.79＜

Table 5 Classification of color and chemical content of high potassium glass and lead-barium glass under weathering conditions

Type	Color	Silica content	Potassium oxide	Calcium oxide	Lead oxide	Barium oxide
High Potassium	Blue-GreenI	91.75	7.36＜	0.43＜		
	Pale BlueI	12.41~92.36			13.77＜	3.25＜
	Dark GreenII	35.77~75.52			16.15~46.56	3.41~10.30
Lead Barium	Blue-GreenIII	37.35~53.34			9.20~24.56	9.22~23.56
	light greenIV	37.62~41.33			39.05~45.68	9.24~12.75
	PurpleV	53.79~59.82			0.34~16.99	0.97~11.87

4. Conclusion

(1) The ARIMA time-series-based chemical composition prediction model established in this paper uses statistical methods to predict the chemical composition content before weathering based on the trend of no-weathering-weathering, which usually conforms to the law of development and the results are more accurate. In addition the established classification model based on XGBoost and principal component-cluster analysis, the use of XGBoost model simplifies the solution problem and avoids the large errors caused by the influence of other indicators.

(2) The article uses principal component-clustering analysis to classify subclasses of glass. The classification method is relatively simple and has certain limitations, which can be improved by using neural networks and machine learning.

The model provided in this paper is also very helpful and widely used in computer, agriculture, medical and ecological fields.

References

[1] Ding Nan. Rationality and difference analysis of China's futures market[J].Journal of Beijing Technology and Business University(Social Science Edition),2009,24(06):38-43.DOI:10.16299/ j.1009-6116.2009.06.023.HOU Zhili.Qinhuangdao Glass Museum Tang and Song Dynasty Glass Cultural Relics Specific Age Study——A Brief Discussion on the Evolution and Development of Sui, Tang and Song Dynasty Glass Cultural Relics[J].Popular Literature and Art,2015(13):47.)

[2] WANG Tiebing,TAI Changsong,WANG Dongwei,SU Zhan,YAO Zhaoji,CHEN Zhiyun. Multivariate conditional logistic regression analysis of Mycobacterium tuberculosis infection in Shenzhen[J].Disease Surveillance and Control,2011,5(07):406-407+405.)

[3] Ouyang Di, Cui Jia. Spatial-temporal dynamics of air pollutants in Harbin under carbon neutrality strategy——Based on Holt-Winters time series model[J/OL].Environmental Protection Science:1-11[2022-10-15]. DOI:10.16803/j.cnki.issn.1004-6216.2022070050.

[4] LU Wanwan,WANG Weifang,MA Yumin. Research on early warning model based on XGBoost algorithm[J].Electronic Design Engineering,2022,30(19):49-54+59.DOI:10.14022/j.issn1674-6236.2022.19.011.)

[5] Cai Hong. Evaluation of urban innovation ecosystem based on global principal component-cluster analysis: A case study of 19 key cities in China[J].Science and Technology and Industry,2022,22(05):229-234.)

[6] Peng Hong, Luo Linkai, Lin Chengde. Parameter optimization of Gaussian functions based on intra-class and inter-class similarity comparison[C]//. Proceedings of 2011 2ND International Conference on Innovative Computing and Communication and 2011 Asia-Pacific Conference on Information Technology and Ocean Engineering(CICC-ITOE 2011 V3).,2011:204-207.

[7] ZHOU Liang,KUANG Huaxing,ZHANG Yutao,DING Chun,WANG Lingling. Sea clutter-modified probability density distribution function based on lognormal distribution[J].Radar and Adversarial,2021,41(01):18-22.DOI:10.19341/j.cnki.issn.1009-0401.2021.01.005.