ISSN 2616-5805 Vol. 3, Issue 2: 57-62, DOI: 10.25236/AJMS.2022.030209

Multiple regression-based glass composition prediction and statistics

Xiaohan Deng[1], Tangliang Wang[2]

[1]College of Science, Chongqing University of Technology, Chongqing, 400054, China
[2]College of Mechanical Engineering, Chongqing University of Technology, Chongqing, 400054, China

Abstract: *The surface weathering of ancient glass can be distinguished from the type of ancient glass by chemical composition as well as color and ornamentation. This paper constructs mathematical models to analyze the relationship between surface weathering and glass type, color and ornamentation based on different chemical compositions and characteristic data of color and ornamentation in ancient glass, explores the statistical law of chemical composition content with and without weathering, and constructs an effective classification model to realize glass type classification. This paper solves the relationship between glass type, decoration and color and their weathering degree; analyzes the statistical law of chemical composition content with and without weathering on the surface of two types of glass; derives the prediction formula and predicts the chemical composition content of two types of glass before weathering.*

Keywords: *Random Forest Model; Multiple Regression Model; High Potassium Glass; Lead Barium Glass*

1. Introduction

Silk Road for ancient China and the Western countries to construct a road between the countries can facilitate trade and exchange of goods. Glass was one of the important commodities along the Silk Road. Archaeological research has uncovered many deeply buried ancient glasses, and the type of glass varies with its chemical composition, so it is necessary to determine the type of glass before more in-depth archaeological research[1].

Wang Chengyun and Tao Ying (2003) mentioned that after a long period of contact and action between glass and the atmosphere, the glass would undergo erosion, that is, the corrosion behavior that occurs after contact with the erosive medium[2]. Under the influence of environmental factors, the internal elements of ancient glass will exchange with external elements to produce weathering[3]. The ratio of its internal chemical composition is severely affected, making it impossible to determine the type of ancient glass[4]. Therefore, it is important to solve the problem based on the current data of ancient glass.

This paper intends to provide an effective mathematical model and method for the identification of ancient glass types and the analysis of chemical composition, which is of great relevance to the archaeological work of ancient glass artifacts. Secondly, we analyze the relationship between surface weathering and glass type, color and decoration based on chi-square test and measure the correlation with φ-column correlation coefficient, which shows that surface weathering is related to glass type but not to color and decoration. Finally, a multiple regression model was developed to predict the content of each chemical composition before weathering at the weathering detection sites.

2. Assumptions and notations

2.1 Assumptions

Use the following assumptions[5].

1) It is assumed that the surface weathering of the artifact is determined only by the type, decoration and color of the glass reflected in the title, regardless of other subjective factors such as environment.

2) It is assumed that all data are calculated correctly or within the error limits.

3) Excluding errors caused by objective influences during data collection

2.2 Notations

The primary notations used in this paper are listed as Table 1.

Table 1: Notations

Symbols	Meaning
f_o	Actual frequency
f_e	Theoretical frequency
x^2	The degree of difference between the actual frequency and the theoretical frequency
m	Number of features selected at any node
Cov	Covariance
Var	Covariance
f	Mean variation of SiO2 in high potassium glass
k	Mean change of K2O in high potassium glass
l	Average variation of Al2O3 in high potassium glass
f_1	Model fitting values of SiO2 in high potassium glass
z_1	Model fitting values of SiO2 in lead-barium glass
z_2	Average variation rate of SiO2 in lead-barium glass
x_2	Average rate of change of PbO in lead-barium glass
y_2	Average rate of change of P2O5 in lead-barium glass

3. Model construction and solving

3.1 Correlation analysis based on chi-square test model

To solve for the relationship between the degree of weathering of glass and its type, decoration, and color. Since these three factors that may affect weathering are inherent properties of glass, we consider them as categorical variables. By using the chi-square test it is possible to find out which factors are correlated with glass weathering. Based on the results obtained from the chi-square test, a column table can be obtained. For the column table, the φ column correlation coefficient is calculated, which is an indicator of the categorical data correlation measure.

(1) Chi-square test

The chi-square test is a statistical method that obeys the standard normal distribution N(0,1). Under a certain confidence level and degree of freedom, the degree of agreement between the theoretical frequency and the actual frequency is judged by comparing the probability values of the chi-square distribution function. According to the degree of agreement, the correlation between two categorical variables can be tested. The chi-square test is calculated as follows.

$$x^2 = \Sigma \frac{(f_o - f_e)^2}{f_e} \tag{1}$$

where fo denotes the actual frequency, fe denotes the theoretical frequency, and x2 is used to measure the degree of difference between the actual and theoretical frequencies. In SPSS software, the default original hypothesis is that there is no difference between theoretical and actual frequencies, i.e., there is no correlation between the two sets of variables. See Table 2 below.

Table 2: P-values for correlation tests of surface weathering with ornamentation, color and glass type

Influencing factors	Ornamentation	Color	Glass Type
P-value	0.056	0.507	0.020

The p-value for surface weathering and decoration is 0.056 greater than 0.05 and the original hypothesis is accepted, therefore for surface weathering and decoration is not relevant. The p-value for surface weathering and glass type was 0.020 less than 0.05, and the original hypothesis was rejected, so for surface weathering and glass type was relevant. For surface weathering and color, the p-value is 0.507 is greater than 0.05, and the original hypothesis is accepted, so for surface weathering and color are not correlated.

(2) φ-column correlation coefficient

The magnitude of the φ-column correlation coefficient indicates the degree of correlation between two variables. If there is correlation between the variables, the more their φ-column correlation coefficients converge to 1. The φ-column correlation coefficients are calculated based on the columns obtained from the chi-square test. φ-column correlation coefficient formula is as follows.

$$\boldsymbol{\varphi} = \sqrt{\frac{\boldsymbol{x^2}}{\boldsymbol{n}}} \qquad (2)$$

When φ is less than 0.3, it means that the correlation between the two variables is weak, and when φ is greater than 0.6, it means that the correlation between the two variables is strong. the larger the absolute value of φ, the stronger the correlation between the two variables. See Table 3 below.

Table 3: Surface weathering and grain, color and glass type φ values

Influencing factors	Ornamentation	Color	Glass Type
φ	0.29234	0.40326	0.30659

For surface weathering and ornamentation, the φ column correlation coefficient is 0.29234, showing a weak correlation. For surface weathering and glass type, the correlation coefficient is 0.30659, showing a weak correlation. For surface weathering and color, the correlation coefficient of φ column is 0.40326, which shows a weak correlation.

Combining the above results, it can be obtained that the correlation between ornamentation, color and glass type and surface weathering of artifacts is weak.

3.2 Analysis of statistical patterns of chemical composition content based on descriptive statistical methods

The statistical law of the content of chemical components with and without weathering on the surface of the artifact samples is solved in conjunction with the glass type. The concept of mean and variance is introduced to quantify each chemical composition in turn. By means and variances, we can draw box plots to obtain the statistical law of the content of chemical components with and without weathering on the surface of the artifact samples.

(1) Numerical characteristics

By distinguishing glass categories to calculate the average value of chemical composition content before and after weathering for each type of glass, the following formula was obtained.

$$A_i = \frac{X_i^d}{P_j} \qquad (3)$$

$X_1{}^d, X_2{}^d, X_3{}^d, \ldots, X_{14}{}^d$ □□ denotes the cumulative sum of the content of each chemical component of each glass weathered or unweathered, and d=1,2,3,4 denotes unweathered high potassium glass, weathered high potassium glass, unweathered lead-barium glass, and weathered lead-barium glass, respectively; □□ denotes the number of each glass weathered or unweathered; □□ denotes the average of unweathered high potassium glass, weathered high potassium glass, unweathered lead-barium glass, and weathered lead-barium glass. weathered lead-barium glass, and the average of four types of weathered lead-barium glass.

(2) Box plot visualization to analyze the statistical law of chemical composition content before and after weathering

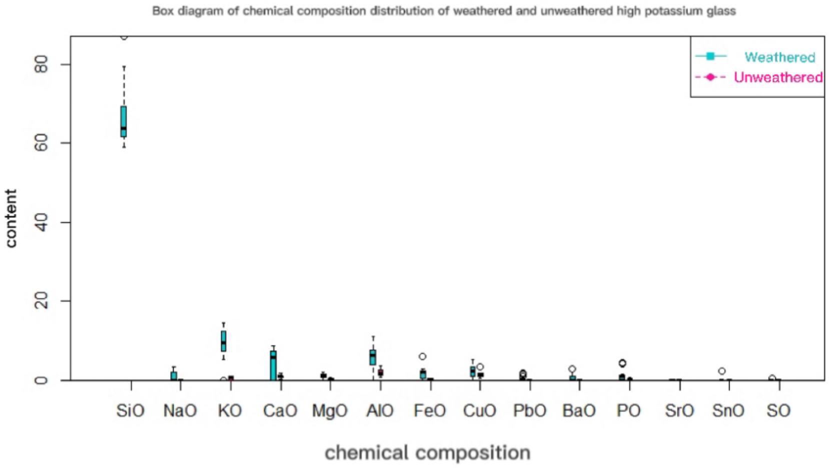

Figure 1: Case diagram of high potassium glass

The chemical composition content of different types of glass before and after weathering can be obtained as a box plot, as Figure 1 and 2.

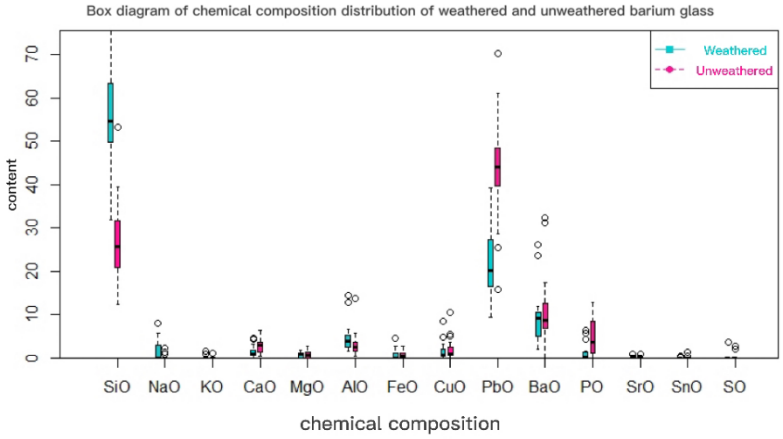

Figure 2: Case diagram of lead barium glass

It can be obtained from the box plot that after weathering of high potassium glass, the median decreases are potassium oxide, calcium oxide and phosphorus pentoxide; silicon dioxide, potassium oxide, iron oxide and sulfur dioxide polarization is enhanced. After weathering of lead-barium glass, the median decreasing chemical composition includes silica, potassium oxide, calcium oxide, iron oxide, copper oxide, lead oxide, and phosphorus pentoxide; silica, potassium oxide, magnesium oxide, copper oxide, barium oxide, and phosphorus pentoxide have enhanced polarization. Comparing before and after weathering, the anomalous values of high potassium glass after weathering are significantly smaller than those of lead-barium glass, indicating that the overall change of chemical composition is more stable during the weathering of high potassium glass.

3.3 Predictive analysis of chemical composition content before weathering based on multiple regression model

The statistical law allows the calculation of the mean percentage of each category before and after weathering, and the top three chemical components in the percentage change are analyzed. A fitted polynomial regression model was established to derive data for each variable regression coefficient equation, and then the P-value obtained by F-test was compared with the fixed parameter 0.05 to find the relationship between the chemical composition content.

(1) Establishment and analysis of multiple regression models

The CFTOOL toolbox in MATLAB was used to analyze and predict the multivariate data, which is essentially the least squares method. When a variable is linked to multiple variables, the optimal combination of multiple independent variables is used together to predict or estimate the dependent variable. However, the multiple regression model is a method of approximating values with polynomials, and MATLAB software for analysis and prediction tends to cause a lack of precision in the predicted values of the data and a large lack of randomness.

The post-weathering chemical composition content was used to predict its pre-weathering chemical composition content using the mean rate of change of the cumulative sum of chemical composition content before and after weathering and the proportion of various chemical composition content in its cumulative sum. Based on the two sets of average values of the pre-weathering and post-weathering chemical composition contents of the two types of glass, the difference between the two sets of average values of the pre-weathering and post-weathering chemical composition contents was found to obtain the percentage change of the various chemical compositions.

The following Figure 3.

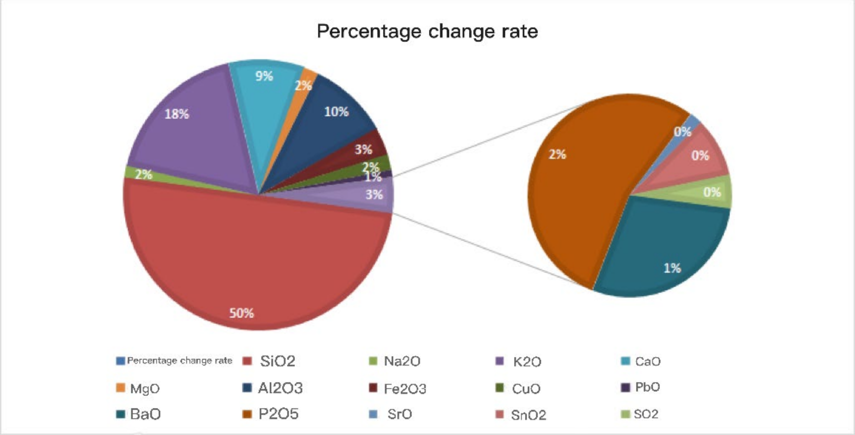

Figure 3: Percentage change of chemical content of high potassium glass before and after weathering

By calculation, the top three chemical components in high potassium glass are: potassium oxide, aluminum oxide, and silicon dioxide. Similarly, the top three chemical components in lead-barium glass are: silica, lead oxide, and phosphorus pentoxide. For the chemical components with large changes in content before and after weathering, the top three chemical components are determined, and the regression error interval model is proposed to make the results more accurate. The relationship of the top three chemical components in high potassium glass with the percentage change was obtained by the following equation.

$$f = P_{00} + P_{10}k + P_{11}l + P_{12}k^2 + P_{13}kl + P_{14}l^2 \tag{4}$$

where k,l,f denote the average rates of change of K2O,Al2O3,SiO2, respectively.

Similarly, the relationship between the chemical components accounting for the top three change rates in lead-barium glass is obtained by the following formula

$$z_2 = P_{22} + P_{21}x_2 + P_{23}y_2 + P_{24}x^2{}_2 + P_{25}x_2y_2 + P_{26}y^2{}_2 + P_{27}x^2{}_2y_2 + P_{28}x_2y^2{}_2 + P_{29}y^3{}_2 \tag{5}$$

where z2,x2,y2 denote the average rate of change of SiO2, PbO, P2O5, respectively.

4. Conclusion

In the multiple regression model, the polynomial does not fit the data well when the order is low, and causes strong shocks when the order is too high. In this paper, the two kinds of glasses obtained in the regression are 2 and 3. When the order is 3, the model does not form overfitting and can obtain good accuracy; when the order is 2, there is an error interval for each chemical content predicted by the fitting function, and the error interval is longer than the error interval when the order is 3. For the errors caused by the order, the model can be improved by regularization. In general, this paper provides effective mathematical models and methods for the identification of ancient glass types and the analysis of chemical composition, which are of great relevance to the archaeological work of ancient glass artifacts.

References

[1] Wang Chengxian, Tao Ying. Weathering of silicate glasses [J]. Journal of Silicates, 2003, 31((1): 78-85.

[2] Lu Yahui. Bayesian classification model based on random forest feature selection and application [C]. North China University of Water Resources and Hydropower, 2017.

[3] Chen, Neng-Mei; Liu, Xiao-Jing. Research on the classification model of channel scene based on random forest algorithm [J]. Journal of Chongqing University of Technology (Natural Sciences), 2017, 31(04):134-140.

[4] Zhu Yixingbo. Research and application of automotive user portrait based on improved K-mean clustering algorithm[C]. Jilin University, 2022.

[5] Gao A-F. Research on tillage depth prediction model based on improved random forest [C]. Changchun University of Technology, 2022.