A Study of The Chemical Composition of Glass Based on Regression Analysis

Xin Dong*, Zekai Wu, Chenxi Liu

School of Economics and Management, China University of Geosciences, Beijing, China

* Corresponding Author Email: dx2123922681@163.com

Abstract. With the continuous archaeological excavations, ancient glass objects have been unearthed in various parts of China, ranging from the Western Zhou Dynasty to the Ming and Qing Dynasties, providing a wealth of physical objects for the study of ancient Chinese glass, which has attracted the general attention of scholars at home and abroad. The study of the chemical composition of ancient glass is one of the most important aspects of ancient glass research. In this paper, a suitable classification model is established from the cluster analysis model to provide the correct classification of glass weathering caused by the external environment. This paper will analyze the relationship between weathering and glass type, decoration, and color. Combined with the type of glass, the statistical law of chemical composition content on the surface of cultural relics is analyzed and described, and the change law between the content of each component and the weathering degree is quantitatively reflected through regression. Based on the data of weathering point detection, inference will be made to predict the chemical composition content before weathering.

Keywords: Chi-square tests, regression fitting, variance analysis, glass artifacts, chemical composition prediction.

1. Introduction

Since the founding of China, with the continuous archaeological excavations, ancient glass artifacts have been unearthed in various parts of China, dating from the Western Zhou Dynasty to the Ming and Qing Dynasties, providing a wealth of objects for the study of ancient Chinese glass, which has attracted the attention of scholars at home and abroad. The study of the chemical composition and physical properties of ancient glass is a very important aspect of ancient glass research, which can provide scientific evidence for archaeological research and contribute to the study of the composition system, manufacturing date, preparation process and technological origin of ancient glass [1,2,9].

However, ancient glass is highly susceptible to weathering by the burial environment. The erosion of glass and atmospheric interaction is called weathering. During the weathering process, a large number of internal elements are exchanged with environmental elements, resulting in a change in the ratio of its composition, which affects the correct determination of its category. [3,4] At the same time, the weathering of glass reduces the surface finish and transparency, and often even forms weathering spots on the surface, which affects the aesthetics of the glass. Therefore, glass weathering has been one of the important topics of research at home and abroad [5,10].

In this paper, we investigate the relationship between surface weathering of glass artifacts and their glass type, decoration and color by first cleaning the data, followed by significance tests using chi-square tests for weathering and glass decoration, weathering and glass type, and weathering and glass color, respectively, with acceptance of the original hypothesis representing consistency between the two, and rejection of the original hypothesis representing variability between the two[6]. The statistical pattern of the content of chemical components with and without weathering on the surface of the samples was analyzed by combining the types of glass, firstly, descriptive statistics were analyzed, and the content of SiO2 was assumed to represent the degree of weathering of the glass, then the changes of the content of each component with the degree of weathering were analyzed qualitatively, and finally, regression analysis was performed, and the components that passed the significance test and the absolute value of the correlation coefficient was greater than 0.8 were selected. The regression equation was established by selecting the components that passed the significance test and the absolute value of correlation coefficient was greater than 0.8, and the regression equation was optimized by circularly eliminating the outliers to obtain a regression equation with a goodness-of-fit of more than 95%, which quantitatively reflected the statistical law between the content of each component and the degree of weathering. The regression equation of each component content can be known from the previous part, and the content of chemical components before weathering can be deduced according to the slope meaning of its linear equation.

2. Model assumptions and notation

2.1. Assumptions [7]

It is assumed that the SiO2 content represents the degree of weathering

Assume that the experimental data allow for a certain margin of error

Assume that the content of components with a correlation coefficient less than 0.8 with the degree of weathering does not change with the degree of weathering.

2.2. Notations

Important notations used in this paper are listed in Table 1.

Table 1 Notations

Symbols	Description
H_j	Original assumption
i	Sampling point number
y_0	Average value of SiO2 content before weathering
Y_{i-sio_2}	SiO2 content of a known weathering sampling point i
k_0	Slope of linear equation
$X_{i\text{-Chemical expression}}$	Content of a component at a weathering sampling point i
$x_{i\text{-Chemical expression}}$	Content of a component at a sampling point i before weathering

2.3. Data sources

After investigation and inquiry of relevant scientific research websites, we have obtained a batch of original data on weathered glass. This batch of data includes the data on the patterns, types, colors and surface weathering of 58 unearthed glass relics, and the proportion of the main components at the sampling points on 58 unearthed glass relics (the blank indicates that the component is not detected). The characteristics of these data are compositional, that is, the cumulative sum of the proportions of each component should be 100%, but the cumulative sum of the proportions of the components may not be 100% due to the detection methods and other reasons, The data between the accumulation of apparent proportion and 85%~105% are regarded as valid data.

3. Model construction and solving

3.1. The relationship between weathering and glass type, decoration and color

Whether weathering, glass type, ornamentation, and color belong to definite categories of data. Generally, when studying the interrelationship between two variables of a certain category, the form of a columnar table is used to construct indicators to measure the correlation between variables based on frequency information, and the induced correlation degree measurement indicators can only indicate the strength of the correlation degree, but not the direction of the correlation [8]. This paper uses SPSS to measure the degree of correlation according to the existing data.

Relationship hypothesis

In order to investigate the relationship between weathering and glass type, decoration and color, this paper will make the hypothesis Hj as follows [9].

H1: There is no significant difference between weathering and glass ornamentation.

H2: There is no significant difference between weathering and glass type.

H3: There is no significant difference between weathering and glass color.

Data were weighted to determine the frequency, and a column-wise cross-tabulation table was constructed.

Cardinality test results

The chi-square test focuses on the variability between the variables of a definite class. The degree of deviation determines the size of the chi-square value; the larger the chi-square value, the larger the deviation, and vice versa, the smaller the deviation.[5] In this paper, the existence of variability between type, ornamentation, color and surface weathering is investigated separately. The results are shown in the following Table 2 to 4.

Table 2 Weathering and glass decoration cardinality analysis results

Title	Name	Surface weathering (%)		Total	χ2	p
		No weathering	weathering			
Ornament	A	11(45.83)	9(30.00)	20(37.04)	5.747	0.056
	B	0(0.00)	6(20.00)	6(11.11)		
	C	13(54.17)	15(50.00)	28(51.85)		
Total		24	30	54		
* p<0.05 ** p<0.01						

Table 3 Weathering and glass type cardinality analysis results

Title	Name	Surface weatherin (%)		Total	χ^2	*p*
		No weathering	weathering			
Type	Lead Barium	12(50.00)	24(80.00)	36(66.67)	5.400	0.020*
	High potassium	12(50.00)	6(20.00)	18(33.33)		
Total		24	30	54		
* *p*<0.05 ** *p*<0.01						

Table 4 Weathering and glass color cardinality analysis results

Title	Name	Surface weatherin (%)		Total	χ^2	*p*
		No weathering	weathering			
Color	Light green	2(8.33)	1(3.33)	3(5.56)	6.287	0.507
	Light blue	8(33.33)	12(40.00)	20(37.04)		
	dark green	3(12.50)	4(13.33)	7(12.96)		
	Dark blue	2(8.33)	0(0.00)	2(3.70)		
	Violet	2(8.33)	2(6.67)	4(7.41)		
	Green	1(4.17)	0(0.00)	1(1.85)		
	Blue-Green	6(25.00)	9(30.00)	15(27.78)		
	Black	0(0.00)	2(6.67)	2(3.70)		
Total		24	30	54		

Based on the above analysis, the following conclusions were obtained.

Different surface weathering samples do not show significance ($p>0.05$) for ornamentation, and the original hypothesis is accepted, implying that different surface weathering samples all show consistency and no variability for ornamentation.

Different surface weathering samples showed significance for type ($p<0.05$), and the original hypothesis was rejected, implying that different surface weathering samples showed variability for type.

Different surface weathering samples did not show significance for color ($p>0.05$), and the original hypothesis was accepted, implying that different surface weathering samples all showed consistency for color and did not differ.

3.2. Analyze the statistical pattern of chemical content of the surface with and without weathering

(1) Data cleaning

Theoretically, the cumulative sum of the chemical composition content of each sampling point should be 100%, but due to the detection means and other external reasons, resulting in problems with the cumulative results, so this paper first initially delete the value of the cumulative sum of components outside 85%~105%, and finally a total of 67 remaining data.

(2) Normalized data processing[7]

The total sum of component percentages of the cleaned data varies, which has some influence on the problem study. Therefore, before the study of the statistical law of chemical composition content, the data will be normalized, the main steps are: first calculate the sampling points of each glass artifacts, that is, the total percentage of each column of data, and then use the specific components divided by the total percentage of components, you can get the normalized percentage data.

(3) Statistical law analysis

surface weathering can not exclude local shallow weathering, while some of the weathered artifacts in the surface of the unweathered area, so the weathering of the artifact sampling points and artifacts may be contradictory surface weathering. However, because the weathering of the artifact sampling points is known and determined, the differentiation of the sampling points based on the sampling points as the basis for the statistical law of chemical composition content.

The sampling points were statistically analyzed, and descriptive statistics for the content of each component before and after weathering were derived using SPSS(Table 5).

Table 5 Average values of chemical compositions in glass of different classes and weathering conditions

		SiO_2	Na_2O	K_2O	CaO	MgO	Fe_2O_3	Al_2O_3	PbO	CuO	BaO	P_2O_5	SrO	SnO_2	SO_2
A	Ⅰ	69%	1%	10%	5%	1%	2%	7%	0%	3%	1%	1%	0%	0%	0%
	Ⅱ	94%	0%	1%	1%	0%	0%	2%	0%	2%	0%	0%	0%	0%	0%
B	Ⅰ	56%	2%	0%	1%	1%	1%	5%	23%	1%	9%	1%	0%	0%	0%
	Ⅱ	26%	0%	0%	3%	1%	1%	3%	45%	2%	12%	5%	0%	0%	1%

(Note:A stands for high potassium glass, B for lead barium glass, Ⅰ means no weathering, Ⅱ means weathering.)

According to the analysis of the table, we can initially obtain the following patterns: first, SiO2 accounts for the largest proportion of each chemical composition and is the main element of glass, and there are more obvious fluctuations before and after weathering; second, the proportion of each component is different for different types of glass, and the components that change before and after weathering and fluctuate mainly are also different. In addition to the SiO2 component, the main components fluctuating in high potassium glass are K2O, CaO, Al2O3, Fe2O3, and lead-barium glass is mainly Na2O, CaO, PbO, BaO.

(4) Single factor variance analysis of SiO2 and weathering degree

In high potassium glass, it can be obtained by one-way ANOVA(Table 6).

Table 6 Single factor equation analysis results of high potassium glass

Source	*SS*	*df*	*MS*	*F*	*P*
Groups	0.25199	1	0.25199	47.61	3.57e-06
Error	0.08469	16	0.00529		
Total	0.33669	17			

The P value is less than 0.01, which rejects the original hypothesis, indicating that weathering or non weathering has a significant impact on SiO2 content in high potassium glass.

In lead barium glass, Table 7 can be obtained through one-way ANOVA:

Table 7 Single factor equation analysis results of lead barium glass

Source	*SS*	*df*	*MS*	*F*	*P*
Groups	1.10389	1	1.10389	88.78	2.107e-12
Error	0.58437	47	0.01234		
Total	1.68825	48			

The P value is less than 0.01, and the original hypothesis is rejected, indicating that weathering or non weathering conditions have a significant impact on SiO2 content in lead barium glass.

(5) Qualitative analysis of SiO2 and the change of content of some components

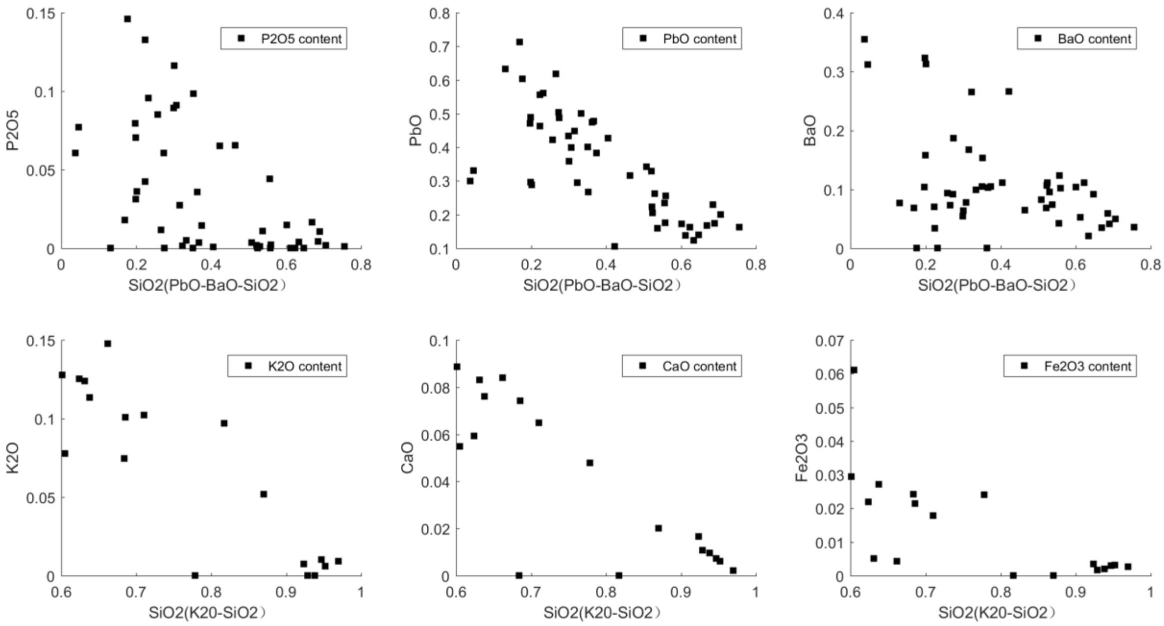

Figure 1 Analysis of SiO2 content and variation of content of some components

The percentage of SiO2 can reflect the size of surface weathering, and the SiO2 content of high potassium glass type increases with the increase of weathering, from 69% to 94%.[8] And the opposite is true for lead-barium glass type, SiO2 content decreases with the increase of weathering degree, from 56% to 26%, so we take SiO2 content percentage in the horizontal coordinate. Meanwhile, according to the preliminary qualitative analysis of correlation according to the chart, we selected three components with greater correlation content as vertical coordinates, P2O5, PbO, BaO in lead-barium glass, and k2O, CaO, Fe2O3 in high potassium glass.

According to Figure 1, the percentage of SiO2 increases with increasing weathering, and the percentage of P2O5, PbO, and BaO decreases in the high potassium glass type. The percentage of SiO2 decreases, and the percentage of k2O, CaO, and Fe2O3 increases with increasing weathering in the lead-barium glass type, and the relationship is approximately linear.

(5) Regression analysis

It was assumed above that the variation of SiO2 content represents the magnitude of weathering, and in order to determine the correlation between each component in high potassium and lead-barium glass and the degree of weathering, Pearson correlation analysis was performed for each component in high potassium and lead-barium glass, respectively, and the results of some high potassium glasses were obtained as shown in Table 8 below.

Table 8 Correlation analysis of some components of high potassium glass

Chemical composition	Correlation coefficient	Significance test
k_2O	-0.873	0.01
CaO	-0.821	0.01
Fe_2O_3	-0.844	0.01
PbO	-0.418	Failed
P_2O_5	-0.457	Failed

The components that pass the significance test and the absolute value of correlation coefficient is greater than 0.8 will show significant changes with the weathering degree, while the other components show smaller changes when the weathering degree changes, so they are not studied here. According to the above rules, three components of K2O, CaO and Fe2O3 were selected in high potassium glass; similarly, three components of P2O5, PbO and BaO were selected in lead barium glass, and the above six components were considered to be correlated with SiO2 content in high potassium and lead barium glass, respectively.

The k2O, CaO, and Fe2O3 components in high potassium glass and P2O5, PbO, and BaO components in lead-barium glass showed significant changes with weathering, while the other components showed less changes with weathering. In order to better present the statistical patterns of the above six components with weathering degree, regression analysis was performed separately.

Step 1: The regression was fitted with the regress function in MATLAB and returned the slope, intercept and goodness of fit of the linear fit function.

Step 2: The rcoplot(r,rint) function is used to plot the residuals of the data after the regression() fit and identify the outliers.[6]

Step 3: Manually eliminate the outliers among them and perform the fitted regression again, cyclically. Until the goodness of fit is greater than 95%, the slope, intercept and goodness of fit of the linear fit function are output (Figure 2).

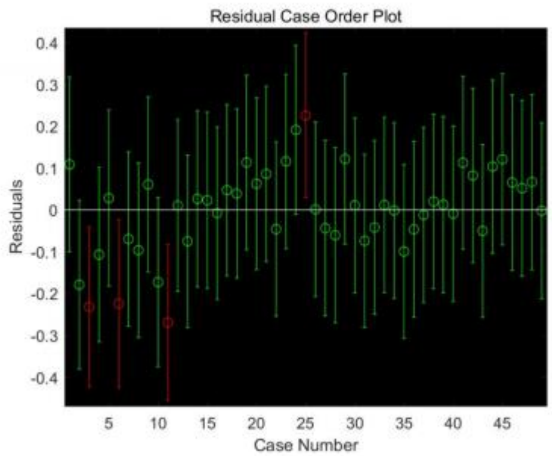

Figure 2 Residual error diagram of fitting data

The specific results are shown in the following Table 9.

Table 9 Six component regression equations

Six component regression equations	
k_2O	$Z_{k_2O} = 0.3403 - 0.3534\, X_{k_2O}$
CaO	$Z_{CaO} = 0.2230 - 0.2271\, X_{CaO}$
Fe_2O_3	$Z_{Fe_2O_3} = 0.0716 - 0.0730\, X_{Fe_2O_3}$
P_2O_5	$Z_{P_2O_5} = 0.0702 - 0.0991 X_{P_2O_5}$
PbO	$Z_{PbO} = 0.7185 - 0.8575\, X_{PbO}$
BaO	$Z_{BaO} = 0.3187 - 0.4447\, X_{BaO}$

The goodness of fit of the above six components is more than 95%, and the fitted regression equation with SiO2 can be derived, which can show the changes of the contents of the above six components under different weathering degrees and reflect their statistical laws.

3.3. Presumed chemical composition content before weathering

The first step because the SiO2 content percentage can reflect the size of the surface weathering, so the SiO2 content percentage before weathering with the average value instead. Set the glass before weathering SiO2 content percentage average y0, according to the statistical analysis of the above question is known high potassium glass y_0 = 69%, lead barium glass y_0 = 56%.

The second step if you want to speculate a sampling point before the weathering chemical composition content percentage, now know a weathering sampling point SiO2 content percentage, set SiO2 content for $Y_{i\text{- SiO2}}$, where i is the serial number of heritage sampling points

The third step to set a weathering sampling point a component content percentage for $X_{i\text{ - chemical expression}}$, where i for the heritage sampling point serial number, chemical expression that a component of the chemical formula, for example, 02 sampling point BaO content for $X_{02\text{ - BaO}}$

The fourth step to set a weathering sampling points a component of the chemical composition before weathering content percentage as

$$x_{i-\text{Chemical expression}} \tag{1}$$

Where i is the serial number of the artifact sampling point, and the chemical expression is the chemical formula of the requested component.

From the above analysis, the regression equation of SiO2 with the percentage content of each component is known. According to the meaning of the slope of the linear equation to infer the percentage content of chemical components before weathering. The chemical equation for the percentage content of a component before weathering that is sought is.

$$x_{i-\text{Chemical expression}} = \frac{y_0 - Y_{i-sio_2}}{k_0} + X_{i-\text{Chemical expression}} \tag{2}$$

(2) Chemical composition content solving

According to the above equation of high potassium glass and lead barium glass components one by one analysis, due to the limited space, only part of the data content is shown, and finally obtained the following Table 10 results.

Table 10 Percentage of chemical content of pre-weathering fraction

Type	Sampling Points	(SiO2)	(Na2O)	(K2O)	(CaO)
High Potassium	07	69%	0%	8.42%	6.50%
High Potassium	09	69%	0%	9.84%	6.58%
High Potassium	10	69%	0%	10.77%	6.56%
High Potassium	12	69%	0%	10.77%	6.56%
High Potassium	22	69%	0%	8.97%	6.96%
Lead Barium	02	56%	0%	1.05%	2.34%
lead barium	08	56%	0%	0%	1.48%
lead barium	08 Severe weathering	56%	0%	0%	3.25%
Lead barium	36	56%	2.27%	0.14%	0.38%
Lead Barium	38	56%	1.40%	0%	0.69%

4. Conclusion

In this paper, the relationship between weathering condition and glass decoration, type and color were analyzed by using chi-square test, respectively, and finally only type showed variability for weathered samples. Descriptive statistics were performed on the data to qualitatively analyze the patterns. Then ANOVA was performed on SiO2 content and weathering degree, assuming that SiO2 content indicates the weathering degree of glass, and six chemical compositions were selected by calculating the correlation coefficient between each chemical composition and SiO2 content, which were P2O5, PbO, BaO selected from lead-barium glass, and k2O, CaO, Fe2O3 selected from high potassium glass. The regression equation of composition and SiO2 content was constructed to quantify the pattern. The regression equation was optimized by circularly eliminating outliers, and the regression equation with a goodness-of-fit of more than 95% was finally obtained. Using the regression equation of the content of each component, the content of chemical components before weathering was deduced according to the meaning of the slope of the linear equation.

In this paper, by using known data and single factor analysis of variance, we creatively propose that sio2 content is used to represent the weathering degree of glass. On this basis, the statistical laws of the main components of high potassium glass and lead barium glass changing with the weathering degree of glass are studied and deduced respectively, and the relevant regression equations are established. It provides new ideas and solutions for related research.

References

[1] Li Fei, Li Qinghui, Gan Fuxi, Zhang Bin, Cheng Huan Sheng. Proton-excited X-ray fluorescence analysis of the chemical composition of a batch of Chinese ancient glasses[J]. Journal of Silicates,2005(05):581-586.

[2] Wang Chengxian, Tao Ying. Weathering of silicate glasses[J]. Journal of Silicate,2003(01):78-85.

[3] Wang Jie, Li Mo, Ma Qinglin, Zhang Zhiguo, Zhang Meifang, Wang Julin. Study on the weathering of an octagonal lead barium glassware in the Warring States Period [J]. Glass and Enamel, 2014, 42 (02): 6-13

[4] Gan Fuxi, Zhao Hongxia, Li Qinghui, Li Ling, Cheng Huansheng. Scientific and technological analysis and research on glass products unearthed in the Warring States Period in Hubei Province [J]. Jianghan Archaeology, 2010 (02): 108-116+151+0

[5] Fang Xiangzhong. Chi Square Distribution and Chi Square Test [J]. China Statistics, 2022 (05): 29-31

[6] Xu Xibao, Deng Yumin, Duan Fangzhen. Residual Analysis of Measurement Data [J]. Science and Technology and Innovation, 2015 (19): 80

[7] Yang Hanyu, Zhao Xiaoyong, Wang Lei. Overview of data normalization methods [J/OL]. Computer Engineering and Application: 1-11 [2022-11-19]

[8] Ruan Zheng. Single factor ANOVA in PSS software operation [J]. China Off campus Education, 2020 (24): 25+101

[9] Li Qinghui, Huang Jiaozhen, Li Fei, et al. Detection of chemical composition of a batch of ancient glass samples of the Warring States Period unearthed in China [J]. Cultural Relics Protection and Archaeological Science, 2006 (02): 8-13

[10] Gan Fuxi. The Development of Chinese Ancient Glass Technology [M]. Shanghai Science and Technology Press, 2005