Prediction of Chemical Composition Ratio of Ancient Glass before Weathering Based on LASSO Regression

Xinyi He *, Dengpan Chang

School of Northwestern Polytechnical University, Xi'an, China

* Corresponding author: hexinyi_npu@163.com

Abstract. Ancient glass is highly susceptible to weathering by the environment in which it is buried. During the weathering process, the internal elements of the glass undergo a large exchange with the environmental elements. This phenomenon leads to a change in the proportion of its composition. By analysing the chemical composition of ancient glass before and after weathering, the relationship between the changes can be identified, which means the original chemical composition can be predicted from the chemical composition of the glass after weathering. This is of great significance to the exploration and study of ancient artefacts. This paper first uses Spearman and Kendall correlation coefficient to identify the most influential factor on the weathering condition of glass from ornamentation, type and color, which eventually points to the glass type. This finding is secondly verified using a chi-square test. So the glass is divided into two types: high-potassium and lead-barium. By using the box plot to compare the chemical composition of high-potassium and lead-barium glass before and after weathering, it is found that there is a certain linear relationship between the chemical composition of glass before weathering and after weathering.LASSO regression was eventually applied to find the correspondence between the chemical composition of glass before and after weathering, thus enabling the prediction of the chemical composition of the glass before weathering.

Keywords: Ancient glassware, Spearman, Kendall, Chi-squaretest, LASSO.

1. Introduction

The Silk Road was a gateway for cultural exchange between China and the West in ancient times. Glass was a valuable material evidence of this early trade. In the West Asian and Egyptian regions, early glass was often made into bead-shaped ornaments and brought to China. After absorbing its techniques, Chinese craftsmen made ancient Chinese glass locally. As a result, its appearance is similar to that of exotic glass objects, but its chemical composition is not the same. The analysis of the physical and chemical properties and types of artefacts is of great significance to the study of this part of history [1-2].

2. The basic funamental of Spearman, Kendall and LASSO

2.1. Introduction to the Spearman and Kendall principle

Spearman correlation coefficient [3-4] is basically defined as the Pearson correlation coefficient between the ranks of two random variables X and Y in fixed order. X and Y are two sets of independent and identically distributed data, both with the number of elements N. The i th value in the two sets of random variables is represented by X_i and Y_i respectively. The two sequences are sorted in descending or ascending order at the same time, and the two element ranking sets x_i and y_i are obtained. The element x_i is the ranking of X_i in X, and the element y_i is the ranking of Y_i in Y. The difference of the corresponding elements in the set x and y forms the ranking difference set d, where $d_i = x_i - y_i (1 \leq i \leq N)$. The Spearman correlation coefficient between X and Y can be calculated by x, y or d. The calculation of r_s is as follows:

$$r_s = \frac{\sum_{i=1}^{n}(r_i-\bar{r})(s_i-\bar{s})}{\sqrt{\sum_{i=1}^{n}(r_i-\bar{r})^2}\sqrt{\sum_{i=1}^{n}(s_i-\bar{s})^2}} \tag{1}$$

In practical applications, the connection between variables is irrelevant. So, it can be calculated by the ranking difference set:

$$r_s = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)} \tag{2}$$

The range of values of Spearman correlation coefficient is $[-1,1]$. The greater the absolute value of r_s, the stronger the correlation. When Spearman similarity coefficient $r_s > 0$, it is considered that there is a positive correlation between the two groups of variables discussed. When the Spearman correlation coefficient $r_s < 0$, it is considered that there is a negative correlation between the two groups of variables discussed.

Multivariate Kendall co-efficiency test is a test method of multivariate co-efficiency proposed by Kendall and Babington in 1939[5].

Suppose there are k variables $X_1, X_2, \cdots, X_n$ and each variable has n observations. Suppose the j th variable $X_j = (X_{1j}, X_{2j}, \cdots, X_{nj})$, the hypothesis testing problem is:

H_0:k unrelated variables $\leftrightarrow$ H_1:k related variables

Let R_{ij} be the rank of X_{ij} in $(X_{1j}, X_{2j}, \cdots, X_{nj})$. Then there is $R_{i\cdot} = \sum_{j=1}^{k} R_{ij}, i = 1,2,\cdots,n$. All ranks sum to $R_{\cdot\cdot} = \sum_{i=1}^{n}\sum_{j=1}^{k} R_{ij} = kn(n+1)/2$. Then available statistics:

$$\sum_{i=1}^{n}\left(R_{i\cdot} - \frac{R_{\cdot\cdot}}{n}\right)^2 \tag{3}$$

After testing the hypothesis, the expression of Kendall synergy coefficient W is given as follows:

$$W = \frac{12\sum_{i=1}^{n} R_{i\cdot}^2 - 3k^2 n(n+1)^2}{k^2(n^3-n)} \tag{4}$$

Where the range of values of W is $0 \leq W \leq 1$.

2.2. Introduction to the cardinality test principle

A chi-square test is a hypothesis test in which the distribution of a statistic approximately follows a chi-square distribution when the null hypothesis holds [6]. The underlying idea is to compare the correlation between two sample rates and two categorical variables. Therefore, the association between two categorical variables can be determined by calculating the cardinality of the two categorical variables. The larger the chi-square value, the greater the association between the two categorical variables and the lower the independence. Conversely, the smaller the association between the two categorical variables and the higher the independence.

The statistical counts of the two characteristics of the sample under different categories form the four-compartment table of the chi-square test, and the four-compartment chi-square formula is shown in equation (5).

$$\chi^2 = \frac{(ad-bc)^2 N}{(a+b)(c+d)(a+c)(c+d)} \tag{5}$$

In the formula, a, b, c, d represent the frequency of each sample in the four grids in the four-grid table. N is the total number, that is, the sum of a, b, c, d frequencies.

2.3. Introduction to the LASSO principle

The LASSO regression method is to add a linear regression model to the L_1 canonical term as a penalty term [7]. The penalty function of the LASSO regression model improves the multicollinearity problem in the regression model by compressing the regression coefficients and shrinking the uncorrelated variables exactly to zero, as defined by [8-9]:

$$\hat{\beta} = arg\ \min_{\beta} \|Y - X\beta\|_2^2 + \lambda \sum_{j=1}^{n} |\beta_j| \tag{6}$$

In equation (6), the β is the n-dimensional parameter vector. X is the matrix. Y is the dependent variable. λ is the contraction parameter. The path of the solution can be effectively estimated by using the minimum angle regression algorithm (Lar's algorithm) in variable selection. Find a new path among the variables that have been selected. When advancing on this path, the correlation coefficients between the current residual and the selected variables are the same until a new variable with a larger correlation coefficient than the current residual is found. The LASSO algorithm is an L1-norm penalized least squares algorithm that solves the following optimization problems:

$$\hat{\beta}_L = arg\ \min_{\beta} (Y - X\beta)(Y - X\beta)' + \lambda \sum_{j=1}^{n} |\beta_j| \tag{7}$$

In equation (7), the hyperparameter $\lambda \in (0, \infty)$ is fixed. As the value of λ tends to ∞, the estimated coefficient narrows to zero. A larger λ indicates a larger penalty. The fewer variables are retained in the model at this point, to the extent that none of the variables in the model are significant at first. However, as the value of λ decreases, the number of variables retained in the model begins to increase and significant variables appear in the regression model in turn. It follows that the LASSO procedure can be interpreted as a stepwise regression with the estimated model parameters shrinking further towards zero relative to the hyperparameter λ[10]. When there are variables in the model that are uncorrelated or have small correlation coefficients, these will be screened out through the LASSO regression, thus improving the accuracy and interpretation of the model.

3. Results

The data used in this paper are from a batch of ancient glass artefacts studied by an archaeological research institute in China, including the basic classification information of glass artefacts, and the ratio of 14 main chemical composition including SiO_2, Na_2O, K_2O, CaO, MgO, Al_2O_3, Fe_2O_3, CuO, PbO, BaO, P_2O_5, SrO, SnO_2, SO_2 etc., of the classified high-potassium and lead-barium glass artefacts before and after weathering. Among these, the basic classification information for glass artefacts contains ornamentation, type, color and weathering condition. By cleaning the sample data, some missing and invalid data were eliminated. Part of the data are shown in Table 1 and Table 2.

Table 1. Part of the basic classification information of glass artefacts

Artefact number	Ornamentation	Type	Color	Weathering Condition
01	A	Lead Barium	Pale blue	Weathered
02	A	High Potassium	Blue and green	Unweathered
03	A	Lead Barium	Black	Weathered
04	B	High Potassium	Blue and green	Weathered
05	C	High Potassium	Blue and green	Unweathered
06	C	Lead Barium	Purple	Weathered
07	C	Lead Barium	Pale blue	Weathered
08	C	High Potassium	Dark green	Unweathered
09	C	Lead Barium	Green	Unweathered

Table 2. Part of chemical composition ratio of classified glass artefacts

Category		SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
High potassium glass	Unweathered point	69.33	0	9.99	6.32	0.87	3.93	1.74	3.87	0	0	1.17	0	0	0.39
	Weathered point	94.29	0	1.01	0.72	0	1.46	0.29	1.65	0	0	0.15	0	0	0
Lead barium glass	Unweathered point	37.36	0	0.71	0	0	5.45	1.51	4.78	9.3	23.55	5.75	0	0	0
	Weathered point	39.57	2.22	0.14	0.37	0	1.6	0.32	0.68	41.61	10.83	0.07	0.22	0	0

3.1. Analysis of factors influencing weathering condition of glass

3.1.1. Correlation analysis of influencing factors based on Spearman and Kendall

Firstly, the weathering condition of glass artefacts was analysed in correlation with glass ornamentation, type and color.

As the four indicators of weathering condition, glass type, ornamentation and color are all dummy variables but not continuous quantitative variables, causing Pearson correlation analysis cannot be used directly. Therefore, Spearman [11] and Kendall correlation coefficient analysis were used. The correlation analysis are performed more visually by plotting Spearman and Kendall correlation coefficient heat map.

For the correlation analysis, Spearman and Kendall correlation coefficients are calculated for the surface weathering condition of the glass artefacts in relation to type, ornamentation and color, and the heat maps are plotted, as shown in Fig.1 and Fig.2 below.

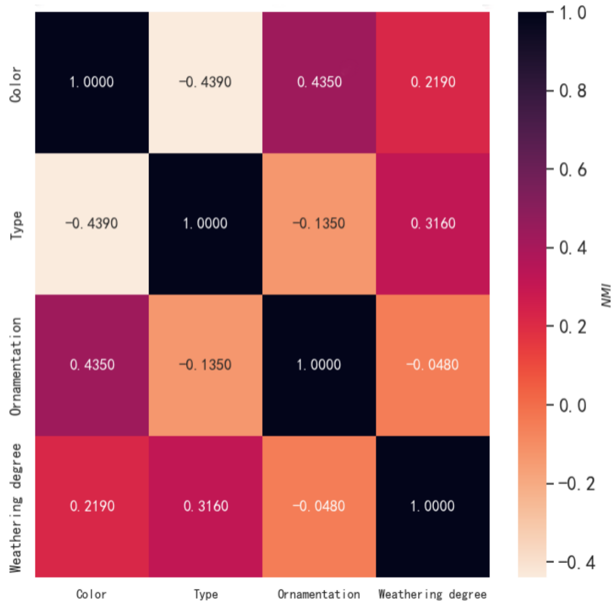

Figure 1. Spearman correlation coefficient heat map

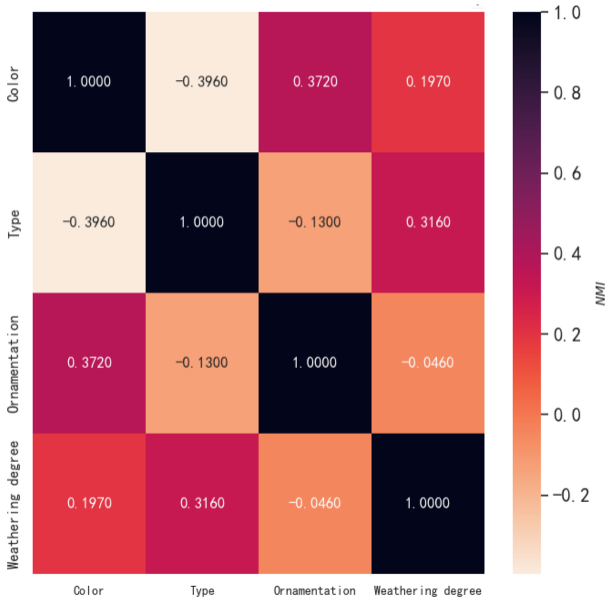

Figure 2. Kendall correlation coefficient heat map

Based on the heat map of the Spearman and Kendall correlation coefficient, it can be seen that glass type has the highest correlation with weathering condition, while other properties such as color and ornamentation do not correlate well with weathering condition.

3.1.2. Difference analysis of influencing factors based on Chi-square test

Secondly, the difference between the weathering condition of glass artefacts and type, ornamentation and color of glass was analyzed. For the difference analysis, Chi-square test was used to calculate the relationship between them, and the results are shown in Table 3 below.

Table 3. Chi-square test results

Title	Name	Weathering condition		Total	X^2	P
		Unweathered	Weathered			
Ornamentation	A	11	11	22	5.747	0.056
	B	0	6	6		
	C	13	15	28		
Type	Lead Barium	12	28	40	5.4	0.020
	High Potassium	12	6	18		
Color	Light green	2	1	3	6.287	0.506
	Pale blue	8	12	20		
	Dark green	3	4	7		
	Deep Blue	2	0	2		
	Purple	2	2	4		
	Green	1	0	1		
	Blue and green	6	9	15		
	Black	0	2	2		

As can be seen from Table 3, the significant p-value between weathering condition and glass type is less than 0.05. Chi-square test results show that there is no significant difference between weathering condition and ornamentation or color, but there is a significant difference between weathering condition and glass type.

The results of the correlation analysis obtained from the Spearman and Kendall coefficients and the difference analysis obtained from Chi-square test both indicate that glass type largely influences glass weathering condition.

3.2. Prediction model of pre-weathering composition ratio based on LASSO regression

Since the type of glass affects the weathering condition of glass to a large extent, the glass was divided into two types: high-potassium glass and lead-barium glass, and the chemical composition ratio of different types of glass before and after weathering was analyzed statistically [12]. The sample size, mean value, standard deviation, interquartile range, variance, minimum and maximum values of each chemical composition ratio before and after weathering were directly calculated according to the definition of each statistic, and the corresponding boxplots were drawn using the above results.

The statistical law of the chemical composition ratio of the high-potassium glass and the lead-barium glass before and after weathering is shown in Table 4.

Table 4. Statistical law of two types of glass

	Weathering	Sample size		Average		Standard deviation		Quartile distance		Variance		Minimum value		Maximum value	
		A	B	A	B	A	B	A	B	A	B	A	B	A	B
SiO_2	Before weathering	12	23	67.984	54.66	8.755	11.829	9.49	13.67	76.652	139.916	59.01	31.94	87.05	75.51
	After weathering	6	26	93.963	24.913	1.734	10.605	2.185	11.41	3.005	112.476	92.35	3.72	96.77	53.33
Na_2O	Before weathering	12	23	0.695	1.683	1.287	2.372	0.525	2.875	1.656	5.625	0	0	3.38	7.92
	After weathering	6	26	0	0.216	0	0.557	0	0	0	0.31	0	0	0	2.22
K_2O	Before weathering	12	23	9.331	0.219	3.92	0.31	4.7	0.275	15.369	0.096	0	0	14.52	1.41
	After weathering	6	26	0.543	0.133	0.445	0.24	0.728	0.24	0.198	0.058	0	0	1.01	1.05
CaO	Before weathering	12	23	5.332	1.32	3.092	1.285	3.535	1.355	9.563	1.65	0	0	8.7	4.49

	After weathering	6	26	0.87	2.695	0.488	1.66	0.393	2.162	0.238	2.755	0.21	0	1.66	6.4
MgO	Before weathering	12	23	1.079	0.64	0.676	0.547	0.978	1.045	0.457	0.299	0	0	1.98	1.67
	After weathering	6	26	0.197	0.65	0.306	0.706	0.405	1.163	0.094	0.499	0	0	0.64	2.73
Al_2O_3	Before weathering	12	23	6.62	4.456	2.492	3.262	2.793	2.87	6.208	10.644	3.05	1.42	11.15	14.34
	After weathering	6	26	1.93	2.97	0.964	2.634	1.022	2.152	0.93	6.939	0.81	0.45	3.5	13.65
Fe_2O_3	Before weathering	12	23	1.932	0.737	1.667	1.155	1.967	1.135	2.778	1.333	0	0	6.04	4.59
	After weathering	6	26	0.265	0.585	0.069	0.737	0.097	0.987	0.005	0.542	0.17	0	0.35	2.74
CuO	Before weathering	12	23	2.453	1.432	1.66	1.97	2.43	1.46	2.756	3.88	0	0	5.09	8.46
	After weathering	6	26	1.562	2.276	0.935	2.821	0.61	2.43	0.874	7.955	0.55	0	3.24	10.57
PbO	Before weathering	12	23	0.412	22.085	0.589	8.215	0.512	10.915	0.347	67.489	0	9.3	1.62	39.22
	After weathering	6	26	0	43.314	0	12.23	0	13.368	0	149.578	0	15.71	0	70.21
BaO	Before weathering	12	23	0.598	9.002	0.982	5.825	1.073	5.36	0.965	33.934	0	2.03	2.86	26.23
	After weathering	6	26	0	11.807	0	9.978	0	7.73	0	99.566	0	0	0	35.45
P_2O_5	Before weathering	12	23	1.403	1.049	1.434	1.847	0.603	1.185	2.056	3.412	0	0	4.5	6.34
	After weathering	6	26	0.28	5.277	0.21	4.197	0.193	7.275	0.044	17.612	0	0	0.61	14.13
SrO	Before weathering	12	23	0.042	0.268	0.048	0.243	0.078	0.28	0.002	0.059	0	0	0.12	0.91
	After weathering	6	26	0	0.418	0	0.265	0	0.365	0	0.07	0	0	0	1.12
SnO_2	Before weathering	12	23	0.197	0.047	0.681	0.127	0	0	0.464	0.016	0	0	2.36	0.44
	After weathering	6	26	0	0.068	0	0.269	0	0	0	0.073	0	0	0	1.31
SO_2	Before weathering	12	23	0.102	0.159	0.186	0.763	0.09	0	0.034	0.582	0	0	0.47	3.66
	After weathering	6	26	0	1.366	0	4.206	0	0	0	17.691	0	0	0	15.95

Annotation: A: high-potassium B: lead-barium

Boxplots of each chemical composition ratio in high-potassium glass and lead-barium glass before and after weathering are shown in Fig.3. The number on the horizontal axis represents weathering condition, that is, "1" represents before weathering, and "2" represents after weathering.

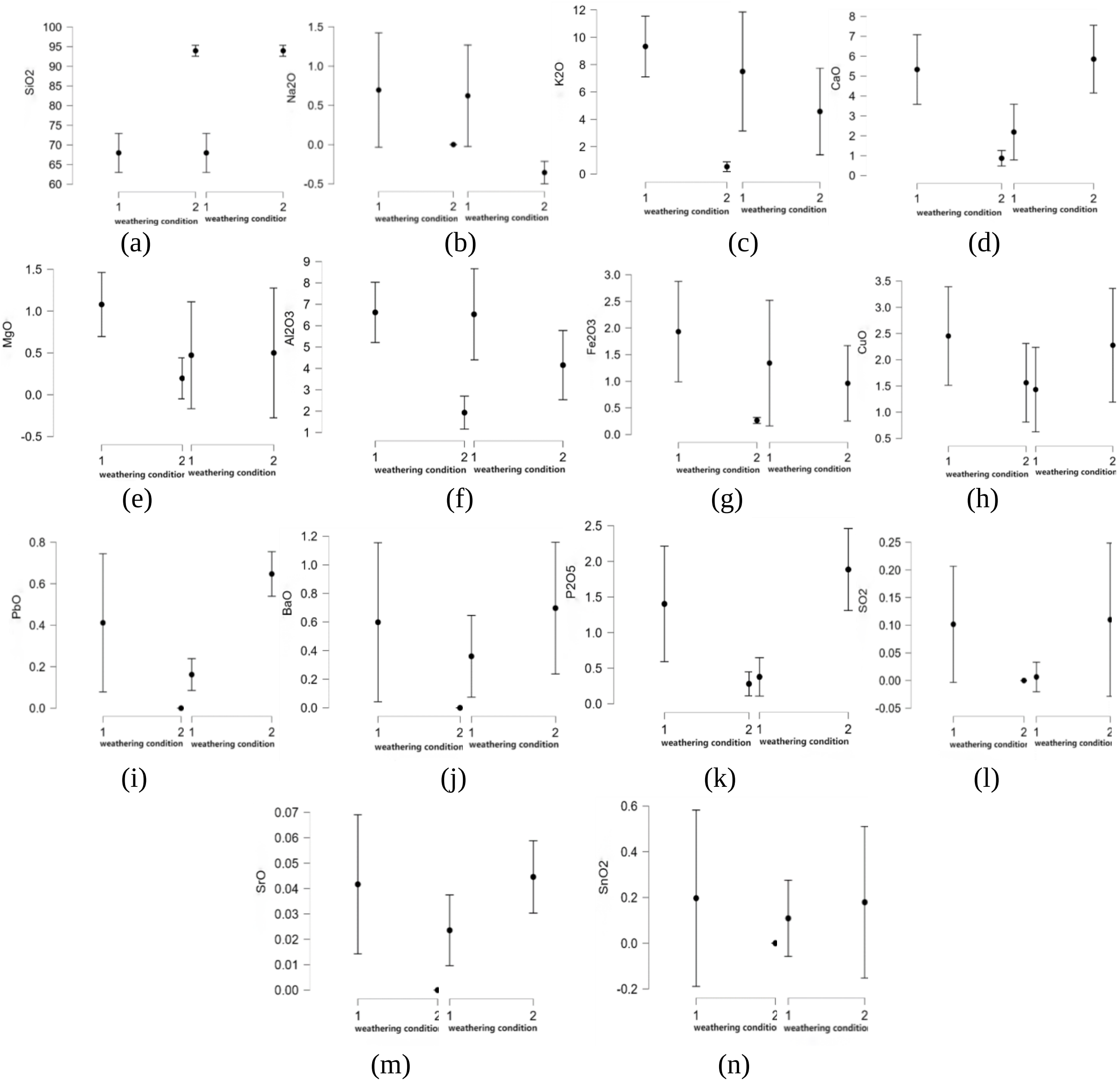

Figure 3. Boxplots of high-potassium glass and lead-barium glass

As can be seen from Fig.3, for high-potassium glass, except for the ratio of SiO_2 rised after weathering, other chemical composition ratios decreased significantly. For lead-barium glass, the ratios of SiO_2, Na_2O, K_2O, Al_2O_3, Fe_2O_3 all decreased after weathering, while the ratio of MgO essentially unchanged, and the ratios of other chemical components increased significantly.

This indicates that the chemical composition ratio of glass changes significantly before and after weathering, and the change direction of different chemical composition ratio varies from one glass type to another. It also confirms the previous results that the type of glass largely influences the weathering condition ratio of glass and that different types of glass need to be studied and analyzed separately.

From the boxplots, it was observed that there seemed to be a certain linear relationship between the chemical composition ratio of different types of glass before and after weathering. Therefore, the ratio of each chemical composition after weathering is taken as the independent variable and the ratio of each chemical composition before weathering is taken as the dependent variable. LASSO regression is used to find out the relationship between the changes in the chemical composition ratio of glass before and after weathering, and to predict the chemical composition ratio of glass before weathering.

LASSO regression analysis was carried out by determining the value of λ through the cross-validation method. λ value is chosen to minimize the mean square error of LASSO model. The λ value and regression coefficient plots were used to determine the variables selected by the model, where variables with zero normalization coefficient could be considered to be eliminated by LASSO regression model. The equations and predictions of the LASSO regression model were obtained, and the retained and removed variables were listed.

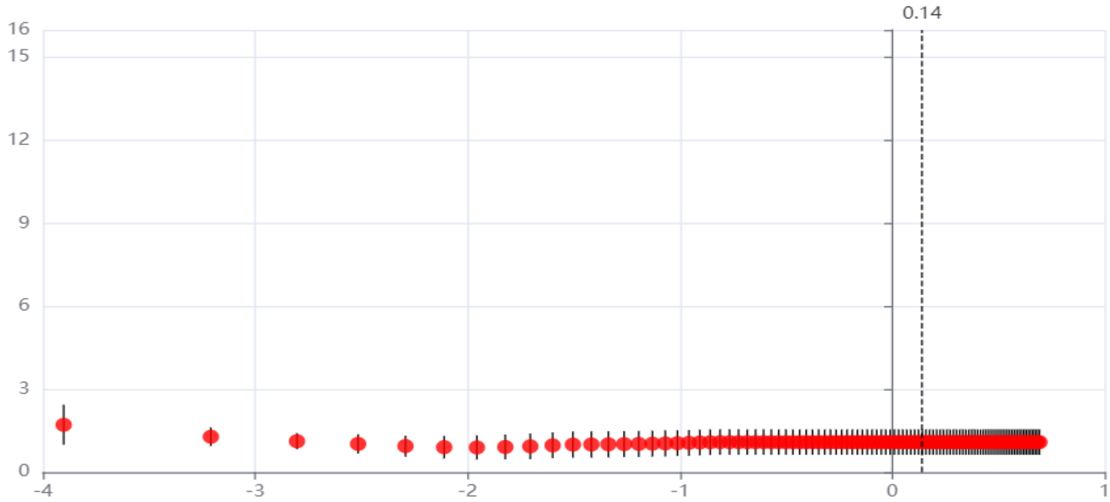

Figure 4. Cross-validation in LASSO regression

The cross-validation adjustment parameters can be obtained from Fig.4. The ordinate value of the above figure is the mean square error of the model, and the abscissa value is the λ logarithm value. The λ value was determined to be 0.14 by cross-validation. The unweathered ratio of SiO_2 was used as the dependent variable, and the remaining weathered components' ratios were used as independent variables for LASSO regression analysis. The following table demonstrates the model coefficients, when the coefficient of the standardized variable in the model is 0, it means that this variable was excluded from the model.

Table 5. LASSO model coefficients for predicting SiO_2 ratio before weathering

Variable name	Standardisation factor	Non-standardized coefficients	R^2
Intercept	52.153	52.356	0.628
SiO_2	0.217	0.225	
Na2O	-10.951	-11.845	
K_2O	-11.207	-13.936	
CaO	-1.217	-1.119	
MgO	-8.385	-9.846	
Al_2O_3	2.14	2.528	
Fe_2O_3	3.04	4.266	
CuO	0.434	0.295	
PbO	-0.055	-0.092	
BaO	0	0	
P_2O_5	0.717	0.54	
SrO	6.391	11.193	
SnO_2	0	-3.37	
SO_2	-0.361	-0.378	

According to the coefficients of the LASSO regression model for unweathered SiO_2 ratio in Table 5, the intercept, SiO_2, Na_2O, K_2O, CaO, MgO, Al_2O_3, Fe_2O_3, CuO, PbO, P_2O_5, SrO, SO_2 variables after weathering were retained ,while the BaO, SnO_2 variables were eliminated.

The unstandardized formula for the model is obtained as

$$y = 52.356 + 0.225 \times C_{SiO_2} - 11.845 \times C_{Na_2O} - 13.936 \times C_{KO} - 1.119 \times C_{CaO} - 9.846 \times C_{MgO} + 2.528 \times C_{Al_2O_3} + 4.266 \times C_{Fe_2O_3} + 0.295 \times C_{CuO} - 0.092 \times C_{PbO} + 0.540 \times C_{P_2O_5} + 11.193 \times C_{SrO} - 3.370 \times C_{SnO_2} - 0.378 \times C_{SO_2}$$

Similarly, regression analysis was conducted on the content of other chemical components according to LASSO regression model, and the non-standardized regression coefficients of each chemical component ratio before weathering were obtained as shown in Table 6 below. The horizontal axis is the chemical composition ratio before weathering to be predicted, and the vertical axis is the coefficient of the chemical composition ratio after weathering obtained by Lasso regression model.

Table 6. LASSO model coefficients for each chemical component ratio before weathering

	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
λ	0.141	2	0.182	0.263	0.081	2	0.182	2	0.081	0.061	0.121	0.141	0.02	2
Intercept	52.356	1.344	-0.008	1.66	0.81	5.199	0.046	1.782	25.968	7.927	1.929	0.15	-0.039	0.139
SiO_2	0.225	0	0.03	0.013	0.002	0	0.005	0	-0.207	-0.053	-0.015	0	-0.001	0
Na2O	-11.845	0	-0.446	-0.664	-0.206	-0.009	0	0	6.191	6.078	-0.089	0.02	0	0
K_2O	-13.936	0	7.233	2.473	0	0	1.266	0	-6.715	0	1.829	0	0	0
CaO	-1.119	0	-0.26	0.077	0	0	0.369	0	1.242	-0.455	0.199	0	-0.019	0
MgO	-9.846	0	-0.632	0	0	0	-0.001	0	7.373	3.961	-0.648	0	0.172	0
Al_2O_3	2.528	0	0	0	-0.06	0	0	0	-1.253	-0.881	-0.023	0	-0.037	0
Fe_2O_3	4.266	0	0.346	0.16	0	0	-0.266	0	-0.139	-0.97	-0.244	0	0.047	0
CuO	0.295	0	0.99	0.289	0.023	0	0.101	0	-1.664	-0.483	0.034	0	0.091	0
PbO	-0.092	0	-0.007	-0.001	0.001	0	-0.007	0	0	0.05	-0.027	0.001	0	0
BaO	0	0	-0.005	-0.056	0	0	0	0	0.169	0.075	0	0	0.007	0
P_2O_5	0.54	0	0.194	0	0.026	0	0.006	0	-1.106	-0.043	0.073	0	-0.02	0
SrO	11.193	0	-2.515	-0.352	-0.228	0	0.038	0	-0.574	-3.28	0	0	0.105	0
SnO_2	-3.37	0	-0.872	-1.212	0	0	-0.229	0	2.867	3.166	0	0	0.141	0
SO_2	-0.378	0	0.081	0.106	0	0	0	0	-0.894	0.402	0.027	0	-0.017	0
Prediction results	well	better	well	well	well	better	well	better	well	well	well	better	well	better

According to the coefficients obtained by Lasso regression in Table 6, a non-standardized prediction equation is established. From the proportion of chemical components after glass weathering in Table 2, the proportion of chemical components before glass weathering is predicted. The predicted results are shown in Table 7.

Table 7. Prediction value before weathering

Category		SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO	P_2O_5	SrO	SnO_2	SO_2
High potassium glass	Actual	69.33	0	9.99	6.32	0.87	3.93	1.74	3.87	0	0	1.17	0	0	0.39
	Predicted	58.837	0.003	8.404	5.166	0.641	5.109	1.641	2.780	0.008	0.130	1.409	0.019	-0.069	0.130
Lead barium glass	Actual	37.36	0	0.71	0	0	5.45	1.51	4.78	9.3	23.55	5.75	0	0	0
	Predicted	35.016	0.134	0.876	0.911	0.966	5.190	1.196	4.782	8.448	19.674	4.820	0.010	0.065	0.109

By comparing Table 2 and Table 7, which contain only partial research data, it can be seen that the relationship between the predicted value and the actual value of SiO2 before weathering is shown in Fig.5 below.

Figure 5. Prediction result of SiO_2

LASSO regression analysis was carried out to determine the weight values of each chemical composition ratio and establish the non-standardized prediction equation, using the ratio of each chemical component after weathering as the independent variable, and the ratio of each chemical component before weathering as the dependent variable. Finally, the prediction accuracy obtained by random sampling verification reached 95.23 %. In summary, this model can better predict the ratio of each component before weathering.

4. Conclusion

In this paper, firstly Spearman and Kendall correlation coefficients were calculated and heat maps were plotted for correlation analysis, concluding that the type of glass has the highest correlation coefficient with weathering condition, while ornamentation and color have little correlation with it. Then Chi-square test was used to carry out for the difference analysis between them, concluding that there is no significant difference between surface weathering condition and ornamentation or color, but there is a significant difference with type. Secondly, the glass was divided into two types, high-potassium and lead-barium, and boxplots were drawn to analyse the statistical patterns of each chemical composition ratio before and after weathering, and then derived the pattern of change in chemical composition ratio before and after weathering. Finally, the LASSO regression method was used to establish a non-standardized prediction equation by the ratio of each chemical component after weathering as the independent variable, and the ratio of each chemical component before weathering as the dependent variable. Through random sampling verification, the prediction accuracy of the LASSO equation reached 95.23 %. Therefore, this model can better predict the ratio of each component before weathering.

References

[1] Gan Fu Xi. The origin and development of glass in ancient China [J]. Journal of Nature, 2006 (04):187 - 193+184.

[2] Chen Shuyu. The origin and development of ancient glass in China [J]. Heritage Identification and Appreciation, 2019 (04): 44 - 45.

[3] Yu Qun, Huo Xiaodong, He Jian, Li Lin, Zhang Jianxin, Feng Yuyao. Trend prediction of power grid outages in China based on Spearman's correlation coefficient and system inertia [J/OL]. Chinese Journal of Electrical Engineering: 1-12 [2023-02-13]. http: //kns.cnki.net/kcms/detail/11.2107.TM.20220824.1625.012.html.

[4] Yang Pan, Lu Lu, Wang Jibao, et al. Application of spearman rank correlation coefficient method based on principal component analysis in water quality analysis of the Yangtze River main stream [J]. Environmental Engineering, 2019, 37 (8): 76 - 80. doi: 10.13205/j.hjgc.201908014.

[5] Liu, Qian-Yu, Tang, Jia-Yin. Statistical test for consistency of product failure mechanism in accelerated testing based on Kendall's coefficient of cooperation [J]. Journal of Hubei University (Natural Science Edition),2021, 43 (6): 671 - 677. doi: 10.3969/j.issn.1000 - 2375.2021.06.011.

[6] Liu YASU, Wang ZH, Li JW, et al. A cardinality test-based method for Android malicious application detection [J]. Journal of Beijing University of Technology,2019,39(3):290-294. doi: 10.15918/j. tbit1001 - 0645.2019.03.011.

[7] Zhang Lijin, Wei Xia Yan, Lu Jiaqi, et al. LASSO regression: from explanation to prediction [J]. Advances in Psychological Science, 2020, 28 (10): 1777 - 1788, back insert 1 - back insert 3.

[8] Zhu, H. L., Li, P. P. Analysis of factors influencing fiscal revenue in Anhui Province based on ridge regression and LASSO regression [J]. Journal of Jiangxi University of Technology, 2022, 43 (1): 59 - 65. DOI: 10.13265/j.cnki.jxlgdxxb.2022.01.009.

[9] Feng, Haiqi, Zhao, Yumei, Wang, Linlin. Analysis of factors influencing the retail sales of social consumer goods in Shandong Province - based on LASSO regression model[J]. Journal of Ludong University (Natural Science Edition), 2021, 37 (4): 310 - 314. doi: 10.3969/j.issn.1673 - 8020.2021.04.005.

[10] Wang Kaiyi, Yang Sheng, Guo Caiyun, et al. Research on spectral variable selection method based on LASSO algorithm [J]. Journal of Analysis and Testing, 2022, 41 (3): 398 - 402, 408. doi: 10.19969/j.fxcsxb.21070503.

[11] Li Jiayi. Applicability of Spearman's rank correlation coefficient method for analyzing automatic water quality monitoring in Jinjiang River basin based on comprehensive water quality pollution index [J]. Journal of Environmental Science, 2022, 41 (4): 88 - 92.

[12] Fang Ni, Ju Guopeng, Hui Jiaojiao, et al. Urban domestic waste generation prediction based on LASSO regression and ARIMA model--a case study of Baoji city [J]. Henan Science, 2022, 40 (1): 98 - 103. doi: 10.3969/j.issn.1004 - 3918.2022.01.015.