Analysis and identification of ancient glass cultural relics based on fuzzy C-mean clustering

Gongcheng He, ZeLin Zhao *, Yujie Lin

Shaoxing University Yuanpei College, Zhejiang, China

* Corresponding Author Email: 2298194363@qq.com

Abstract. Glass is a valuable physical evidence of ancient Silk Road trade, and ancient glass in China was made locally by absorbing its technology, and it is difficult to distinguish from foreign glass products in appearance, but the internal chemical composition of the two is different. Therefore, this paper focuses on the chemical composition data of the collected ancient glass artifacts, and uses fuzzy C-means aggregation to analyze and identify the composition of ancient glass. The results show that the classification of ancient glass artifacts using fuzzy C-mean clustering is reasonable. The experimental results and experimental data processing methods provide a new way to study ancient Chinese glass artifacts.

Keywords: Fuzzy C-mean clustering, Ancient glass, Chemical composition.

1. Introduction

Ancient Chinese glass artifacts are a testimony to history and a cultural legacy. During the evolution of history, the production process and raw materials of glass relics have been improved and enhanced, which has led to great changes in the chemical composition of glass relics. Based on the chemical composition of glass artifacts, their degree of differentiation can be analyzed and classified [1-4].

Through chemical analysis, we can determine the main raw materials of glass artifacts, such as elements of silicon, carbon, and oxygen. Also, we can analyze trace elements in glass artifacts, such as aluminum, manganese, and barium, which can provide us with information about when, where, and how the glass artifacts were made. Depending on the chemical composition of glass artifacts, they can be classified into different types. For example, carbonate glass is made of calcium carbonate, magnesium, sodium, silicate and other raw materials, has a high hardness and low temperature melting point, and is usually used to make vessels, bottles and other containers. Borosilicate glass, on the other hand, is made of boric acid, silicate and other raw materials, has a lower hardness and higher temperature melting point, and is often used to make glass fibers and optical components, etc. As well as the analysis of the chemical composition of glass artifacts, the degree of differentiation can also be determined [5-9]. The degree of differentiation refers to the degree of mixing of raw materials in a glass artifact and the uniformity of its internal structure. During the production of a glass artifact, the raw materials are subjected to high temperatures that gradually mix and blend together to form a homogeneous structure. Therefore, the higher the degree of differentiation, the more uniform the chemical composition, the more stable the structure, and the greater the durability of the glass artifact.

In conclusion, the chemical composition of ancient Chinese glass artifacts can be analyzed to understand the degree of differentiation and the source of raw materials, and to classify glass artifacts based on this information This is of great significance for the conservation and study of ancient Chinese glass artifacts. In this paper, we will analyze the chemical composition data of ancient glass artifacts from the perspective of multivariate statistics using fuzzy c-mean clustering [10], which provides an important reference for us to better understand the history and culture of ancient Chinese glass artifacts.

2. Sample and chemical composition

The chemical composition data of some ancient high-potassium glass and lead-barium glass products were collected in this paper, and the samples are described in Table 1. Table 1 Description of glass artifact samples. And to facilitate the discussion, the compositional data of the samples were selected and pooled in Table 2.

Table 1. Description of glass artifact samples

Artifact Number	Ornamentation	Color	Sample Type
1 ~ 12	A C	Blue green, light blue, dark green	High potassium unweathered glass
13 ~ 18	B	Blue-Green	High Potassium Weathered Glass
19 ~ 46	A C	Light green, light blue, purple, dark green, blue-green, black, none	Lead barium weathered glass
47 ~ 58	A C	Dark green, light blue, green, purple	Lead barium unweathered glass

Table 2. List of chemical composition of glass artifact sampling (%)

Heritage Sampling Sites	SiO_2	Na_2O	K_2O	CaO	MgO	Al_2O_3	Fe_2O_3	CuO	PbO	BaO
1	69.33	0	9.99	6.32	0.87	3.93	1.74	3.87	0	0
2 parts2	61.71	0	12.37	5.87	1.11	5.5	2.16	5.09	1.41	2.86
2 parts1	87.05	0	5.19	2.01	0	4.06	0	0.78	0.25	0
3	65.88	0	9.67	7.12	1.56	6.44	2.06	2.18	0	0
4	61.58	0	10.95	7.35	1.77	7.5	2.62	3.27	0	0
5 parts1	67.65	0	7.37	0	1.98	11.15	2.39	2.51	0.2	1.38
5 parts2	59.81	0	7.68	5.41	1.73	10.05	6.04	2.18	0.35	0.97
6	59.01	2.86	12.53	8.7	0	6.16	2.88	4.73	0	0
7	62.47	3.38	12.28	8.23	0.66	9.23	0.5	0.47	1.62	0
19	36.28	0	1.05	2.34	1.18	5.73	1.86	0.26	47.43	0
20 Severe weathering points	4.61	0	0	3.19	0	1.11	0	3.14	32.45	30.62
20	20.14	0	0	1.48	0	1.34	0	10.41	28.68	31.23
21	33.59	0	0.21	3.51	0.71	2.69	0	4.93	25.39	14.61
22	29.64	0	0	2.93	0.59	3.57	1.33	3.51	42.82	5.35
23 Unweathered spots	53.79	7.92	0	0.5	0.71	1.42	0	2.99	16.98	11.86
24 Unweathered spots	50.61	2.31	0	0.63	0	1.9	1.55	1.12	31.9	6.65
25 Severe	3.72	0	0.4	3.01	0	1.18	0	3.6	29.92	35.45
25	19.79	0	0	1.44	0	0.7	0	10.57	29.53	32.25
26	68.08	0	0.26	1.34	1	4.7	0.41	0.33	17.14	4.04

2.1. Data pre-processing

According to the collected chemical composition data, the data are compositional in nature, i.e., the sum of the component ratios should be 100%, but due to the detection means, the data with the sum of the component ratios between 85% and 105% are considered as valid data, i.e., the data with the sum of the component ratios deviating from this range are excluded. The proportional sum of sampling point 12 was 79.47, and the proportional sum of sampling point 14 was 71.89, which deviated from the interval to be excluded, while the rest of the sampling points met the requirements of valid data and were selected for retention.

As some of the artifacts in Table 1 have missing colors, the missing artifacts are numbered 22, 32, 37 and 46. Observations show that the weathering of the surface of such artifacts are all weathered, so it is presumed that the missing colors are due to the severe weathering of the surface of the artifacts and the existence of a large weathered layer resulting in the inability to observe the color of the glass artifacts themselves. Therefore, such artifacts are classified as a special category.

From the data in Table 2, it can be seen that for lead-barium glass, some of the heritage sampling points are not indicated whether they are weathering points, so this paper first uses the logistic classification model based on the regularization of the elastic network to determine the weathering of the heritage sampling points.

2.2. Classification rules for basic information of glass artifacts

In order to explore the classification law of the basic information of glass artifacts on glass types, firstly, high potassium and lead-barium glass were taken as the major categories, and then subdivided into three categories of decoration, color and surface weathering, subdivided each subcategory of information, made information tables, and finally, the frequency percentages of each information in the subcategories were statistically analyzed, and the frequency percentage distribution was obtained as shown in Figure 1.

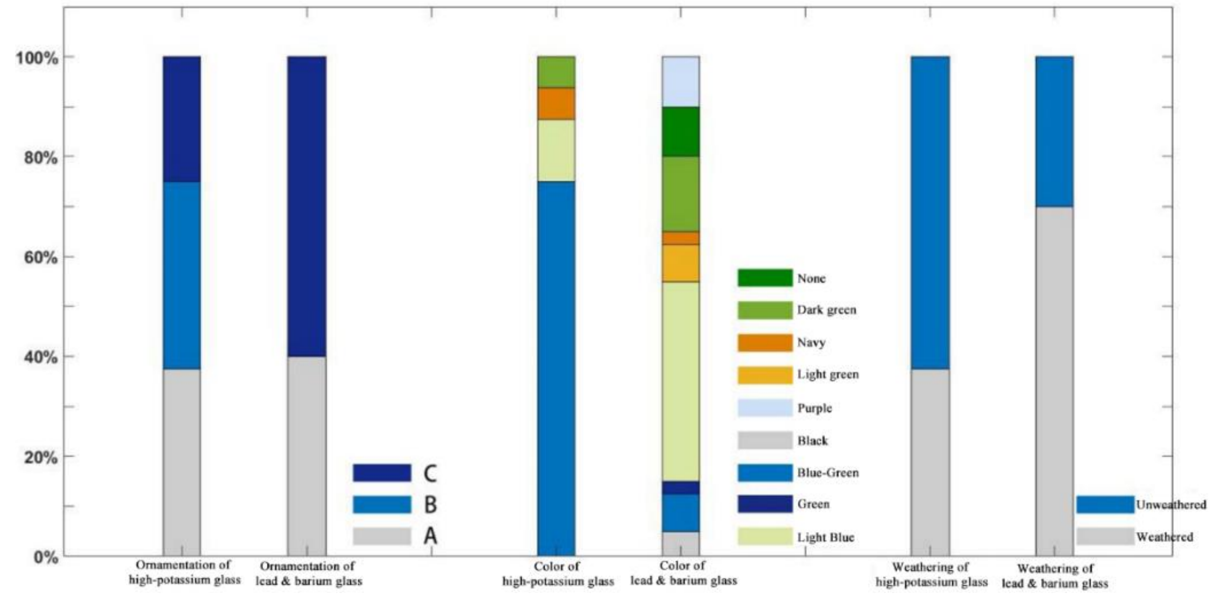


Figure 1. Frequency percentage distribution

From the frequency distribution of high potassium glass and barium lead glass, the following classification patterns are obtained: from the decoration, the decoration types of high potassium glass are divided into A, B and C, but there is no decoration type B for barium lead glass; from the color, there are more color types of barium lead glass than high potassium glass, and high potassium glass is mainly blue-green, while light blue lead barium glass appears more frequently and there are differences; from the surface weathering, high potassium glass has more no weathering, and barium lead glass has more serious surface weathering. The no-weathering situation is more frequent, and the surface weathering situation of lead-barium glass is more serious

3. Fuzzy C-means clustering

3.1. Principle of fuzzy C-means clustering

Fuzzy C-Means (FCM) is a common fuzzy clustering algorithm. It is somewhat similar to the traditional c-means clustering algorithm, but has some differences. The basic principle of FCM is to assign each data point to several clustering centers, and each clustering center is responsible for generalizing some data points. The relationship between each data point and the clustering centers is represented by a probability value indicating under the influence of which clustering center the data point is more important.

In the FCM algorithm, each data point has a probability distribution vector indicating the probability that it belongs to each clustering center. The weight of each clustering center is also a probability distribution vector indicating the probability of the data points it is responsible for generalizing. As shown in Figure 2.

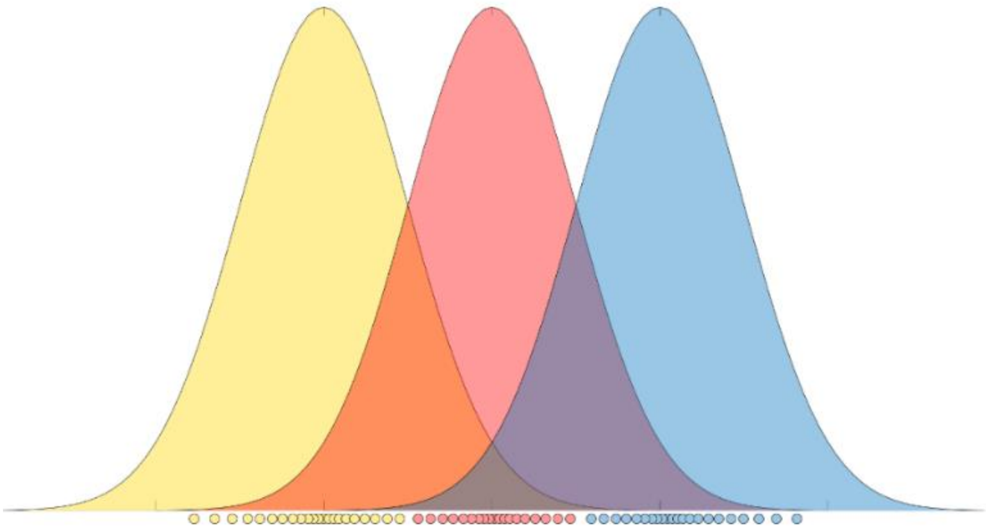


Figure 2. Fuzzy C-mean clustering schematic

The goal of the FCM algorithm is to minimize the distance between each data point and the clustering center responsible for generalizing it. This distance is computed by a distance function, usually using Euclidean distance or cosine similarity. During the calculation, the probability distribution vector of data points and the weight vector of clustering centers need to be adjusted iteratively until a certain termination condition is satisfied. The termination condition can be that the position of the cluster centers no longer changes, or the value of the distance function no longer changes, or the number of iterations reaches an upper limit.

In general, the advantage of FCM algorithm is that it can handle different types of data and get more accurate results. However, it has a high computational complexity and can be slow when dealing with large data sets.

3.2. Algorithm steps for fuzzy C-mean clustering

Step1. Initialization: First, choose K initial clustering centers, where K is the number of clusters.

Step2. Calculate the distance: Calculate the distance from each sample point to the cluster center.

Step3. Calculate fuzzy factor: Calculate the fuzzy factor of each sample point. The fuzzy factor indicates the probability that the sample points belong to each cluster.

Step4. Update the cluster center: update the cluster center using fuzzy factors.

Step5. Repeat steps 2-4 until the cluster centers no longer change or the maximum number of iterations is reached.

Step6. Output result: output the clusters to which each sample point belongs.

4. Results

The different glass sampling points were divided into two categories of high potassium and lead-barium, and then subdivided into whether the weathering situation, and the chemical compositions of the four groups were obtained as the sets of n_1, n_2, n_3 and n_4 samples as the input, and the clusters of the sample sets as the output, and the cluster centers of were randomly selected.

4.1. The objective function is as follows

$$K_m = \min \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| n_i - c_j \right\|^2, \quad 1 \le m < \infty \tag{1}$$

Where, m is the number of classification clusters i.e. subclassifications, N is the number of samples, C is the cluster class center, c_j indicates the j th cluster center, n_i represents the i th glass sampling point, u_{ij} represents the affiliation of the glass sampling point n_i to the cluster center c_j

4.2. Cluster center c_j Calculation

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot n_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{2}$$

4.3. Calculating affiliation u_{ij}

The degree of affiliation indicates the degree to which the chemical composition of the glass sampling site belongs to the subclassification classification, and the more the affiliation is close to 1, the more likely it is to be classified as a class. The calculation formula is as follows.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left\| n_i - c_j \right\| n^{\frac{2}{m-1}}}{\left\| n_i - c_k \right\|)^{\frac{2}{m-2}}} \right.} \tag{3}$$

For each glass sampling point n_i , the sum of its affiliation for the cluster is 1. Therefore, the termination condition of the iteration is

$$\max_{ij} \left\{ \left| u_{ij}^{k+1} - u_{ij}^{k} \right| \right\} < \varepsilon \tag{4}$$

Where, k is the number of iteration steps and ε is the error threshold. After iterating the above equation, the degree of affiliation gradually stabilizes and reaches the optimal state, which is the local minimum of the objective function K_m . Python software, the cluster center matrix E_{ij} is obtained by clustering the high potassium glass weathering group as an example.

$$E_{ij} = \begin{bmatrix} 77.84544 & 94.32666 & 84.32079 \\ 0.012566 & 0.548603 & 7.486909 \\ 4.774147 & 0.873191 & 0.982442 \\ 1.237202 & 0.19708 & 0.804284 \\ 6.278144 & 1.937748 & 3.587511 \\ 2.401179 & 0.265566 & 9.37E-05 \\ 3.324014 & 1.56909 & 0.381343 \\ 1.11619 & 0.281772 & 1.037318 \end{bmatrix} \tag{5}$$

Where, i denotes the eight chemical components and j denotes the number of clusters. The solution yields the affiliation degree u_{ij}

$$u_{ij}=\begin{bmatrix} 0.000476 & 0.997582 & 0.001942 \\ 1.39E-05 & 0.999929 & 5.74E-05 \\ 3.94E-06 & 0.999978 & 1.78E-05 \\ 1.26E-05 & 0.999939 & 4.86E-05 \\ 0.000402 & 0.998178 & 0.00142 \\ 0.00116 & 0.993383 & 0.005457 \\ 0.011215 & 0.032454 & 0.956331 \\ 0.007431 & 0.002959 & 0.989609 \\ 1 & 2.09E-12 & 1.15E-11 \end{bmatrix} \tag{6}$$

According to the affiliation magnitude, four groups of glass sub-classification results were obtained, and the cluster analysis tree diagram was drawn as shown in Figure 3.

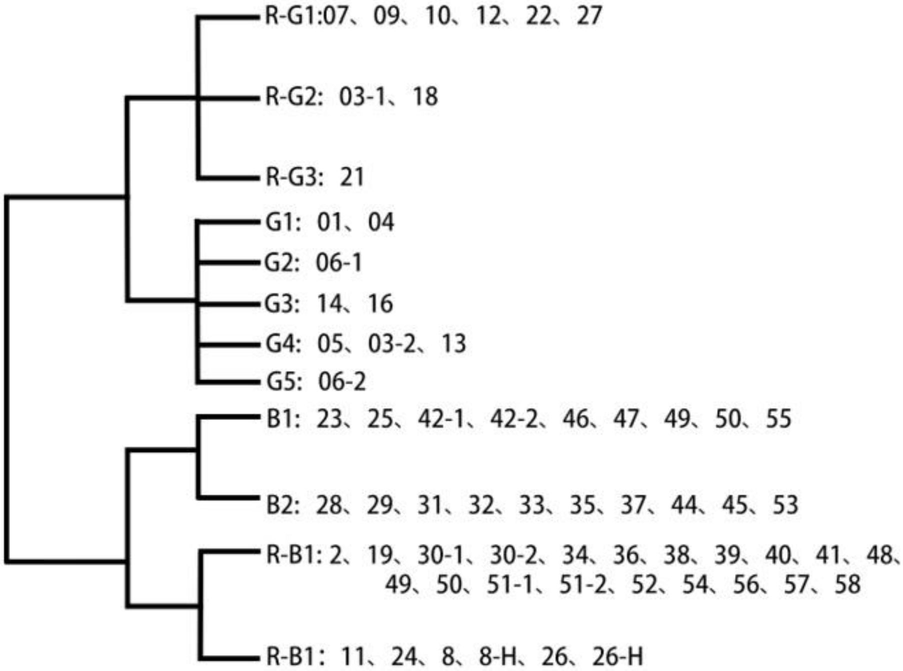
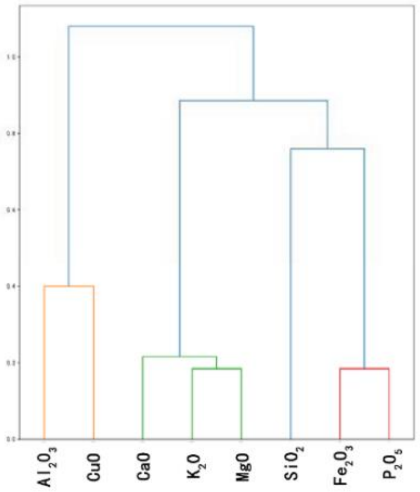


Figure 3. Cluster analysis tree diagram

In the above figure, $R-G$ indicates weathered high-potassium glass, G indicates unweathered high-potassium glass, B indicates unweathered lead-barium glass, and R B indicates weathered lead-barium glass. It is observed that the clustering method classifies weathered high-potassium glass into 3 classes; unweathered high-potassium glass into 5 classes; and unweathered and weathered lead-barium glass into 2 classes.

4.4. Analysis of the rationality of classification results

To verify that the above subclassification of chemical components is justified, the chemical components with significant differences were extracted from the classification results of the four groups of glasses. Hierarchical clustering of the components by their inter-component Spearman coefficients was considered, and the retention of individual features from each cluster was restricted by selecting thresholds. Thus the importance of each chemical component can be filtered by adjusting the threshold value. The data in the high-potassium weathered glass group were substituted into the hierarchical clustering model for solution, and the clustering tree diagram of chemical components and the heat map of Spearman's correlation coefficient were obtained as shown in Figure 4.



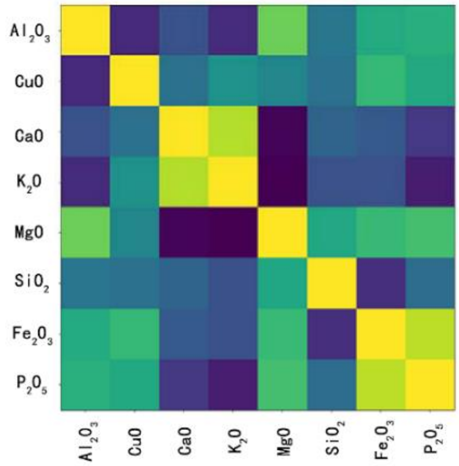


Figure 4. chemical composition clustering tree diagram and Spearman's correlation coefficient heat map

In the above figure, the variability of Al_2O_3 and CuO is small, i.e., the correlation is high; the correlation between CaO, K_2O and MgO is high; and the correlation between Fe_2O_3 and P_2O_5 is high. Then adjusting the threshold value, the chemical components with less variability were grouped into one category, and the hierarchical diagram of the importance of chemical components was obtained as Figure 5.

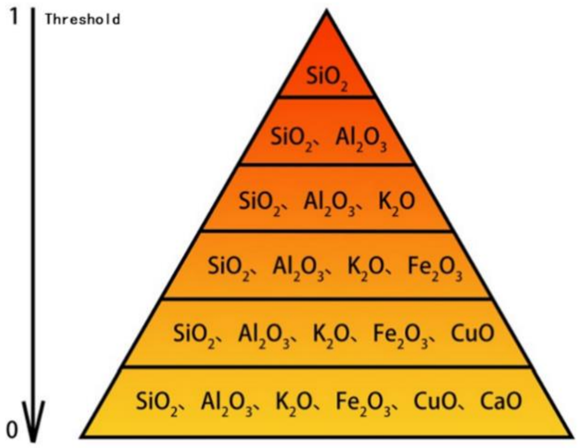


Figure 5. chemical composition importance level chart

In the pyramid, the number of occurrences decreases in importance from most to least. Among them, SiO_2 is the most important, and the others are Al_2O_3 , K_2O , Fe_2O_3 ,and CuO in that order. The above selected chemical composition data were extracted and the classification results of subclassification were used to do one-way ANOVA to determine whether the different classes of high-potassium weathered glass were different, as shown in Table 3.

Table 3. ANOVA for High Potassium Weathering Glass of Different Classes

Chemical composition	SiO_2	Al_2O_3	K_2O	Fe_2O_3	CuO
P value	0.043	0.149	0.021	0.001	0.047

As judged from the above table P 0.05, the above chemical compositions are significantly different except for Al_2O_3 , probably because the relative amount of variation in Al_2O_3 content is small, so the classification results are still considered reasonable. The naming of the glass clustering results based on the selected chemical compositions is shown in Table 4.

Table 4. Naming table for glass clustering results

Classification	Name
$R-G1$	$MgO(L)-SiO_2$ System Weathered high potassium glass
$R-G2$	$K_2O(H)-Fe_2O_3-SiO_2$ System Weathered high potassium glass
$R-G3$	$K_2O(H)-CaO-Fe_2O_3(H)-SiO_2$ System Weathered high potassium glass
$G1$	$SiO_2(H)$ System Unweathered high potassium glass
$G2$	$Al_2O_3(H)-SiO_2$ System Unweathered high potassium glass
$G3$	$K_2O(H)-CuO(L)-SiO_2$ System Unweathered high potassium glass
$G4$	$K_2O(H)-CuO(H)-SiO_2$ System Unweathered high potassium glass
$G5$	$Fe_2O_3(H)-SiO_2$ System Unweathered high potassium glass
$B1$	$PbO(H)-BaO(H)-SiO_2$ System Unweathered lead & barium glass
$B2$	$PbO(L)-BaO(L)-SiO_2$ System Unweathered lead & barium glass
$R-B1$	$Fe_2O_3(H)-PbO(H)-SiO_2$ System Weathered lead & barium glass
$R-B2$	$CuO(H)-BaO(H)-SiO_2$ System Weathered lead & barium glass

*(H) indicates high chemical content in the glass, (L) indicates low chemical content in the glass

5. Conclusion

The analysis of the chemical composition data of ancient high-potassium glass and lead-barium glass products and the analysis of the results of one-way ANOVA on fuzzy C-mean clustering showed that it is feasible to use fuzzy C-mean clustering to analyze the chemical composition data of ancient glass artifacts.

This cluster analysis allows for the continued subdivision of lead-barium silicate glass and potassium silicate glass into 12 major systems - $R-G1$, $R-G2$, $R-G3$, $G1$, $G2$, $G3$, $G4$, $G5$, $B1$, $B2$, $R-B1$, and $R-B2$. The results of the research and data analysis used in this paper help to further develop the culture and communication of ancient glass.

References

[1] Wang Y, Liu W, Lin Z. Component Analysis and Identification Model of Ancient Glass Products Based on Correlation Analysis[J]. Analytical Chemistry: A Journal, 2022, 1(1): 75-83.

[2] Yao S, Yang Nie W Z, Yang X. Study on Composition Analysis and Identification of Ancient Glass Products[J].

[3] Yifan Yuan, Shitong Li, Yiqing Huang. Study the Chemical Composition Content of Ancient Glass Products before Weathering[J]. Analytical Chemistry A Journal,2022,1(1).

[4] Origin of Chinese ancient glasses——study on the earliest Chinese ancient glasses[J]. Science in China (Series E: Technological Sciences),2006(06):701-713.

[5] FU Xiu feng, GAN Fu xi. Chemical composition analysis of some ancient glasses unearthed from southern and southwestern China [J]. Sciences of Conservation and Archaeology, ,2006(04): 6-13.DOI: 10.16334/j.cnki.cn31-1652/k.2006.04.002.

[6] ZHANG Yao ting, FANG Kai tai. Introductory of multivariate statistical analysis method [M]. Beijing: Science Press, 1982.

[7] Bethany Matthews, Bruce Arey, Carolyn Pearce, Albert Kruger. Characterization of Glass Alterations in Ancient Glass from Various Environments from Broborg, a Vitrified Swedish Hillfort[J]. Microscopy and Microanalysis,2020,26(S2).

[8] Green Lorna R., Alan Hart F.. Colour and chemical composition in ancient glass: An examination of some roman and wealden glass by means of ultraviolet-visible-infra-red spectrometry and electron microprobe analysis[J]. Journal of Archaeological Science,1987,14(3).

[9] Li Q, Dong J, Gan F. Research and discussion on chemical composition and technics of the early faience and glass artifacts unearthed from China[J]. J Guangxi University for Nationalities (Natural Science Edition), 2009, 15(4): 31-41.

[10] Askari Salar. Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: review and development[J]. Expert Systems with Applications,2020,165(prepublish).