# Composition Analysis and Subclass Division of Ancient Glass Artifacts Based on R-Q Type Clustering

Boyao Yang[1, #], Yue He[2, #], Ruijuan Wang[2, *]

[1]International College, Zhengzhou University, Henan, China, 450002

[2]School of Cyber Science and Engineering, Zhengzhou University, Henan, China, 450002

*Corresponding author: rjwang@zzu.edu.cn

[#]These authors contributed equally.

**Abstract.** The Silk Road served as a conduit for cultural exchange between ancient East and West, with glass being a valuable material evidence of early trade interactions. Archaeologists have classified ancient glassware primarily into two categories: high-potassium glass and lead-barium glass. This study focuses on the sub-classification of high-potassium glass. Firstly, the high-potassium glass is divided into two major categories, weathered and unweathered, based on their weathering properties. Subsequently, the R-Q clustering algorithm is employed to further subdivide the unweathered high-potassium glasswhile the same approach is applied to the weathered high-potassium glass. The results are validated through rational analysis and can be applied to other types of ancient glassware.

**Keywords:** Ancient Glass Artifacts, R-type Clustering, Q-type Clustering, Subclass Division.

## 1. Introduction

Glass is primarily composed of silica, with its main raw material being quartz sand. Due to the high melting point of silica, the addition of flux is necessary during the refining process. The choice of flux affects the chemical composition of the glass. For instance, potassium glass is often produced by firing with plant ash as a flux, resulting in a higher potassium content. Ancient glass, when preserved to the present day, undergoes weathering due to the influence of its storage environment, leads to changes in its chemical composition. By studying the rule of chemical composition content of ancient glass, it is helpful for experts to classify glass relics more accurately. It can improve the scientific nature and accuracy of cultural relics identification and has certain practical significance to some extent. Before our study, some scholars used X-ray diffraction and other technologies to study glass components [1-2], and some scholars also used various cluster analysis methods to classify glass products [3-5].

The R-Q clustering algorithm is a clustering algorithm for high-dimensional datasets. It exhibits good robustness when dealing with datasets containing abnormal data and noise [6-7].This study takes unweathered high-potassium glass as an example, utilizing the R-Q clustering model [8]to select appropriate chemical composition indicators and conduct sub-classification, yielding specific classification results. The sub-classification of other types of glass follows a similar approach. The data in this paper comes from Google Scholar. Ten glass samples, each of which has 14 chemical components, are selected for analysis.

## 2. R-Q Type Clustering

### 2.1. R-Type Clustering Method

(1): Similarity Measurement of Variables

The R-Type clustering method aggregates variables into several clusters based on their similarity, aiming to identify the primary factors influencing the system. The correlation coefficient is utilized to measure the similarity of chemical composition indicators [9-10].

Let $x_j$ denote the values of the variable $\left[x_{ij}, x_{2j}, \cdots, x_{mj}\right]^{\top} \in R^m \left(j = 1, 2, \cdots, p\right)$. It takes the number of sample points *m* and takes the number of chemical composition indicators *p*, with a maximum value of 14. Equation (1) represents the correlation coefficient between the variable $x_j$ and $x_k$, which serves as the similarity measurement for the chemical composition indicators.

$$r_{ij} = \frac{\sum_{i=1}^{m}\left(x_{ij} - \overline{x}_j\right)\left(x_{ik} - \overline{x}_k\right)}{\left[\sum_{i=1}^{m}\left(x_{ij} - x_j\right)^2 \sum_{i=1}^{m}\left(x_{ik} - x_k\right)^2\right]^{\frac{1}{2}}} \quad j, k = 1, 2, \cdots, p \tag{1}$$

In Equation (1), $\overline{x}_j = \frac{1}{m}\sum_{i=1}^{m} x_{ij}$

(2): Variable Clustering

In this paper, the function $linkage\left(X, 'average'\right)$ [11] in MATLAB is directly used to realize the clustering of chemical composition indexes.

### 2.2. Q-Type Clustering

(1): Similarity Measurement of Samples

Let $\Omega$ be the set of artifact sample points, and let distance $d\left(\cdot,\cdot\right)$ be a function of $\Omega \times \Omega \to R^+$. Let $a = \left[a_1, a_2, \cdots, a_p\right]^{\top}$, $b = \left[b_1, b_2, \cdots, b_p\right]^{\top}$ and use the Minkowski distance to calculate the similarity between each pair of sampled points [12]. This can be represented as:

$$d_q\left(a, b\right) = \left[\sum_{k=1}^{p}\left|a_k - b_k\right|^q\right]^{\frac{1}{q}}, q > 0 \tag{2}$$

We utilize the commonly employed Euclidean distance, which is a popular metric in Minkowski distance, to measure the similarity between samples. The formula for calculating this similarity is represented as Equation (3).

$$d_2\left(a, b\right) = \left[\sum_{k=1}^{p}\left|a_k - b_k\right|^2\right]^{\frac{1}{2}} \tag{3}$$

(2): Similarity Measurement between Clusters

Similar to the R-Type clustering method, the Q-Type clustering method also employs the class average method to calculate the distance between clusters. The formula for calculating this distance is represented as Equation (4).

$$D\left(G_1, G_2\right) = \frac{1}{n_1 n_2}\sum_{x_i \in G_1}\sum_{x_j \in G_2} d\left(x_i, x_j\right) \tag{4}$$

In Equation (4), $n_1$ and $n_2$ respectively represent the number of sample points in $G_1$ and $G_2$.

## 3. Results

The unweathered high-potassium glass exhibits 14 chemical composition indicators with a total of 10 sample points. The sample points are numbered from 01 to 10 in sequential order, with the respective numbers assigned as follows: 01, 03, 04, 05, 06, 13, 14, 16, 18, and 21. The correlation between the 14 chemical composition indicators is analyzed using the R-Type clustering method. Initially, the correlation coefficients are calculated for the proportions of each chemical composition indicator. Table 1 presents the correlation coefficient matrix for the second to eighth chemical composition indicators.

**Table 1.** Partial Correlation Coefficient Matrix

| | Na2O | K2O | CaO | MgO | Al2O3 | Fe2O3 | CuO | PbO |
|---|---|---|---|---|---|---|---|---|
| Na2O | 1.00 | -0.56 | -0.65 | -0.91 | 0.33 | -0.70 | -0.32 | -0.29 |
| K2O | -0.56 | 1.00 | 0.54 | 0.56 | -0.81 | 0.52 | -0.27 | -0.16 |
| CaO | -0.65 | 0.54 | 1.00 | 0.46 | -0.36 | 0.19 | -0.30 | -0.22 |
| MgO | -0.91 | 0.56 | 0.46 | 1.00 | -0.43 | 0.73 | 0.34 | 0.35 |
| Al2O3 | 0.33 | -0.81 | -0.36 | -0.43 | 1.00 | -0.08 | 0.09 | -0.19 |
| Fe2O3 | -0.70 | 0.52 | 0.19 | 0.73 | -0.08 | 1.00 | 0.17 | -0.04 |
| CuO | -0.32 | -0.27 | -0.30 | 0.34 | 0.09 | 0.17 | 1.00 | 0.91 |
| PbO | -0.29 | -0.16 | -0.22 | 0.35 | -0.19 | -0.04 | 0.91 | 1.00 |

The correlation coefficient matrix reveals strong correlations among certain chemical composition indicators. Consequently, the R-Type clustering method is applied to classify the 14 chemical composition indicators, selecting representative ones from each cluster.

The proportion data for each variable (chemical composition) is standardized. The similarity between each pair of chemical composition indicators is calculated using the correlation coefficient, while the similarity between clusters is measured using the class average method. A clustering dendrogram is generated using MATLAB, as depicted in Figure 1.

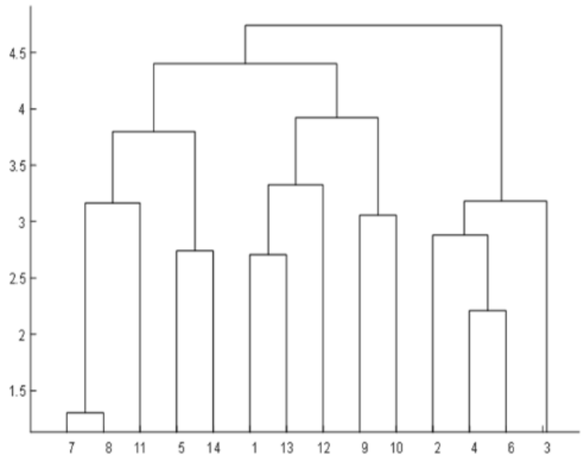


**Figure 1.** Clustering of Chemical Composition Indicators

According to the dendrogram-based clustering, the 14 chemical composition indicators can be divided into five clusters as shown in Table 2:

**Table 2.** Classification Diagram of Chemical Composition Indicators

| Category | Chemical Composition Indicators |
|---|---|
| Class 1 | 1 Silica dioxide、12 Strontium oxide、13 Tin oxide |
| Class 2 | 5 Magnesium oxide、14 Sulfur dioxide |
| Class 3 | 9 Lead oxide、10 Barium oxide |
| Class 4 | 7 Iron oxide、8 Copper oxide、11 Phosphorus pentoxide |
| Class 5 | 2 Sodium oxide、3 Potassium oxide、4 Calcium oxide、6 Aluminum oxide |

From these five categories of chemical composition indicators, one indicator is selected from each category for subsequent sample clustering analysis. In this study, the chosen indicators are 1, 14, 9, 7, and 3. Using these five chemical composition indicators, a Q-type clustering analysis is conducted on the 10 samples of high-potassium unweathered artifacts.

Initially, The proportion data of the chemical composition for each sample is standardized. The similarity between samples is measured using the Euclidean distance, while the distance between clusters is calculated using the class average method. A clustering dendrogram is generated by MATLAB, as depicted in Figure 2.

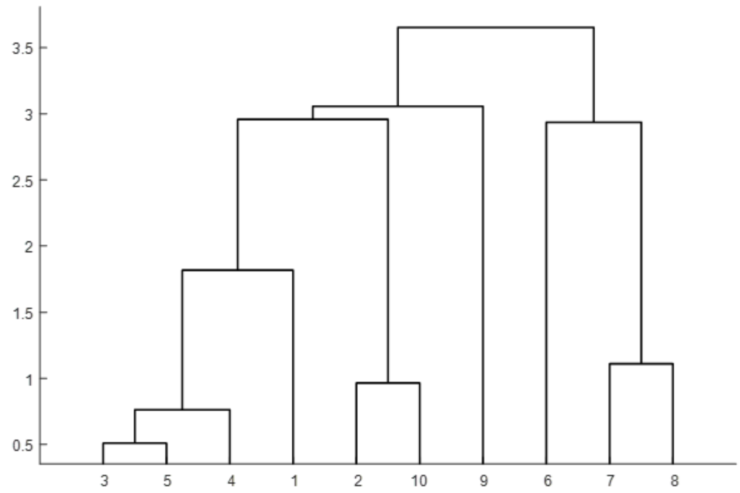


**Figure 2.** Dendrogram of Clustering Analysis for High-Potassium Unweathered Samples

Given the limited number of samples (only 10), the high-potassium unweathered glass is divided into three subclasses in this study. The classification of high-potassium unweathered artifacts is presented in Table 3:

**Table 3.** Clustering Results of High-Potassium Unweathered Artifacts

| R-Clustering Indicator Chosen | | Q-Clustering (Sample IDs) | |
|---|---|---|---|
| Class 1 | 1. Silica dioxide | Class 1 | 03、21 |
| Class 2 | 3.Sodium oxide | | |
| Class 3 | 7.Magnesium oxide | Class 2 | 01、04、05、06、13、16、18 |
| Class 4 | 9.Copper oxide | | |
| Class 5 | 14. Barium oxide | Class 3 | 14 |

Based on the clustering results of the high-potassium unweathered artifacts, the chemical composition indicators for each subclass are organized in Table 4. According to Table 4, it can be observed that the samples within each subclass exhibit small differences in the content of each chemical composition indicator, indicating high similarity. Therefore, the classification results can be considered reasonable.

**Table 4.** Chemical Composition Indicator Content for Each Subclass of Samples

| Category | ID | SiO2 | K2O | Fe2O3 | PbO | SO2 |
|---|---|---|---|---|---|---|
| Class1 | 03 | 74.38 | 8.78 | 2.16 | 0.83 | 0 |
| | 21 | 76.68 | 0 | 2.37 | 1 | 0 |
| Class2 | 01 | 69.33 | 9.99 | 1.74 | 0 | 0.39 |
| | 04 | 65.88 | 9.67 | 2.06 | 0 | 0.36 |
| | 05 | 61.58 | 10.95 | 2.62 | 0 | 0.47 |
| | 06 | 63.73 | 10.31 | 2.34 | 0 | 0.415 |
| | 13 | 59.01 | 12.53 | 2.88 | 0 | 0 |
| | 16 | 65.18 | 14.52 | 0.42 | 0.11 | 0 |
| | 18 | 79.46 | 9.42 | 0 | 0 | 0 |
| Class3 | 14 | 62.47 | 12.28 | 0.5 | 1.62 | 0 |

## 4. Conclusion

The R-Q clustering model was used for composition analysis and subclassification of ancient glass artifacts. By selecting appropriate chemical composition indicators using R-type clustering, the high-potassium unweathered glass was further classified into three subclasses using Q-type clustering. The final classification results were validated for their reasonability, demonstrating the robustness of the R-Q clustering model [13]. The next step is to add more data to refine our model so that it can be used to classify other chemical composition types of glass. Based on the model in this paper, we can continue to study the prediction model to predict the category of glass based on its chemical composition in the future. In view of the increasing amount of data, we also need to consider the efficiency and accuracy of the model, so that the model can quickly produce better classification results even if the input data is large.

## References

[1] Bouhenguel, R., Bounakhla, M., Draa, S. A., et al. (2018). Analytical investigations of ancient glasses from Morocco. Journal of Saudi Chemical Society, 22(8), 994-1001.

[2] Kassouf, R., & Yacoub, S. (2020). Analysis of glass and pottery shards by X-ray fluorescence spectroscopy in cultural heritage studies. X-Ray Spectrometry, 49(1), 1-11.

[3] Chen Zhihao, Ji Jingmin. Component analysis and identification of ancient glass products based on the hierarchical clustering model [J]. Modern Information Technology, 2023,7(08):122-125.

[4] Kaczmarczyk, A., Szczerba, M., Trenczek-Zajac, A., et al. (2018). Chemical composition and cluster analysis of roman glass bracelets from the lublin region (SE Poland). Archaeometry, 60(2), 251-266.

[5] Soliman, M. A., & Ali, M. A. (2019). Comparative Study of Different Clustering Algorithms for Analyzing Ancient Glass Data. International Journal of Computer Science and Information Security, 17(5), 75-82.

[6] Lei, J., Zhao, X., Wu, J., & Xu, R. (2019). Robust clustering with high breakdown point based on RQ criterion. Computational Statistics & Data Analysis, 130, 30-47.

[7] Du, Y., Chen, L., & Lv, G. (2020). A novel method for outlier detection in high-dimensional data based on RQ clustering. IEEE Access, 8, 35572-35581.

[8] Lai, J., & Meng, J. (2022). Application of RQ Clustering Algorithm for the Study of Ancient Glass Technology. Journal of Archaeological Science: Reports, 47, 103252.

[9] Conticelli, S., D'Orazio, M., & Massanisso, P. (2021). R-mode cluster analysis of rare earth and trace elements in glass from Roman archaeological sites in central Italy. Journal of Archaeological Science, 132, 105423.

[10] Touhami, L., Mountasser, H., & Essaoudi, I. (2021). Application of R-mode cluster analysis to determine the characteristics of sand and its impact on groundwater in the Taourirt basin, eastern Morocco. Journal of African Earth Sciences, 179, 104195.

[11] Si, S. K., & Sun, X. J. (2016). Mathematical modeling algorithms and applications. Beijing: National Defense Industry Press.

[12] Crispin, K. L., Bernasconi, S. M., & Spencer, E. C. (2020). Q-type cluster analysis of compositions of slagged copper alloys from the Wadi Faynan region. Archaeometry, 62(4), 798-816.

[13] Li, S., Wu, X., Wang, Y., Li, R., & Wu, X. (2020). Robust RQ clustering algorithm for analyzing ancient Chinese ceramics. Journal of Cultural Heritage, 45, 1-8.