Analysis of the composition of ancient glass objects based on cluster analysis and Bayesian discriminant methods

Junjian Wang [1, *], Chenghao Cao [2], Wenqing Zhang [3]

[1] School of Materials Science and Engineering, University of Jinan, Jinan, China

[2] School of Mathematical science, University of Jinan, Jinan, China

[3] School of Information Science and Engineering, University of Jinan, Jinan, China

* Corresponding author: junjianwang_ujn@163.com

Abstract. The aim of this study is to investigate the open problem of classification laws and results. First, a comprehensive research method of quantitative analysis-Bayesian discrimination is used to reasonably and accurately investigate the classification laws of high potassium glass and lead-barium glass, obtaining the Bayesian discriminant for high potassium glass $F1 = 2.909x1 - 0.117x2 + 4.487\ x3 + 2.432x4 + 1.097x5 + 3.352x6 + 3.403x7 + 6.303x8 + 2.669x9 + 0.897x10 + 1.882x11 - 4.543x12 + 5.002x14 - 150.381$ with the lead-barium glass Bayesian discriminant $F2 = 2.447x1 - 1.195x2 + 42.807x3 + 2.117x4 + 1.0971.534x5 + 3.221x6 + 3.299x7 + 4.830x8 + 2.513x9 + 1.226x10 + 1.590x11 - 5.628x12 + 3.087x13\text{-}111.481$. Then, using the clustering process-Bayesian discrimination method, the number of sample clusters was roughly determined using systematic clustering, and the data were categorized in detail using K-Means clustering, resulting in the possibility of classifying each of the high-potassium and lead-barium glasses into Three subclasses were identified, and Bayesian discriminant functions were derived for each subclass, and the data were tested to justify the subclasses.

Keywords: Glass, Component Analysis, Cluster Processing, Bayesian Discriminant, Quantitative Analysis.

1. Introduction

Our ancient glass in the absorption of foreign technology, and foreign glass chemical composition is not the same [1, 2].

The main raw material for glass is quartz sand, the main chemical composition of which is silicon dioxide, and the stabilizer is limestone, which is converted to calcium oxide after calcination. Commonly used fluxes include grass ash, natural bubble soda, saltpeter and lead ore, etc. Different fluxes will produce different main chemical compositions [3]. Lead barium glass is fired with lead ore as a flux, which has a high content of lead oxide and barium oxide, while potassium glass is fired with substances containing high levels of potassium such as grass wood ash as a flux [4-6].

Weathering due to the burial environment can affect the judgement of the glass by altering the proportions of the elements in its composition. The colour and decoration of the artefacts is evident on the surface of the unweathered artefacts, and the surface weathered artefacts contain large grey-yellow areas of obvious weathering and purple areas of general weathering [7-9]. Unweathered areas also exist on weathered surfaces. Scholars have classified our samples of ancient glass artefacts into two types, high potassium glass and lead-barium glass, based on their chemical composition and other means [10].

This study analyses the classification patterns of high potassium glass (K2O-CaO-SiO2) and lead-barium glass (PbO-BaO-SiO2); for each category the appropriate chemical composition is chosen to classify them into subclasses, and specific classification methods and results are given, followed by an analysis of the reasonableness and sensitivity of the classification results.

2. An integrated research approach to quantitative analysis- Bayesian discriminant

2.1. Qualitative and quantitative analysis of data

Based on the data in Form 2, the artefact samples were divided into two types of glass, high potassium glass and lead-barium glass, and their chemical content was characterized. Preliminary observations of the data revealed that the weathering of the same type of glass had a significant effect on the chemical content of some of the components, and it was decided to use average values to further characterize the differences in chemical content between the high potassium glass and the lead-barium glass. The results are shown in Figure 1.

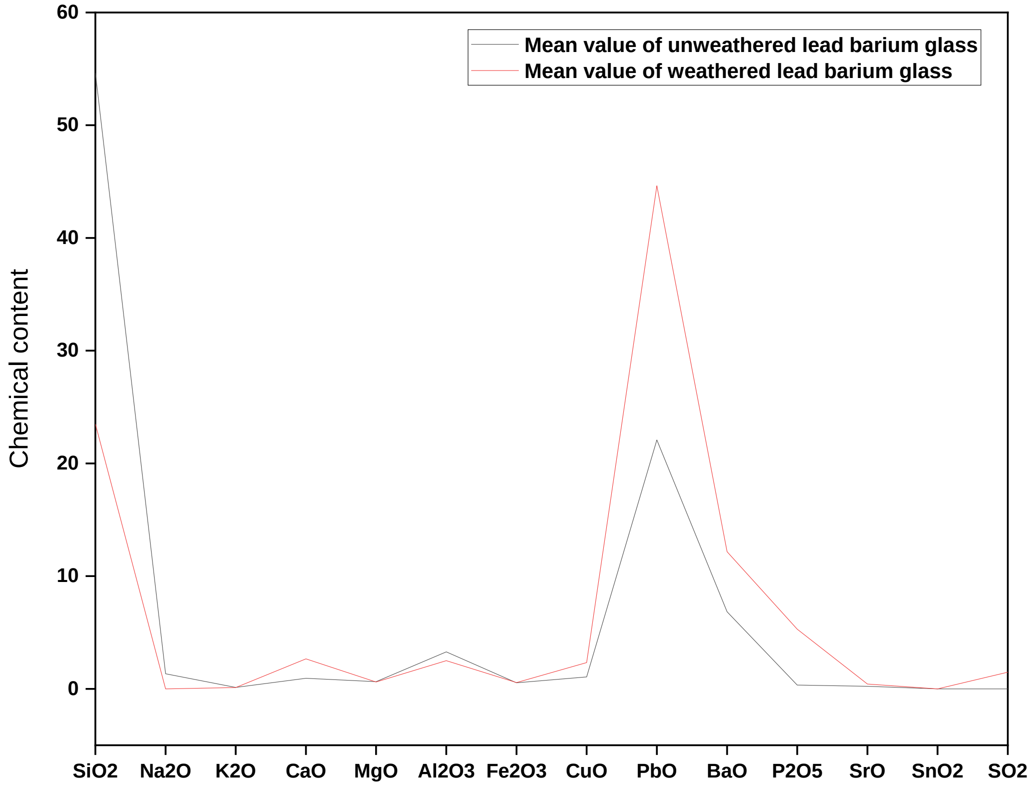

Figure 1. Average values for each chemical composition of the two types of glass

By looking at the line graph it can be concluded that the SiO2 content of the high potassium glass is much higher than that of the lead barium glass, but the PbO content of the lead barium glass is much higher than that of the high potassium glass. This can be explained as being mainly due to the different raw materials used in the preparation of the two types of glass, such as high potassium glass (K2O-CaO-SiO2), which was made by selecting appropriate oxalic ash, potassium-containing ores (potassium feldspar, saltpeter, potassium mannite), with quartz sand and other raw materials, and firing; whereas the production of lead-barium glass (PbO- BaO-SiO2) was based on the early use of lead for smelting bronze, and the ancients in In the production of silicate glass, lead ore, such as lead ore (PbS), and barium ore, such as barite (BaSO4), were used as fluxes, resulting in a higher PbO and BaO content in lead-barium glass than in high-potassium glass.

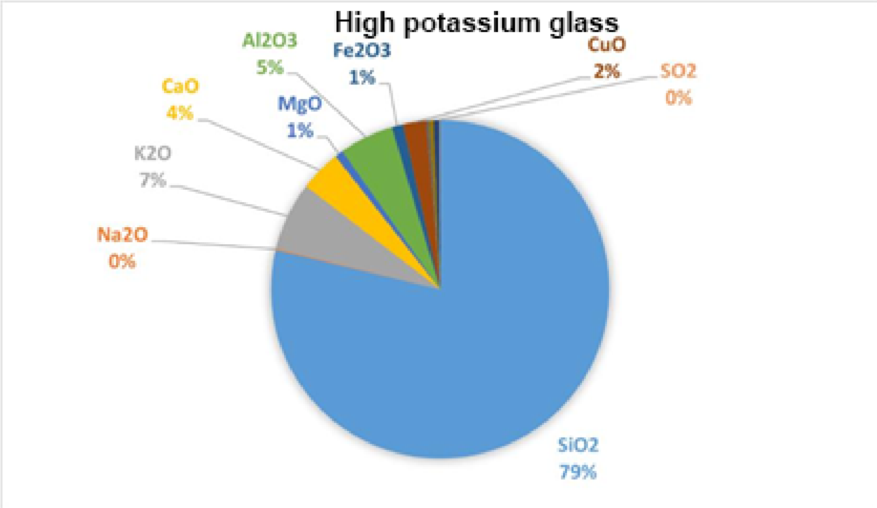

Figure 2. Chemical composition content of high potassium glasses

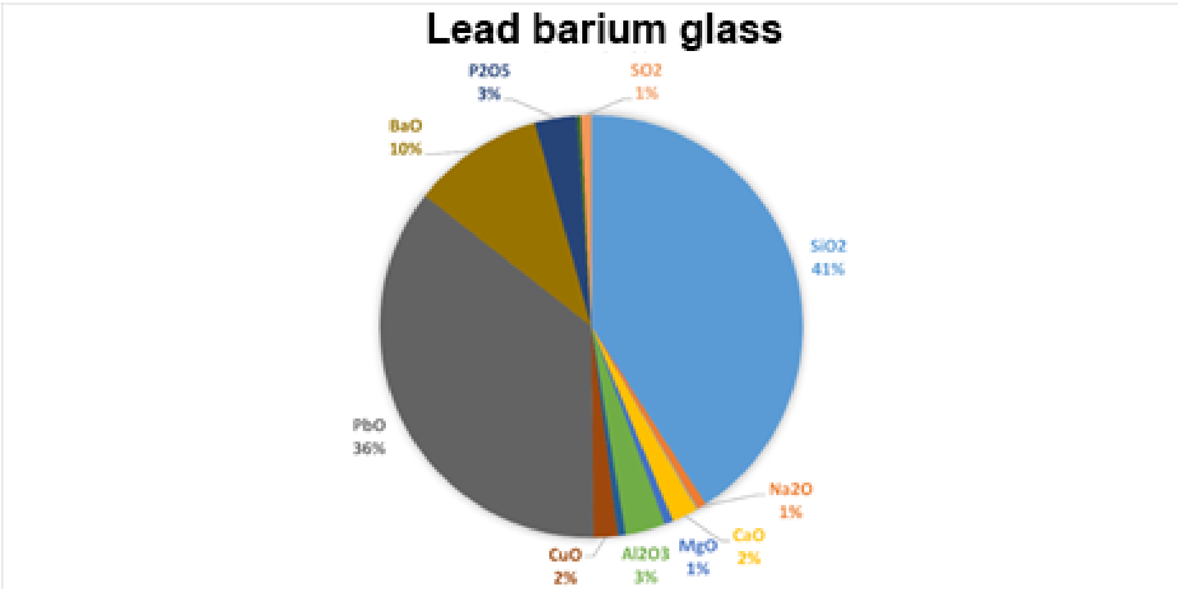

Figure 3. Chemical composition content of lead-barium glass

The results obtained from the pie chart show that in the high potassium glass, SiO2 accounts for the largest proportion, followed by K2O and Al2O3. This difference in content is ultimately due to the different raw materials and production processes of the two types of glass.

2.2. Solving for Bayesian discriminant functions

In order to investigate the classification of high potassium glass and lead-barium glass in more depth, it was decided to give a Bayesian discriminant function that could classify the samples more specifically and accurately. The Bayesian discriminant function was solved using SPSS software. The Bayesian discriminant function can be used to classify samples of unknown types of glass artefacts in a more accurate, reliable and reasonable way.

Using SPSS software, the results of the Bayesian discriminant function were solved as follows.

High Potassium Glass:

$$\begin{aligned} F_1 = {} & 2.909x_1 - 0.117x_2 + 4.487x_3 + 2.432x_4 + 1.097x_5 + \\ & 3.352x_6 + 3.403x_7 + 6.303x_8 + 2.669x_9 + 0.897x_{10} \\ & +1.882x_{11} - 4.543x_{12} + 5.002x_{14} - 150.381 \end{aligned} \tag{1}$$

Barium lead glass:

$$F_2 = 2.447x_1 - 1.195x_2 + 42.807x_3 + 2.117x_4 + 1.0971.534x_5 + 3.221x_6 + 3.299x_7 + 4.830x_8 + 2.513x_9 + 1.226x_{10} + 1.590x_{11} - 5.628x_{12} + 3.087x_{13} - 111.481 \quad (2)$$

The chemical content of each glass artefact is brought into the function to obtain the data. If F1> F2, then the glass artefact is a high potassium glass; if not, then it is a lead-barium glass.

After the SPSS software calculations, the Wilke Lambda coefficient test was tabulated:

Table 1. Wilke Lambda coefficient test table

Wilke Lambda				
Function Check	Wilke Lambda	Card Parties	Freedom	Significance
1	0.159	107.663	13	0.000

The p-values given in Table 1, p=0.00<0.05, indicate that there is reason to reject the original hypothesis that there is a significant difference between the mean Bayesian discriminant function values of the two groups at the 0.05 level of significance. This indicates that the Bayesian discriminant function equation derived above for high potassium glass and lead-barium glass meets the significance test. Moreover, the Wilke Lambda is small, demonstrating that the group means between the two classifications are unequal and widely different, proving that there is some difference between the two types of data. In summary, it can be shown that the above Bayesian discriminant function is reasonable.

3. Cluster processing - a Bayesian discriminant approach

3.1. Analysis of the problem

The two glasses were divided into subclasses by selecting the appropriate chemical composition as an indicator for the two glasses. We first continued with the first question and initially divided the two glasses into four sections: high potassium weathered glass, high potassium unweathered glass, lead-barium weathered glass and lead-barium unweathered glass. However, the observed data revealed that some of the unweathered glass contained certain sample indicators similar to weathered glass, or that there were outliers in the weathered glass. This led to a systematic clustering process for both glass types, which was judged to be classifiable into subclasses into several classes, the results of which were used as a reference for determining the number of classes in the K-Means clustering analysis. This was followed by K-Means clustering analysis to analyse the samples for appropriate subclasses and to present data to demonstrate the reliability and accuracy of the classification.

In order to test the reasonableness of the classification, the data was processed and the Bayesian discriminant function was derived using SPSS software and tested for significance. This would justify the resulting classification results. The analysis flow chart is shown in Figure 4.

Figure 4. Analysis flow chart

3.2. Systematic clustering analysis algorithms

Systematic cluster analysis: consists of Q-type cluster analysis and R-type cluster analysis. Q-type cluster analysis is the clustering of samples and is based on the following algorithm.

For a group of sample points to be classified that need to be described by p variables, each sample point can be considered as a point in Rp space. Then for samples x1, x2, x3, ..., xn, each of their corresponding p variables is represented as a matrix $\begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}$. For Q-type cluster analysis, the essence is to use distance distance to measure the degree of similarity between sample points. In this paper, we use the squared Euclidean distance, which is the most commonly used distance between samples x_i and x_j.

Squared Euclidean distance.

$$d_{i,j} = \sum_{i=1}^{p}(x_{il} - x_{jl})^2 \tag{3}$$

Class-to-class similarity between classes is measured using the class averaging method.

$$D(G_1 G_2) = \frac{\sum_{xi\in G1}\sum_{xj\in G2} d(x_i, x_j)}{n1n2} \tag{4}$$

For any two samples, the final sample clustering results are derived from the squared Euclidean distance between the samples for all variables, combined with a similarity measure function between each category.

3.3. Bayesian discriminant algorithm

Let there be k p-dimensional aggregates G_1, G_2..., G_K, each with a different probability density function, and suppose that the k The probability of occurrence of each of the k aggregates is q_1, q_2..., q_k (a priori probability), satisfying $q_i \geq 0$, i=1, 2..., k, $\sum_{i=1}^{k} q_i = 1$. If the loss caused by misclassifying a sample belonging to aggregate G_I to aggregate G_j is known to be C(j|i), i, j=1, 2..., k. For any i, j=1, 2..., k, we have C(i|j)=0, C(i|j)≥0. Let k aggregates G_1, G_2..., G_K. The corresponding p-dimensional space of G_K is divided into R_1, R_2..., R_K, which can be abbreviated as a discriminant rule R=(R_1, R_2,..., R_K).

In such a case the following discriminant criterion is established to minimise the total average loss, where the total average loss is

$$g(R) = \sum_{j=1}^{k}\int_{R_J}(\sum_{I=1}^{K} q_i C(j|i)\, f_i(X))dX \tag{5}$$

(R_1, R_2,..., RK) is $R_t = \{X: ht(X) < hJ(X), (j \neq t), \{j = 1,2, \dots, k\}, \{t = 1,2, \dots, k\}$ (6)

Where $h_i(X) = \sum_{i=1}^{k} q_i C(t|i) f_i(X)$.

3.4. Q-Type Cluster Analysis

The data were systematically clustered using SPSS, with the clustering method being "component linkage" and the measurement method being "squared Euclidean distance". The results are shown in Figures 5 and 6.

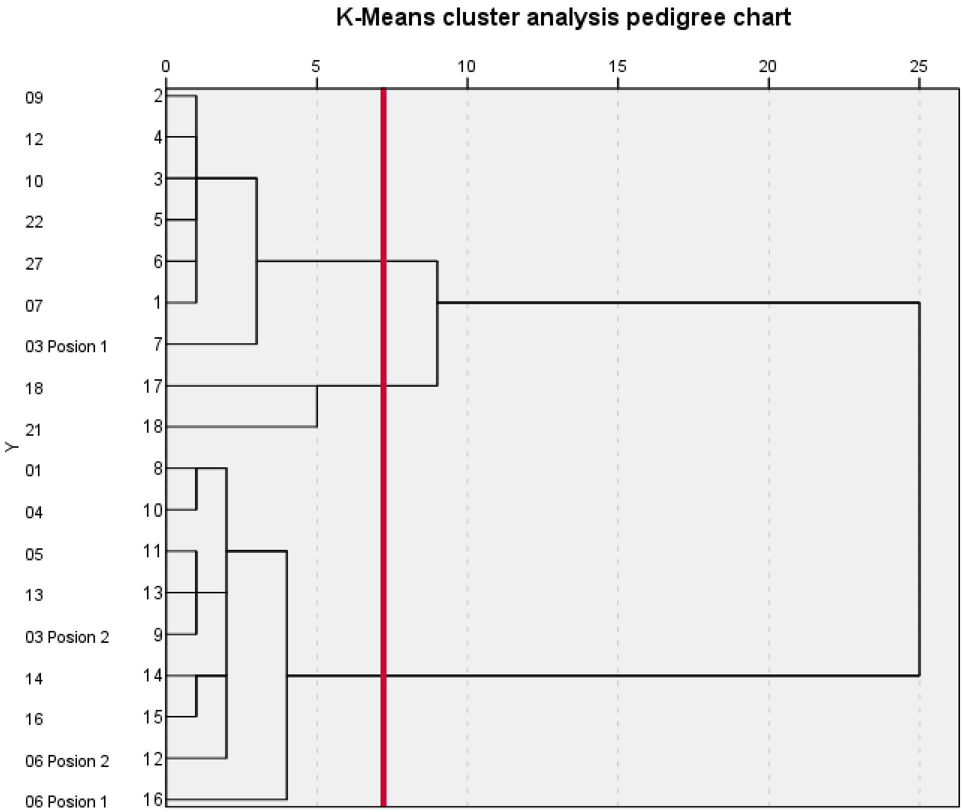

Figure 5. All high potassium glass samples plotted on a spectrum

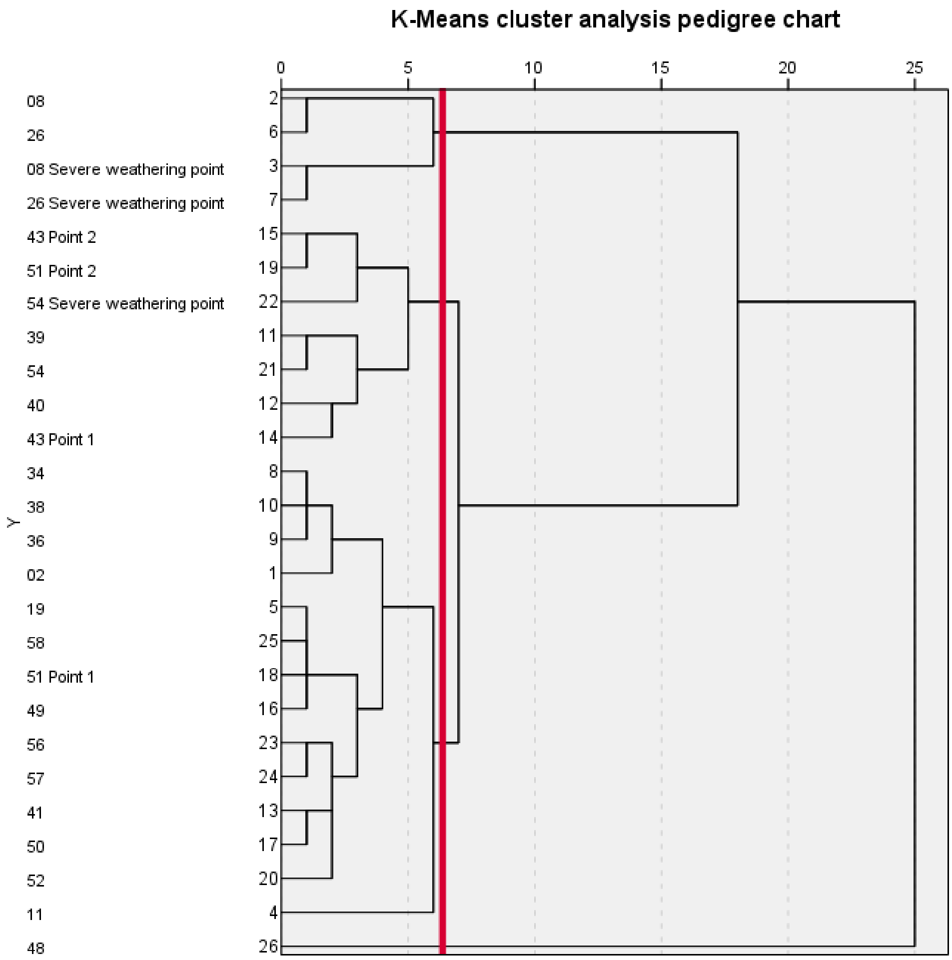

Figure 6. Pedigree diagram of all lead-barium glass samples

K-Means clustering analysis

From the above spectral diagram, it is possible to first classify the high potassium glass into 3 classes and the lead-barium glass into 4 classes for K-Means clustering analysis. In this paper, K-Means clustering analysis was applied to SPSS software, using an iterative and classification method with 100 iterations, and the results were obtained as follows.

For the high-potassium glass samples, considering the small amount of data given in the title, the results were combined with the results obtained from the systematic cluster analysis and divided into three subclasses, which we named: the first subclass, the second subclass and the third subclass. The degree of weathering of the artefact samples included in each subclass is shown in Figure 7 below.

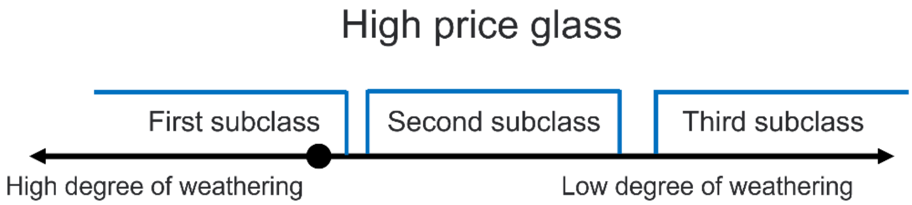

Figure 7. High potassium glass subclass weathering degree diagram

Interpretation of the image: it extends infinitely from the origin to the left and right. The sample is located in the area to the right of the origin, the further it is from the origin, the more weather-free it is; the closer it is to the origin, the less weather-free it is

(i. e. biased towards weathering). The sample is located to the left of the origin and the further it is from the origin, the more weathered it is; the closer it is to the origin, the less weathered it is. It is suggested that the first sub-category contains some of the less weathered and weathered samples; the second sub-category contains the moderately weathered samples; and the third sub-category contains the more weathered samples.

The following are the sample numbers of the artefacts included in each subcategory.

Subcategory I: 07, 09, 10, 12, 22, 27, 03 Part 1, 7 samples in total.

Subcategory II: 01, 03 part 2, 04, 05, 06 part 2, 13, 14, 16, 8 samples in total.

Subclass III: 06 part 1, 18, 21, 3 samples in total.

The above classifications show that the same sample was taken from different parts of the sample, resulting in different subclasses. It is reasonable to assume that there is no problem with these classifications, given the different external environments surrounding the same sample before it was excavated.

For the lead-barium glass samples, the results of the systematic cluster analysis were combined to classify them into four subclasses, and during the qualitative analysis of the subclasses, it was found that there was a problem with sample number 48, which was defined as a weathered sample in the title, but the content of its chemical indicators tended to be more consistent with an unweathered sample. The final number of data involved in the K-Means clustering analysis was 48. The results of the systematic clustering analysis were then reapplied to the K-Means clustering analysis and we decided to classify the lead-barium glass into three subclasses, each containing the degree of weathering of the samples as shown in Figure 8 below.

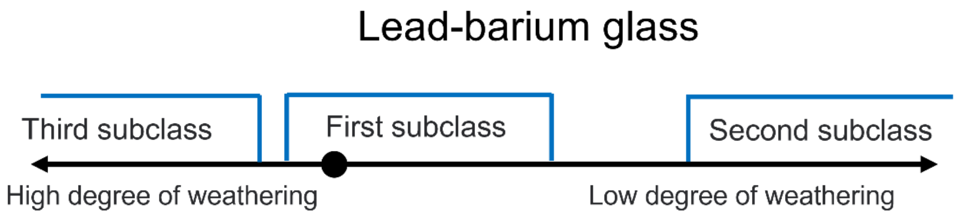

Figure 8. Diagram showing the degree of weathering of subclasses of lead barium glass

Image interpretation: Interpretation of the axes is as above, for lead-barium glass, subcategory I contains samples with low and no weathering; subcategory II contains samples with moderate weathering and subcategory III contains samples with high weathering.

Category I Weathered plus unweathered: 02, 11, 19, 34, 36, 38, 39, 40, 41, 43 parts

1, 43 Part 2, 49, 50, 51 Part 1, 51 Part 2, 52, 54, 54 Severe weathering, 56, 56,57, 24, 30 part 1, 30 part 2, 24 samples in total.

Category II Unweathered: 23 Unweathered, 25 Unweathered, 28 Unweathered, 29 Unweathered, 42 Unweathered Point 1, 42 Unweathered Point 2, 44 Unweathered, 49 Unweathered, 50 Unweathered, 53 Unweathered, 20, 31, 32, 33, 35, 37, 45, 46, 47, 55, 20 samples in total.

Category III: Severe weathering: 08 Severe weathering, 08, 26, 26 Severe weathering.

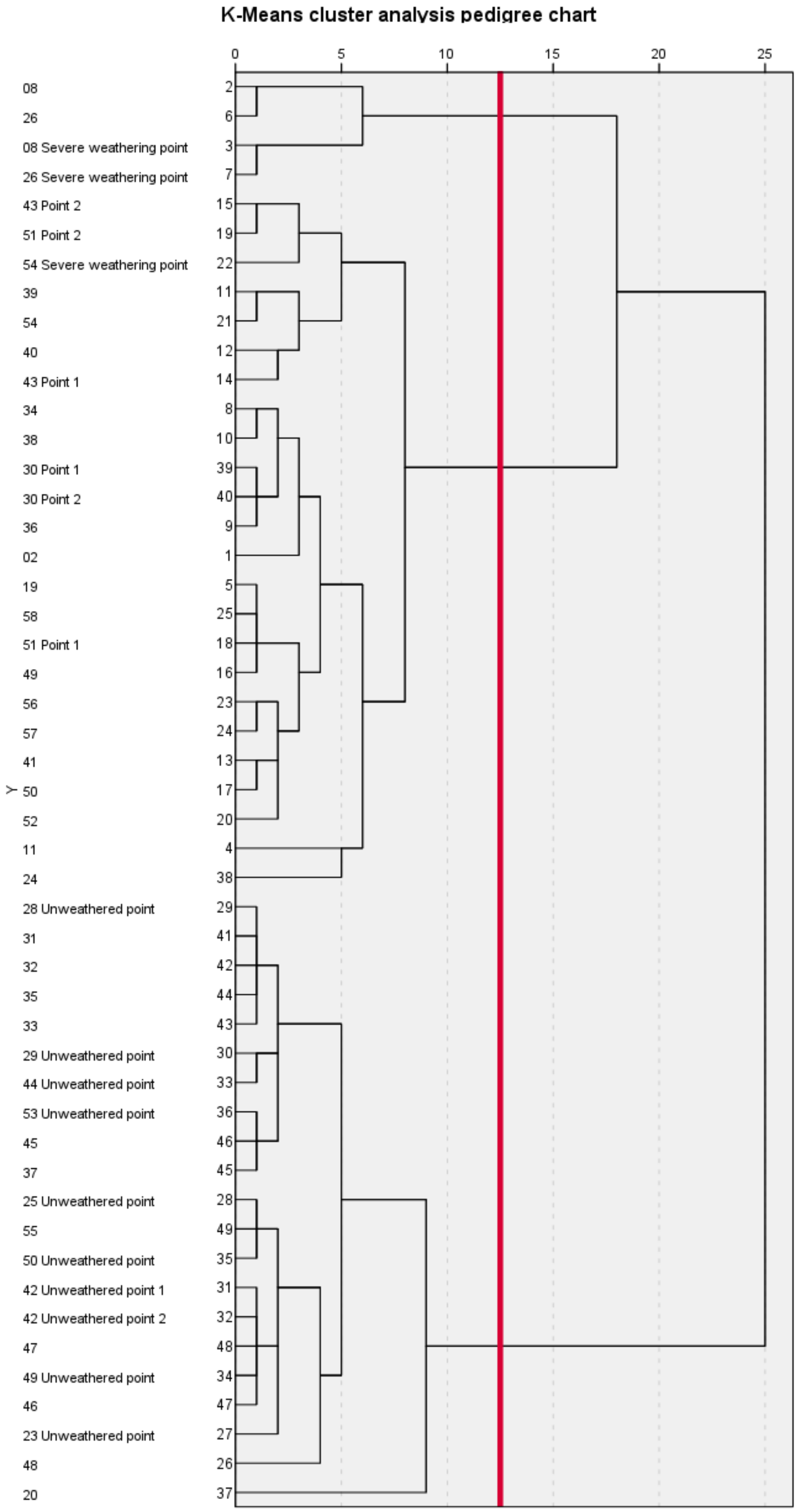

Figure 9. Lead and barium glass spectrum after deletion of 48 samples

3.5. Solving for Bayesian discriminant functions

Based on the above classification results, the specific Bayesian discriminant function for each subclass was derived using SPSS software and the results were as follows.

For the high potash glass divided into three subclasses, the discriminant equations are as follows.

Subclass I (partially unweathered and weathered samples).

$$F_1 = 1685.968x_1 - 3007.421x_2 + 1384.698x_3 - 391.764x_4 + 875.859x_5 +3218.252x_6 - 2304.473x_7 + 1215.858x_8 -8279.822x_9 + 6261.979x_{10} + 2752.188x_{11} - 111054.215x_{12} -23894.229x_{13} - 83612.595 \quad (7)$$

Subclass II (moderately unweathered samples)

$$F_2 = 1367.447x_1 - 2940.042x_2 + 1184.480x_3 + 41.619x_4 +848.969x_5 +2483.997x_6 - 2005.838x_7 + 1139.325x_8 -6700.699x_9 + 4857.099x_{10} + 1507.068x_{11} - 81425.203x_{12} -17961.372x_{13} - 56337.186 \quad (8)$$

Subclass III (samples without high levels of weathering)

$$F_3 = 1553.867x_1 - 2711.992x_2 + 1268.667x_3 - 382.393x_4 +782.948x_5 + 2977.255x_6 - 2095.528x_7 + 1102.415x_8 -7592.317x_9 + 5764.572x_{10} + 2623.856x_{11} - 102060.579x_{12} - 21997.747x_{13} - 71058.079 \quad (9)$$

4. Conclusion

The aim of this study is to investigate the open problem of classification laws and results. First, a comprehensive research method of quantitative analysis-Bayesian discrimination is used to reasonably and accurately investigate the classification laws of high potassium glass and lead-barium glass, obtaining the Bayesian discriminant for high potassium glass $F1 = 2.909x1 - 0.117x2 + 4.487\ x3 + 2.432x4 + 1.097x5 + 3.352x6 + 3.403x7 + 6.303x8 + 2.669x9 + 0.897x10 + 1.882x11 - 4.543x12 + 5.002x14 - 150.381$ with the lead-barium glass Bayesian discriminant $F2 = 2.447x1 - 1.195x2 + 42.807x3 + 2.117x4 + 1.0971.534x5 + 3.221x6 + 3.299x7 + 4.830x8 + 2.513x9 + 1.226x10 + 1.590x11 - 5.628x12 + 3.087x13 - 111.481$. Then, using the clustering process-Bayesian discrimination method, the number of sample clusters was roughly determined using systematic clustering, and the data were categorized in detail using K-Means clustering, resulting in the possibility of classifying each of the high-potassium and lead-barium glasses into Three subclasses were identified, and Bayesian discriminant functions were derived for each subclass, and the data were tested to justify the subclasses.

References

[1] Li Huizhen, Li Yunhui, Gui Cenlin, Chen Daqi, Chen Lanfang, Luo Le, Huang Guobao, Yuan Yang, He Rong, Xia Fan, Wang Jiahai. Bare glassy nanopore for length-resolution reading of PCR amplicons from various pathogenic bacteria and viruses [J]. Talanta, 2023, 256.

[2] Zhou Q M, Ding R Y, Li L, Bai W, Hu J Y. [Influence of paper mixing pads thickness on the compressive strength of glass ionomer cement]. [J]. Beijing da xue xue bao. Yi xue ban = Journal of Peking University. Health sciences, 2023, 55 (1).

[3] Skwira Adrianna, Szewczyk Adrian, Sądej Rafał, Prokopowicz Magdalena. Bioglass obtained via one-pot synthesis as Osseo integrative drug delivery system [J]. International Journal of Pharmaceutics, 2023, 633.

[4] Scott Douglas M, Prud'homme Robert K, Priestley Rodney D. Effects of the polymer glass transition on the stability of nanoparticle dispersions. [J]. Soft matter, 2023.

[5] Costello Mark A, Liu Joseph, Wang Yan, Qin Bin, Xu Xiaoming, Li Qi, Lynd Nathaniel A, Zhang Feng. Reverse engineering the Ozurdex dexamethasone intravitreal implant. [J]. International journal of pharmaceutics, 2023.

[6] Alharbiy N., Khattari Z. Y., Rammah Y. S., Saleh Abdelmoneim. Role of Al2O3, WO3, Nb2O5, and PbO on the physical, elasto-mechanical and radiation attenuation performance of borotellurite glasses [J]. Journal of Materials Science: Materials in Electronics, 2023, 34 (3).

[7] Alrowaili Z. A., Khattari Z. Y., Alsaif Norah A. M., Shams M. S., El Refaey Adel M., Elsad R. A., Al Buriahi M. S., Rammah Y. S. Synthesis, physical properties, neutron, and gamma-ray shielding competence of borate-based glasses reinforced with erbium (III) oxide: a closer-look on the impact of Eu2O3 [J]. Journal of Materials Science: Materials in Electronics, 2023, 34 (3).

[8] Xia Pengfei, Liu Fan, Duan Yuru, Hu Xuefang, Lu Changgui, Xu Shuhong, Wang Chunlei. Glass-compatible and self-powered temperature alarm system by temperature-responsive organic manganese halides via backward energy transfer process [J]. Journal of Energy Chemistry, 2023, 78.

[9] Dziadek Michal, Dziadek Kinga, Checinska Kamila, Salagierski Szymon, Choinska Emilia, Szatkowski Piotr, Wajda Aleksandra, Kopec Aneta, Cholewa-Kowalska Katarzyna. Polyphenolic compounds affect the long-term degradation behaviour of polymer and composite materials based on PCL, PLGA, and bioactive glass [J]. Sustainable Materials and Technologies, 2023, 35.

[10] Siekkinen Minna, Karlström Oskar, Hupa Leena. Dissolution of bioactive glass S53P4 in a three-reactor cascade in continuous flow conditions [J]. Open Ceramics, 2023, 13.