SORT 47 (2) July-December 2023, 1-18 DOI:

Interpretation of coal compositional data on whole-coal versus ash bases through the weighted symmetric pivot coordinates method

Na Xu[1,*], Ru Wang[1], Mark A. Engle[2], Wei Zhu[1], Qiang Li[1] and Zhiwei Wang[1]

Abstract

In addition to approaches based on a number of physical and chemical analyses, statistical methods have been commonly used for determining the modes of occurrence of elements in coal. The Pearson correlation coefficient of element concentrations vs. ash yields is the simplest method that has been widely used. Concentrations of elements in coal are usually reported on two bases: whole-coal and ash bases. Coal compositional data on whole-coal basis can be converted back to ash basis. However, in many cases, the correlation between corresponding pairs of elements in coal is inconsistent when reported on whole-coal versus ash bases. Therefore, traditional statistical methods, such as correlation analysis, based on whole-coal and ash bases can sometimes lead to misleading or confusing results. Previous investigations have suggested using logratio variance or related parameters (i.e., stability) to examine these data, as they provide consistent results regardless of the sample basis. However, logratio variance based approaches are unable to distinguish the inverse relationships between parts. To provide more clarity on the relationships between parts, weighted symmetric pivot coordinates are used to analyze the correlation between elements in coal on whole-coal basis and ash basis. To illustrate this approach, 106 late Paleozoic coal samples from the Datanhao and Adaohai coal mines, Daqingshan Coalfield, northern China, are used for performance evaluation. Experimental results show that the weight symmetric pivots method is more effective than the stability method in predicting the modes of occurrence of elements in coal for these samples, providing deeper insight than logratio variance based approaches.

MSC: *62P12, 92E99, 00A69.*

Keywords: *Whole-coal basis, Ash basis, Correlation, WSPC method.*

[1] College of Geoscience and Survey Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China. * Corresponding author: xuna1011@gmail.com

[2] Department of Earth, Environmental and Resource Sciences, University of Texas at El Paso, El Paso, Texas 79968, USA.

Received: November 2022
Accepted: May 2023

1. Introduction

Coal elemental data describing the abundance of major and trace elements (Dai et al., 2010) are both compositional (composed of relative, strictly positive data known as parts) and very complex. Determining modes of occurrence of elements in coal is an important research area (Dai et al., 2021; Zhou et al., 2022), because the geological process of coal formation can be inferred and the potential sources of critical metals present in the coal can be understood (Dai et al., 2010; Hou et al., 2023). Traditional physical and chemical methods (e.g., density separation, chemical leaching, in situ micro-analyses) and a number of statistical methods have been employed to analyze the modes of occurrence of elements in coal (Dai et al., 2021; Moore et al., 2022; Nechaev, et al., 2022). Among them, the Pearson correlation coefficient between elements in coal is one of the simplest and the most common methods in such studies. Coal elemental data can be either reported on ash basis or on whole-coal basis. Coal elemental data can be converted between two bases using the formula:$[E_i]_{\text{coal}} = ([E_i]_{\text{ash}} \times \text{ ash yield})$, where $[E_i]_{\text{coal}}$ denotes the elemental concentration on whole-coal basis and $[E_i]_{ash}$ denotes the elemental concentration on ash basis (Geboy et al., 2013; Dai et al., 2020).

However, previous studies found that the Pearson correlation coefficient provides inconsistent results between coal elemental data reported on the two bases (whole-coal basis and ash basis) (Dai et al., 2021; Geboy et al., 2013; Xu et al., 2020). The modes of occurrence of elements in coal inferred from the correlation between elements in coal on two bases may be different because they are compositional data and are prone to issues of subcompositional coherence (Pawlowsky-Glahn and Egozcue, 2022). For example, Pearson correlation coefficients between pairs of elements in coal from the Adaohai and Datanhao coal mines (Daqingshan Coalfield, northern China) are inconsistent between the two bases, leading to the differences in the interpretation of the modes of occurrence of elements in coal (Xu et al., 2020). Geboy et al. (2013) also made a similar observation in the coal samples analysis and concluded that such inconsistency was a mathematical artifact due to the influence of constant sum of the compositional data.

Traditional statistical analysis methods applied in coal elemental data relies on typical Euclidean space. Because most coal elemental data exist in non-typical Euclidean space, wrong results may be obtained when using these conventional tools (Xu et al., 2020, 2022). To fully understand the modes of occurrence of elements in coal, the consistence of results when analyzing coal elemental data on the whole-coal basis and ash basis is necessary (Xu et al., 2020, 2022; Dai et al., 2021). Geboy et al. (2013) proposed to overcome the problems with bivariate interpretation of coal elemental data by examining a scaled version of the logratio variance of elements of interest (known as stability which is independent of sample basis). The logratio variance approach works well when the logratio variance is small (i.e., the pair of elements are strongly proportional) but when it is large (i.e., the elements are not proportional) it is difficult to differentiate inverse relationship between the elements (e.g., variation along a solid solution of a mineral system). As an alternative, we explored the application of weighted symmetric pivot coordinates (WSPC) to allow for direct determination of correlation coefficient between two logratio variables using coal elemental data on both bases from the Adaohai and Datanhao coal mines, China. Both WSPC and pairwise logratios are in essence log-contrasts and the two methods are consistent due to the property of log-contrast scale invariance. The aim of this paper is to provide interpretation of coal compositional data on whole-coal versus ash bases through the WSPC method.

2. Methods for consistent interpretation of correlation

In this section, we focused on the two methods which provide interpretation of coal compositional data on whole-coal versus ash bases: stability method and WSPC method.

2.1. Correlation by weighted symmetric pivot coordinates method

Compositional data analysis can treat non-typical space properly. Before analysing compositional data, a family of logratios will be used. The common logratios are additive logratio (alr) (Aitchison, 1986, 2008), centered logratio (clr) (Aitchison, 1986) and orthonormal logratio (olr) (Egozcue et al., 2003; Egozcue and Pawlowsky-Glahn, 2019). The olr-variables exhibit orthogonal properties while the alr-variables and clr-variables do not.

2.1.1. Orthonormal logratio coordinates

To establish an orthogonal basis, olr coordinates were proposed by Egozcue et al.(Egozcue et al., 2003; Egozcue and Pawlowsky-Glahn, 2019). The olr coordinates can be defined as follows:

$$\mathrm{olr}_i(\mathbf{x}) = \left(\sqrt{\frac{n-i}{n-i+1}}\ln\frac{x_i}{\sqrt[n-i]{\prod_{k=i+1}^{n} x_k}}\right), i = 1\ldots,n-1 \tag{1}$$

where $\mathrm{olr}_i(\mathbf{x})$ is the i^{th} olr-coordinate of the composition $\mathbf{X}$, n represents the total groups of parts. Olr coordinates can be developed in such a way to make them successful in correlation and regression analysis, though the interpretation of the results could be challenging.

Olr coordinates have been shown to be an effective compositional logratio and have been widely used in various fields (Razum et al., 2021).

2.1.2. Weighted symmetric pivot coordinates method

Hron et al. (2017) proposed symmetric pivot coordinates, a particular form of olr coordinates, to measure the correlation strength of components by selecting correlation coefficients from specific orthogonal coordinates. Symmetric pivot coordinates can be used to identify relations between two parts based on their degree of dominance over other parts in the compositions (i.e., other elements are not directly being correlated in the case of whole-coal basis). Therefore, the symmetric pivot coordinates method aggregates the related logarithms of the parts. However, if any of the parts (e.g., elements) contain erroneous results or exhibit large logratio variance, the calculation results are not desirable because it can affect the correlation between any other pairs of symmetric pivot coordinates. Hron et al. (2021) proposed the weighted symmetric coordinates method based on pair-wise correlation, which can reduce the impact from noisy or problematic parts in the composition.

The WSPC method is applied to coal elemental data to construct an orthogonal coordinates system for processing the relative weight information of two different compositional parts. To study the correlation between two different components, two parts of interest in the correlation are moved to the two first positions (x_1 and x_2) in the full composition X_n. The key difference between WSPC and SPC are that the impact of particularly noisy parts (i.e., those with high logratio variances) can be reduced through the implementation of weights. The variation matrix T of composition X_n is used as the basis for constructing the weight of the WSPC method. The variation matrix is defined as (Aitchison, 1986):

$$\mathbf{T} = \left\{t_{i,j}\right\}_{i,j} = \left\{\mathrm{Var}\left(\ln\frac{x_i}{x_j}\right)\right\}_{i,j} \tag{2}$$

where T is a matrix containing the logratio variance for each pair in the composition (e.g., between all the elements in the coal samples).

For x_1 and x_2, normalized weight α^* is calculated as follows: (Hron et al., 2021)

$$\widetilde{\alpha_j} = \frac{1}{\left(t_{1,j}\right)^2}; \alpha_j^* = \frac{\widetilde{\alpha_j}}{\sum_{k=3}^{n}\widetilde{\alpha_k}}, j = 3,\ldots,n \tag{3}$$

and

$$\widetilde{\beta_j} = \frac{1}{\left(t_{2,j}\right)^2}; \beta_j^* = \frac{\widetilde{\beta_j}}{\sum_{k=3}^{n}\widetilde{\beta_k}}, j = 3,\ldots,n \tag{4}$$

Values for parameters γ and C are determined from the normalized weights α^* and β^* (Hron et al., 2021):

$$\gamma^* = \frac{\alpha_3^*\beta_3^* + \ldots + \alpha_n^*\beta_n^*}{2}; C = \frac{-1+\sqrt{1+4\gamma^*}}{2\gamma^*}; \gamma = C^2\gamma^* \tag{5}$$

The weight calculation results are shown as follows (Hron et al., 2021):

$$\alpha_i = C\alpha_i^*, \beta_i = C\beta_i^*, i = 3,\ldots,n \tag{6}$$

For the parts x_1 and x_2 the weighted symmetric pivot coordinates (Hron et al., 2021) are:

$$W_1 = \frac{1}{\sqrt{1+\gamma^2+\alpha_3^2+\ldots+\alpha_n^2}} \ln\frac{x_1}{x_2^{\gamma}x_3^{\alpha_3}\ldots x_n^{\alpha_n}} \tag{7}$$

$$W_2 = \frac{1}{\sqrt{1+\gamma^2+\beta_3^2+\ldots+\beta_n^2}} \ln \frac{x_2}{x_1^{\gamma} x_3^{\beta_3} \ldots x_n^{\beta_n}} \tag{8}$$

The correlation between x_1 and x_2 in the m samples is defined as $CC_m(x_1, x_2)$ which is calculated by W_1 and W_2 as follows:

$$CC_m(x_1, x_2) = \frac{\sum_{i=1}^{m} \left(W_{i,1} - \overline{W_1}\right)\left(W_{i,2} - \overline{W_2}\right)}{\sqrt{\sum_{i=1}^{m} \left(W_{i,1} - \overline{W_1}\right)^2} \sqrt{\sum_{i=1}^{m} \left(W_{i,2} - \overline{W_2}\right)^2}} \tag{9}$$

Where $W_{i,1}, W_{i,2}$ is W_1, W_2 in the i^{th} sample. Pearson correlation is used in WSPC method as a measure of similarity, which has proven to work well in predicting modes of occurrence for coal elemental data (Xu et al., 2022). Finally, the above process to calculate $CC_m(x_i, x_j)$ is repeated for each pair of parts in the composition (e.g., pairs of elements in the coal samples).

This paper applies the WSPC method to examine associations in coal elemental data to determine the modes of occurrence for elements in a manner that is consistent between the whole-coal basis and ash basis. We found that the results based on the WSPC method are internally consistent with findings from previous investigations. The WSPC approach has the advantage that calculated correlation coefficients remain the same on the whole-coal basis and ash basis and there is an ability to distinguish between a lack of relationship between elements (near zero correlation) versus inverse relationship (negative correlation). However, WSPC correlation is not invariant under subcompositions while the stability method is.

2.2. Correlation replaced by stability method

Geboy et al. (2013) proposed to overcome the problems with bivariate interpretation of coal elemental data by examining the stability metric calculated as:

$$\mathrm{stab}(x_1, x_2) = \exp(\mathrm{var}(\mathrm{olr}(x_1, x_2))) = \exp\left(-\mathrm{var}\left(\frac{1}{\sqrt{2}} \ln \frac{x_1}{x_2}\right)\right) \tag{10}$$

where (x_1, x_2) is the pair of parts in a two-part composition. In the interpretation of stability or logratio variances, the concept of association is replaced by that of proportionality, where stability values near zero indicate lack of proportionality while values near 1 are associated with strong proportionality.

One advantage of the stability method is that the results are independent of the sample bases, providing consistent interpretation of coal compositional data on whole-coal versus ash basis. As noted above, the method is relatively easy to interpret when the logratio variance is small but not when they are large (i.e., lack of a relationship between two elements versus antithetical behavior which produces the similar stability values).

2.3. Hierarchical clustering algorithm to predict the modes of occurrence of elements in coal

In this paper, hierarchical clustering algorithm (Brian and Everitt., 2001) is used to interpret the modes of occurrence for coal elemental data. Among many hierarchical clustering algorithms, the most common clustering algorithm is average-linkage which calculates the average dissimilarity between the compositional data in two different clusters (Xu et al., 2022). Data matrices consisting of pairwise values of stability measures (ranging from 0 to +1) and correlation between WSPC (ranging -1 to +1) were used as the input data to the clustering algorithm, and results between the two inputs were compared. In both cases, the measures (9) and (10) were transformed to dissimilarities by subtracting their values from 1.

3. Results and discussion

To compare the performance of different association methods for coal elemental data, coal elemental data from the Adaohai and Datanhao coal mines were used for performance evaluation (Dai et al., 2012; Zhao et al., 2016). 63 elements of Adaohai coal mine and 58 elements of Datanhao coal mine have been included in this investigation. To facilitate visualization, only part of bivariate association (top 20 elements) of coal elemental data using WSPC method and stability method of coal elemental data from the Adaohai and Datanhao coal mines on the whole-coal and ash basis are shown in Figure 1. The tables and figures with all the elements are given in the Supplementary Material.

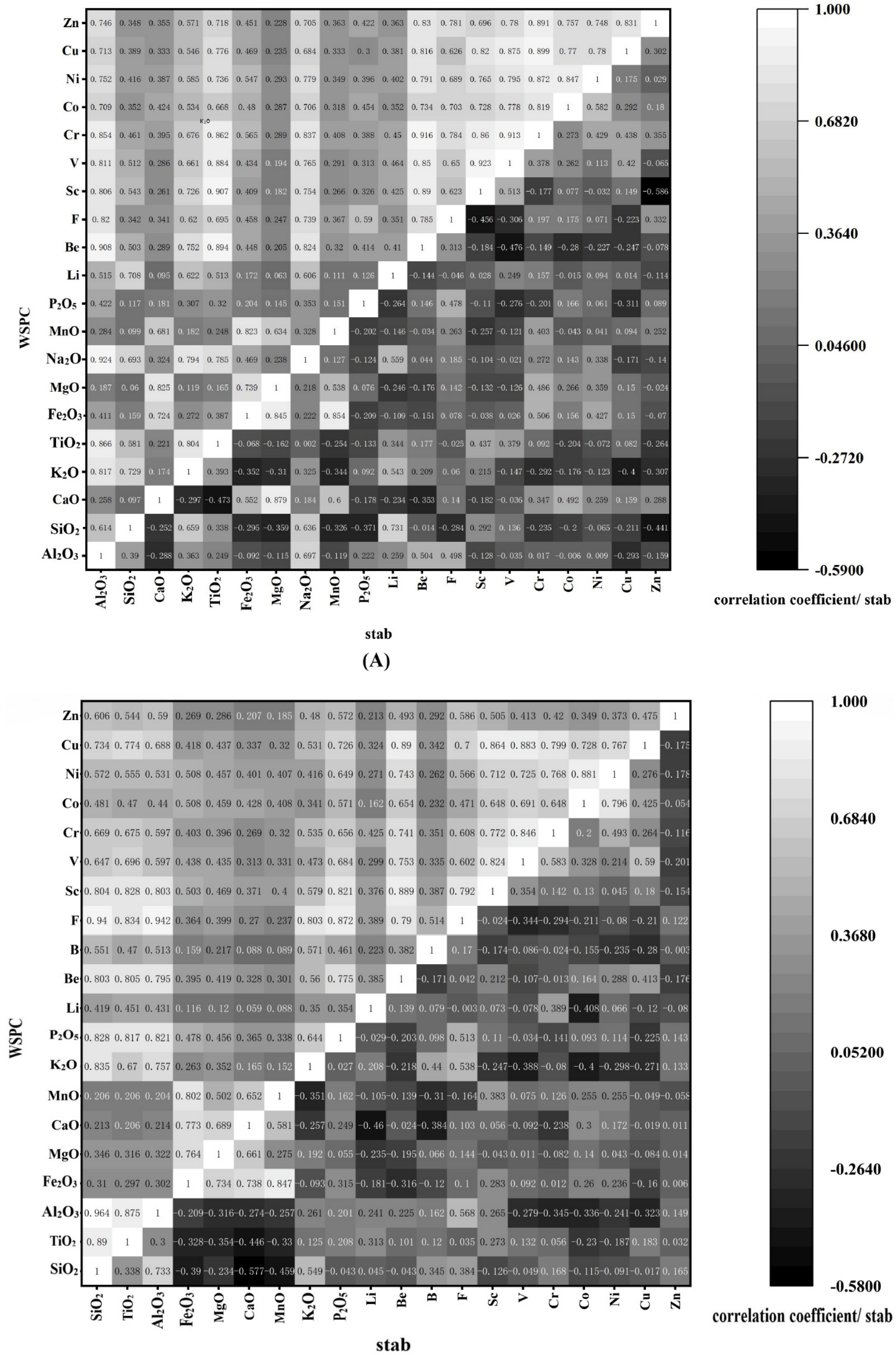
On the basis of trace-elements' geochemical nature and on the investigations by Zhao et al. (2016) and Dai et al. (2012) using direct analysis of coal samples, such as Scanning Electron Microscope/Energy-dispersive X-ray spectroscopy (SEM/EDS) and X-ray diffraction (XRD), the several pairs of elements are known to be strongly co-associated in individual modes (e.g., elements, macerals, or other phases), for example, Sr and Ba, Sn and Hg, Cd and Zn, and Nb and Ta (Dai et al., 2005, 2021). Additionally, the major elements which compose carbonate phases, including Ca, Mg, Mn and Fe (Liu et al., 2021a, b; Dai et al., 2021) are expected to be clustered together. Furthermore, Al and Si are both largely associated with aluminosilicate minerals and should be expected to be strongly associated (Templ, Filzmoser and Reimann, 2008; Arbuzov et al., 2019). In addition, rare earth elements and yttrium (REY) in coal should generally be proportional with one another (Seredin and Dai, 2012). Relationship between these groups of elements, which has previously been demonstrated to be related in individual phases (e.g., Seredin and Dai, 2012), are examined in our statistical methods to assess the difference in performance and interpretation between various clustering methods.

As shown in Figures 2 and 3, results from the hierarchical clustering analysis using WSPC method pairwise correlations vary notably from those based on the stability measure in coal elemental data from both mines. In general, many clusters with high similarity in samples from the Datanhao mine (Figure 2) were observed using both stability and correlation of WSPC input data (e.g., K_2O and Rb, Nb and Ta, Zr and Hf, Co and Ni, and Al_2O_3 and SiO_2) and are expected given that these elements often substitute for one another in minerals or are stoichiometrically linked in the same minerals (Dai et al., 2013, 2021). Differences in the cluster analysis results are generally more distinct for the Adaohai mine samples (Figure 3), where pairs of elements that are very similar in the WSPC method while not similar when using the stability method (e.g., Ca and Mg, Fe and Mn, Sn and Te, and P_2O_5 and Sr) suggesting more disparity between these methods when applied to this dataset.

In evaluating the ability of the cluster analyses to link elements which are known to coexist we explore how these elements behaved in various clustering methods. In the pairs of elements known to co-exist (noted above) most of them appear strongly associated using both data input schemes for the Datanhao mine, including Sr and Ba, Nb and Ta, elements which are common in carbonate minerals (Ca, Mg, Mn and Fe), and Al and Si. In other cases, notable difference between methods were observed. For example, Dai et al. (2021) found using direct analysis that Cd and Zn were strongly associated with sulfide minerals in these coals and thus were expected to be closely clustered, but the results from the WSPC method showed a much stronger relationship between the element pairs than the stability-based approach. Conversely, the stability-based approach generally showed closer clustering among REY than the clustering of correlations using WSPC method, where groups of various elements in the range (e.g., Tm, Yb, and Lu vs. La and Ce vs. Pr, Nd, Sm, Eb, Tb, and Ho vs. Eu and Sc) (Dai, Graham and Ward, 2016; Seredin and Dai, 2012) clustered separately. Another notable difference is that the elements which appeared least compatible with the other elements in the suite (e.g., those elements with the tallest vertical bars in the dendrogram plots) were notably different between the two approaches. Elements exhibiting this behavior in the WSPC correlation-based cluster analysis including Tm, Dy, and Y compared to Bi, Li, B, As, and Zn for the stability-based analysis.

As shown in Figure 3, many pairs of elements known to be co-associated also show high similarity in samples from the Adaohai mine, for both methods; namely, Sr and Ba; Cd and Zn (through the association in the WSPC method is stronger); Nb and Ta, and carbonate sourced elements (Ca, Mg, Mn, and Fe). In a manner similar to the data from the Datanhao mine, the REY data are more closely clustered in the stability-based clustering than the WSPC correlation-based clustering. Moreover, elements which exhibit the poorest clustering results were also markedly different between the two methods for the Adaohai mines coal samples. From the clustering analysis of the WSPC correlation, Er and As exhibited a relatively weak association with other elements, and for the clustering of the stability method Cl showed the weakest clustering by far.

The results show that, generally the two different methods showed relatively similar behaviors for elements that are strongly co-associated but strongly very different responses for elements that are not proportional. Comparison of the results between



(B)

Figure 1. *Association between two logratio variables using WSPC method and stab on whole-coal and ash basis in (A) Adaohai coal mine and (B) Datanhao coal mine (less elements in the correlation matrices to facilitate visualization for the figures).*

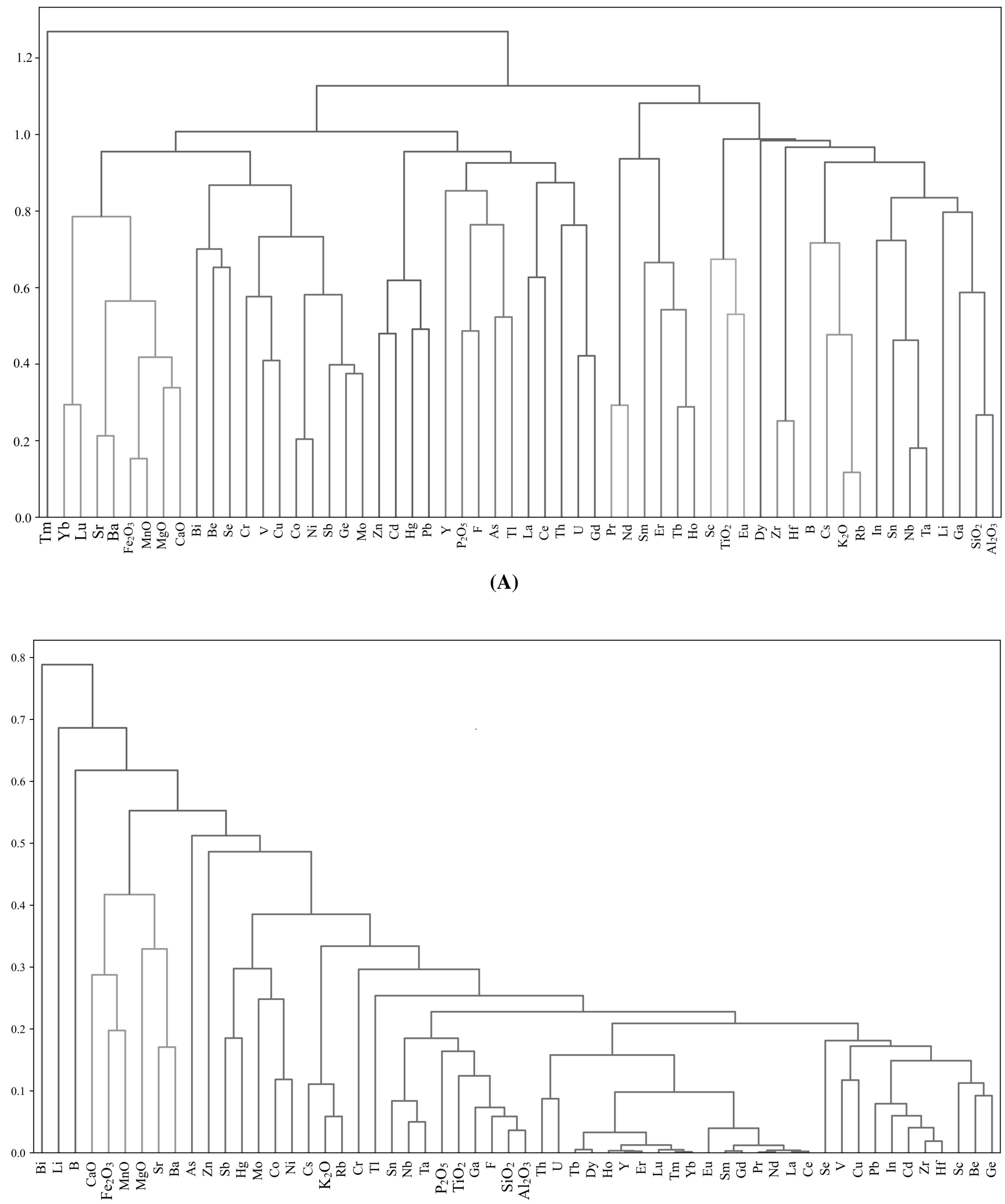
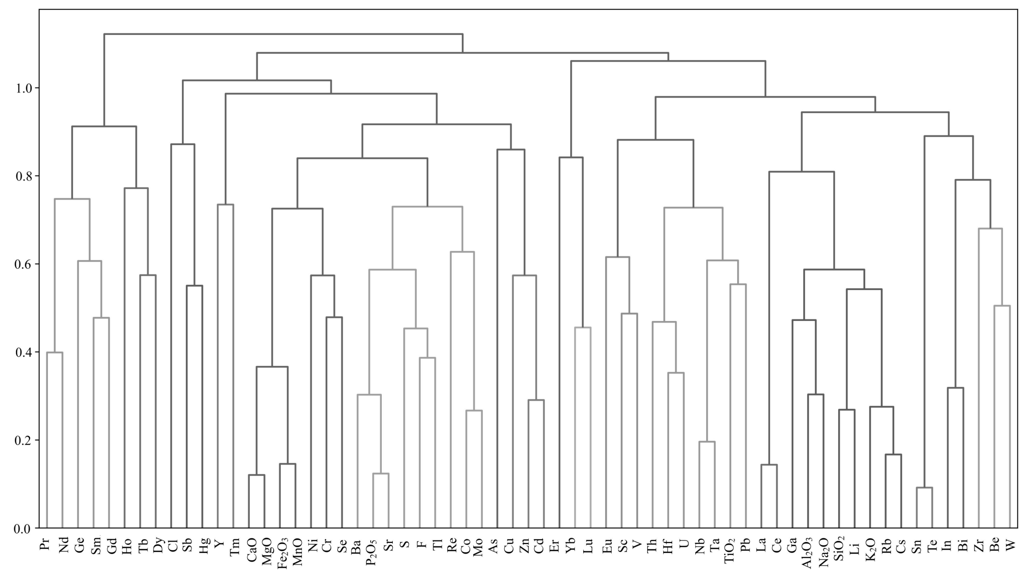
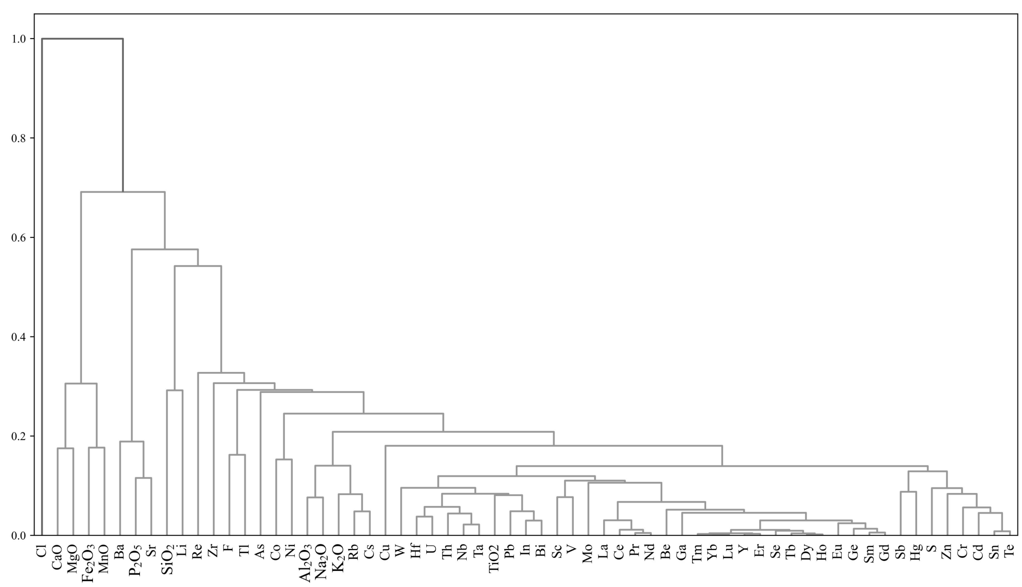


Figure 2. *Clustering analysis of coal element data in Datanhao coal mine. (A) correlation from WSPC on whole-coal basis and ash basis; (B) stability values from whole-coal basis and ash basis. In both cases the measures have been subtracted from 1, for clustering of dissimilarities.*



(A)



(B)

Figure 3. *Clustering analysis of coal element data in Adaohai coal mine using (A) correlation using WSPC on whole-coal basis and ash basis; (B) stability values on whole-coal basis and ash basis. In both cases the measures have been subtracted from 1, for clustering of dissimilarities.*

the two approaches against known results suggest that the WSPC method is better in predicting the modes of occurrence of elements in coal than the stability for these data.

In addition, the performance evaluation of association between WSPC method versus stability method can also be analysed based on the plots of the pairwise relationship between the two. As shown in the Figure 4, we can roughly see that the results of stability method and WSPC method show an approximate linear relationship. For the most part, pairs of elements which exhibit strong WSPC correlation, also exhibit high (near 1) stability values. However, data distribution patterns of the two methods diverge from there.

Examining results for the Datanhao coal mine, it is clear that for element pairs where correlation coefficients using WSPC method are near-zero or negative values, the corresponding stability values range greatly, though some important observations emerge (Figure 4A). For example, from analysis of coal samples, it is known that Hg occurring in sulfides is correlated weakly with Si and Al, which occur in clay minerals (Zhao et al., 2016). In correlation between WSPC, Hg is negatively correlated with Si and exhibits a weak positive correlation with Al which is to be expected. Comparatively, the stability values for Hg versus Si and Al are high. suggesting disparity between observations from other techniques and results using this approach.

For Adaohai coal mine sample data, Si is mainly enriched in clay minerals. Gorceixite and fluorapatite are rich in P(Dai et al., 2012). Thus, based on the analysis of the known modes of occurrence of P and Si, they should exhibit negative correlation (Dai et al., 2012). In correlation analysis of WSPC method , P and Si indicate negative correlation while in stability method they indicate low proportionality. Likewise, aluminosilicate affinity elements (K_2O, Be, Ga, Rb, Cs, W) mainly occur in diaspore and clay minerals. Thus, this group of elements should exhibit a weak positive correlation with P_2O_5 and negative correlation with barium.In correlation results from the WSPC method the aluminosilicate affinite elements indicate negative correlation with barium and weak positive correlation with P_2O_5.Conversely, from the results of stability method a strong degree of proportionality is observed with barium and P_2O_5. In another example, S has a strong negative correlation coefficient with ash yield which indicates a dominant organic affinity (Dai et al., 2012).

Results from the WSPC method between S and inorganic affinity elements should be weak. However, high stability values for S with other elements in stability method, which is not consistent with the modes of occurrence of S in the Adaohai coal.

Although WSPC method produces better geological results than stability method, we acknowledge that there are still some apparent discrepancies between the results and known modes of occurrence. For example, Zr-Hf and Al-Si do not cluster together quickly in the Adaohai coal mine, but they do in the Datanhao coal mine. Also, a number of element pairs exhibit high stability values but low or negative correlation, which is seemingly paradoxical. However, it must be considered that stability is based solely on the relationship between the elements in the pair, while correlation using WSPC examines the covariance of two elements relative to all of the other elements in the com-

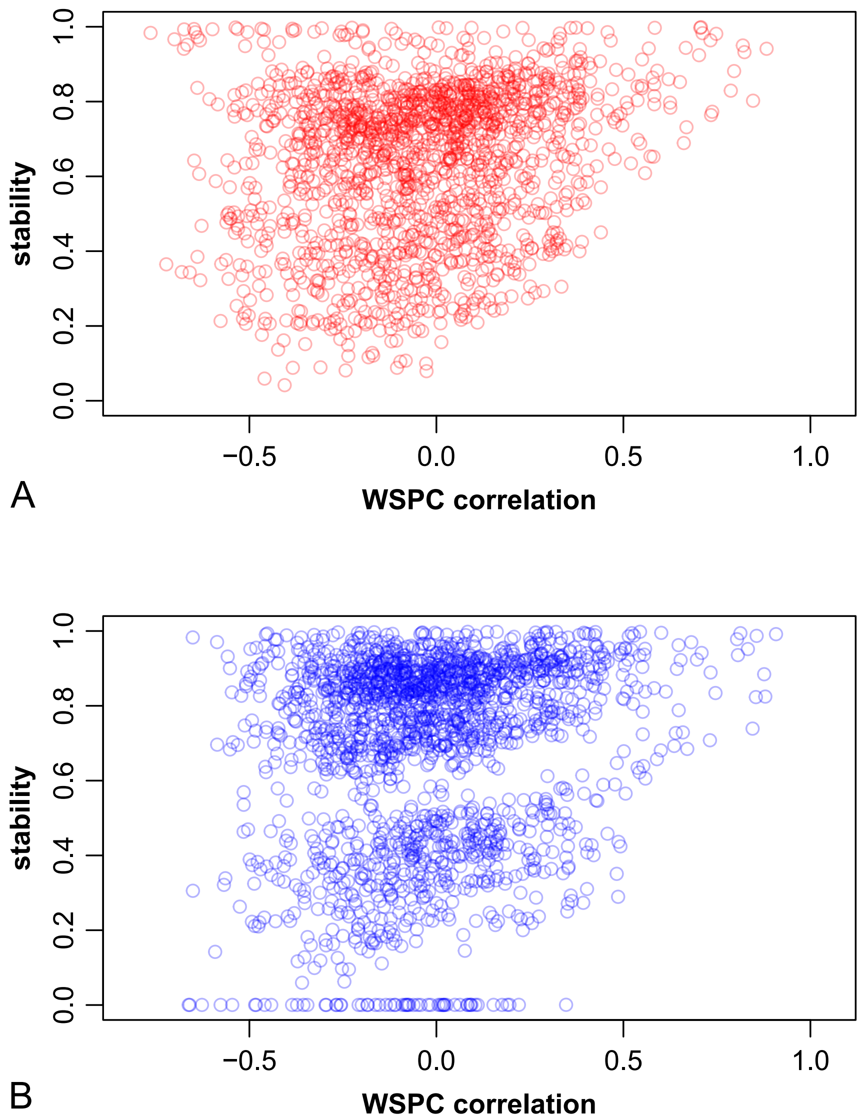


Figure 4. *Scatterplot of stability and WSPC correlation in Adaohai and Datanhao coal mine. (A) Comparison of all pairwise values for stability versus WSPC correlations on whole-coal basis and ash basis (Datanhao). (B) Comparison of all pairwise values for stability versus WSPC correlations on whole-coal basis and ash basis (Adaohai).*

position. That is, when a large number of elements are highly co-varying, correlation using WSPC may reduce the effect of that relationships as all the variables are moving in tandem, while stability value between corresponding pairs of elements would be universally high.

Another key takeaway from this analysis is that we can identify samples that exhibit relatively low logratio variance (i.e., stability) and attempt to parse them into element pairs that are not proportional (WSPC correlation coefficient near zero) versus those that are inversely related (negative WSPC correlation coefficient). This ability to discriminate between these two behaviors is well exhibited in the data for the Adaohai mine (Figure 4B), where there are two clusters, largely divided by a stability ratio of 0.5. Those element pairs in the cluster with the lower stability are primarily constrained to correlation using WSPC range of -0.6 to +0.5. It appears that there is a large fraction of pairs constrained to the range of -0.2 to +0.2, indicating that these elements are simply not related or proportional. However, a smaller subset of element pairs in this cluster exhibit much more negative WSPC correlation coefficients, suggesting that these pairs may represent antithetical relationships which we previously could not identify. However, further research is needed to clarify these distinctions.

4. Conclusions

The problem of inconsistencies in the bivariate statistical analysis between the coal elemental data on ash basis and whole-coal basis has been recognized as a long-term problem in the compositional data analysis community. In this study, we show that (1) the correlation between logratio variables using WSPC method is the same on both ash basis and whole-coal basis;(2) The stability method can also successfully solve the consistency problem, and the association is the same regardless of the bases;(3) In order to verify the performance of WSPC method and stability method, hierarchical clustering algorithm can be used to accurately establish the prediction model of the modes of occurrence of coal elemental data. The prediction results show that the WSPC method produced results more in-line with known modes of occurrence than the stability method.

All in all, the WSPC method appears to provide more insights than the previously suggested stability method. However, the WSPC method utilizes the geometric mean of all parts in the composition and can be impacted by subcompositional incoherence. For the unsolved problems in this study, we believe that some mineral forming factors in coal can be introduced in future studies to explain the early clustering of some coal elemental data. Of course, interpretation of whole-coal basis and ash basis coal elemental data through deep learning is still a hot research direction.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (No. 2021YFC2902005). Thanks are given to anonymous reviewers for their careful reviews and detailed comments.

Conflict of interest

The authors declare that there are no conflicts of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Supplementary Material

This article contains supplementary material consisting of

- The correlation matrices based on the WSPCs for each coal mine
- The matrices of stabilities for each coal mine
- Heat maps of the correlations and stabilities

References

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. London: Chapman & Hall. *Reprinted by the Blackburn Press, London, UK: Chapman & Hall*.

Aitchison, J. (2008). The single principle of compositional data analysis, continuing fallacies, confusions and misunderstandings and some suggested remedies. *In: Daunis-i-Estadella, J. and Martín-Fernández, J.A. (Eds.), Proceedings of CODAWORK'08, The 3rd Compositional Data Analysis Workshop*. May 27-30, University of Girona, Girona (Spain), CD-ROM (ISBN:978-84-8458-272-4), 28p.

Arbuzov, S. I., Chekryzhov, I. Y., Finkelman, R. B., Sun, Y. Z., Zhao, C. L., Ilénok, S. and Zarubina, N. V. (2019). Comments on the geochemistry of rare-earth elements (La, Ce, Sm, Eu, Tb, Yb, Lu) with examples from coals of north Asia (Siberia, Russian far East, North China, Mongolia, and Kazakhstan). *International Journal of Coal Geology*, 206, 106-120. https://doi.org/10.1016/j.coal.2018.10.013.

Brian, S. and Everitt, B. (2001). Applied multivariate data analysis. *1st edn.John Wiley & Sons*, 126-160. https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118887486.

Dai, S., Finkelman, R. B., French, D., Hower, J. C., Graham, I. T. and Zhao, F. (2021). Modes of occurrence of elements in coal: A critical evaluation. *Earth-Science Reviews*, 222, 103815. https://doi.org/10.1016/j.earscirev.2021.103815.

Dai, S., Graham, I. T., and Ward, C. R. (2016). A review of anomalous rare earth elements and yttrium in coal. *International Journal of Coal Geology*, 159, 82-95. https://doi.org/10.1016/j.coal.2016.04.005.

Dai, S., Hower, J. C., Finkelman, R. B., Graham, I. T., French, D., Ward, C. R. and Zhao, L. (2020). Organic associations of non-mineral elements in coal: A review. *International Journal of Coal Geology*, 218, 103347. https://doi.org/10.1016/j.coal.2019.103347.

Dai, S., Ren, D., Tang, Y., Yue, M. and Hao, L. (2005). Concentration and distribution of elements in Late Permian coals from western Guizhou Province, China.*International Journal of Coal Geology*, 61, 119-137. https://doi.org/10.1016/j.coal.2004.07.003.

Dai, S., Zhang, W., Ward, C. R., Seredin, V. V., Hower, J. C., Li, X. and Zhou, D. (2013). Mineralogical and geochemical anomalies of late Permian coals from the Fusui Coalfield, Guangxi Province, southern China: Influences of terrigenous materials and hydrothermal fluids. *International Journal of Coal Geology*, 105, 60-84. https://doi.org/10.1016/j.coal.2012.12.003.

Dai, S., Zhao, L., Peng, S., Chou, C. L., Wang, X., Zhang, Y. and Sun, Y. (2010). Abundances and distribution of minerals and elements in high-alumina coal fly ash from the Jungar Power Plant, Inner Mongolia, China. *International Journal of Coal Geology*, 81, 320-332. https://doi.org/10.1016/j.coal.2009.03.005.

Dai, S., Zou, J., Jiang, Y., Ward, C. R., Wang, X., Li, T. and Zhou, D. (2012). Mineralogical and geochemical compositions of the Pennsylvanian coal in the Adaohai Mine, Daqingshan Coalfield, Inner Mongolia, China: Modes of occurrence and origin of diaspore, gorceixite, and ammonian illite. *International Journal of Coal Geology*, 94, 250-270. https://doi.org/10.1016/j.coal.2011.06.010.

Egozcue, J.J. and Pawlowsky-Glahn, V. (2019). Compositional data: the sample space and its structure. *TEST*, 28(3), 599-638.

Egozcue, J. J., Pawlowsky-Glahn, V., . Mateu-Figueras, G. and Barceló-Vidal, C. (2003). *Isometric logratio transformations for compositional data analysis. Mathematical geology*, 35(3), 279-300. https://doi.org/10.1023/A:1023818214614.

Hou, Y., Dai, S., Nechaev, V. P., Finkelman, R. B., Wang, H., Zhang, S. and Di, S. (2023). Mineral matter in the Pennsylvanian coal from the Yangquan Mining District, northeastern Qinshui Basin, China: Enrichment of critical elements and a Se-Mo-Pb-Hg assemblage. *International Journal of Coal Geology*, 266, 104178. https://doi.org/10.1016/j.coal.2022.104178.

Hron, K., Engle, M., Filzmoser, P. and Fišerová, E. (2021). Weighted symmetric pivot coordinates for compositional data with geochemical applications. *Mathematical Geosciences*, 53, 655-674. https://doi.org/10.1007/s11004-020-09862-5.

Hron, K., Filzmoser, P., de Caritat, P., Fišerová, E. and Gardlo, A. (2017). Weighted pivot coordinates for compositional data and their application to geochemical mapping.*Mathematical Geosciences*, 49, 797-814. https://doi.org/10.1007/s11004-017-9684-z.

Liu, J., Dai, S., Song, H., Nechaev, V. P., French, D., Spiro, B. F. and Zhao, J. (2021a). Geological factors controlling variations in the mineralogical and elemental compositions of Late Permian coals from the Zhijin-Nayong Coalfield, western Guizhou, China.*International Journal of Coal Geology*, 247, 103855. https://doi.org/10.1016/j.coal.2021.103855.

Liu, J., Spiro, B. F., Dai, S., French, D., Graham, I. T., Wang, X., Zhao, L., Zhao, J. and Zeng, R. (2021b). Strontium isotopes in high- and low-Ge coals from the Shengli Coalfield, Inner Mongolia, northern China: New indicators for Ge source. *International Journal of Coal Geology*, 233, 103642. https://doi.org/10.1016/j.coal.2020.103642.

Moore, T. A., Dai, S., Huguet, C., Pearse, J., Liu, J., Esterle, J. S. and Jia, R. (2022). Petrographic and geochemical characteristics of selected coal seams from the Late Cretaceous-Paleocene Guaduas Formation, Eastern Cordillera Basin, Colombia. *International Journal of Coal Geology*, 259, 104042. https://doi.org/10.1016/j.coal.2022.104042.

Nicholas, J., Geboy, M. A. E. and Hower, J. C. (2013). Whole-coal versus ash basis in coal geochemistry: A mathematical approach to consistent interpretations. *International Journal of Coal Geology*, 113, 41-49. https://doi.org/10.1016/j.coal.2013.02.008.

Nechaev, V. P., Dai, S., Chekryzhov, I. Y., Tarasenko, I. A., Zinḱov, A. V. and Moore, T. A. (2022). Origin of the tuff parting and associated enrichments of Zr, REY, redox-sensitive and other elements in the Early Miocene coal of the Siniy Utyes Basin, southwestern Primorye, Russia. *International Journal of Coal Geology*, 250, 103913. https://doi.org/10.1016/j.coal.2021.103913.

Pawlowsky-Glahn, V. and Egozcue, J. J. (2022). Notes on the space of parts and subcompositional coherence. In: Thomas-Agnan, C. and Pawlowsky-Glahn, V.(Eds.)*Proceedings of CODAWORK2022, The 9th Compositional Data Analysis Workshop*. 28p. 28 June - 1 July, Toulouse (France), (ISBN: 978-84-947240-4-6).

Razum, I., Lužar-Oberiter, B., Zaccarini, F., Babić, L., Miko, S., Hasan, O. and Pawlowsky-Glahn, V. (2021). New sediment provenance approach based on orthonormal log ratio transformation of geochemical and heavy mineral data: Sources of eolian sands from the southeastern Adriatic archipelago.*Chemical Geology*, 583, 120451. https://doi.org/10.1016/j.chemgeo.2021.120451.

Seredin, V. V. and Dai, S. (2012). Coal deposits as potential alternative sources for lanthanides and yttrium. *International Journal of Coal Geology*, 94, 67-93. https://doi.org/10.1016/j.coal.2011.11.001.

Templ, M., Filzmoser, P. and Reimann, C. (2008). Cluster analysis applied to regional geochemical data: Problems and possibilities.*Applied geochemistry*, 23(8), 2198-2213. https://doi.org/10.1016/j.apgeochem.2008.03.004.

Xu, N., Peng, M., Li, Q. and Xu, C. (2020). Towards consistent interpretations of coal geochemistry data on whole-coal versus ash bases through machine learning.*Minerals*, 10(4), 328. https://doi.org/10.3390/min10040328.

Xu, N., Xu, C., Finkelman, R. B., Engle, M. A., Li, Q., Peng, M. and Yang, Y. (2022). Coal elemental (compositional) data analysis with hierarchical clustering algorithms. *International Journal of Coal Geology*, 249, 103892. https://doi.org/10.1016/j.coal.2021.103892.

Zhao, L., Sun, J., Guo, W., Wang, P. and Ji, D. (2016). Mineralogy of the Pennsylvanian coal seam in the Datanhao mine, Daqingshan Coalfield, Inner Mongolia, China: Genetic implications for mineral matter in coal deposited in an intermontane basin. *International Journal of Coal Geology*, 167, 201-214. https://doi.org/10.1016/j.coal.2016.10.006.

Zhou, M., Dai, S., Wang, X., Zhao, L., Nechaev, V. P., French, D., Graham, I.T., Zheng, J., Wang, Y. and Dong, M. (2022). Critical element (Nb-Ta-Zr-Hf-REE-Ga-Th-U) mineralization in Late Triassic coals from the Gaosheng Mine, Sichuan Basin, southwestern China: Coupled effects of products of sediment-source-region erosion and acidic water infiltration. *International Journal of Coal Geology*, 262, 104101. https://doi.org/10.1016/j.coal.2022.104101.