The impact of the compositional nature of data on coal reserve evaluation, a case study in Parvadeh IV coal deposit, Central Iran

Hossein Molayemat[1], Farhad Mohammad Torab[1], Vera Pawlowsky-Glahn[2],
Amin Hossein Morshedi[1], Juan José Egozcue[3]

Abstract:

Coal proximate analysis is a form of typical compositional data, commonly represented with constant sum. Although the direct geostatistical modeling of compositional data provides apparently reasonable outputs, the results are always exposed to inconsistency and non-optimality. In this paper, we compare the compositional and noncompositional approaches to assess the problems caused by neglecting the compositional nature of data. The ultimate goal is to attain an accurate approach for coal reserve evaluation. The presented compositional approach was executed and validated on field data from the Parvadeh IV coal deposit in Central Iran. The comparison of sum of values maps, Aitchison's distance, actual vs. estimated cross-plots and correct classification rate of the mentioned approaches illustrate that the compositional approach is notably more accurate and reliable. To compare tonnage – grade curves, new formulations are proposed to calculate T and T_+, two moments of total tonnage, defined based on the random variables of stochastic simulation. Tonnage – grade and mean grade – cut-off grade curves showed that the noncompositional approach has overestimated ash values and underestimated carbon content. This can lead the analyst to misinterpretation and underrating the deposit. Quantitative comparison of tonnage – grade curves of the approaches revealed that at cut-offs of 31% for ash and 52% for fixed carbon, nearly 5 and 3 million tons of coal are being considered as waste by the noncompositional approach. Consequently, neglecting the compositional nature of data will result in deviated outputs, unrealistic models and unreliable evaluations and finally lead to financial losses. Thus, it is strongly recommended to consider the compositional nature of data in reserve evaluations.

Keywords: Compositional modeling; ilr coordinates; tonnage – grade curves; coal quality estimation; Tabas coalfield.

---

[1] Department of Mining and Metallurgical Engineering, Yazd University, Yazd, Iran.
[2] Department of Informatics, Applied Mathematics and Statistics, University of Girona, Girona, Spain.
[3] Department of Civil and Environmental Engineering, Technical University of Catalonia, Barcelona, Spain.

1. Introduction

Mineral reserve evaluation is the most important and critical stage of exploration which directly affects the economy of the reserve. Tonnage – grade and mean grade – cut-off grade diagrams are common tools in reserve evaluation. These diagrams are usually computed based on geostatistical estimation and simulation of attributes such as grade, thickness, etc. Almost all types of data which are subjected to geostatistical modeling have compositional nature (Pawlowsky-Glahn and Egozcue, 2016). Compositional data or *compositions* consist of observations which are parts of a whole and carry relative information (Pawlowsky-Glahn et al., 2015a).

It is proven that straight application of standard statistical and geostatistical methods to compositional data will expose the results to inconsistencies (Pawlowsky-Glahn et al., 2015b; Reimann et al., 2008; Pawlowsky-Glahn and Egozcue, 2006). The main problem of geostatistical analysis of regionalized compositions (*r-compositions* for short) in their raw form can be stated in terms of covariances which are subjected to nonstochastic controls (Pawlowsky-Glahn, 1984).

In the 80's Aitchison developed compositional data analysis (CoDa) and introduced two log-ratio transformations: additive log-ratio (alr) and centered log-ratio (clr) transformations. Further researches illustrated that alr-transformation may lead the geostatistical estimations to inconsistent results (Pawlowsky-Glahn and Olea, 2004; Pawlowsky-Glahn et al., 1995). On the other hand, because the covariance matrix of a clr transformation is singular, it is difficult to use clr with geostatistical methods. In 2003, Egozcue et al. (2003) introduced the family of isometric log-ratio (ilr) transformations and later its advantage was revealed as a proper transformation for combined application with geostatistical methods (Tolosana-Delgado and Boogaart, 2013; 2014).

In the last few years, compositional data analysis became a point of interest in geosciences. Many studies were performed in the fields of soil geochemistry (Levitan et al., 2015; Sun et al., 2014; Palarea-Albaldejo et al., 2014; Zhang et al., 2013; Reimann et al., 2012; Filzmoser et al., 2010; 2009; Filzmoser and Hron, 2009) and sedimentology (Scealy et al., 2015; Grunsky et al., 2014; Engle and Rowan, 2013; Lark et al., 2012; Tolosana-Delgado, 2012). Some other studies used CoDa for lithogeochemical data (Montreuil et al., 2013; Martin-Fernandez et al., 2005) as well as mineralogy, genesis and quality modelling of ore deposits (Makvandi et al., 2016; Engle and Blondes, 2014; Jelsema and Paul, 2013; Grunsky et al., 2008; Drew et al., 2008). A wide application of CoDa can be found in hydrochemistry and hydrogeology (Olea et al., 2017; Owen et al., 2016; Pardo-Iguzquiza et al., 2015; Buccianti et al., 2014). Recently, CoDa was used as data processing tool in mineral exploration (Moeini and Torab, 2017; Zuo et al., 2013). Combination of CoDa and geostatistical methods is discussed by Olea et al. (2016), Olea and Luppens (2015), Pawlowsky-Glahn et al. (2015c), Tolosana-Delgado and Van Den Boogaart (2014) and Karacan and Olea (2018) but none of them dealt with reserve evaluation in terms of cut-off based curves.

In this study, for the first time, the impact of the compositional nature of data on coal reserve evaluation is discussed. The main goal of this paper is to evaluate a deposit using the conventional and CoDa based approaches, for comparison and to provide reliable reserve evaluation.

2. Case study

Parvadeh coalfield is the largest coking coal deposit in Iran, and Parvadeh IV (PIV) is the largest sub-region of the Parvadeh deposit. Considering the fault structures, PIV can be divided into northern and southern parts (Fig. 1). Regarding the extent, continuity and quality, B_2 seam is the most favorable coal seam in PIV and it is the main source of bituminous coal and coalbed methane (CBM) in the deposit (Molayemat and Torab, 2017).

In the northern PIV, 94 vertical boreholes were drilled in various mineral exploration phases. The drill core samples of B_2 seam were analyzed for ash (%), moisture (%) and volatile matter (%) on the basis of proximate analysis routine (ASTM, 2013). Fixed carbon (carbon for short) concentration is not obtained by measurement; it is calculated as the difference of 100 to the sum of ash, moisture and volatile matter. In Fig. 2, proximate analysis reports 4 positive values with a constant sum of 100%.

In this paper, we aimed to propose an approach to model the spatial dependency of regionalized compositional variables of a coal proximate analysis data. CoDa is used to prevent our model from falsification due to spurious spatial correlations.
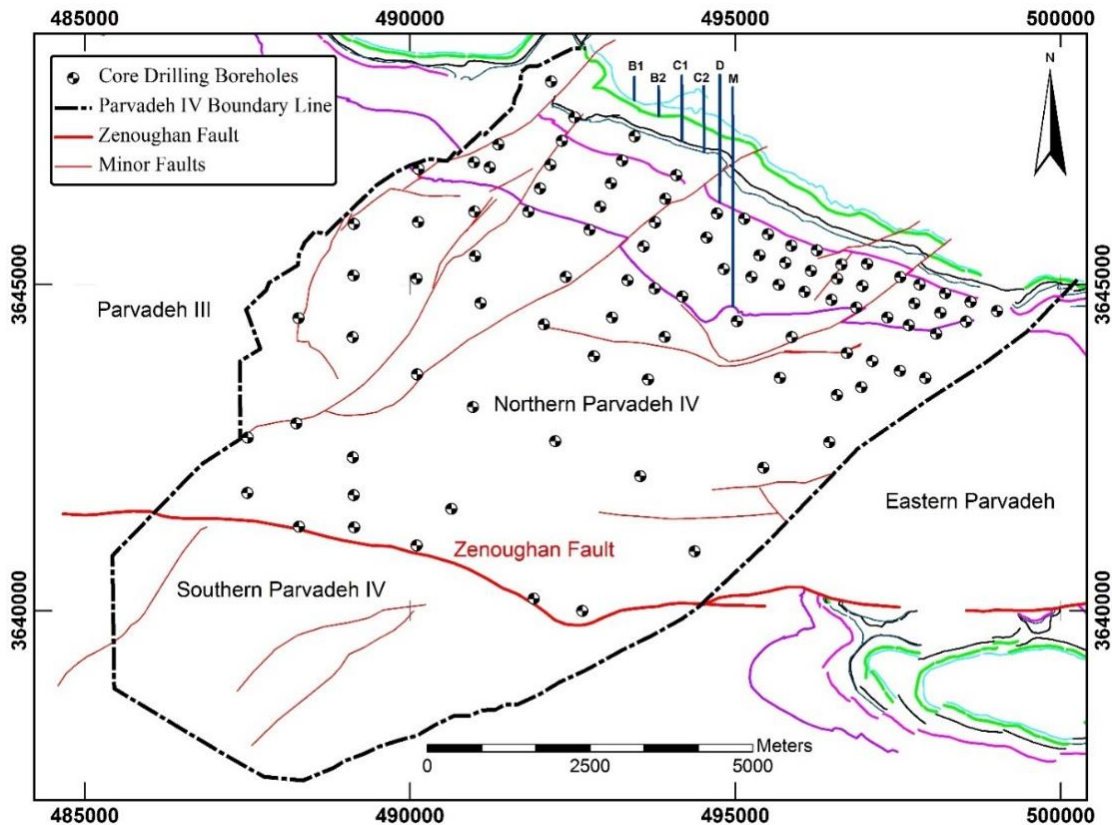

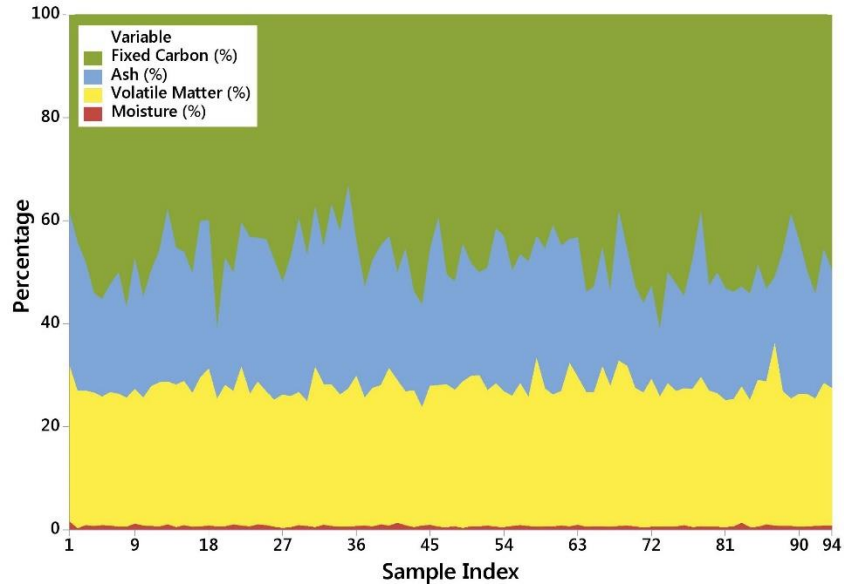**Fig. 1. Basemap of the northern Parvadeh IV in the upper part of Zenoughan fault with the layout of exploratory boreholes.**

Fig. 2. Area graph of 94 coal samples with constant sum of 100%.

3. Methodology

To attain the main goal of this paper, two approaches were followed:

- First approach (noncompositional): geostatistical estimation, simulation and evaluation based on the raw data – which is common in exploration projects (Fig. 3).
- Second approach (compositional): geostatistical estimation, simulation and evaluation based on CoDa of r-compositions (Fig. 4).

The main objectives of this paper are summarized as follow:

- Combining CoDa with geostatistical tools and comparing the outputs with conventional geostatistical modeling;
- Analyzing and investigating the effect of CoDa on optimality and validity of geostatistical models;
- Comparing tonnage – grade and mean-grade – cut-off grade curves of the two mentioned approaches.

The first step in data modeling is studying the descriptive statistics. As a part of noncompositional approach, normal probability plot and basic statistics of four variables are shown in Fig. 5. As variables are constrained to be positive and smaller than 100, they cannot be normal; this is a theoretical impossibility. Any assessment in favor of a normal distribution is just an incoherent approximation. The empirical distributions of ash and carbon are close to symmetrical distributions. Volatile matter and moisture percentages were transformed into ~~the normal~~ distribution for further Gaussian simulations. As there are no censored and negative values in the raw data, it is easy to execute further calculations.

4

To perform geostatistical modeling, both kriging and stochastic simulation are executed on the PIV dataset. Kriging is a method with minimum estimation error, but it is conditionally biased and also not suitable for uncertainty modeling (Olea and Lupens, 2015). Moreover, kriging results are exposed to the smoothing effect and the problematic different spatial statistics, such as different variograms of estimated values and samples (Olea, 2009). On the other hand, stochastic simulation fills the gap and weaknesses of kriging.
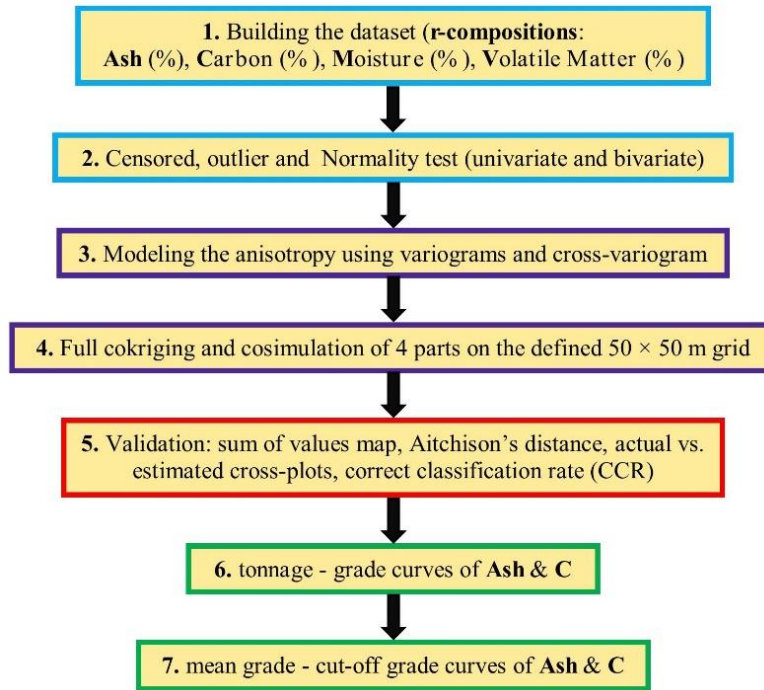
1. Building the dataset (**r-compositions**: **Ash** (%), **Carbon** (%), **Moisture** (%), **Volatile Matter** (%))

2. Censored, outlier and Normality test (univariate and bivariate)

3. Modeling the anisotropy using variograms and cross-variogram

4. Full cokriging and cosimulation of 4 parts on the defined 50×50 m grid

5. Validation: sum of values map, Aitchison's distance, actual vs. estimated cross-plots, correct classification rate (CCR)

6. tonnage - grade curves of **Ash** & **C**

7. mean grade - cut-off grade curves of **Ash** & **C**

Fig. 3. Flowchart of the first approach (noncompositional)

1. Building the dataset (**r-compositions**: **Ash** (%), **Carbon** (%), **Moisture** (%), **Volatile Matter** (%))

2. Defining SBP matrix based on clr-biplots of Ash, C, M, V.

3. ilr coordinates

4. Normality test of balances (univariate and bivariate)

5. Modelling the anisotropy (semivariograms and cross-variograms)

6. Full cokriging and cosimulation of 3 balances on the defined 50 × 50 m grid

7. Calculating E-types

8. ilr back-transformation

9. Validation: sum of values map, Aitchison's distance, actual vs. estimated cross-plots, correct classification rate (CCR)

10. tonnage - grade curves of **Ash** & **C**

11. mean grade - cut-off grade curves of **Ash** & **C**

Fig. 4. Flowchart of the second approach (compositional)

Fig. 5. Normal probability plot of the proximate analysis data: (A) ash; (B) moisture; (C) volatile matter; (D) fixed carbon; (E) transformed moisture; (F) transformed volatile matter.

3.1 Logratio approach for geostatistical modeling

When using alr, clr and ilr as coordinates, only ilr is free of problems, as they are Cartesian coordinates of the compositions. Problems that can appear are singular covariance matrices, subcompositional incoherence and oblique coordinates (Egozcue et al., 2003; Filzmoser and Hron 2009; Olea and Lupens, 2015), which are critical for the combined application of CoDa and geostatistics. In this paper, the unconstrained ilr-coordinates known as *balances* are produced by defining a sequential binary partition (SBP) for the 4-part r-compositions. Here we concentrate on the application of balances in the compositional approach and we do not go deep into mathematical details of the ilr representation; readers may refer to Egozcue et al. (2003).

The full closed system with D parts (in our case D = 4) is transformed into D-1 balance-coordinates and the relationship of these balances with the parts is through log-ratios of geometric means of parts. In practice, a proper SBP is defined based on the mutual relationship of parts. Considering

a compositional dataset with D parts, an SBP matrix will have D columns and D-1 rows; each row indicates a group of parts. The participation of a part in a group is coded by +1, -1 and 0; where, by convention, +1 indicates the parts in the numerator, -1 the parts in the denominator, and 0 the parts not participating in the partition. The code does not refer to any specific relationship between the parts, except that the researcher thinks they should be in a group.

Drawing clr biplots is a graphical technique to visualize the behavior of parts and fill the SBP matrix with +1, -1 and 0 values. In this study there are 4 parts including ash, carbon (C), moisture (M) and volatile matter (V). The ilr transformation leads to 3 balances denoted by ilr_1, ilr_2 and ilr_3. These balances are the regionalized variables in the second approach. It is notable that all E-types of the second approach are computed before ilr backtransformation ~~to avoid artifacts.~~

==The clr biplot of proximate analysis data of PIV explains 95.0% of the variance and it is shown in Fig. 6. Based on inverse relationship of ash with the other parts in clr-biplot of Fig. 6 (A) and considering that the clr (Ash (%)) has the longest ray, it has been decided to mark Ash (%) with -1 at the first row of the SBP matrix. This procedure was repeated for M (%) based on clr biplot of Fig 6 (B).==

The SBP matrix is defined by Eq. 1:

$$SBP = \begin{array}{c} \\ ilr_1 \\ ilr_2 \\ ilr_3 \end{array} \begin{bmatrix} Ash & C & M & V \\ -1 & +1 & +1 & +1 \\ 0 & +1 & -1 & +1 \\ 0 & +1 & 0 & -1 \end{bmatrix}. \tag{1}$$

Based on SBP matrix, ilr balances are given by Eq. 2:

$$ilr_i = \sqrt{\frac{r_i s_i}{r_i + s_i}} \ln \left[\frac{g(z_+)}{g(z_-)} \right], \qquad i = 1, 2, \ ..., \ D\text{-}1 \tag{2}$$

where r_i is the number of +1 values in i-th row of SBP, s_i is the number of -1 values in the same row, $g(z_+)$ and $g(z_-)$ are the geometric mean of parts which corresponding values in the SBP are +1 and -1 respectively.
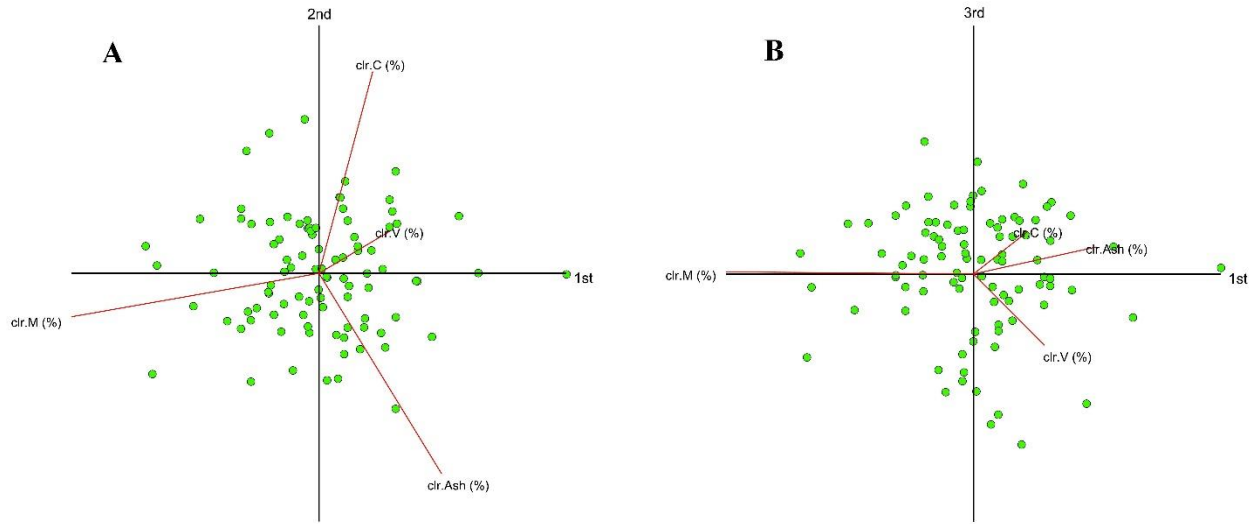
3.2 Geostatistical modeling of case study

Considering the flowcharts of Figs. 4 and 5, approaches are fed with different input datasets which contain ash, C, M and V for the first approach and ilr_1, ilr_2 and ilr_3 for the second one. Performing preliminary statistical tests revealed no censored and outlier data. To perform the Gaussian simulation in both approaches, the distribution function of all variables must follow a normal distribution. Anderson-Darling and Kolmogorov-Smirnov normality tests showed that ash, C and all three balances can be assumed to follow symmetrical distribution (Fig. 5(A)-(B), Fig. 7). V and M variables are transformed with sinus hyperbolic functions to follow a symmetric distribution and backtransformed into their original domains after estimation and simulation (Fig. 5(E)-(F)).
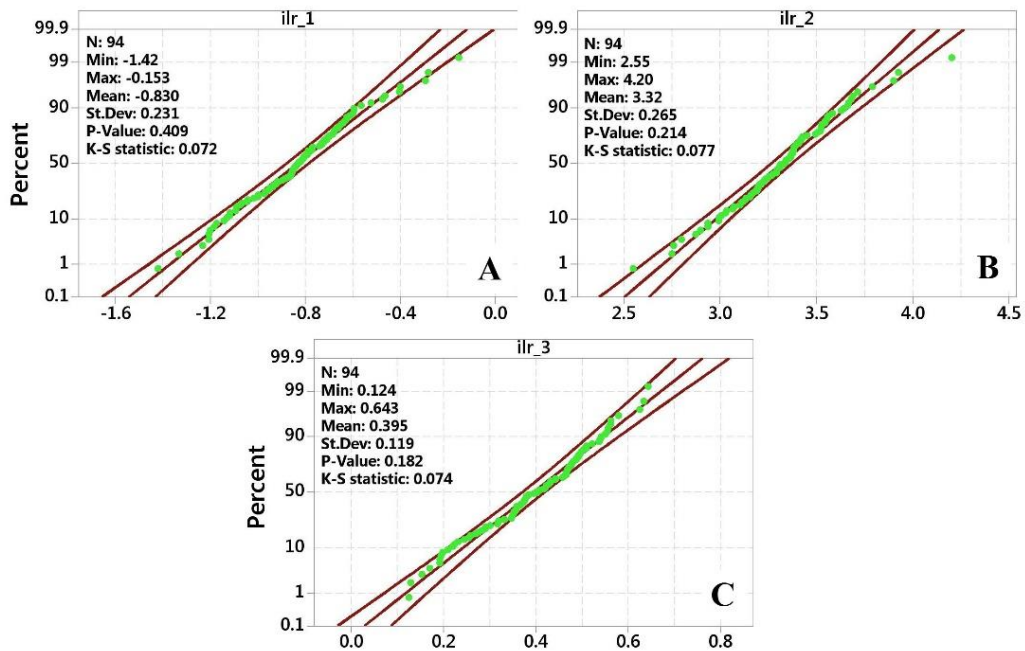

Fig. 7. Normal probability plot of ilr balances: (A) ilr_1; (B) ilr_2; (C) ilr_3.

9

The next step is the structural analysis of spatial dependency via modeling experimental variograms and cross-variogarms for both approaches. Because they indicated no significant anisotropy, experimental variograms were modeled by isotropic models (Figs. 8 and 9). Paired variables show spatial cross correlation, comprising ash-c, M-V, ilr_1-ilr_2 and ilr_1-ilr_3. Because there are cross correlations among variables, cokriging and cosimulation are employed to perform geostatistical modeling with both approaches.

For interpolation of variables, a two dimensional grid is defined with the cell size of 50×50 m, concerning longwall mining plans in PIV. Taking the deposit boundary into account, the grid contains a total of 25756 cells in the area of 64.39 km^2. By performing both approaches, there will be 714 estimated/simulated values for every single cell (400 values of cosimulated ash, C, M, V; 4 values of their E-types; 4 values of cokriging results and 300 values of cosimulation of ilr_1, ilr_2, ilr_3; 3 values of their E-types and 3 of cokriging results).
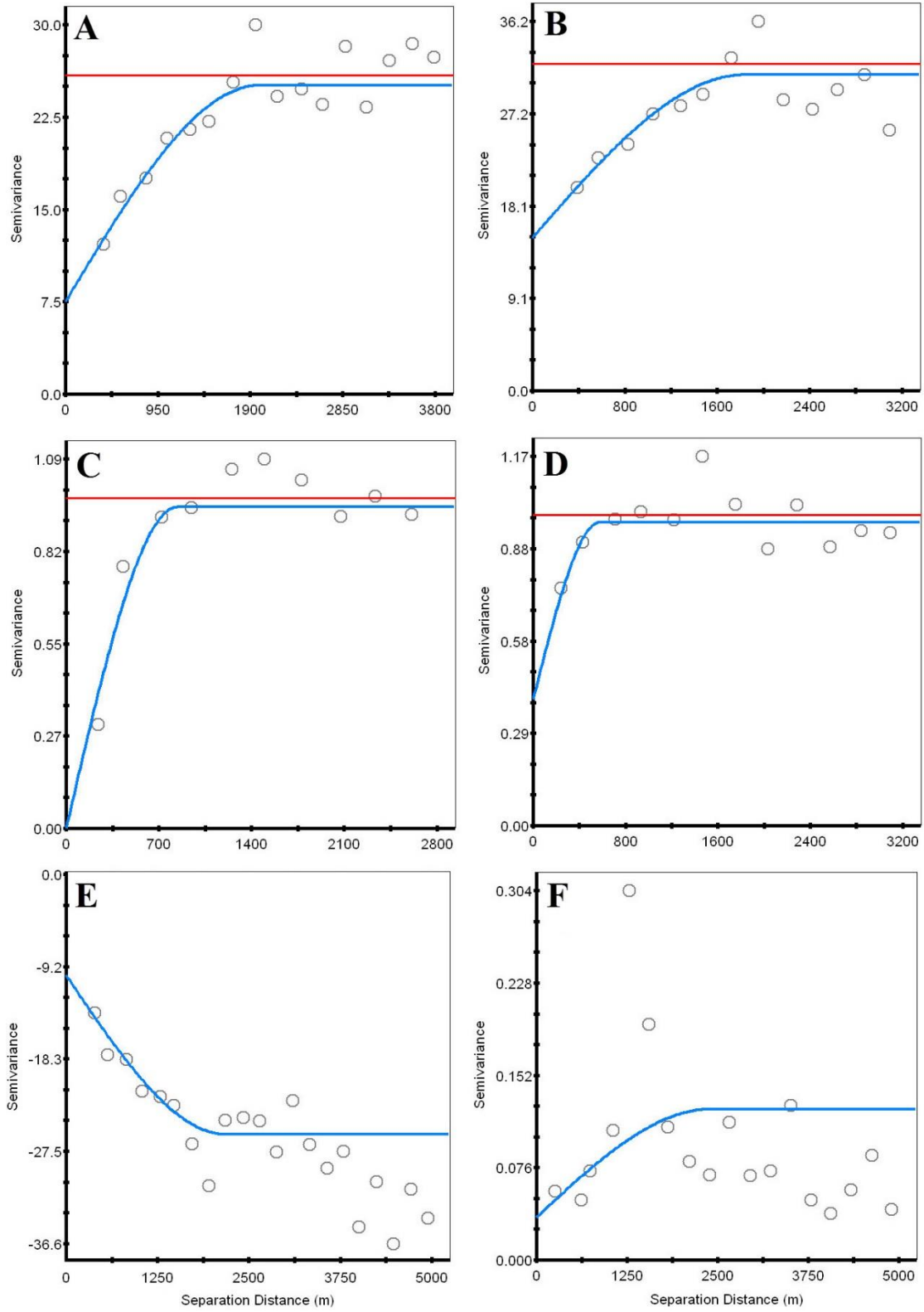
Fig. 8. Spatial structure analysis of raw data using variogram and cross-variogram: (A) ash; (B) carbon; (C) transformed moisture; (D) transformed volatile matter; (E) ash-carbon; (F) transformed moistre – transformed volatile matter
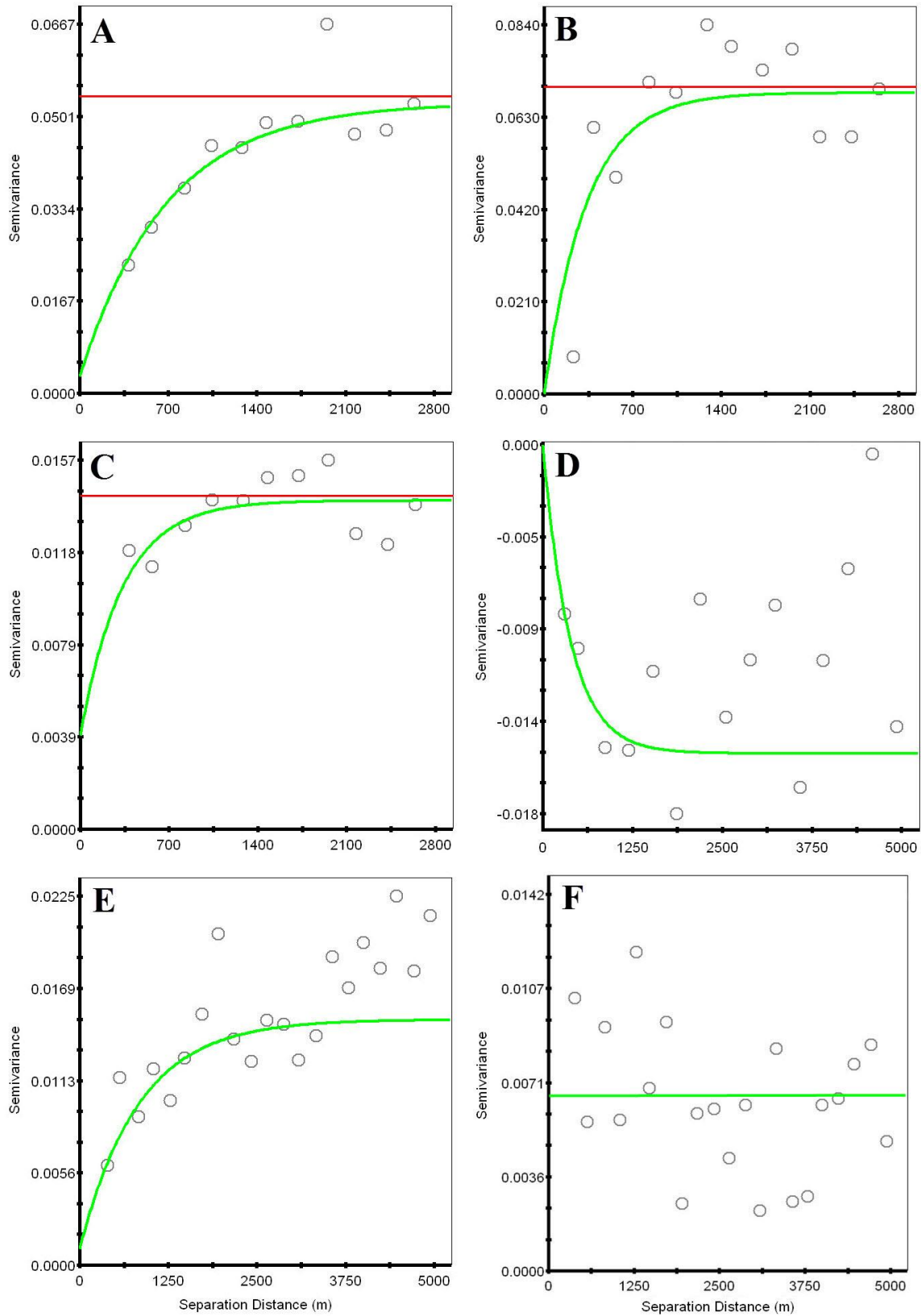
Fig. 9. Spatial structure analysis of ilr balances using variogram and cross-variogram: (A) ilr$_1$; (B) ilr$_2$; (C) ilr$_3$; (D) ilr$_1$-ilr$_2$; (E) ilr$_1$-ilr$_3$; (F) ilr$_2$-ilr$_3$.

In order to compare the approaches and calculate cut-off based curves, it is necessary to backtransform estimated/simulated ilrs to the original data space with ilr inverse function. By defining clr transformation (Eq. 3) and ilr transformation in matrix form (Eq. 4), the ilr inverse function is given by Eq. 5 (Pawlowsky-Glahn et al., 2015a).

$$clr\left(\mathbf{Z}(x_i)\right) = \ln\frac{x_i}{g_m\left(\mathbf{Z}(x_i)\right)}, \qquad i = 1, 2, ..., N \tag{3}$$

$$ilr\left(\mathbf{Z}(x_i)\right) = \boldsymbol{\psi} \cdot clr\left(\mathbf{Z}(x_i)\right), \qquad i = 1, 2, ..., N \tag{4}$$

$$\mathbf{Z}(x_i) = c \cdot exp\left(\boldsymbol{\psi}^T \cdot ilr\left(\mathbf{Z}(x_i)\right)\right), \qquad i = 1, 2, ..., N \tag{5}$$

where $clr(\mathbf{Z}(x_i))$ is the set of transformed composition by clr transformation, $g_m(\mathbf{Z}(x_i))$ is the componentwise geometric mean of the composition, $ilr(\mathbf{Z}(x_i))$ is the set of transformed balances, $\boldsymbol{\psi}$ is the contrast matrix, which is defined by Eq. 6, $\mathbf{Z}(x_i)$ is the set of compositions at location x_i and c is the closure constant one over the sum of elements in the vector $\boldsymbol{\Psi}^T \cdot ilr$.

$$\psi_{i,j} = \begin{cases} 0 \ \text{if} \ v_{ij} = 0, \\[2mm] \sqrt{\dfrac{s_i}{r_i \cdot (s_i + r_i)}} \quad \text{if} \ v_{ij} = +1, \\[2mm] -\sqrt{\dfrac{r_i}{s_i \cdot (s_i + r_i)}} \quad \text{if} \ v_{ij} = -1, \end{cases} \tag{6}$$

where v_{ij} denotes the element of i-th row and j-th column in the SBP matrix. r_i and s_i were explained in section 3.1. Figs. 10-13 display the E-type maps of ash and carbon for both approaches. SGeMS software (Remy et al., 2009) was used for sequential Gaussian co-simulation and E-type calculations based on mean values of simulated realizations for every cell in the grid. Depending on the geometry of the sample space, mean values are calculated by arithmetic means. General spatial variability of ash and carbon are nearly similar but there are differences in the shape and extent of areas with high and low values (e.g. low ash bearing area in east and high carbon bearing area in southern parts of the deposit).
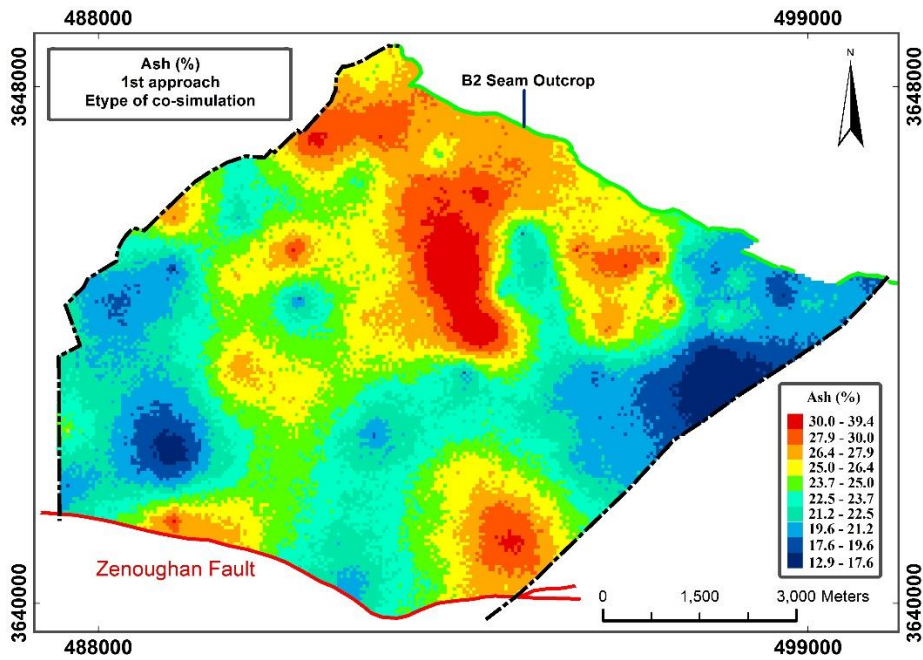
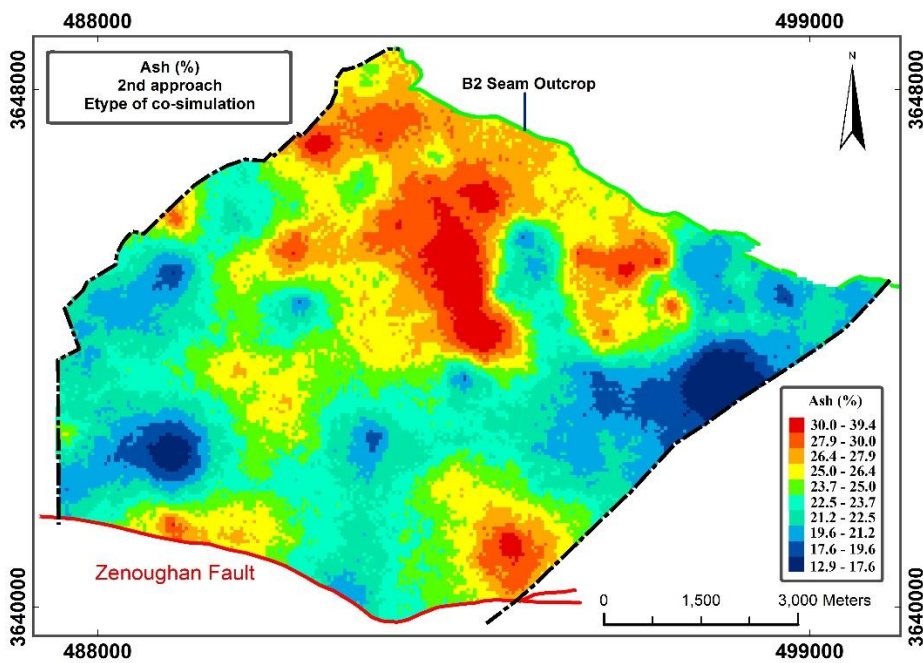**Fig. 10. E-type map of ash (noncompositional approach)**

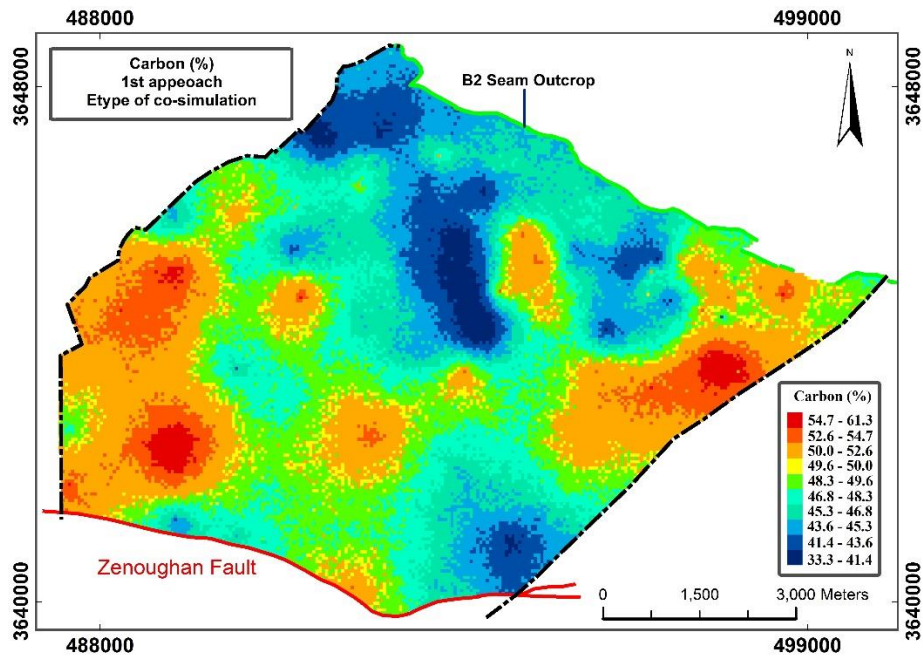**Fig. 11. E-type map of ash (compositional approach)**
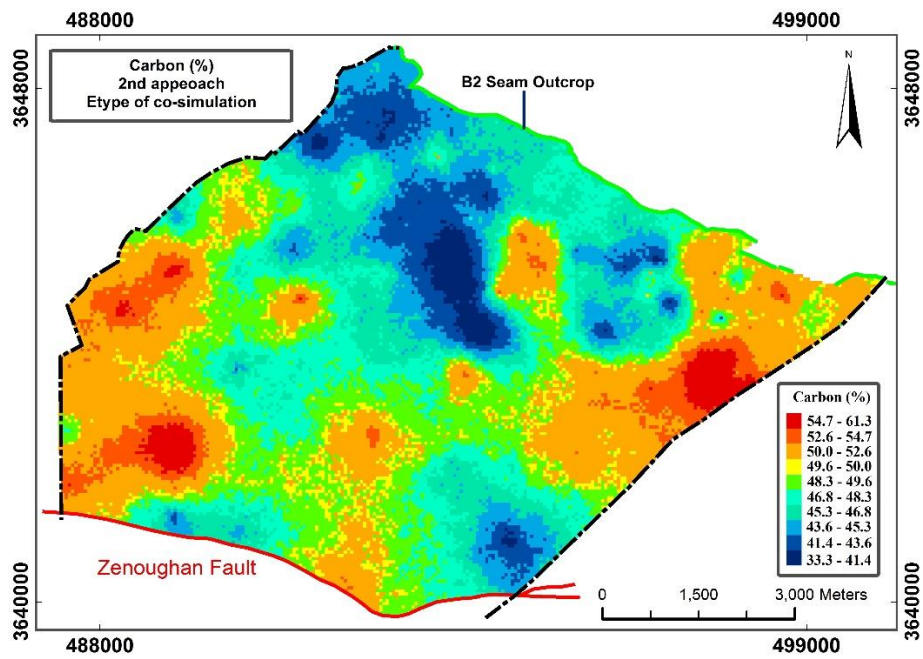
Fig. 12. E-type map of carbon (noncompositional approach)

Fig. 13. E-type map of carbon (compositional approach)

15

3.3 Reserve evaluation

3.3.1 Tonnage – grade diagrams and random compositions

Tonnage – grade curves are useful tools in reserve evaluation, which enable the modeler to summarize estimations by reporting tonnage values in various possible cut-offs. One possible form of tonnage – grade curves are those which calculate *contained tonnage*. In such cases, if an estimated grade of a block (cell) is greater than the cut-off grade, the estimated grade multiplies the block tonnage. Repeating this procedure for all blocks and summing all the tonnages will give the total contained tonnage. This method is based on the assumption that estimated grade of a block is unique. In this paper, we propose a way of tonnage-grade calculation, based on random compositions.

Consider the compositional random field, characterized by a 4-part r-compositions (ash, C, M, V), $\mathbf{z}(\mathrm{x})$ at each coordinate x. Assume that each point x_i is on a regular grid and $\mathbf{z}(\mathrm{x}_i)$ has been estimated or simulated applying cokriging or cosimulation. This means that the mean and the variance of the ilr-coordinates (balances) are known. Let the coordinates be $\mathbf{W}(\mathrm{x}_i) = \mathrm{ilr}(\mathbf{z}(\mathrm{x}_i))$, which are real random variables; their mean is denoted as $\mu_w(\mathrm{x}_i) = E(\mathbf{W}(\mathrm{x}_i))$ and the covariance matrix of the ilr-coordinates as $C(\mathrm{x}_i)$. Consider carbon as the first component of $\mathbf{z}(\mathrm{x})$, the traditional tonnage estimation consists in adding the contribution of those blocks with centers in x_i which estimated value is above the cut-off. Assuming interpolation done on the ilr-coordinates, the tonnage would be

$$T_{tr} = \sum_{i=1}^{N} M_i \cdot S_Z^1(\mathrm{x}_i) \cdot \mathrm{I}\{S_Z^1(\mathrm{x}_i) \geq t_1\}, \tag{7}$$

where T_{tr} is the contained tonnage which is calculated in traditional reserve evaluations, M_i is the tonnage of the block, $S_Z^1 = ilr_1^{-1}(S_W(\mathrm{x}_i))$ is the carbon component of the composition and the 1's are referred to the first component. The indicator function $\mathrm{I}\{S_Z^1(\mathrm{x}_i) \geq t_1\}$ takes the value 1 when estimated carbon $S_Z^1(\mathrm{x}_i)$ is greater than the cut-off t_1 and 0 otherwise. $S_Z(\mathrm{x}_i)$ in T_{tr} was obtained by E-type calculation of 100 simulated realizations and E-type values of the second approach were calculated before ilr backtransformation. In this approach, the variability of estimated components is ignored.

T_{tr} is based on the idea that $M_i \cdot S_Z^1(\mathrm{x}_i)$ is the mass of carbon in a block centered at x_i, and that the estimated value for the block is constant. This approach neglects the grade uncertainty of a block and calculates the tonnage based on conditionally biased values. To solve this limitation, it seems reasonable to consider random variables to account for variability with a simulation technique.

The tonnage of coal in a block, M_i, is calculated by multiplying Specific Gravity (SG) to seam thickness of the block multiplying by block area. After the required preprocessing, structural analysis of SG and seam thickness was performed by generating experimental variograms.
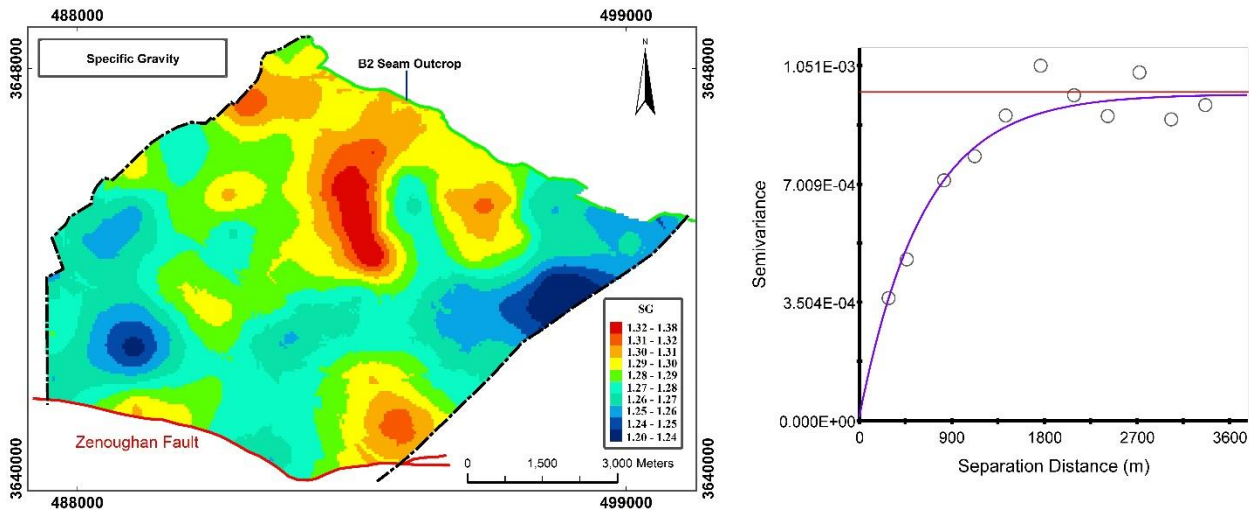
Fig. 14. Omni-directional variogram and distribution map of Specific Gravity in the northern PIV.
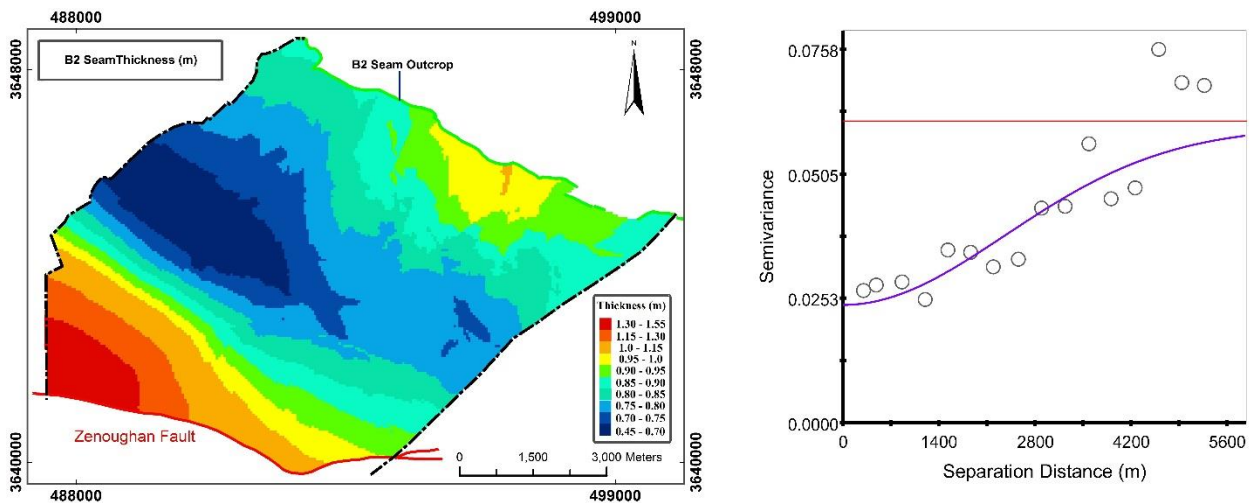

Fig. 15. Omni-directional variogram and distribution map of seam thickness in the northern PIV.

17

Let $\mathbf{w}^{(k)}(\mathbf{x}_i)$ be the k-th simulation of $\mathbf{W}(\vec{\mathbf{x}}_i)$ following the normal distribution with known mean and covariance matrix. We assume k as the number of realization in each block, $k = 1, 2, \ldots, K$, with K a large number. With 3 ilr balances, 2572 blocks in PIV deposit and $K = 100$, the total number of simulated values is 772200. The simulated ilr-coordinates (balances) $\mathbf{w}^{(k)}(\mathbf{x}_i)$ can be backtransformed to a composition $\mathbf{z}^{(k)}(\mathbf{x}_i) = ilr^{-1}(\mathbf{w}^{(k)}(\mathbf{x}_i))$, from which we are interested in the carbon component $\mathbf{z}_1^{(k)}(\mathbf{x}_i)$. Assuming that tonnage is a real variable, T is defined as:

$$T = \frac{1}{K}\sum_{i=1}^{N} M_i \sum_{k=1}^{K} z_1^{(k)}(\mathbf{x}_i) \cdot \mathrm{I}\left\{z_1^{(k)}(\mathbf{x}_i) \geq t_1\right\}. \tag{8}$$

T is the total tonnage, which is calculated by considering grade uncertainty via simulation. The expression modifies to a geometric mean if total tonnage (T_+) is taken as a variable in $\mathbb{R}_+$, but block masses are still thought as something to sum up.

$$T_+ = \exp\left[\frac{1}{K}\sum_{k=1}^{K} \ln\left(\underbrace{\sum_{i=1}^{N} \underbrace{M_i z_1^{(k)}(\mathbf{x}_i) \cdot \mathrm{I}\left\{z_1^{(k)}(\mathbf{x}_i) \geq t_1\right\}}_{\text{(k)-th simulated mass in i-th block}}}_{\text{(k)-th simulated tonnage}}\right)\right]. \tag{9}$$

T is always greater than T_+, as they are an arithmetic and a geometric mean of simulated masses in blocks.

Whereas ash has a negative influence on coal quality, it has to be performed in the reverse manner of imposing cut-offs, hence, the indicator function of ash is defined as $\mathrm{I}\left\{z_2^{(k)}(\mathbf{x}_i) \leq t_1\right\}$, which takes the value 1 when simulated ash $z_2^{(k)}(\mathbf{x}_i)$ is less than the cut-off t_1 and 0 otherwise. This made us to choose descending x axis in ash tonnage – grade diagrams in order to draw strictly descending curves. As seen in Figs. 16 and 17, there are differences between the tonnage – grade curves of the two approaches, especially in the ash curves. At small cut-off values of carbon, T_{tr} is greater than T and T_+, but the reverse behavior is observed for cut-offs greater than 48% (Table 1). It is the same for ash curves and the cut-off of 23.5% (Table 2). This trend may occur as a result of *cell-based* and *simulation-based* calculation of T_{tr} and T, T_+ respectively.

In case of ash, T_{tr} values for both approaches are almost the same for grades less than 30%. For cut-offs greater than 22%, T values of the first approach are greater than those of the second approach, but this relationship is reversed for cut-offs less than 22% (Table 2). This manner is observed as well for T_+ curves (Fig. 16 and Table 2). Tonnage – grade curves of ash show that the common geostatistical modeling of raw data overestimated ash content rather than those estimated by the compositional method. In the next section, we will show that based on several criteria, the compositional approach results are more accurate in comparison with results obtained by the noncompositional approach. Thus, the deviated results of the first approach can decreases the

18

To overcome these problems, the second approach is recommend to the modelers, which is presented in section 3.

In case of carbon, there is no significant difference between the tonnage – grade curves of both approaches (Fig. 17 and Table 1). It seems that carbon tonnages are ~~free of~~ compositional constraint. This is probably due to the range of values of carbon content. The problem is, one never knows in advance if both approaches will give the same or similar results. Therefore, to be on the safe side, the second approach is recommended.

For detailed quantitative comparison, tonnage – grade curves of both approaches are analyzed based on two cut-offs of 31% of ash and 52% of carbon. Studying T diagrams and tables 1 and 2 demonstrated that the contained ash from the first approach is 1.55 million tons greater than the one obtained by the second approach. This amount is equal to 5.0 million tons of coal. On the other side, contained carbon of the first approach is 1.48 million tons less than that of the second approach, which is equal to 2.85 million tons of coal. It can be summarized that if ash content is considered as the coal quality criterion, the first approach misclassifies 5.0 million tons of coal as waste at the cut-off of 31%. This amount is 2.85 million tons of coal if carbon content is considered as coal quality criterion. It is noteworthy that the total coal tonnage of northern PIV is 71.1 million tons, thus, these biased estimates can cause considerable financial losses.
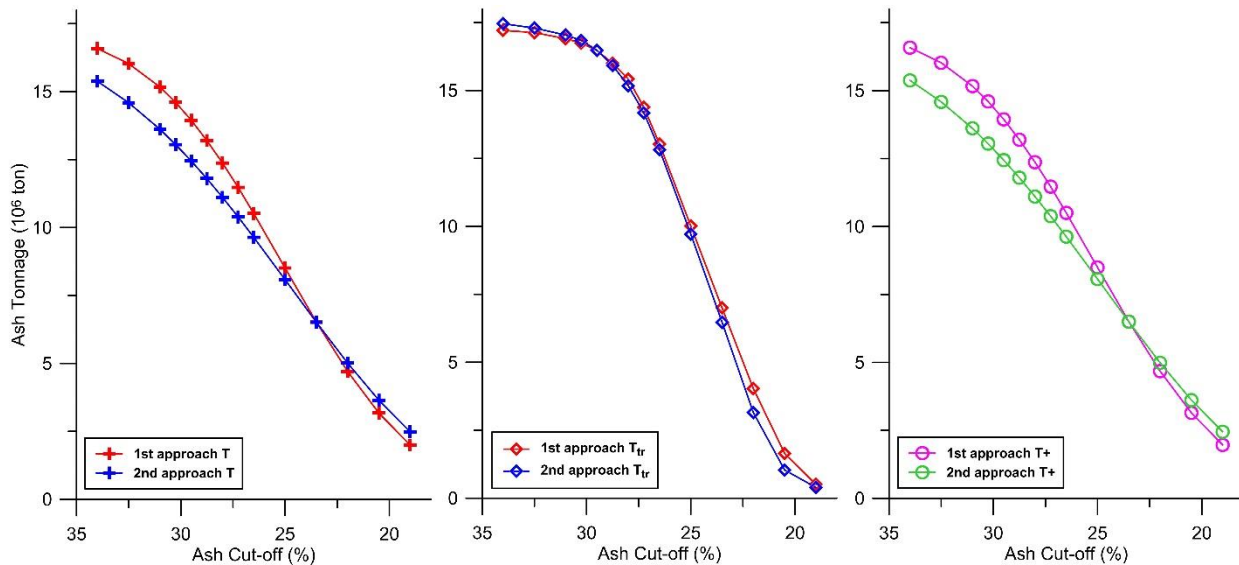

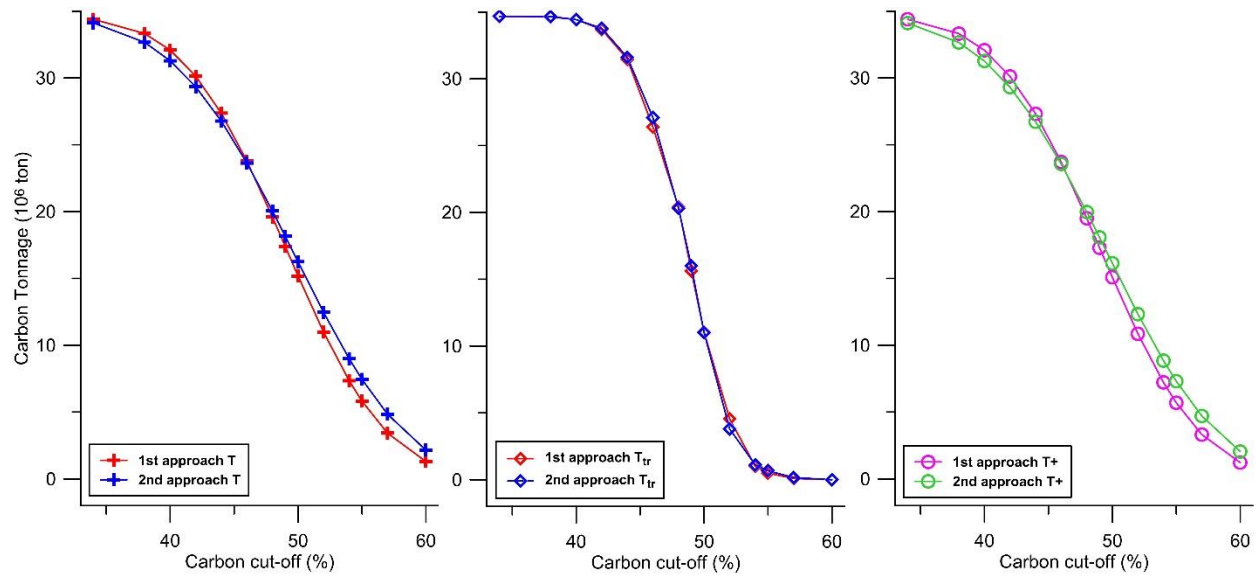**Fig. 16. Ash tonnage – grade curves for both approaches**

Fig. 17. Carbon tonnage – grade curves for both approaches

Table 1. Contained tonnage of carbon for both approaches (in million tons).

Cut-off (%)	1st app. T_{tr}	2nd app. T_{tr}	1st app. T	2nd app. T	1st app. T_+	2nd app. T_+
34	34.66	34.68	34.39	34.12	34.39	34.11
38	34.63	34.66	33.33	32.67	33.32	32.65
40	34.42	34.42	32.09	31.29	32.07	31.27
42	33.69	33.78	30.14	29.34	30.11	29.31
44	31.46	31.61	27.36	26.77	27.32	26.73
46	26.39	27.10	23.77	23.63	23.71	23.57
48	20.40	20.31	19.59	20.05	19.50	19.97
49	15.63	16.00	17.39	18.16	17.30	18.07
50	11.05	11.01	15.19	16.25	15.09	16.14
52	4.56	3.80	10.98	12.46	10.86	12.33
54	1.01	1.12	7.34	8.99	7.22	8.85
55	0.49	0.67	5.82	7.45	5.70	7.31
57	0.09	0.16	3.42	4.84	3.31	4.71
60	0.00	0.00	1.30	2.15	1.22	2.05

20

Table 2. Contained tonnage of ash for both approaches (in million tons).

Cut-off (%)	1st app. T$_{tr}$	2nd app. T$_{tr}$	1st app. T	2nd app. T	1st app. T$_+$	2nd app. T$_+$
34	17.21	17.46	16.59	15.38	16.58	15.38
32.5	17.13	17.30	16.03	14.59	16.03	14.59
31	16.91	17.04	15.17	13.62	15.17	13.61
30.25	16.74	16.84	14.60	13.05	14.60	13.05
29.5	16.47	16.48	13.95	12.45	13.94	12.44
28.75	16.01	15.92	13.20	11.80	13.19	11.79
28	15.43	15.17	12.37	11.11	12.37	11.10
27.25	14.38	14.18	11.48	10.39	11.47	10.38
26.5	13.03	12.82	10.51	9.64	10.50	9.63
25	10.01	9.72	8.51	8.08	8.49	8.07
23.5	7.00	6.47	6.52	6.52	6.50	6.49
22	4.03	3.15	4.70	5.01	4.68	4.98
20.5	1.65	1.04	3.17	3.64	3.14	3.61
19	0.50	0.40	1.99	2.48	1.97	2.45

3.3.2 Mean grade – cut-off grade diagrams

Mean grade – cut-off grade (MG-CG) curves are also common tools in reserve evaluations. These curves are similar to tonnage – grade curves, but instead of contained tonnage, mean grade is calculated for those parts of the deposit which are larger or smaller than a cut-off.

The MG-CG curves of ash and carbon for both approaches are calculated using the same cut-offs as in section 3.3.1. To take random variables into account, the output realizations of cosimulation are employed to compare the uncertainty of mean grade for compositional and noncompositional approaches. There are 100 realizations of ash and carbon per each approach. For better illustration, it is decided to calculate and display the MG-CG curves of only 10 realizations from minimum to maximum values of mean grade. In case of carbon, mean grade of each realization is defined as:

$$MG_{carbon} = \frac{\sum_{i=1}^{N} z_1(\mathrm{x}_i) \cdot \mathrm{I}\{z_1(\mathrm{x}_i) \geq t_1\}}{\sum_{i=1}^{N} \mathrm{I}\{z_1(\mathrm{x}_i) \geq t_1\}}, \tag{10}$$

where $z_1(\mathrm{x}_i)$ is the simulated carbon grade and $\mathrm{I}\{z_1(\mathrm{x}_i) \geq t_1\}$ is the indicator function which takes the value 1 when carbon is greater than cut-off t_1 and 0 otherwise. N is 25756.

As seen in Fig. 18, carbon mean grade uncertainty (the gap between highest and lowest mean grades) of both approaches are almost identical in all cut-offs. It seems that ilr transformation does not influence the mean grade uncertainty of carbon. Generally, mean grades of the second approach are greater than the ones of the first approach. It means that if the analyst does not perform ilr

transformation, it may lead him or her to evaluate the deposit incorrectly and increase the risk of financial loss.

As previously mentioned, ash has negative influence on coal quality, so its mean grade is defined as:

$$MG_{ash} = \frac{\sum_{i=1}^{N} z_2(x_i) \cdot I\{z_2(x_i) \leq t_1\}}{\sum_{i=1}^{N} I\{z_2(x_i) \leq t_1\}} , \qquad (11)$$

where $z_2(x_i)$ is the simulated ash grade and $I\{z_2^{(k)}(x_i) \leq t_1\}$ is the indicator function which takes the value 1 when simulated ash $z_2(x_i)$ is less than the cut-off t_1 and 0 otherwise.

As seen in Fig. 19, ash MG-CG curves of the first approach are generally higher than ~~the ones~~ of the second approach. This deviation causes overestimation of ash mean grades and, consequently, the deposit will be underrated. Both carbon and ash MG-CG curves demonstrate that geostatistical modeling of raw data lead the whole process to evaluate the PIV deposit with a lower quality than real. In the present study, the differences of MG-CG curves of compositional and noncompositional approaches are minor, but this difference may be remarkable for other types of deposits.
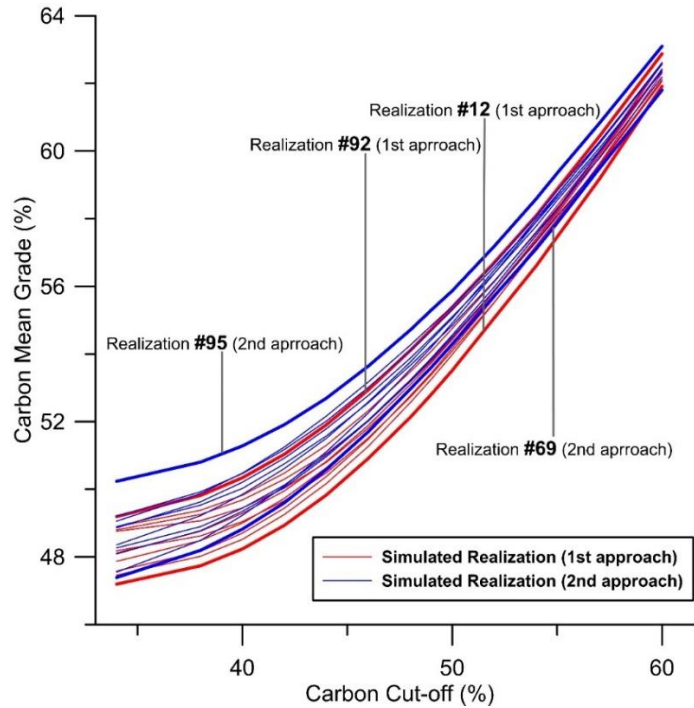

Fig. 18. MG-CG curves of multiple realizations of carbon for both approaches. Although mean grade uncertainty (the gap between highest and lowest mean grades) of both approaches are almost identical, in general, mean grades of the second approach are greater than the first approach.
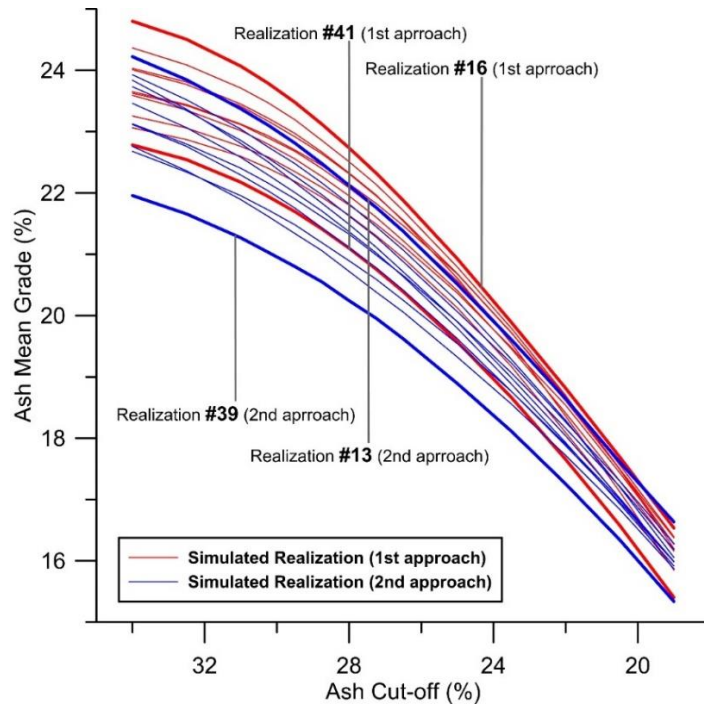
Fig. 19. MG-CG curves of multiple realizations of ash for both approaches. Although mean grade uncertainty (the gap between highest and lowest mean grades) of both approaches are almost identical, in general, mean grades of the first approach are greater than the second approach.

4. Discussion

In this section, the impact of the logratio approach on the optimality and validity of geostatistical estimation in PIV is discussed. Here we use four criteria to compare the results of compositional and noncompositional approaches:

- *Sum of values* maps,
- Aitchison's distance,
- Actual vs. estimated cross-plots,
- Correct classification rate (CCR).

Last three criteria are only available for cokriging because conditional cosimulation is required to honor input values at all sample location. It means that simulated point and measured data are equal at sample locations for both first and second approaches. Thus, it is meaningless to perform the last three criteria for cosimulation results.

4.1 Sum of values maps

These maps can be simply generated by summing up the 4 estimated/simulated parts of each cell. Figs. 20 and 21 show that the first approach fails to add up to 100 for almost all cells, which is

inconsistent. It can be concluded that the reliability of geostatistical modeling of raw values without considering their compositional nature is doubtful, both for cokriging and cosimulation. On the other hand, the second approach is free of this problem because the structure of the ilr inverse function assures returning the exact sum of 100 for all cells after ilr backtransformation.
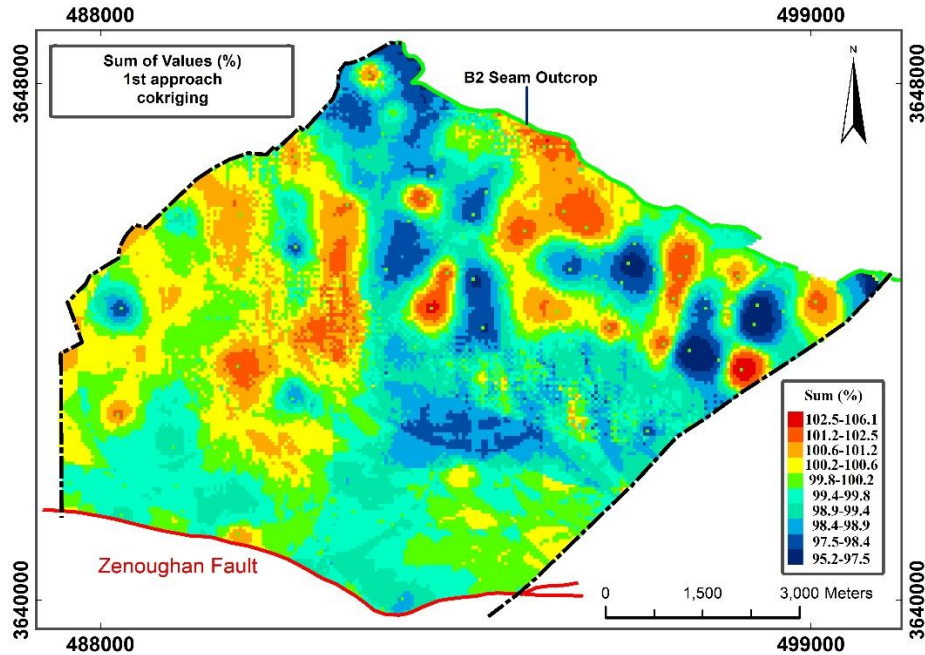

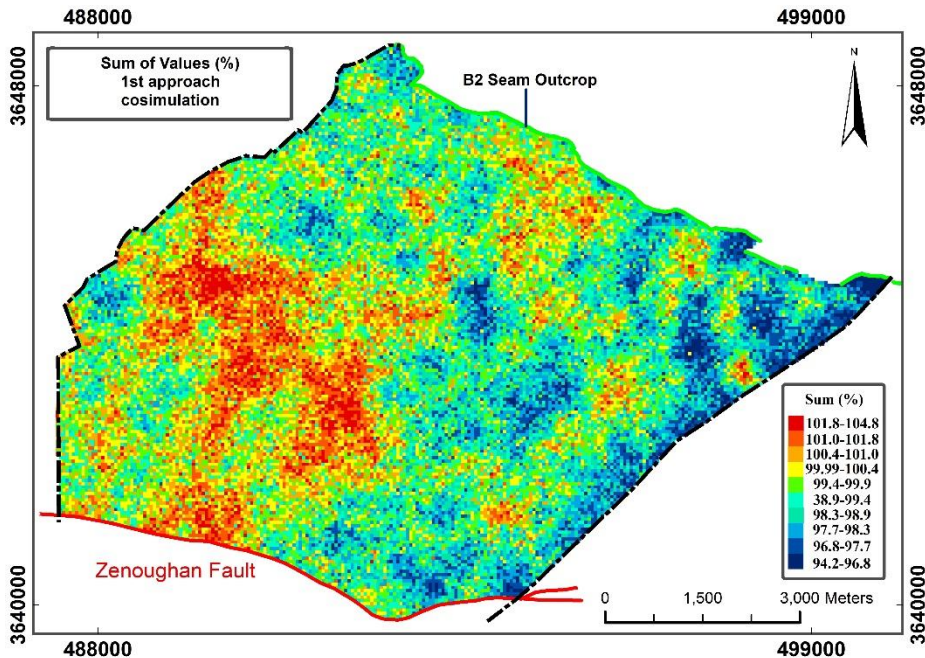**Fig. 20. Sum of values map for cokriging of 4 raw components (noncompositional approach)**

Fig. 21. Sum of values map for cosimulation of 4 raw components (noncompositional approach)

4.2 Aitchison's distance

Aitchison's distance (*A.D*) is a measure of difference between two vectors of multivariate data (i.e. two compositions). Aitchison's distance is scale invariant, permutation invariant, perturbation invariant and subcompositionally dominant (Palarea-Albaladejo et al., 2012). Thus, it is a good criterion to calculate the dissimilarity of two compositions. For two 4-part compositions V (v_1, v_2, v_3, v_4) and W (w_1, w_2, w_3, w_4), the Aitchison's distance is defined by Eq. 12.

$$A.D(V,W) = \left[\sum_{i=1}^{4} [\log(\frac{v_i}{g(V)}) - \log(\frac{w_i}{g(W)})]^2 \right]^{\frac{1}{2}}, \tag{12}$$

where v_i and w_i are actual and estimated vectors of compositions at the sampling points, $g(V)$ and $g(W)$ represents the geometric mean of actual and estimated compositions. This distance is equal to the Euclidean distance between the transformed compositions by clr or ilr transformation.

Considering the actual and estimated datasets of ash, C, M and V, *A.D* can be calculated between actual and corresponding estimated values. Computing and comparing *A.D* allows to make a decision about the realism of the approaches. When we want to compare two estimated compositions, smaller *A.D* between actual and estimated values refers to better estimation. This comparison could be visually summarized as a distance-distance cross-plot (Fig. 22) and a map of the difference between *A.D*s of the two approaches (Fig. 23).

It is clear that in almost all samples, estimated values by the second approach have a smaller *A.D* than the first approach. Although there is no evident regionalization in the map of Fig. 21, it can be found that the noncompositional approach has much larger *A.D*s in the area of the deposit with highest ash and lowest carbon contents, i.e., the part of the deposit with lowest coal quality (orange and red squares in central and northern parts of Fig. 23), while the compositional approach has provided a more accurate estimation in this very part of the deposit. Thus, using the compositional approach is critical for modeling the poor parts of the deposit to avoid further misclassification. This result agrees with the results by Karacan and Olea (2018). Based on *A.D* criterion, it is more reliable to use the logratio transformation when we want to estimate compositional data with cokriging.
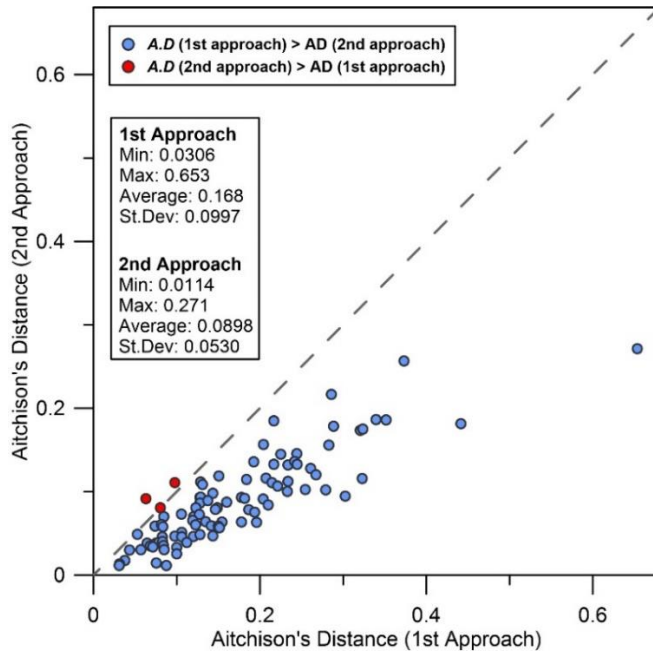
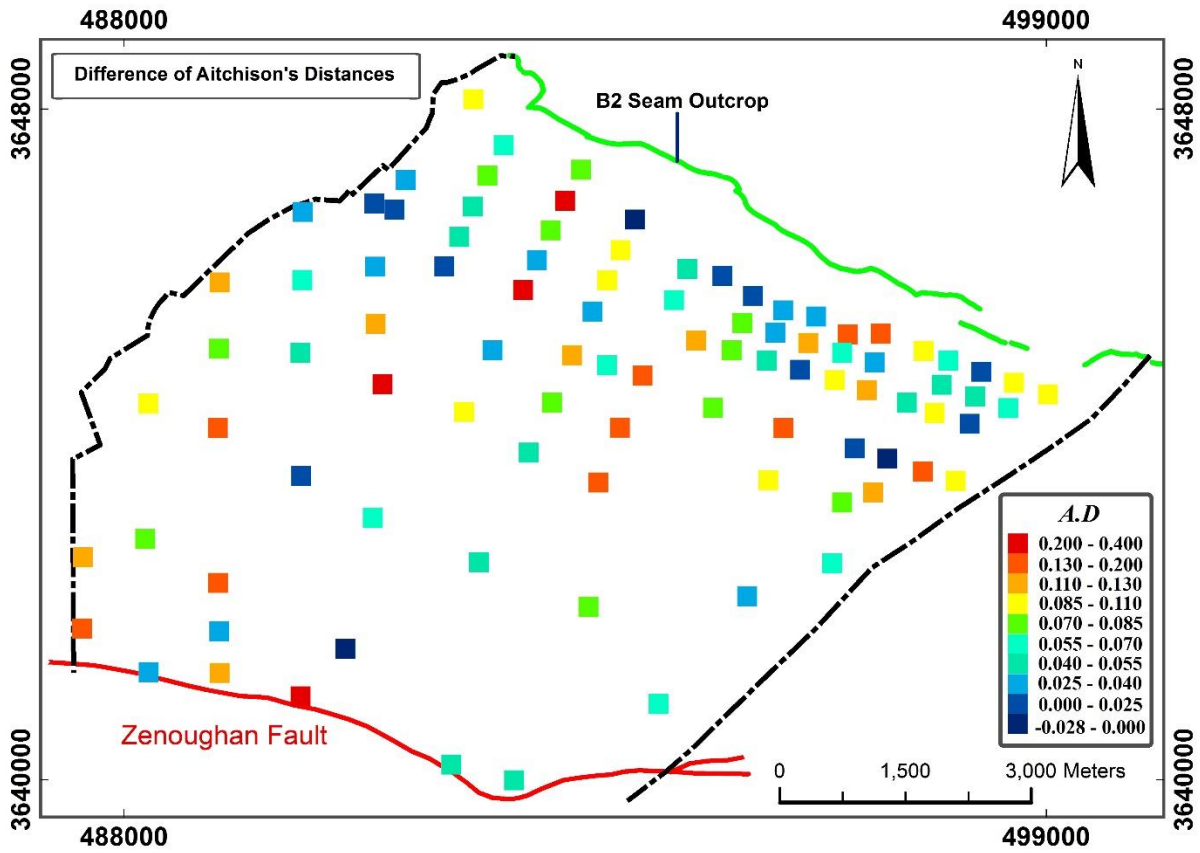**Fig. 22. Distance – distance plot of *A.D* for both approaches**

Fig. 23. Map showing the Differences between Aitchison's distances of the two approaches
($A.D$ first approach $- A.D$ second approach) **in the northern PIV.**

4.3 Actual vs. estimated cross-plots

Here we discuss about the effect of CoDa on validity of cokriging results in terms of cross-plots as those which are frequently used in geostatistical modeling projects. Comparing the best fit equations of the first and second approaches, we can understand and compare their ability to establish accurate estimations.

Fig. 24 show actual vs. estimated cross-plots for ash, C, M and V; with and without using logratio approach. It is obvious that the compositional approach provides more accurate results rather than the noncompositional approach. The superiority of the compositional approach is extremely distinctive in high and low grades, namely, third and first quantiles. This is important because biased estimation of carbon in its third quantile may lead to loss of high-quality coal. It is also generalizable for ash and its first quantile.

It is well-known that kriging based estimations fall into smoothing effect which leads to overestimation in the first and under-estimation in the third quantiles. In Fig. 24 it is shown that application of ilr balances with cokriging reduced the inherent and undesirable smoothing effect of kriging, especially in high and low grades.

In the following, we discuss about validation of ash and carbon; because basically, ash and carbon are principal variables in bituminous coal quality evaluation.
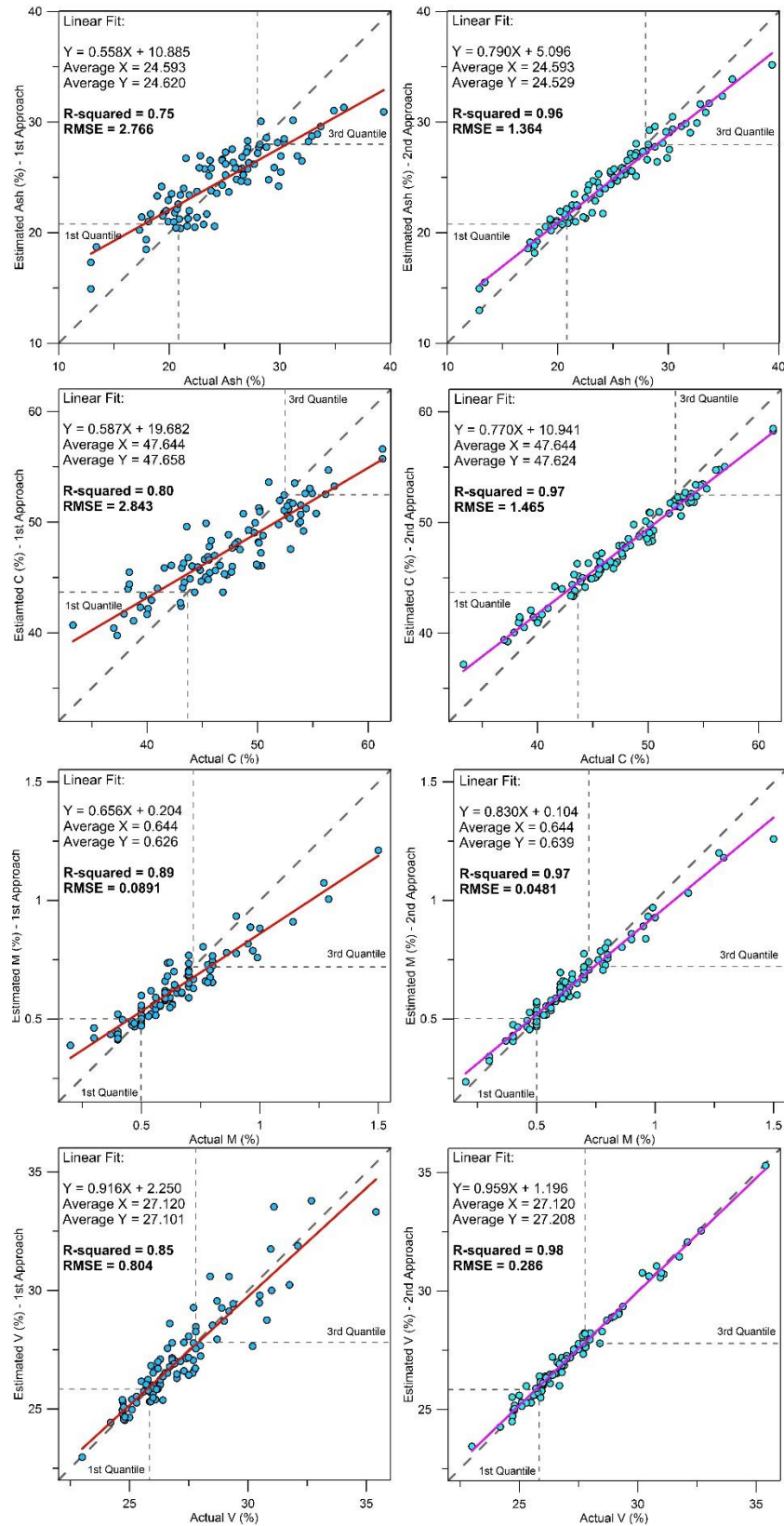
Fig. 24. Actual vs. estimated cross-plots of ash, carbon, moisture and volatile matter for first (left) and second (right) approaches.

4.4 Correct Classification Rate (CCR)

CCR is a measure of goodness of classification, frequently reported in percentage. CCR is calculated by counting the estimated samples of correct classes divided by the number of all samples. Waste-ore classes are defined based on a cut-off. CCR is defined as:

$$CCR\,(\%) = \frac{\sum_{i=1}^{N} I\{g_a(x_0) \geq c\} \cdot I\{g_e(x_0) \geq c\} + \sum_{i=1}^{N} I\{g_a(x_0) \leq c\} \cdot I\{g_e(x_0) \leq c\}}{N} \times 100, \qquad (13)$$

where N is the number of ~~samples at location x_i~~, I$\{g_a\}$ and I$\{g_e\}$ are indicator functions taking the value of 1 when the condition in the argument is fulfilled and 0 otherwise, c is a defined cut-off and ~~x_0~~ is a sampling point

CCR is a proportion and the scale of proportion is not absolute, thus, it is convenient to enhance the differences by using a logit transform:

$$logit\,(x) = \ln(\frac{x}{100 - x}), \qquad (14)$$

where x is the only measured or calculated percentage. Logit(CCR) values of first and second approaches are calculated for 11 predefined cut-offs of ash and C (Table 3, Fig. 25). Logit(CCR) values of the second approach are always greater than those obtained from the first approach, both for ash and C. It demonstrates that the second approach clearly outperforms the first approach in terms of classification concept.

Regarding section 4 and on the basis of better results obtained by the compositional approach in the cross validation stage (sections 4.1 to 4.4), it can be concluded that the joint application of CoDa and geostatistics can improve the validity of models in comparison to modelling raw data of PIV coal deposit.

Table 3. CCR (%) and logit(CCR) values of the first and second approaches

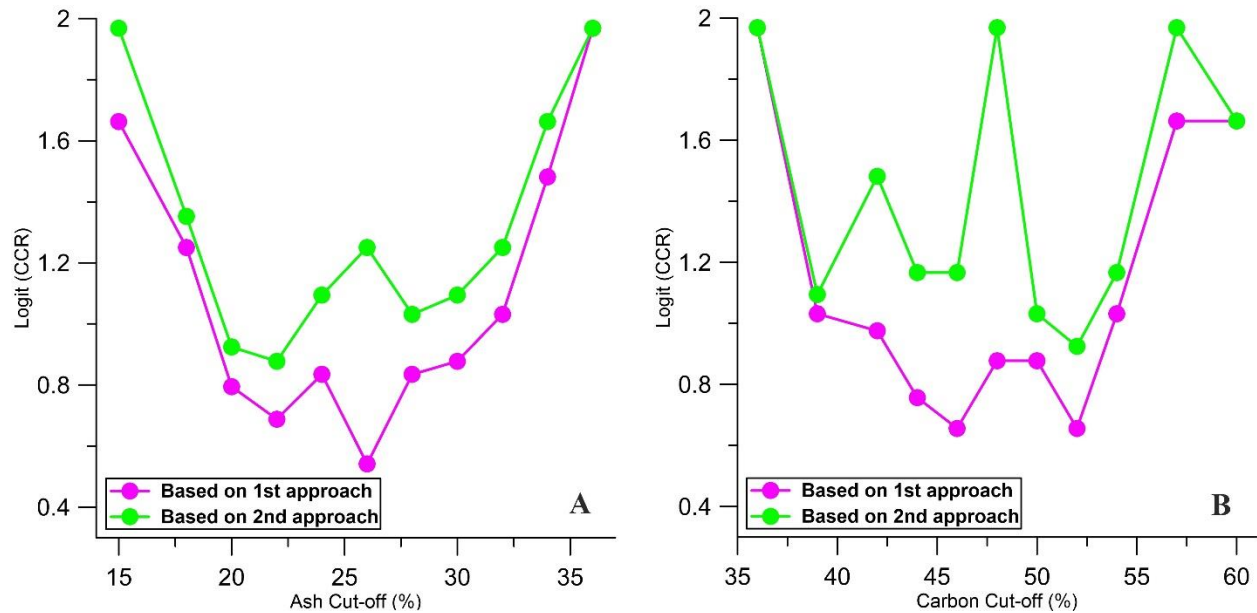| ash cut-off (%)	CCR (%)		Logit(CCR)		C cut-off (%)	CCR (%)		Logit(CCR)	
	1st app.	2nd app.	1st app.	2nd app.		1st app.	2nd app.	1st app.	2nd app.
15	97.9	98.9	1.66	1.97	36	98.9	98.9	1.97	1.97
18	94.7	95.7	1.25	1.35	39	91.5	92.6	1.03	1.09
20	86.2	89.4	0.79	0.92	42	90.4	96.8	0.98	1.48
22	83.0	88.3	0.69	0.88	44	85.1	93.6	0.76	1.17
24	87.2	92.6	0.83	1.09	46	81.9	93.6	0.66	1.17
26	77.7	94.7	0.54	1.25	48	88.3	98.9	0.88	1.97
28	87.2	91.5	0.83	1.03	50	88.3	91.5	0.88	1.03
30	88.3	92.6	0.88	1.09	52	81.9	89.4	0.66	0.92
32	91.5	94.7	1.03	1.25	54	91.5	93.6	1.03	1.17
34	96.8	97.9	1.48	1.66	57	97.9	98.9	1.66	1.97
36	98.9	98.9	1.97	1.97	60	97.9	97.9	1.66	1.66

Fig. 25. Logit(CCR) of (A) ash and (B) carbon for both approaches. In all defined cut-offs, second approach shows higher values in logit scale, which means better classification.

5. Conclusions

Coal proximate analysis data is a typical compositional data, including 4 parts: ash, carbon, moisture and volatile matter, which are strictly positive and have a constant sum of 100%. In such cases, data values are constrained and do not vary over the whole real space. Thus, traditional geostatistical methods such as kriging and stochastic simulation are not applicable to directly model the regionalized compositions (r-compositions). However, direct modeling of compositional data is popular in almost all fields of geostatistical studies. Although frequently this approach provides apparently acceptable outputs, the results are always exposed to incoherence and non-optimality.

In this paper, Parvadeh IV coal deposit was selected to practically assess the incoherence and non-optimality of the noncompositional approach and to investigate the negative effects of the direct modeling of r-compositions in the mining projects. Results are valid for all similar cases. The goal of this paper is to analyze the impact of the compositional nature of data on reserve evaluation of a coal deposit and to propose a reliable and accurate approach to deal with regionalized compositions. For this purpose, we studied the differences between the results obtained by the compositional and noncompositional approaches. Sum of values maps, Aitchison's distance, actual vs. estimated cross-plots, correct classification rate (CCR), tonnage – grade and mean grade – cut-off grade diagrams were computed and compared for both approaches as follows:

- Performing the noncompositional approach is a little bit easier and faster and needs less mathematical understanding. But this approach is exposed to incoherence, e.g., sum of

values does not add up to 100%. This unacceptable inconsistency represents the intrinsic bias of direct geostatistical modeling of regionalized compositions.

- A compositional approach provides more accurate and valid estimations than a noncompositional approach. This is remarkable when we compare Aitchison's distance, cross-plots and CCR of the approaches.
- Comparison of ash T_{tr}, T and T_+ curves of the two approaches implies that the noncompositional approach has overestimated the ash content. This makes the analyst underrate the deposit in terms of coal quality; which is an evident misinterpretation.
- Detailed comparison of tonnage – grade curves showed that the noncompositional approach causes a coal loss of nearly 5 and 3 million tons, based on certain cut-offs of ash and carbon respectively.
- Study of mean grade – cut-off grade curves of both approaches confirmed that the noncompositional approach has underestimated contained carbon and overestimated contained ash. Again, this makes the deposit to be underrated, which is another kind of dangerous misinterpretation.

The whole procedure of Parvadeh IV reserve evaluation demonstrated that results of the noncompositional approach is biased and this leads the analyst to underrate the deposit. In summary, what we want to warn is that neglecting the compositional nature of data can result in non-optimal modeling, unreliable evaluation and considerable economical loss. Thus, it is strongly recommended to consider the compositional nature of data in reserve evaluations.

References

American Association for Testing of Material (ASTM), 2013. Standard Practice for Proximate Analysis of Coal and Coke. ASTM Standard D3172–13 (2 pp.)

Buccianti, A. Egozcue, J.J., Pawlowsky-Glahn, V., 2014. Variation diagrams to statistically model the behavior of geochemical variables: Theory and pplications. Journal of Hydrology. 519, 988–998. Available at: http://dx.doi.org/10.1016/j.apgeochem.2013.02.005.

Drew, L., Grunsky, E., Schuenemeyer, J., 2008. Investigation of the Structure of Geological Process through Multivariate Statistical Analysis—The Creation of a Coal. Mathematical Geosciences. 40(7), 789–811.Available at: http://dx.doi.org/10.1007/s11004-008-9176-2.

Egozcue, J.J., Pawlowsky-Glahn, V., Mateu-Figueras, G., Barceló-Vidal, C., 2003. Isometric Logratio Transformations for Compositional Data Analysis. Mathematical Geology. 35(3), 279–300.

Engle, M.A., Blondes, M.S., 2014. Linking compositional data analysis with thermodynamic geochemical modeling: Oilfield brines from the Permian Basin, USA. Journal of Geochemical Exploration. 141, 61–70. Available at: http://dx.doi.org/10.1016/j.gexplo.2014.02.025.

Engle, M.A., Rowan, E.L., 2013. Interpretation of Na-Cl-Br Systematics in Sedimentary Basin Brines: Comparison of Concentration, Element Ratio, and Isometric Log-ratio Approaches. Mathematical Geosciences. 45(1), 87–101.

Filzmoser, P., Hron, K., 2009. Correlation Analysis for Compositional Data. Mathematical Geosciences. 41(8), 905–919. Available at: http://link.springer.com/article/10.1007/s11004-008-9196-y\nhttp://link.springer.com.ezproxy.cul.columbia.edu/article/10.1007/s11004-008-9196-y.

Filzmoser, P., Hron, K., Reimann, C., 2009. Principal components analysis for compositional data with outliers. Environmetrics. 20(6), 621–632.

Filzmoser, P., Hron, K., Reimann, C., 2010. The bivariate statistical analysis of environmental (compositional) data. Science of the Total Environment. 408(19), 4230–4238. Available at: http://dx.doi.org/10.1016/j.scitotenv.2010.05.011.

Grunsky, E.C., Kjarsgaard, A., Egozcue, J.J., Pawlowsky-Glahn, V., Thio, S., 2008. Studies in Stoichiometry with Compositional Data. In International Geological Congress. Oslo. 1–7.

Grunsky, E.C., Mueller, U.a., Corrigan, D., 2014. A study of the lake sediment geochemistry of the Melville Peninsula using multivariate methods: Applications for predictive geological mapping. Journal of Geochemical Exploration. 141, 15–41. Available at: http://dx.doi.org/10.1016/j.gexplo.2013.07.013.

Jelsema, C., Paul, R., 2013. Spatial mixed effects model for compositional data with applications to coal geology. International Journal of Coal Geology. 114, 33–43. Available at: http://dx.doi.org/10.1016/j.coal.2013.04.004.

Karacan, C.Ö., Olea, R.A., 2018. Mapping of compositional properties of coal using isometric logratio transformation and sequential Gaussian simulation—A comparative study for spatial ultimate analysis data. Journal of Geochemical Exploration, vol. 118, p. 36−49.

Lark, R.M., Dove, D., Green, S.L., Richardson, a.E., Stewart, H., Stevenson, A., 2012. Spatial prediction of seabed sediment texture classes by cokriging from a legacy database of point observations. Sedimentary Geology. 281, 35–49. Available at: http://dx.doi.org/10.1016/j.sedgeo.2012.07.009.

Levitan, D.M., Zipper, Carl E., Donovan, P., Schreiber, M.E., Seal, R.R., Engle, M.A., Chermak, J.A., Bodnar, R.J., Johnson, D.K., Aylor, J.G., 2015. Statistical analysis of soil geochemical data to identify pathfinders associated with mineral deposits: An example from the Coles Hill uranium deposit, Virginia, USA. Journal of Geochemical Exploration. 154, 238–251. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0375674214004038.

Makvandi, S., Ghasemzadeh-Barvarz, M., Beaudoin, G., Grunsky, E.C., Beth McClenaghan, M., Duchesne, C., 2016. Principal component analysis of magnetite composition from volcanogenic massive sulfide deposits: Case studies from the Izok Lake (Nunavut, Canada) and Halfmile Lake (New Brunswick, Canada) deposits. Ore Geology Reviews. 72, 60–85. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169136815001778.

Martín-Fernández, J.A., Barceló-Vidal, C., Pawlowsky-Glahn, V., Kovács, L.Ó., Kovács, G., 2005. Subcompositional patterns in Cenozoic volcanic rocks of Hungary. Mathematical Geology. 37(7), 729–752.

Moeini, H., Torab, F.M., 2017. Comparing compositional multivariate outliers with autoencoder networks in anomaly detection at Hamich exploration area, east of Iran. Journal of Geochemical Exploration. 180, 15–23.

Molayemat, H., Torab, F.M., 2017. Evaluation of coalbed methane potential in Parvadeh IV coal deposit in central Iran using a combination of MARS modeling and Kriging. Journal of Mining and Environment. Vol. 8, Issue 2, 307–319. Available at: http://dx.doi.org/10.22044/jme.2017.874

Montreuil, J.-F., Corriveau, L., Grunsky, E.C., 2013. Compositional data analysis of hydrothermal alteration in IOCG systems, Great Bear magmatic zone, Canada: to each alteration type its own geochemical signature. Geochemistry: Exploration, Environment, Analysis. 13(4), 229–247. Available at: http://geea.lyellcollection.org/content/13/4/229.short.

Olea, R.A., 2009. A Practical primer on geostatisitcs. U.S. geological Survey, Open-File Report 2009-1103, 346 p. http://pubs.usgs.gov/of/2009/1103/ofr2009-1103-rev-jan2010.pd

Olea, R.A., Luppens, J.A., 2015. Mapping of coal quality using stochastic simulation and isometric logratio transformation with an application to a Texas lignite. International Journal of Coal Geology, 152(PA), 80–93. Available at: http://www.sciencedirect.com/science/article/pii/S0022169414006386.

Olea, R.A., Luppens, J.A., Egozcue, J.J., Pawlowsky-Glahn, V., 2016. Calorific value and compositional ultimate analysis with a case study of a Texas lignite. International Journal of Coal Geology. 162, 27–33. Available at: http://dx.doi.org/10.1016/j.coal.2016.05.005.

Olea, R.A., Raju, N.J., Egozcue, J.J., Pawlowsky-Glahn, V., Singh, S., 2017. Advancements in hydrochemistry mapping: methods and application to groundwater arsenic and iron concentrations in Varanasi, Uttar Pradesh, India. Stochastic Environmental Research and Risk Assessment. Feburary, 1–19. doi:10.1007/s00477-017-1390-3.

Owen, D.D.R., Pawlowsky-Glahn, V., Egozcue, J.J., Buccianti, A., Bradd, J.M., 2016. Compositional data analysis as a robust tool to delineate hydrochemical facies within and between gas-bearing aquifers. Water Resources Research. 52(8), 5771–5793.

Palarea-Albaladejo, J., Martín-Fernández, J.A., Buccianti, A., 2014. Compositional methods for estimating elemental concentrations below the limit of detection in practice using R. Journal of Geochemical Exploration. 141, 71–77. Available at: http://www.sciencedirect.com/science/article/pii/S0375674213001866.

Palarea-Albaladejo, J., Martín-Fernández, J.A., Soto, J.A., 2012. Dealing with Distances and Transformations for Fuzzy C-Means Clustering of Compositional Data. Journal of Classification. 29(2), 144–169.

Pardo-Igúzquiza, E., Chica-Olmo, M., Luque-Espinar, J.A., Rodríguez-Galiano, V., 2015. Compositional cokriging for mapping the probability risk of groundwater contamination by nitrates. Science of the Total Environment. 532, 162 – 175.

Pawlowsky-Glahn, V., Egozcue, J.J., Tolosana-Delgado, R., 2015a. Modeling and Analysis of Compositional Data. Statistics in practice. John Wiley & Sons, Cichester, UK, 272 pp.

Pawlowsky-Glahn, V., Egozcue, J.J., Lovell, D., 2015b. Tools for compositional data with a total. Statistical Modelling. 15(2), 175–190. Available at: http://www.scopus.com/inward/record.url?eid=2-s2.0-84926148988&partnerID=40&md5=7384168d65cc077624e60a5215cc6f8a.

Pawlowsky-Glahn, V. Egozcue, J.J., Olea, R.A., Pardo-Igúzquiza, E., 2015c. Cokriging of compositional balances including a dimension reduction and retrieval of original units. The Journal of The Southern African Institute of Mining and Metallurgy. 115.

Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional data and their analysis: an introduction. Geological Society, London, Special Publications. 264(1), 1–10. Available at: http://sp.lyellcollection.org/lookup/doi/10.1144/GSL.SP.2006.264.01.01.

Pawlowsky-Glahn, V., Egozcue, J.J., 2016. Spatial analysis of compositional data: A historical review. Journal of Geochemical Exploration. 164, 28–32.

Pawlowsky-Glahn, V., 1984. On spurious spatial covariance between variables of constant sum. Science de la Terre, S´er Informatique. 21:107–13.

Pawlowsky-Glahn, V., Olea, R.A., Davis, J.C., 1995. Estiamtion of regionalized compositions: a comparison of three methods. Mathematical Geology. 27 (1).

Pawlowsky-Glahn, V., Olea, R.A., 2004. Geostatistical Analysis of compositional data. Number 7 in Studies in Mathematical Geology. Oxoford University press.

Reimann, C. Filzmoser, P., Garrett, R.G., Dutter, R., 2008. Statistical Data Analysis Explained. Statistics, p.359. Available at: http://doi.wiley.com/10.1002/9780470987605.

Reimann, C. Filzmoser, P., Fabian, K., Hron, K., Birke, M., Demetriades, A., Dinelli, E., Ladenberger, A., Albanese, S., Andersson, M., Arnoldussen, A., Baritz, R., Batista, M.J., Bel-lan, A., Cicchella, D., De Vivo, B., De Vos, W., Duris, M., Dusza-Dobek, A., Eggen, O. a., Eklund, M., Ernstsen, V., Finne, T. E., Flight, D., Forrester, S., Fuchs, M., Fugedi, U., Gilucis, A., Gosar, M., Gregorauskiene, V., Gulan, A., Halamic, J., Haslinger, E., Hayoz, P., Hobiger, G., Hoffmann, R., Hoogewerff, J., Hrvatovic, H., Husnjak, S., Janik, L., Johnson, C. C., Jordan, G., Kirby, J., Kivisilla, J., Klos, V., Krone, F., Kwecko, P., Kuti, L., Lima, A., Locutura, J., Lucivjansky, P., Mackovych, D., Malyuk, B. I., Maquil, R., McLaughlin, M. J., Meuli, R. G., Miosic, N., Mol, G., Négrel, P., O'Connor, P., Oorts, K., Ottesen, R. T., Pasieczna, A., Petersell, V., Pfleiderer, S., Ponavic, M., Prazeres, C., Rauch, U., Salpeteur, Schedl, A., Scheib, A., Schoeters, I., Sefcik, P., Sellersjö, E., Skopljak, F., Slaninka, I., Šorša, A., Srvkota, R., Stafilov, T., Tarvainen, T., Trendavilov, V., Valera, P., Verougstraete, V., Vidojevic, D., Zissimos, a. M., Zomeni, Z., 2012. The concept of compositional data analysis in practice - Total major element concentrations in agricultural and grazing land soils of Europe. Science of the Total Environment. 426, 196–210. Available at: http://dx.doi.org/10.1016/j.scitotenv.2012.02.032.

Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics With SGeMS, A User's Guide. Cambridge University Press, Cambridge, United Kingdom (264 pp.).

Scealy, J.L., de Caritat, P., Grunsky, E. C., Tsagris, M. T., Welsh, a. H., 2015. Robust Principal Component Analysis for Power Transformed Compositional Data. Journal of the American Statistical Association.110(509), 136–148. Available at: http://dx.doi.org/10.1080/01621459.2014.990563.

Sun, X.-L., Wu, Y.J., Wang, H.L., Zhao, Y.G., Zhang, G.L., 2014. Mapping Soil Particle Size Fractions Using Compositional Kriging, Cokriging and Additive Log-ratio Cokriging in Two Case Studies. Mathematical Geosciences. 46(4), 429–443. Available at: http://link.springer.com/10.1007/s11004-013-9512-z.

Tolosana-Delgado, R., 2012. Uses and misuses of compositional data in sedimentology. Sedimentary Geology. 280, 60–79. Available at: http://dx.doi.org/10.1016/j.sedgeo.2012.05.005.

Tolosana-Delgado, R., Van Den Boogaart, K.G., 2013. Joint Consistent Mapping of High-Dimensional Geochemical Surveys. Mathematical Geosciences. 45(8), 983–1004.

Tolosana-Delgado, R., Van Den Boogaart, K.G., 2014. Towards compositional geochemical potential mapping. Journal of Geochemical Exploration. 141, 42–51.

Zhang, S., Shen, C.Y., Chen, X.Y., Ye, H.C., Huang, Y.F., Lai, S., 2013. Spatial Interpolation of Soil Texture Using Compositional Kriging and Regression Kriging with Consideration of the Characteristics of Compositional Data and Environment Variables. Journal of Integrative Agriculture. 12(9), 1673–1683.

Zuo, R., Xia, Q., Wang, H., 2013. Compositional data analysis in the study of integrated geochemical anomalies associated with mineralization. Applied Geochemistry. 28, 202–211. Available at: http://dx.doi.org/10.1016/j.apgeochem.2012.10.031.