Studia Quaternaria, vol. 39, no. 2 (2022): 83–93 DOI: 10.24425/sq.2022.140885

APPLYING GRAIN-SIZE AND COMPOSITIONAL DATA ANALYSIS FOR INTERPRETATION OF THE QUATERNARY OXBOW LAKE SEDIMENTATION PROCESSES: EASTERN GREAT HUNGARIAN PLAIN

Abdelrhim Eltijani*, Dávid Molnár, László Makó, János Geiger, Pál Sümegi

University of Szeged, Department of Geology, 2-6 Egyetem u., H-6722, Szeged, Hungary;
e-mails: Abdelrhim Eltijani abdelelt@geo.u-szeged.hu, ORCID iD: 0000-0002-6952-158X;
Dávid Molnár molnard@geo.u-szeged.hu; László Makó makol@geo.u-szeged.hu;
János Geiger matska@geo.u-szeged.hu; Pál Sümegi sumegi@geo.u-szeged.hu

** corresponding author*

Abstract:
Grain size distribution is one of the paleoenvironmental proxies that provide insight statistical distribution of size fractions within the sediments. Multivariate statistics have been used to investigate the depositional process from the grain size distribution. Still, the direct application of the standard multivariate methods is not straightforward and can yield misleading interpretations due to the compositional nature of the raw grain size data. This paper is a methodological framework for grain size data characterization through the centered log ratio transformation and euclidean data, coupled with principal component analysis, cluster analysis, and linear discriminant analysis to examine Quaternary sediments from Tövises bed in the southeast Great Hungarian Plain. These approaches provide statistically significant and sedimentologically interpretable results for both datasets. However, the details by which they supplemented the conceptual model were significantly different, and this discrepancy resulted in a different temporal model of the depositional history.

Key words: Grain size distribution, log ratio transformation, Multivariate statistics, Tövises bed, Great Hungarian Plain

Manuscript received 25 November 2021, accepted 19 January 2022

INTRODUCTION

Grain Size Distribution (GSD) is a paleoenvironmental proxy that provides information on depositional environments and processes (He *et al*., 2015; Zhang *et al*., 2018). GSD firstly evolves during transport and deposition (Reading, 1996; McLaren *et al*., 2007). Numerous works have interpreted the depositional conditions based on GSDs by applying univariate statistical parameters. For instance, the median, mean, sorting, and skewness of the distribution (e.g., Folk and Ward, 1957; Visher, 1969; Blott and Pye, 2001; Fournier *et al*., 2014). The CM patterns where (C) is one percentile and (M) is the median grain size can help analyze the ancient and recent depositional processes (Passega, 1957, 1964, 1977). One of the methods to determine the grain size fractions susceptible to environmental changes is classifying the GSDs using the standard deviation of the distribution (Boulay *et al*., 2003). These methods, however, depend on median diameter instead of the whole distribution, revealing relative and restricted information on the distribution (Zhang *et al*., 2018).

The polymodal GSDs indicate the presence of individual subpopulations (Folk and Ward, 1957; Ashley, 1978; Flemming, 1988). Moreover, the polymodality also suggests that the sediments were not well mixed in the suspension. Multivariate approaches to distinguish the subpopulations of GSDs include, among others, cluster analysis (CA) and principal component analysis (PCA) (Sarnthein *et al*., 1981; East, 1985, 1987). However, the application of cluster analysis on GSDs focused on provenance studies and stratigraphic analysis. Furthermore, these studies used a few parameters, e.g., mean and standard deviation. These parameters provide limited information on the depositional conditions (Donato *et al*., 2009; Zhang *et al.,* 2018).

Although statistical analysis is a useful tool for GSDs description, and PCA represents a crucial dimensionality reduction method (Palazón and Navas, 2017; Katra and Yizhaq, 2017). The direct application of such statistical techniques for analyzing GSDs is challenging because the grain size data is a typical compositional set (e.g., Aitchison, 1982; Flood *et al*., 2015). The sum of weight percentages of the size fractions is 100%. Consequently, they do not form

an independent system (Aitchison, 1986; Flood *et al.*, 2015). The definition of compositional data has gradually evolved from vectors of positive components adding to a given constant (e.g., Aitchison, 1982) to a new general definition based on equivalence classes (Egozcue *et al.*, 2018). A composition is a set of multivariate vectors that vary by a scalar factor and have nonnegative components. Accordingly, the composition can be regarded as a vector of proportions with nonnegative components constrained to a K constant.

One of the crucial questions in statistical analysis is euclidating the compositional constraint. A simple approach ignores compositional constraints (i.e., Euclidean data analysis approach) and treats the data as Euclidean (Tsagris *et al.*, 2016). On this topic, there is another school following the work of Aitchison (1982, 1983, 1992). The followers of this school have suggested several transformations using the logarithms of ratios, or log ratios, to get solutions for the compositional constraints (e.g., Aitchison, 1986; Aitchinson *et al.*, 2002; Pawlowsky-Glahn and Egozcue, 2001; Egozcue *et al.*, 2018).

This paper aims to study the utility and geological interpretability of the Euclidean and centered log ratio (clr) transformation approaches coupled with cluster analysis, principal component analysis, and Linear Discriminant Analysis through investigating the GSD of oxbow lake sediments of Tövises bed from eastern Great Hungarian Plain.

MATERIAL AND METHODS

Tövises bed is located in the Pocsaj "gate" geological and geomorphological system and Érmellék region in the eastern Great Hungarian Plain (Fig. 1). The topmost Holocene sequence of this region is characterized by loess sequences overlaid by alluvial fan deposits (Szöőr *et al.*, 1991; Sümegi and Vissi, 1991; Sümegi, 1993). Szöőr *et al.* (1991) suggested an age range of (40,000–45,000) BP.

Core samples

The core from the Tövises bed core exhibits bimodal and polymodal GSDs. The entire sequence is characterized by periodic intercalation of the sediments, albeit the core is composed of fine materials (clay to very fine sand) (Fig. 2). The upper part of the core contains coarse and medium silts intercalated with fine silt at the middle part. The second section contains medium silt intercalated with coarse silt at the top and fine silt at the bottom. The middle part comprises fine and very fine silty medium silt followed by the alternating medium and fine silts at the bottom. The lower part consists of alternating coarse, medium, and fine silts. The very fine sand and the medium silty coarse silt fractions change parallelly in the vertical compositional diagram. The very fine sand fraction forms five peaks in the sequence (Fig. 2).

Grain Size Analysis

Undisturbed 346 cm core sediments were stored at a constant temperature of 4ºC and sectioned into 346 discrete 1 cm subsamples. Grain size analysis was completed at 1 cm intervals, and the measurements procedure followed Konert and Vandenberghe (1997). First, the samples were

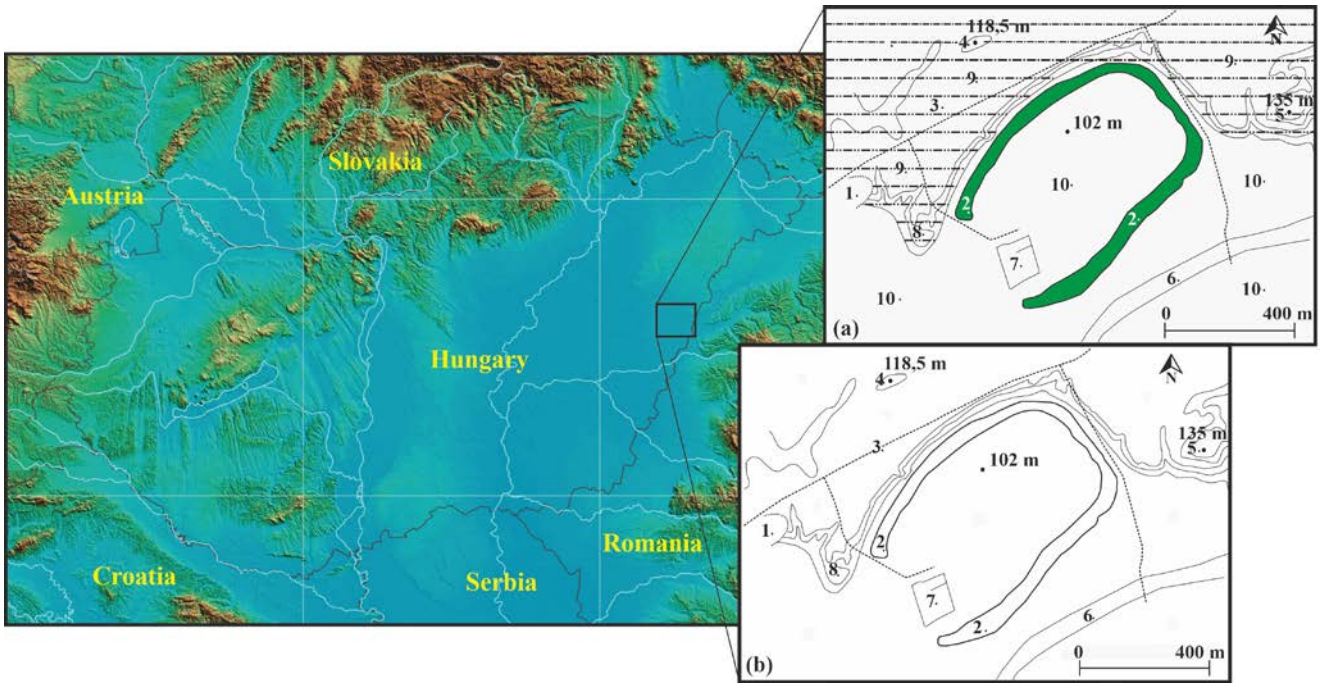

Fig. 1. The location, geologic, and topographic features of the Tövises paleochannel and vicinity. (a): 1. Sandpit, 2. Peat, 3. Loess covered Pleistocene alluvial fan, 4. Ebéd-hill (Late Copper Age kurgan), 5. Laponya-halom (kurgan), 6. Canalized bed of Ér creeck, 7. Leányvár, Late Neolithic and Middle Bronze Age tell, 8. Loess, 9. Alluvia sediments); (b): 1. Sandpit, 2. Tövises bed (paleochannel), 3. Dirty roads, 4. Ebéd Hill (Late Copper Age kurgan) 5. Laponya Hill (Late Copper Age kurgan), 6. Canalized recent Ér creeck, 8. Leányvár, Late Neolithic and Middle Bronze Age).

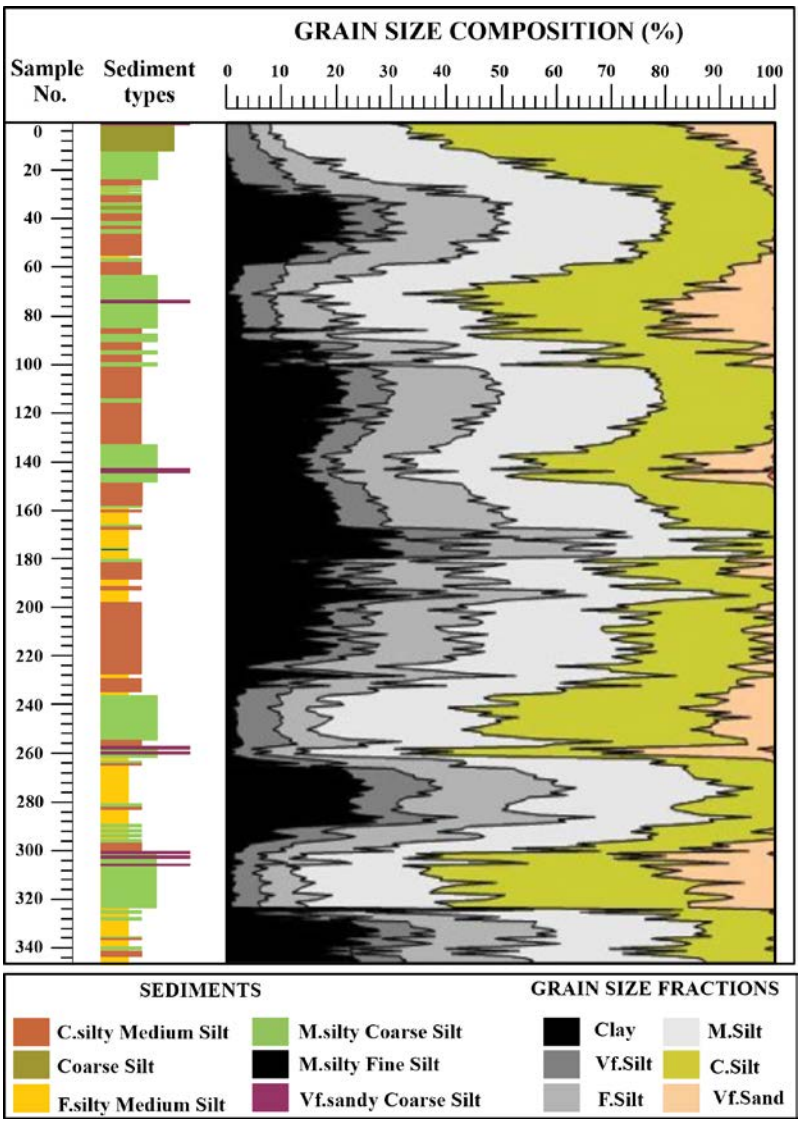

Fig. 2. Graphical depiction of the core description, along with the compositional chart of the cumulative percentages of the grain size fractions.

dried at 55ºC. Then 30 ml $Na_2P_6O_{18}$ solution was added to 0.6 g of the sample to disperse the particles. The grain size analysis was carried out in the Department of Geology and Paleontology at the University of Szeged, Hungary, using the Easysizer20 laser particle sizer instrument of OMEC company; with a measuring range: 0.1 to 500 µm and a repeatability error of less than 3%. The device uses 54 built-in detectors based on the Mie scattering. After measurements, the GSDs were decomposed into grain size fractions following the Udden Wentworth scale (Udden, 1914; Wentworth, 1922). The (D5), (D50), and (D95) parameters were determined to emphasize the finest, average, and coarsest sizes. The used nomenclature is: fine silty medium silt (= fine silt "less dominant" + medium silt); medium silty fine silt (= medium silt "less dominant" + fine silt); medium silty coarse silt (= medium silt "less dominant" + coarse silt); coarse silty medium silt (= coarse silt 'less dominant + medium silt); very fine sandy coarse silt (= very fine sand "less dominant" + coarse silt).

Compositional Data and Log ratio Transformation

Compositions describe parts of a whole that contain relative information. Aitchison (1986) introduced compositional data analysis for variables in closed number systems. The technique aims to avoid misleading interpretations based on spurious correlations. If there are p variables in the closed data, the log ratio transformation can be operational in $(p - 1)$ dimensional space, allowing unconstrained multivariate analysis (Aitchison, 1986; Pawlowsky-Glahn and Buccianti, 2011). For a dataset consisting of J compositional parts, various log ratio transformations have been suggested, including the additive log ration (alr) transformation, which has been used since the early work of Aitchison (Greenacre *et al.*, 2019). The centered log ratio (Aitchison, 1986) is the log ratio between a part (x) and the geometric mean of the entire set ($g_D(x)$):

$$y = clr(x) = ln \frac{x}{g_D(x)} \qquad (1)$$

Regarding the applicability of the standard multivariate methods in the clr-transformed compositional data, Buccianti *et al*. (2006) were aware that the covariance matrix for clr-transformed data is singular. However, Egozcue and Pawlowsky-Glahn (2016) showed that the clr transformation does not require an appropriate statistical method to evaluate and interpret the data. Following this view, this study applies standard cluster analysis and principal component analysis for the clr-transformed data.

Applied multivariate statistical methods

Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique that summarises information from a large dataset into small variables that retain the most information from large datasets. Detailed descriptions of PCA can be found in several textbooks (e.g., Agterberg, 1974; Davis, 2002). PCA allows consideration of the relationships (illustrated by the correlation matrix) among variables. The process generates new variables called principal components (PCs), and each PC describes a percentage of the total variance in the data. This percentage is interpreted as the portion explained by the process represented by the PC. The first PC is highly related to the original variables than the second, and the second PC is more related to them than the third component. To determine the number of PCs to be considered, we used the Kaiser criterion (i.e., retention of PCs whose eigenvalue is greater than 1), and scree plot criteria.

The crucial point of PCA is that the PCs can be linked with geological processes, showing groups of variables that might not be observed by other means. The interpretation of the PCA results aims to reveal processes (in the present case, the transportation processes) that cause the correlations between the PCs and the original variables (Davis, 2002; Szilágyi and Geiger, 2012).

Cluster Analysis

Cluster analysis aims to identify group structure amongst the cases. The fundamental criteria applied in the partitional clustering are homogeneity and separation. Homogeneity means that the two arbitrary objects that belong to a cluster are sufficiently similar, while separation of clusters means; that the two arbitrary objects that belong to different groups are sufficiently different. Gordon (1999) gives a comprehensive general reference. The clustering algorithms are categorized as hierarchical (agglomerative or divisive) and nonhierarchical (optimal partitioning) methods. In the present work, the hierarchical agglomerative algorithm was applied, in which the objective function was defined by Ward's minimum variance method (Ward, 1963). In this method, the sum of squared distances between objects and the cluster's center, to which the objects belong, is minimized. The similarity coefficient measuring the closeness between any two input sample points was the squared euclidean distance. This study applies hierarchical clustering because the dendrogram could depict the hierarchy of the fluvial system sub environment. The validation of the clustering results poses some difficulties as the clustering itself (Pfitzer *et al*., 2009). There are 'internal' and 'external' approaches to the evaluations, but they have a lot of uncertainty (Feldman and Sanger, 2007). That is why a geological criterion (the Passage's CM diagram) is applied to check whether the result can be interpreted from a sedimentological point of view.

Linear Discriminant Analysis (LDA)

The discriminant analysis aims to help distinguish between two or more groups of data based on observed quantitative variables. The LDA model was developed in 1936 by Fisher (1936) for categorizing objects from a set of independent variables in one or more sets of mutually exclusive groups. This model is robust, easy to use, and has high predictive accuracy. There are two objectives of discriminant analysis: One is to separate the samples into groups as well as possible; The second is to classify new observations as belonging to one group or another by using the classification functions (Mishra and Datta-Gupta, 2018).

The larger the standardized coefficient indicates a more significant contribution of the corresponding variable to the discrimination between groups. Therefore, the variable with the highest (regression) coefficient contributes most to predicting group membership. The analysis calculates one discriminant function for each group, and these functions are independent by construction, so the discrimination between groups is not overlapping. The classification table shows the result of assigning observed and new cases to a group using derived classification rules.

The CM diagram

The CM diagram (Passega, 1957, 1964) is used to establish the relationships between the sediment textures and processes of deposition. Passega (1957) defined M and C as the median and the one percentile of the cumulative GSDs, respectively. These values can readily be obtained as the grain diameters (in mm or micron) belonging to the 50 and 95 percentiles of the cumulative distribution functions. Sometimes it is hard to determine the diameter belonging to one percentile based on the laboratory analysis; therefore, the D95 percentile is used instead.

The workflow

The workflow involves the calculation of the D50, D95, and D5 percentiles from the cumulative grain size distributions, then the decomposition of GSD into clay, very fine silt, fine silt, medium silt, coarse silt, and very fine sand fractions. These fractions are then transformed by applying the additive log ratio transformation (clr) in the open source CoDaPack, version 2.02.21. (Aitchison, 1986; Egozcue and Pawlowsky-Glahn, 2005). Subsequently, two

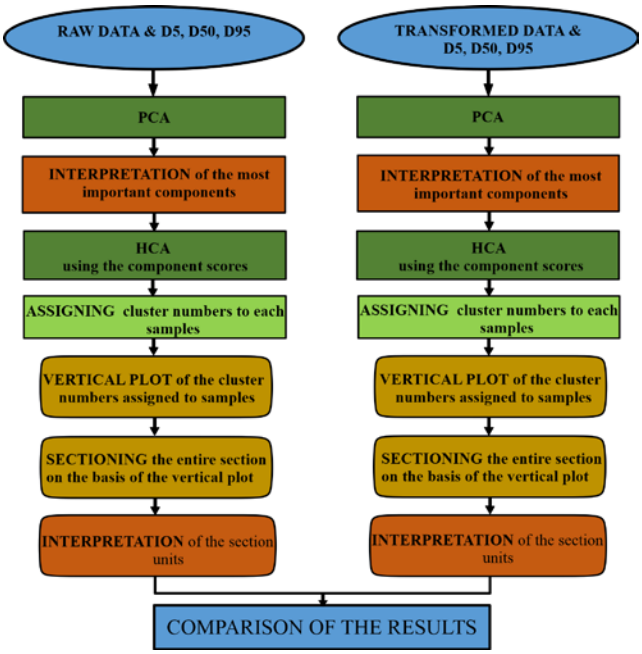

Fig. 3. The workflow of the analyses.

datasets were compiled for further use. The first dataset contained the frequency percentages of the different grain size fractions and the diameters belonging to the D5, D50, and D95 cumulative percentages of the grain size distributions.

The second contained the clr-transformed percentages of the grain size fractions and the D5, D50, and D95 diameters (Fig. 3). The PCA was performed on both datasets. The transporting processes were interpreted using the high loadings of the most important components. The component scores belonging to each sample were used as variables in the hierarchical cluster analysis (HCA) applying Ward's method (Ward, 1963). Since the clustering method relied on the component scores, the resulting clusters are supposed to correspond to the different transport processes of the fluvial system. This thought was checked by depicting the points of the clusters in the CM diagram (Fig. 5). Parallelly, the statistical reality of the clusters was tested by the discriminant analysis.

RESULTS AND INTERPRETATION

Results of Principal Component Analysis

The PCA for clr-transformed data has resulted in two PCs. The first component described 60.038% of the total variance, while the second component accounted for 23.182%. In the case of the non-transformed data, the first two components could describe 87.198% of the total variance (Table 1). The variance explained by the first two PCs is large enough for both datasets to base the interpretations only on them.

In the case of the non-transformed dataset, the grain size fractions belonging to the very fine suspension have high positive loadings in PC1. All the coarser grain size fractions and the M and C parameters appeared with strong negative component loadings in the first PC. The PC2, uncorrelated with PC1, has one variable (e.i., medium silt fraction (Table 1). The clr-transformed dataset showed that the grain sizes from clay to medium silt showed high positive loadings. In contrast, the coarsest grain size fractions showed significantly negative PC loading. The M and C change parallelly with the coarse silt fraction in the PC2 (Table 1).

Table. 1 The results of the PCA for the clr-transformed and non-transformed datasets. The component loadings larger than |0.7| are highlighted.

		clr-Transformed		Non-Transformed	
		PC 1	PC 2	PC 1	PC 2
Fractions	Clay	0.929	-0.397	0.955	0.073
	Very Fine Silt	0.746	0.288	0.924	0.005
	Fine Silt	0.941	0.003	0.997	0.045
	Medium Silt	0.290	0.943	0.897	0.236
	Coarse Silt	-0.934	0.199	-0.448	0.739
	Very Fine Sand	-0.902	-0.269	-0.898	-0.222
Parameters	D5	-0.847	0.418	-0.781	-0.268
	M (= D50)	-0.982	0.049	-0.533	0.794
	C (= D95)	-0.910	-0.185	-0.006	0.914
% of variance		60.038	23.182	71.786	15.412
Cumulative % of variance		60.038	83.22	71.786	87.198

Results of Cluster Analysis

Both datasets could be subdivided into four clusters. Their average compositions are summarized in Table 2.

The average compositional data showed that the fine silt medium silty prevails in cluster 4 in both sample sets; the medium silt in cluster 3 in the clr-transformed set and cluster 2 in the non-transformed set; the coarse silt in cluster 2 of the clr-transformed and cluster 1 of the non-transformed sets; and in the cluster 1 groups of the clr-transformed data and cluster 3 subset of the non-transformed sets medium and coarse silts (Table 2).

According to the CM pattern, the studied sediments were deposited partly from the RS (Fig. 4a, b) and partly from the QR (Fig. 4a, b). Within the QR, three parts can be described from fine to coarse; fine-grained, medium-grained, and coarse-grained QR. In Fig. 4, four colors code is used to assign the four clusters of the clr-transformed and non-transformed sets. This method was effective in the identification of transport processes. The results showed that the deposits of RS belonged to cluster 4 in the case of clr-transformed and cluster 3 in the case of non-transformed datasets. The sediments of the fine-grained graded suspension were represented by cluster 3 and cluster 2 in the clr-transformed and the non-transformed samples, respectively. Cluster 1 of both datasets contained the medium-grained QR sediments, while the

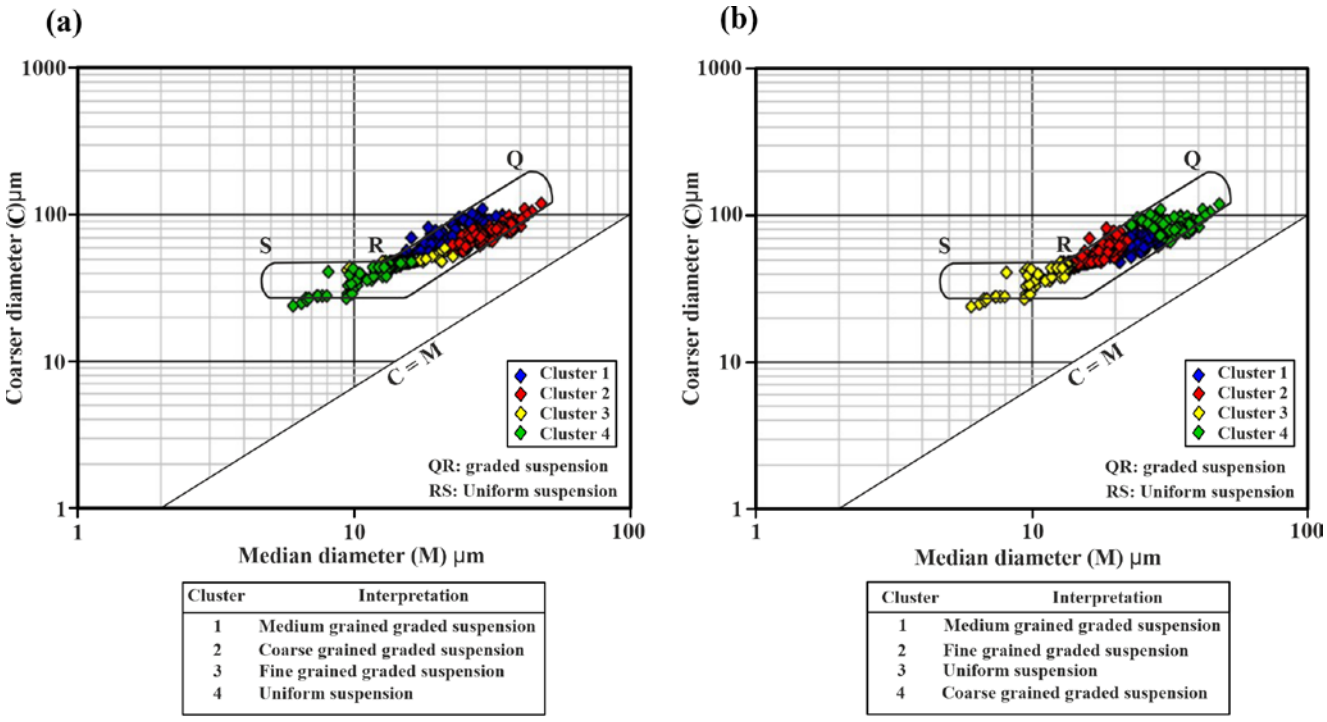

Fig. 4. Relations between the clusters and the CM diagrams in (a) the clr-transformed and (b) the non-transformed datasets.

Table 2. The average composition of the clusters in the clr-transformed and the non-transformed datasets. The dominant fractions are highlighted.

	Cluster	No. of samples	Clay	very Fine Silt	Fine Silt	Medium Silt	Coarse Silt	Very Fine Sand	
clr-Transformed	Cluster 1	114	11.7	6.9	16.62	29.22	27.01	8.4	Coarse Silty Medium Silt
	Cluster 2	81	1.47	6.74	7.69	29.04	41.95	13	Coarse Silt
	Cluster 3	98	18.52	8.19	20.11	30.92	21.87	0.38	Medium Silt
	Cluster 4	53	25.13	11	22.33	29.22	12.3	0.02	Fine Silty Medium Silt
Non-Transformed	Cluster 1	60	1.38	6.53	7.27	28.89	43.58	12.33	Coarse Silt
	Cluster 2	79	9.52	7.6	17.18	32.28	27.64	5.77	Medium Silt
	Cluster 3	39	5.65	5.63	10.52	24.38	32.84	20.32	Medium Silty Coarse Silt
	Cluster 4	168	21.1	8.97	20.64	29.93	19	0.36	Fine Silty Medium Silt

Table. 3. The discriminant analysis results of the clr-transformed and non-transformed datasets.

	clr-Transformed					Non-Transformed				
Class	Percent	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Percent	Cluster 1	Cluster 2	Cluster 3	Cluster 4
	Correct	p = .1850	p = .4509	p = .1503	p = .2139	Correct	p = .1850	p = .4509	p = .1503	p = .2139
Cluster 1	99.1228	113	1	0	0	95.3125	61	3	0	0
Cluster 2	100	0	81	0	0	100	0	156	0	0
Cluster 3	100	0	0	98	0	73	0	14	38	0
Cluster 4	69.8113	0	0	16	37	95	2	2	0	70
Total	95.0867	113	82	114	37	94	63	175	38	70

coarse-grained QR was shown by cluster 2 of the clr-transformed and cluster 4 of the non-transformed datasets (Fig. 4a, b). It showed that similar suspensions could be identified in both the clr-transformed and the non-transformed sets. In both cases, cluster1 contained the medium-grained QR (Fig. 4).

Results of Discriminant Analysis

The relationship between the two classifications and the reliability of the clustering is further explained by discriminant analysis results (Table 3). The success of classifications in both the clr-transformed and non-trans-

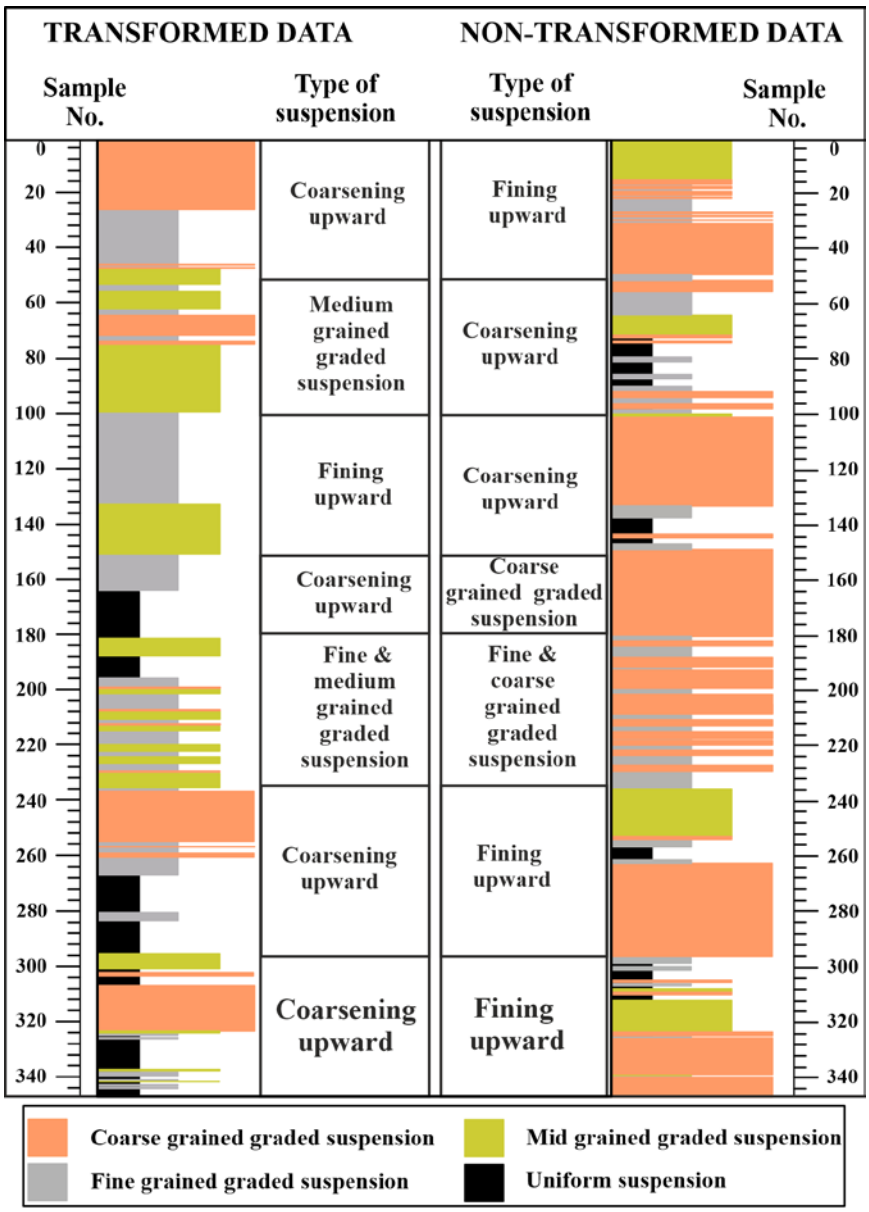

Fig. 5. Vertical sequences of the samples belonging to the different clusters and the identified vertical units of the transport processes.

formed sets was high, 95% and 94%, respectively. That is, both classifications were valid from the statistical point of view.

Vertical sequence of the identified types of suspensions

In Fig. 5, each sample is assigned to the suspension type suggested by Fig. 4 and put back to the actual stratigraphical order. In that way, the vertical pattern of the temporal change of the suspension type is established. In Fig. 4, the pattern of QR (Fig. 4a, b) is subdivided into; fine, medium, and coarse-grained QR. So, the fining upward (FU), and coarsening upward (CU) (Fig. 5) are implied. In that way, both vertical sets described a seven step temporal evolution. In the case of clr-transformed data, the CU sequences suggested flooding conditions with gradually increasing transport energy. In contrast, these flooding sequences were described with relatively thick beds of coarse-grained graded suspension (Fig. 5).

Interpretation

The most striking feature of the Tövises bed core is the periodic intercalation of the sediments, albeit the core is composed of fine grains (clay to very fine silt). This situation is well expressed in the compositional chart (Fig. 2). The CM pattern of the samples suggests that the cyclicity can be connected to the intermittently increasing transport energy when the coarser grained QR can also appear. In these periods, sediments of traction carpet origin were de-

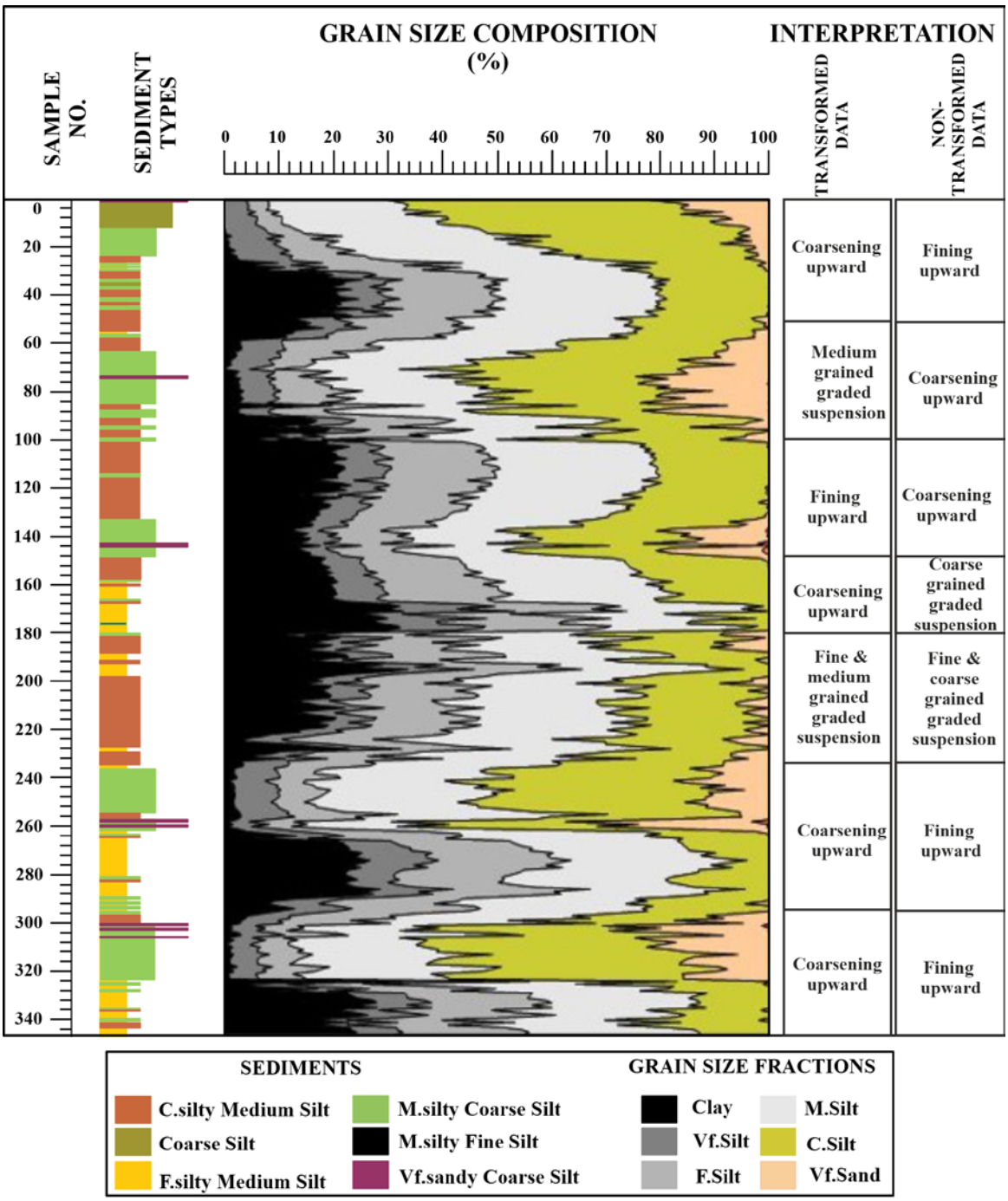

Fig. 6. A general summary of the compositional charts and the depositional histories derived from the cluster analysis of the clr-transformed and non-transformed datasets.

posited. So, an oxbow lake environment is probably the depositional site. This oxbow lake intermittently could get sediment influx from the nearby main channel in the form of QR.

In the case of clr-transformed samples, the volume of the deposited RS increases with the increasing first background process. This process decreases the volume of the deposited QR (Table 1, "Transformed," PC1). This process is probably (current free) sedimentation in an oxbow lake environment. The second principal component represents such a process. Increasing the M and C increases the coarse silt fraction (Table 1 "Transformed," PC1). In the CM diagram, the joint increase of M and C describes the QR of the traction carpet (Fig. 4). In the clr-transformed dataset, this process affects only the coarse silt fraction (Table 1 "Transformed," PC1). In an oxbow lake, two independent depositional processes can be outlined. The principal one was the quiet water sedimentation of the RS. This sedimentation was temporarily interrupted by the loads with traction carpet origin. In these periods, coarse silt grains were deposited. This deposition

model adds essential detail to the conceptual model: the oxbow lake had weak periodic connections with the channel.

For non-transformed data, the first PC describes that the decrease of M and C increases the RS deposits and decreases the coarser grain sizes (Table 1 "Non-transformed," PC1). PC1 describes the sedimentation from a traction carpet under decreasing mean flow velocity as the principal process of the deposition. There is only one grain size class in the second PC with large positive loading, i.e., the medium silt fraction is (Table 1 "Non-transformed," PC2). The critical consequence of this depositional model is that the deposition was controlled by traction flows with decreasing energy. This model supplement assumes that the oxbow lake had a permanent but weak connection with the main channel. During the flooding periods, this weak connection became stronger.

The applied CA gave a possibility to define units of the depositional history. Seven depositional events could be described (Fig. 5). The results showed that in the case of the clr-transformed dataset, the seven units corresponded to seven mainly CU cycles. This situation may suggest that the units can be connected to the periodic connection between the oxbow lake and the main channel. For the non-transformed dataset, there are seven units characterized by FU cycles (Fig. 5). In this case, the cycles can be drawn back to the periodically decreasing energy of the almost permanent weak traction flows in the oxbow lake system. The situation assumes a permanent but weak connection with the adjacent channel. Fig. 6 connects these two results with the compositional diagram of the cores.

DISCUSSION

This paper aims to explain the utility of centered log ratio transformation, euclidean data analysis, and multivariate methods in characterizing the compositional grain size data. The advantage of using clr-transformed and non-transformed datasets coupled with multivariate statistics (PCA, CA, and LDA) is that the entire GSD is considered in the analysis, unlike applying only the individual size fractions.

The first sediments to be deposited after cut off from the main channel are a plug of channel sands at each end of the oxbow lake. The exchange of water and sediments between the oxbow lake and the active channel is maintained through a narrow channel (Rowland and Dietrich, 2006). These narrow channels (known as the Tie channel) develop during lake formation (Blake and Ollier, 1971). Oxbow lake can receive relatively fine sediment transported as suspension onto the floodplain (Allen, 1970). Therefore, the oxbow lake sediments consist of silt, clay, and coarser sediments. The water enters the lake when the river is active, and the lake level is below the river and during a flood. Characterizing this variability and processes through data decomposition into dimensionally reduced components enables maximal variance to be retained using PCA. PCA applied to the clr-transformed data discovered that the first and second PCs characterize 60% and 23% of the variance, respectively (Table 1). Correspondingly, for non-transformed data, 71% and 15% of the variance are explained by the first and second PC variances, respectively (Table 1).

The following substantial question is how this conceptual model interprets the results from analyzing the clr-transformed and the non-transformed datasets. In the interpretation of PCA, the PCs are regarded as independent background processes. A particular background process significantly influences the variables with high component loadings (high positive or small negative numbers). The sign of the component loadings describes whether the increasing background process increases (positive sign) or decreases (negative sign) of the affected variables. In the case of the clr-transformed data, the very fine suspension sediments (clay, very fine silt, and fine silt) have high positive loadings in PC1. All the coarser grains sizes and the M and C parameters have high negative loadings in the first PC. The PC2, which is uncorrelated with PC1, had only one important variable, the medium silt fraction (Table 1) suggested that the oxbow lake had periodic (during floods) connections with the main channel. The non-transformed data showed that the grain sizes from clay to medium silt had high positive loadings, while the coarser grain sizes showed a significantly negative PC loading. The M and C parameters and coarse silt fraction show positive loadings in PC2 (Table 1), indicating that the oxbow lake had a permanent but weak connection with the main channel, probably, through a Tie channel. During the flooding periods, this weak connection became strong.

The CA applied for the principal component scores of both datasets revealed four groups. The validation of the clustering results poses some difficulties as the clustering itself (Pfitzer *et al*., 2009). There are 'internal' and 'external' methods to the evaluations, but they have a lot of uncertainty (Feldman and Sanger, 2007). However, the classification is statistically valid as the LDA indicates a high total percent corrects for clr-transformed and non-transformed datasets, 95% and 94%, respectively (Table 3). The sedimentological criteria used to judge the efficacy of these models is the CM diagram (Fig. 3). Accordingly, the clustering of the clr-transformed data was inefficient in generating a distinct boundary between deposition by QR and the deposition by the RS (Fig. 4a), as cluster 4 is presented as deposition of RS and QR. Therefore, the two approaches can be applied in similar situations where the sedimentological and geological criteria can assess the validity and test their efficacy.

CONCLUSIONS

This paper represents a methodology framework for characterizing the GSDs through the centered log ratio transformation, euclidean data analysis approaches, and multivariate statistics. The presented methodologies eliminate challenges caused by the closed dataset with PCA extracting the maximum variance present. This variance was studied through CA and LDA. The study revealed that the

deposition occurred as RS and QR loads in multiple stages interrupted with bottom current loads.

The CA applied for the PC scores of both datasets revealed fours groups of sequences. The vertical sequence can be subdivided into seven genetic units with similar boundaries. However, in the case of the clr-transformed dataset, they corresponded to mainly CU, suggesting the periodic connection between the oxbow lake and the main channel. Contrary to the clr-transformed dataset, the seven units of non-transformed data are characterized by FU cycles, indicating the permanent weak traction flows with periods of decreasing energy in the oxbow lake system. The situation assumes a permanent but weak connection with the main channel.

This finding demonstrates a great potential for applying clr transformation, and Euclidean data analysis approaches coupled with PCA, CA, and LDA to characterize the GSD and interpret oxbow lakes' deposition and sedimentation processes. The results are statistically significant and sedimentologically interpretable for both datasets. However, the details by which they supplemented the conceptual model are significantly different, resulting in a different temporal model of the depositional history. Therefore, the reliability of the models derived from such methods must be cross-checked with sedimentological and geological criteria as they cannot guarantee a meaningful result.

ACKNOWLEDGMENT

This research was supported by the Hungarian Government, Ministry of Human Capacities (20391-3/2018/FEKUSTRAT). The authors would like to thank the members of the Department of Geology and Paleontology for helping in the laboratory works and fruitful discussions and efforts to make the manuscript better and reaching its final form.

COMPETING INTERESTS

The authors declare no known conflicts of interest associated with this publication.

REFERENCES

Agterberg, F.P., 1974. Geomathematics. Elsevier Publ. Co, 125–148 pp.

Aitchison, J., 1982. The Statistical Analysis of Compositional Data. Journal of the Royal Statistical Society. Series B (Methodological) 44 (2), 139–177.

Aitchison, J., 1983. Principal component analysis of compositional data. Biometrika 70 (1), 57–65.

Aitchison, J., 1986. The Statistical Analysis of Compositional Data: Monographs on Statistics and Applied Probability. Chapman & Hall Ltd. London 436.

Aitchison, J., 1992. On criteria for measures of compositional difference. Mathematical Geology 24 (4), 365–379.

Aitchison, J., Barceló-Vidal, C., Pawlowsky-Glahn, V., 2002. Some comments on compositional data analysis in archaeometry, in particular the fallacies in Tangri and Wright's dismissal of log ratio analysis. Archaeometry 44 (2), 295–304.

Allen, J.R.L., 1970. Physical Processes of Sedimentation: An Introduction. Earth Science Series. Allen & Unwin Ltd., London, 248 pp.

Ashley, G.M., 1978. Interpretation of polymodal sediments. Journal of Geology 86, 411–421.

Blake, D.H., Ollier, C.D., 1971. Alluvial plains of the Fly River, Papua. Zeitschrift für Geomorphologie 12, 1–17.

Blott, S.J., Pye, K., 2001. Gradistat: A Grain Size Distribution and Statistics Package for the Analysis of Unconsolidated Sediments. Earth Surface Processes and Landforms 26 (11), 1237–1248.

Boulay, S., Colin, C., Trentesaux, A., Pluquet, F., Bertaux, J., Blamart, D., Buehring, C., Wang, P., 2003. Mineralogy and Sedimentology of Pleistocene Sediment in the South China Sea (ODP Site 1144). Proceedings of the Ocean Drilling Program: Scientific Results 184.

Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V., 2006. Compositional data analysis in the geosciences: from theory to practice. Geological Society, London, Special Publications 264, p. 212.

Davis, J.C., 2002. Statistics and Data Analysis in Geology, 3ed. John Wiley & Sons Inc., New York 550.

Donato, S.V., Reinhardt, E.G., Boyce, J.I., Pilarczyk, J.E, Jupp, B.P., 2009. Particle-Size Distribution of Inferred Tsunami Deposits in Sur Lagoon, Sultanate of Oman. Marine Geology 257 (1–4), 54–64.

East, T.J., 1985. A factor analytic approach to the identification of geomorphic processes from soil particle size characteristics. Earth Surface Processes and Landforms 10, 441–463.

East, T.J., 1987. A multivariate analysis of the particle size characteristics of regolith in a catchment on the Darling Downs, Australia. Catena 14, 101–118.

Egozcue, J.J., Pawlowsky-Glahn, V., 2005. Groups of Parts and Their Balances in Compositional Data Analysis. Mathematical Geology 37 (7), 795–828.

Egozcue, J.J., Pawlowsky-Glahn, V., 2016. What are compositional data and how should they be analyzed? Boletín de Estadíıstica e Investigacíon Operativa 32 (1), 5–29.

Egozcue, J.J., Pawlowsky-Glahn, V., Gloor, G. B., 2018. Linear association in compositional data analysis. Austrian Journal of Statistics 47 (1), 3–31.

Feldman, R., Sanger, J., 2007. The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press.

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Human Genetic 7.2, 179–188.

Flemming, B.W., 1988. Process and Pattern of Sediment Mixing in a Microtidal Coastal Lagoon along the West Coast of South Africa. In: de Boer, P.L., van Gelder, A., Nio, S.D. (Eds), Tide-Influenced Sedimentary Environments and Facies 275–288.

Flood, R.P., Orford, J.D., McKinley, J.M., Roberson, S., 2015. Effective Grain Size Distribution Analysis for Interpretation of Tidal-Deltaic Facies: West Bengal Sundarbans. Sedimentary Geology 318, 58–74.

Folk, R.L., Ward, W.C., 1957. Brazos River Bar: A Study in the Significance of Grain Size Parameters. Journal of Sedimentary Petrology 27 (1), 3–26.

Fournier, J., Gallon, R. K., Paris, R., 2014. G2Sd: a new R package for the statistical analysis of unconsolidated sediments. Geomorphologie: relief, processus, environnement 1, 73–78.

Gordon, A.D., 1999. Classification. Second Edition. Chapman & Hall, London, 272 pp.

Greenacre, M., Grunsky, E., Bacon-Shone, J., 2019. A comparison of amalgamation log ratio balances and isometric log ratio balances in compositional data analysis. In revision at Computers & Geosciences, 1–38.

He, Y., Zhao, C., Song, M., Liu, W., Chen, F., Zhang, D., Liu, Z., 2015.

Onset of Frequent Dust Storms in Northern China at ~AD 1100. Scientific Reports 5, 1–7.

Katra, I., Yizhaq, H., 2017. Intensity and degree of segregation in bimodal and multimodal grain size distributions. Aeolian Research 27, 23–34. https://doi.org/10.1016/j.aeolia.2017.05.002.

Konert, M., Vandenberghe, J., 1997. Comparison of Laser Grain Size with Pipette and Sieve Analysis. Sedimentology 44, 523–535.

McLaren, P., Hill, S.H., Bowles, D., 2007. Deriving Transport Pathways in a Sediment Trend Analysis (STA). Sedimentary Geology 202 (3), 489–498.

Mishra, S., Datta-Gupta, A., 2018. Applied Statistical Modeling and Data Analytics. A Practical Guide for the Petroleum Geosciences. Elsevier Inc. 97–118.

Palazón, L., Navas, A., 2017. Variability in Source Sediment Contributions by Applying Different Statistic Test for a Pyrenean Catchment. Journal of Environmental Management 194, 42–53.

Passega, R., 1957. Texture as Characteristic of Clastic Deposition 41 (9) 1952–1984.

Passega, R., 1964. Grain Size Representation by CM Patterns as a Geologic Tool. Journal of Sedimentary Research 34 (4), 830–847.

Passega, R., 1977. Significance of CM Diagrams of Sediments Deposited by Suspensions. Sedimentology 24 (5), 723–733.

Pawlowsky-Glahn, V., Egozcue, J.J., 2001. Geometric approach to statistical analysis on the simplex. Stochastic Environmental Research and Risk Assessment 15 (5), 384–398.

Pawlowsky-Glahn V., Buccianti A., 2011. Compositional data analysis – theory and applications. Wiley, New York, 400 pp.

Pfitzner, D., Richard, L., David, P., 2009. Characterization and evaluation of similarity measures for pairs of clusterings. Knowledge and Information Systems. Springer 19 (3), 361–394.

Reading, H.G., 1996. Sedimentary Environments: Processes, Facies and Stratigraphy. Sedimentary Environments: Processes, Facies and Stratigraphy, 688 pp.

Rowland, J.C., Dietrich, W.E., 2006. The Evolution of a Tie Channel. River, Coastal and Estuarine Morphodynamics: RCEM 2005 – Proceedings of the 4th IAHR Symposium on River, Coastal and Estuarine Morphodynamics 2 (Mossa 1996), 725–736.

Sarnthein, M., Tetzlaff, G., Koopmann, B., Wolter, K., Pflaumann, U., 1981. Glacial and Interglacial Wind Regimes over the Eastern Subtropical Atlantic and North-West Africa. Nature 293 (5829), 193–196.

Sümegi, P., Vissi, E., 1991. A pocsaji láp kialakulása és fejlődéstörténete. Calandrella 5, 15–27.

Sümegi, P., 1993. Sedimentary geological and stratigraphical analysis made on the material of the Upper Paleolithic Settlement at Jászfelsõszentgyörgy-Szúnyogos. Tisicum 8, 63–70.

Szilágyi, S.Sz., Geiger, J., 2012. Sedimentological study of the Szőreg-1 reservoir (Algyő field, Hungary): a combination of traditional and 3D sedimentological approaches. Geologia Croatica 65/1, 77–90.

Szöőr, Gy., Sümegi, P., Balázs, É., 1991. Sedimentological and geochemical analysis of Upper Pleistocene paleosols of the Hajdúság region, NE Hungary. In: Pécsi, M., Schweitzer, F. (Eds), Quaternary environment in Hungary, 26. Akadémiai Kiadó, Budapest, 47–60.

Tsagris, M., Preston, S., Wood, A.T., 2016. Improved classification for compositional data using the α-transformation. Journal of classification 33 (2), 243–261.

Udden, J.A., 1914. Mechanical Composition of Clastic Sediments. Geological Society of America Bulletin 25, 655–744.

Visher, G.S., 1969. Grain Size Distributions and Depositional Processes. SEPM Journal of Sedimentary Research 39, 1074–1106.

Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function 58 (301), 236–244.

Wentworth, C.K., 1922. A scale of grade and class terms for clastic sediments. The journal of geology 30 (5), 377–392.

Zhang, X., Zhou, A., Wang, X., Song, M., Zhao, Y., Xie, H., Russell, J.M., Chen, F., 2018. Unmixing Grain-Size Distributions in Lake Sediments: A New Method of Endmember Modeling Using Hierarchical Clustering. Quaternary Research (United States) 89 (1), 365–373.