Modern Environmental Science and Engineering (ISSN 2333-2581)
December 2016, Volume 2, No. 12, pp. 784-791
Doi: 10.15341/mese(2333-2581)/12.02.2016/004
Academic Star Publishing Company, 2016
www.academicstar.us

A Mathematical Procedure to Estimate Variations between Grain-Size Distributions

Laura Calcagni, Bianca Di Lorenzo, Rossella Di Mento, Ornella Nonnis, and Anna Maria Cicero

Institute for Environmental Protection and Research (ISPRA), Via Vitaliano Brancati, Rome 48-00144, Italy

Abstract: Grain-size is the most fundamental property of sediment particles, its analysis is essential to understand sediment provenance, transport history and depositional conditions. In the ambit of environmental monitoring a key point is often the study of the variations of sediments composition in terms of variation of their grain-size distributions. In this work it is presented a procedure based on the computing of easy mathematical indexes useful to perform in an effective way the comparison of grain-size distributions. The developed procedure consists in two steps of analysis to quantify the dissimilarities between grain-size distributions and to characterize the typology of occurred variations. A validation process is executed to verify the proper work of the procedure, using a large dataset of grain-size distributions.

Two possible applications of the procedure are presented, one to study spatial alterations of sediments composition and another one to analyze temporal changes of sediment in the same sampling station.

The proposed procedure allows to analyze in a quick way large datasets and it is flexible tool to be adapted to the peculiarities of the analyzed data, in order to optimize the achievable results.

Key words: grain-size distributions, Fréchet distance, environmental monitoring

1. Introduction

Grain-size distribution is one of the most important characteristics of sediment; since it characterizes the physical properties of sediment, it determines its provenance, transport history and depositional conditions. In the ambit of environmental monitoring, a key point is often the study of the variations of sediments composition in terms of variation of their grain-size distributions, useful, for example, to model spatial or temporal changes of sediments. In this work it is presented a procedure to perform in an effective way the comparison of grain-size distributions.

2. Fundamentals of Grain-Size Analysis

A grain-size distribution consists in an ordered sequence of values that represent the percentage quantities of granulometric composition of a sediment, expressed by mm (diameter of individual grains) or Φ ($-\log_2$ diameter of individual grains). The comparison of different sediments is dealt with different criteria [1]. In some cases standard statistics of the grain-size distributions such as mean, standard deviation, skewness and a range of cumulative percentile values (e.g., D_{50}) are used as indexes to perform the study of the grain-size variations [2]; these approaches present the limit to not consider all the distribution but only some its aspects, losing the whole characterization of sediments composition. Other authors suggest to analyze the deviation of a grain-size distribution from a prescribed ideal distribution (typically Normal distribution) but in this case it is observed the problem to weight in a different way the different particles diameters [3].

Corresponding author: Laura Calcagni, Ph.D., research areas/interests: numerical modelling; statistical modelling; environmental awareness. E-mail: laura.calcagni@isprambiente.it.

3. Materials and Methods

The aim of this work is to set up a procedure to perform the comparison of grain-size distributions, using parameters easy to compute, without a target ideal distribution in order to not influence the results by a prior fixed model and considering the whole grain-size distribution to have a complete characterization of the sediment.

Let f, g be two grain-size distributions, the first index considered to describe their similarity is the maximum difference between all couples of homologous values, that is the points characterized by the same x-coordinate:

$$IM = d_{dM} = max|f_i - g_i| \leq \alpha \qquad (1)$$

where i = 1,…,n represents the number of ordinate particles diameters of the grain-size distribution and α is the threshold to discriminate similar or dissimilar distributions; when IM index is higher than the threshold α means that grain-size distributions differ of this quantity in one of the analyzed particles diameters. In this way IM index is a measure of the variation between two grain-size distributions but it does not provide an indication of the typology of the variation occurred. In the sediments analysis, nevertheless, often it is necessary also to understand the possible cause that determines a grain-size variation, especially if there is a shift of the granulometric peak that is the signal of a deep change in sediments composition. It occurs, for example, in the marine sedimentation strongly influenced by river floods that periodically introduce particles of different diameter. There is the need to deal the study of sediment variations following other approaches, as it is performed also in other ambits of environmental modeling to compare different types of data series by the use of proper metrics [4].

Let now consider two grain-size distributions as two curves, to compute their mutual distance it is possible to apply the Fréchet distance equation that allows to compute the maximal distance between two oriented lines [5]: if f, g are the two curves (grain-size distributions) such as $f, g: [0,1]^k \rightarrow \mathbb{R}^d, k = 1{,}2 \;\; k \leq d$, their Fréchet distance is defined as:

$$d_F(f,g) = \inf_{\sigma\, hom} \max_{t \in [0,1]^k} d(f(\sigma(t)), g(\sigma(t))) \qquad (2)$$

where $\sigma:[0,1]^k \rightarrow [0,1]^k$ ranges over all orientations preserving homeomorphisms and d is the Euclideian distance. An approximation of the Fréchet distance considered by several authors is the *discrete Fréchet distance* d_{dF}, where the curves are modeled as the ordered sequences of their vertices. Let P and Q be two polygonal curves given by the ordered sequences of their end points $< p_1 \dots p_m >$ and $< q_1 \dots q_n >$ and let $C =< c_1, \dots, c_k >$ be a coupling of P and Q in an ordered sequences of pairs of vertices, where each c_i has the form $c_i = (p, q)$ with $p \in P$ and $q \in Q$, the discrete Fréchet distance (Eiter et al. (1994)) is defined as:

$$IF = d_{dF}(P,Q) = \min_{Ccoupling} \max_{(p_i, q_i) \in C \| p_i - q_i \|} \qquad (3)$$

Also if IF is a useful tool to compare two curves measuring their mutual distance, by applying it to quantify the differences between two grain-size distributions, it is pointed out that it does not work properly in some cases, where it does not represent the effective maximum distance between two distributions. In order to understand this, it is necessary to consider carefully its computing process. The Eq. (3) is computed recursively as follows [6]:

$$d_{Fd}(P,Q) = max\begin{pmatrix} d_E(P_n, Q_m) \\ min\begin{pmatrix} d_{Fd}(< P_1 \dots P_{n-1} >, < Q_1 \dots Q_m >) \forall n \neq 1 \\ d_{Fd}(< P_1 \dots P_n >, < Q_1 \dots Q_{m-1} >) \forall m \neq 1 \\ d_{Fd}(< P_1 \dots P_{n-1} >, < Q_1 \dots Q_{m-1} >) \forall n \neq 1, \forall m \neq 1 \end{pmatrix} \end{pmatrix} \qquad (4)$$

where d_E is an Euclidean distance and d_{dF} is a Fréchet distance. Applying recursively the d_{dF} process with parameters $<P_1 \dots P_{n-1}>$ and $<Q_1 \dots Q_{m-1}>$, the process ends when the two lines are reduced to two single points $<P_1>$ and $<Q_2>$. The recursive computing is performed setting up two matrices whose dimension is $m \; x \; n$, that is the number of vertices of P and Q. These matrices are MD, the matrix of Euclidean Distance and MF, the Fréchet matrix, thus the formula to compute $d_{dF}(P,Q)$ can be rewritten as:

$$d_{dF}(P,Q) = \max(d_E(P_i,Q_j), \min(MF_{i-1,j}, MF_{i,j-1}, MF_{i-1,j-1})) \quad (5)$$

The above formula works fixing each time the curve segment to analyze and computing *MD* and *MF* by coupling the segments vertices. Dividing the curves in a sequence of segments, the computing is performed for each couple of segments and the value obtained at the end of the iterative process represents the Fréchet distance. The Fréchet distance thus is the result of a minimization/maximization process of the distances between the vertices of partial segments in which the curves are divided, but this not guarantee that it represents a distance between two homologous points, as it is shown in Fig. 1. This is a limit for the application of Fréchet distance to the study of variations of sediments composition where the comparison has to be limited only to the same particle diameter.

The limit of the application of Fréchet distance to study the variations between grain-size distributions, it is solved through the joint usage of *IM* and *IF* indexes, if they differ there is a shift of granulometric peak. Hence, another index is introduced to study grain-size variations:

$$ID = \frac{1}{\alpha}(IM - IF) = \begin{cases} \in [0,1] \\ > 1 \end{cases} \quad (6)$$

when $ID \in [0,1]$ the variation between two grain-size distributions is due to a percentage change in the granulometric peak or moderate variation in all analyzed fractions; when ID > 1 there is a shift of granulometric peak between the two analyzed distributions. The parameter α is the threshold value

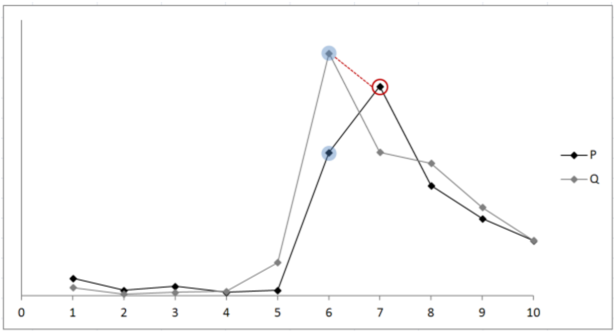


Fig. 1 The Fréchet distance computing respect to the homologous points (same x-coordinate).

used in Eq. (1). The proposed procedure is based on a two steps process, represented in the diagram of Fig. 2. In the first step *IM* index (1) allows to distinguish similar by dissimilar grain-size distributions; in the second step *ID* index (6) provides an indication of the typology of variation.

4. Validation of the Procedure

The procedure is implemented in a code developed in R language using the package "longitudinal Data", in order to analyze in a quick way also large datasets. The procedure is tested on 180 couples of real grain-size distributions; the threshold α is fixed equal to 10%, since lower variations are considered admissible. In Table 1 are summarized the results obtained at the first step of the procedure, applied to discriminate similar by dissimilar distributions by the use of the IM index, almost 50% of the analyzed curves present significant differences. On these curves the second step of procedure is applied, with the results shown in Table 2, in the 92 analyzed curves, 10 present a shift of granulometric peak.

Once completed the procedure for the whole dataset, the obtained results are checked by a visual inspection of the analyzed curves. The possible records of the

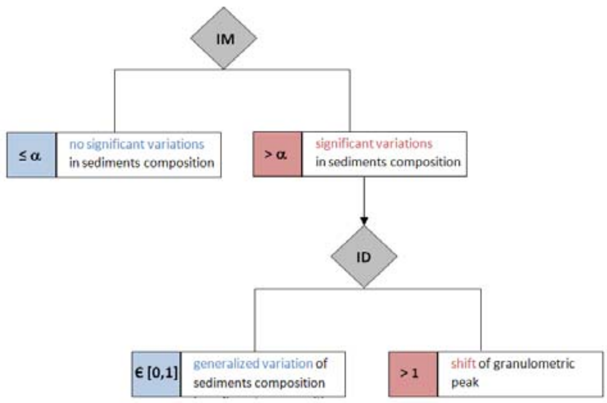


Fig. 2 Conceptual scheme of the procedure.

Table 1 Validation results of the first step of procedure — *IM* index.

couples of grain-size distributions	180
couples with no significant variations ($IM \leq \alpha$)	**88**
couples with significant variations ($IM > \alpha$)	**92**

Table 2 Validation results of the second step of procedure — *ID* index.

curves with significant variations	92
Percentage change of the peak or moderate variation in all analyzed diameters (ID∈[0, 1])	**82**
Shift of granulometric peak (*ID* > 1)	**10**

comparison between couples of grain-size distributions are represented in the Figs. 3-5. In the Fig. 3 it is represented the case of two grain-size distributions that not present significant variations, this is properly signaled by *IM* index that present values lower than the threshold α. The Fig. 4 shows a case where it is obtained a *IM* index greater than the threshold α and a *ID* index equal to 0.23, in the range [0, 1] that represents generalized variations of sediment composition. The graphs of Fig. 5, finally, show the case of the shift of granulometric peak, underlined by a *IM* index greater than the threshold α and the index *ID* > 1, in particular equal to 2.42.

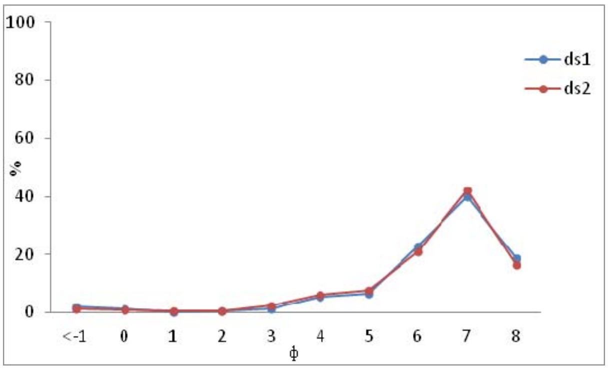


Fig. 3 Case of IM < α, the procedure does not identify significant variations between grain-size distributions.

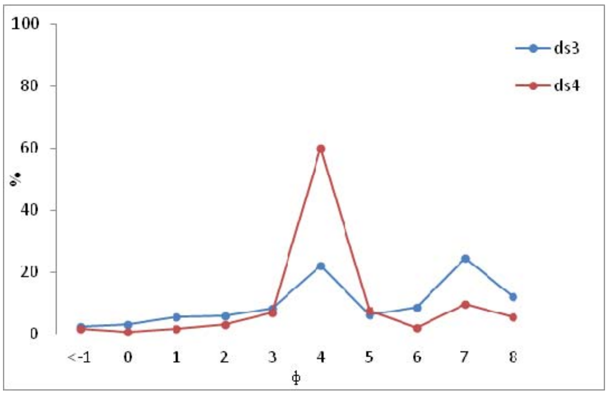


Fig. 4 Case of IM > α and ID∈[0,1], the procedure identifies a generalized variation between the two grain-size distributions.

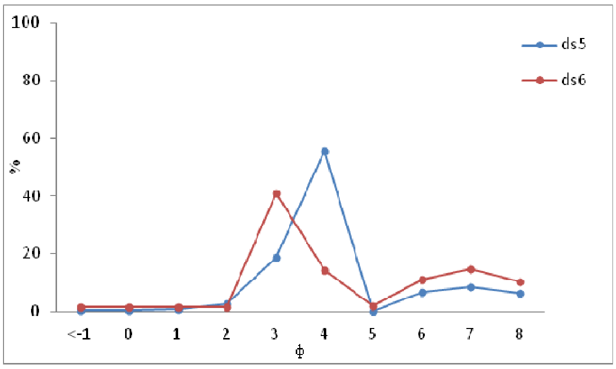


Fig. 5 Case of IM > α and ID > 1, the procedure identifies a shift of granulometric peak.

For all the analyzed grain-size distributions the correct outcome of the procedure is verified, this validation process also allowed to validate the threshold value α, since the chosen value of 10% worked proficiency to discriminate in a proper way similar by dissimilar distributions. This threshold value is independent from the considered dataset.

5. Applications of the Procedure

Two possible applications of the proposed procedure are presented, connected with two different needs of environmental monitoring.

The first application regards the study of spatial variations of sediments characteristics along monitoring transepts, where four stations (P1, P2, P3, P4) are located at different distances, as shown in Fig. 6, an amount of 30 transepts are considered.

The variations of sediments are studied considering the grain-size distributions relative to each monitoring station and coupling them two by two, in this way for each transept 6 couples of distributions are obtained. Fixing a threshold value α equal to 10%, in accord with the results described in the previous paragraph, the *IM* index is computed to each couple of grain-size distributions, and when *IM* > α also the *ID* index is computed.

The obtained results are summarized in Table 3; for each monitoring transept there are two lines, the first one with the values of *IM* index and the second one with the *ID* index. The values of *IM* index greater than α are underlined by gray color, the red color indicates the cases where *ID* > 1;; when *IM* index is lower than α the *ID* index is not computed, in accord with the procedure (Fig. 2), and the corresponding cells are empty.

Analyzing the table, since for each transept there are 4 grain-size distributions corresponding to the monitoring stations, each distribution is involved in 3 couples, if all the corresponding *IM* indexes are greater the threshold α means that this distribution differs in significant way by the others, in this way the procedure allows not only to understand that the monitoring transept is characterized by changes in the sediments composition but also to identify the station (or the stations) mainly anomalous. For example, in the transept no. 9 of the Table 3, the station number 1 presents a sediments composition highly different by the others, as it is shown in Fig. 7a.

Furthermore, examining the *ID* index values it also possible to characterize the typology of the variation and if *ID* > 1, as for example in the transept no. 15 of the Table 3, means that a shift of granulometric peak is occurred, in particular in the station number 3 as it is represented in Fig. 7b, it is possible to suppose a different origin of the sedimentation process.

The second application concerns the study of temporal variations of sediments composition. An environmental monitoring plan is performed by periodic surveys of the same area where there are several monitoring points; the data relative to each survey are compared respect to the previous ones in order to underline possible changes.

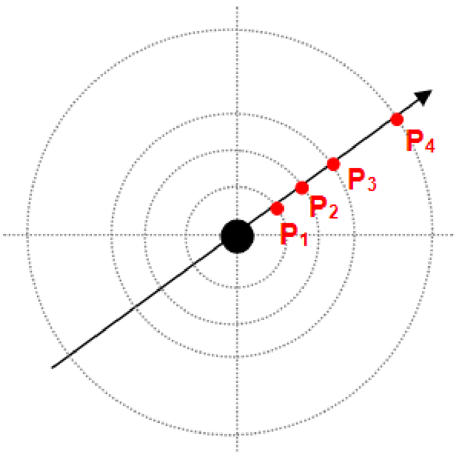
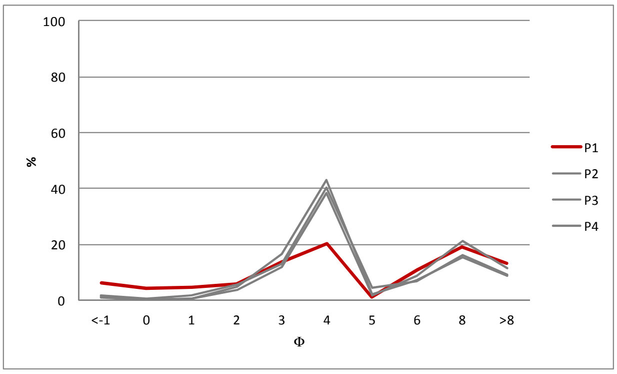


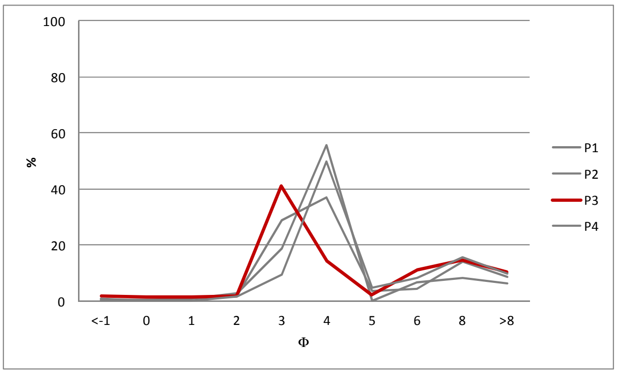
Fig. 6 An example of the considered monitoring transepts.

For example, considering to have for each monitoring point a temporal series of 7 grain-size distributions ($c_1, c_2, c_3, c_4, c_5, c_6, c_7$), obtained during the surveys of the last seven years, the goal is to compare the grain-size distribution of the current year (c_8) respect to the previous ones. Coupling two by two the eight grain-size distributions it is possible to obtain 28 couples, in seven of them the c_8 distribution is involved, in the 21 remaining it does not.

The *IM* index is computed for all couples of grain-size distributions, in this case it is not compared to a threshold value (1), instead two average values are computed, one relative to the 7 couples that involve the c_8 distribution (*avg_check*) and the other one relative to all other couples (*avg_target*). This is due to the different goal to reach, that in this case is not a



(a)



(b)

Fig. 7 An example of a two monitoring transepts, one with a station characterized by a different sediments compositions (a) and another one of a transept where a shift of granulometric peak occurred (b).

Table 3 Results of the application of the procedure to the study of spatial variation of sediments composition.

TRANSEPT	INDEX	P1-P2	P1-P3	P1-P4	P2-P3	P2-P4	P3-P4
1	IM	8.72	5.00	5.74	11.13	6.56	8.15
	ID				0.00		
2	IM	18.19	18.10	28.96	6.98	10.77	13.35
	ID	0.00	2.49	0.00		0.00	0.00
3	IM	15.49	5.05	2.08	12.91	15.28	3.88
	ID	0.00			0.00	0.00	
4	IM	4.96	8.32	6.85	9.51	3.42	11.61
	ID						0.00
5	IM	7.72	5.43	6.51	4.95	5.59	2.22
	ID						
6	IM	28.38	28.43	27.47	3.78	10.43	6.77
	ID	0.00	0.00	0.00		0.00	
7	IM	32.12	13.54	7.00	18.58	29.40	10.82
	ID	0.00	0.00		0.00	0.00	0.00
8	IM	8.34	8.76	6.13	7.76	5.52	4.31
	ID						
9	IM	20.10	17.82	22.58	5.12	3.50	5.93
	ID	0.00	0.00	0.00			
10	IM	7.38	31.09	9.70	38.09	14.55	23.54
	ID		13.27		24.87	0.00	14.14
11	IM	18.98	19.56	22.92	2.58	6.37	6.07
	ID	0.00	0.00	0.00			
12	IM	3.21	22.93	8.79	26.14	8.28	20.47
	ID		0.00		0.00		0.00
13	IM	34.23	26.69	27.59	18.11	18.44	5.98
	ID	20.04	16.68	13.07	0.00	0.00	
14	IM	4.47	9.24	23.49	7.75	24.26	16.51
	ID			0.00		0.00	0.00
15	IM	18.41	22.84	19.76	41.25	9.57	35.37
	ID	0.00	10.92	0.00	24.15		25.83
16	IM	12.13	16.80	20.56	5.12	8.43	5.47
	ID	0.00	0.00	0.00			
17	IM	11.52	11.09	12.20	22.61	23.72	5.23
	ID	0.00	0.00	0.00	0.00	0.00	
18	IM	7.91	4.72	13.72	4.84	12.47	12.45
	ID			0.00		0.00	0.00
19	IM	16.67	13.02	13.60	4.55	7.24	3.92
	ID	4.60	0.00	0.00			
20	IM	10.39	6.32	7.61	4.07	4.32	1.45
	ID	0.00					
21	IM	37.84	37.96	27.17	7.60	11.04	10.79
	ID	2.31	2.32	2.26		0.00	0.00
22	IM	19.03	17.25	13.44	12.64	17.37	5.67
	ID	5.41	9.82	4.37	0.00	6.53	
23	IM	8.41	22.63	21.80	17.25	16.42	3.98
	ID		0.00	0.00	0.00	0.00	
24	IM	24.96	22.45	23.79	7.81	11.73	4.93
	ID	0.00	5.30	6.61		3.96	
25	IM	14.52	13.01	12.42	2.17	7.86	6.03
	ID	0.00	0.00	2.35			
26	IM	5.51	12.75	5.94	16.16	10.46	6.88
	ID		0.00		2.35	3.54	
27	IM	3.72	3.50	8.88	4.48	10.80	7.80
	ID					3.59	
28	IM	18.88	9.52	6.94	9.35	14.77	5.42
	ID	12.39				6.42	
29	IM	11.44	23.64	18.27	12.92	7.07	8.84
	ID	0.00	0.00	0.00	0.72		
30	IM	9.83	4.39	2.66	14.23	12.08	2.15
	ID				0.00	0.00	

comparison between two grain-size distributions but is a check of the general trend of a distribution respect to a temporal series of other grain-size distributions. The results are summarized in the Table 4 for each monitoring point.

As it is possible to observe, for each point the *avg_check* are greater than the *avg_target*, this means than the couples that involve the grain-size distribution c_8 present an higher value of *IM* index, thus this distribution differs in a significant way from all the other ones, this means that for this survey a significant change of the sediments is occurred. In Fig. 8 it is represented the situation of one of the monitoring point, as it is possible to observe the grain-size distribution c_8 presents a trend very different from the others.

6. Conclusions

The study of the variations of sediments composition in terms of variation of their grain-size distributions plays a key role in the ambit of environmental monitoring, it is thus fundamental to compare in an effective way different grain-size distributions to model spatial or temporal changes of a sediment.

Table 4 Results of the application of the procedure to a temporal analysis of grain-size data.

Point	avg_target	avg_check
1	21.57	40.98
2	22.02	43.24
3	25.12	37.93
4	22.91	28.38
5	15.78	31.31
6	24.97	35.42
7	25.29	35.57
8	14.34	28.67
9	22.84	29.27

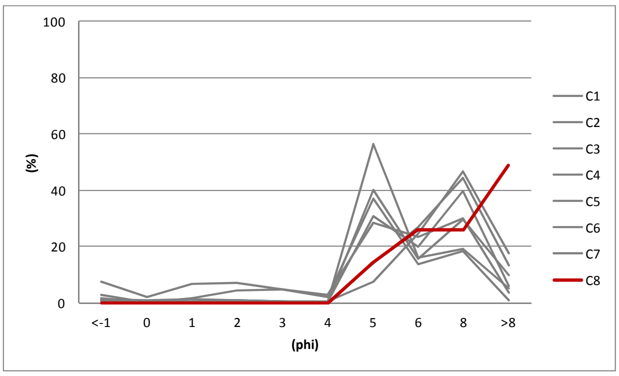


Fig. 8 An example of the temporal variation of grain-size distributions.

The developed procedure consists in two steps of analysis to quantify the dissimilarities between grain-size distributions and to characterize the typology of occurred variations; the procedure is tested on a large dataset of grain-size distributions.

Two possible applications of the procedure are presented, in two different ambits. It is shown its application to analyze spatial alterations of the sediments composition in environmental monitoring studies, the procedure indexes allow to understand if the investigated transepts are characterized by changes in the sediments composition but also to identify the monitoring stations mainly anomalous. It is also described the usage of the procedure in a temporal analysis of sediments changes, it permits to compare the trend of a grain-size distribution respect to a temporal series of other grain-size distributions relative to the same monitoring point.

The proposed procedure, being based on indexes, is standardized and repeatable; the used indexes are also easy to compute. It allows to analyze in a quick way large datasets and it works properly as shown in the different proposed applications. It is a flexible tool that can be adapted to the peculiarities of the analyzed data to optimize the achievable results.

References

[1] J. S. Blott and K. Pye, Gradistat: A grain size distribution and statistics package for the analysis of unconsolidated sediments, *Earth Surface Processes and Landforms* 26 (2001) 1237-1248.

[2] R. L. Folk and W. C. Ward, Brazos River bar: A study in the significance of grain size parameters, *Journal of Sedimentary Petrology* 27 (1957) 3-26.

[3] J. McManus, *Grain size Determination and Interpretation: Techniques in Sedimentology*, Tucker M. (Ed.), Blackwell: Oxford, 1988, pp. 63-85.

[4] U. Ehret and E. Zehe, Series distance — An intuitive metric to quantify hydrograph similarity in terms of

occurrence, amplitude and timing of hydrological events, *Hydrology and Earth System Sciences* 15 (2011) 877-896.

[5] H. Alt and M. Godau, Computing the Fréchet distance between two polygonal curves, *International Journal of Computational Geometry & Applications* 5 (1995) (1 & 2) 75-91.

[6] A. Mascrete, T. Devogele, I. Le Berre and A. Hénaff, Coastline matching process based on the discrete Fréchet distance, *Progress in Spatial Data Handling*, 2006, pp. 383-400.