Sedimentary Geology 398 (2020) 105590

Contents lists available at ScienceDirect

Sedimentary Geology

journal homepage: www.elsevier.com/locate/sedgeo

# CFLab: A MATLAB GUI program for decomposing sediment grain size distribution using Weibull functions

Li Wu [a,1], Wout Krijgsman [b,1], Jian Liu [a,c,*,1], Chaozhu Li [c,1], Rujian Wang [a,1], Wenshen Xiao [a,1]

[a] State Key Laboratory of Marine Geology, Tongji University, 200092 Shanghai, China
[b] Department of Earth Sciences, Utrecht University, 3584, CD, Utrecht, The Netherlands
[c] Institute of Geomechanics, Chinese Academy of Geological Sciences, Ministry of Natural Resources, Beijing, 100081, China

ARTICLE INFO

*Article history:*
Received 27 November 2019
Received in revised form 28 December 2019
Accepted 4 January 2020
Available online 9 January 2020

Editor: Dr. J. Knight

*Keywords:*
Sediment Grain Size Distribution
Matlab
Weibull Probability Distribution functions
Curve Fitting Technique

ABSTRACT

Curve fitting is a powerful tool that has been widely used to explore sediment grain size distributions. Open-source software designed for realization of the technique, however, is scarce. Here we provide a flexible and efficient MATLAB® GUI (Graphic User Interface) program CFLab (Curve Fitting Lab) to perform curve fitting on sediment grain size distributions using Weibull Probability Distribution Functions. CFLab deals with one grain size distribution each time. It considers the curve fitting problem as a problem of constrained nonlinear programming. In CFLab, an initial solution for this problem can be set either by designating a series of numbers for the undetermined parameters in the problem or by using a novel interactive strategy. The trust-region-reflective algorithm is used to solve the problem. Major results generated by CFLab include the percentages and grain size distributions of subpopulations fitting the raw grain size distribution data and many statistic parameters of the subpopulations calculated using two different methods. CFLab can be used to study a wide range of eolian and water-deposited sediments. A short case study is presented to demonstrate that the evolution of sediment grain size distributions in a specific sedimentary environment can be explored using CFLab.



## 1. Introduction

Sediment grain size distributions contain rich information on sediment sources, transport and depositional dynamics (e.g., Folk and Ward, 1957; Wu et al., 2018; McCave and Andrews, 2019). Parameters describing a grain size distribution (GDS), including mean, median, sorting, skewness and kurtosis etc., follow the strategy to describe the shape of a unimodal statistical Probability Distribution Function (PDF) (i.e., a normal PDF curve) (Blott and Pye, 2001). In reality, however, sediment GSDs are usually polymodal due to sorting and mixing of unimodal sediment end members from different sources and/or by variable dynamical processes (Folk and Ward, 1957; Tanner, 1964; Middleton, 1976; Ashley, 1978; Seidel and Hlawitschka, 2015). This calls for effective techniques to analyze polymodal sediment GSDs.

For this purpose, a series of numerical methods have been designed, such as various kinds of End Member Modeling Algorithms (EMMA) (Weltje, 1997; Dietze et al., 2012; Paterson and Heslop, 2015; Seidel and Hlawitschka, 2015; Yu et al., 2016), and Curve Fitting Technique (CFT) (e.g., Sun et al., 2002; Xiao et al., 2009). EMMA deals with a matrix dataset. It expresses the input dataset as mixtures of a limited number of end-members and utilizes an inversion algorithm to solve the mixing problem (Weltje, 1997; Dietze et al., 2012; Paterson and Heslop, 2015). CFT, by contrast, is designed for processing a single polymodal GSD. It considers the GSD as being composed of a series of unimodal subpopulations that can be described with specific PDFs. The idea of CFT is derived from Curray (1960), who manually decomposed polymodal GSDs into a set of lognormal PDF curves by a subtractive method. Afterwards, Sheridan et al. (1987) and Shyuer-Ming and Komar (1994) provided numerical versions of this procedure.

In its early form, CFT adopts a set of normal/lognormal PDFs to approximate a GSD curve (Clark, 1976; Bagnold and Barndorff-Nielsen, 1980; Sheridan et al., 1987; Shyuer-Ming and Komar, 1994). However, shapes of normal and lognormal PDF curves are either symmetric or right-skewed, unsuitable for fitting a GSD curve composed of asymmetric or left-skewed components (e.g., Sun et al., 2002; Yi et al., 2012). After comparing different types of statistical PDFs, Sun et al. (2002) optimized CFT by using a Weibull (Rosin) PDF as the elemental PDF, since a Weibull PDF curve offers great flexibility in terms of shape. Namely, it can be either left- or right-skewed or even symmetric (Weibull, 1951).

Nonetheless, disagreement exists on the validity of CFT itself, as fitting well does not necessarily mean significance geologically (Weltje and Prins, 2007). Applying both EMMA and CFT to the same

* Corresponding author at: Institute of Geomechanics, Chinese Academy of Geological Sciences, Ministry of Natural Resources, Beijing, 100081, China.
*E-mail address:* jianliu686@163.com (J. Liu).

[1] Author contributions: L.W. made the software and wrote the manuscript. W.K. and J.L. initiated the idea and designed the study. C.L., R.W. and W.X. contributed to the writing.

https://doi.org/10.1016/j.sedgeo.2020.105590
0037-0738/

GSD data set, Weltje and Prins (2007) and Paterson and Heslop (2015) suggested that caution must be paid when using CFT to decompose a sediment GSD. Challenges come mainly from two aspects: (1) the difficulty in deciding how many end-members are sufficient to represent a given GSD, and (2) the lack of geological context to constrain the general behavior of the end-members (i.e., using a single specimen as opposed to the entire data set). Nevertheless, unimodal GSDs, which are produced due to persistent selection during long-term sediment transport regardless of initial conditions (Folk and Ward, 1957; Tanner, 1964; Middleton, 1976; Ashley, 1978; Weltje and Prins, 2007), can be reasonably and empirically approximated by parametric distributions in many cases, and this fact makes CFT very popular among the earth science community (e.g., Sun et al., 2002, 2004; Weltje and Prins, 2007; Yi et al., 2012; McGee et al., 2013; Li et al., 2014, 2017; Osorio et al., 2014; Serno et al., 2014; Menéndez-Aguado et al., 2015; Paterson and Heslop, 2015; Wang et al., 2015; Zhang et al., 2015), albeit with those challenges.

Practicing CFT is computationally-consuming. Available software designed for CFT realization, however, is scarce. In this study, our main goal is to provide a flexible MATLAB® GUI (Graphic User Interface) program CFLab (Curve Fitting Lab) to facilitate curving fitting work on sediment GSDs using Weibull PDFs. We provide a case study to show its application in exploring sediment GSD evolution in an estuary environment. Then, we discuss its scope of applicability and compare it to the AnalySize software that integrates the curve fitting function (Paterson and Heslop, 2015) to show its merits in curve fitting.

## 2. Rationale of the Weibull CFT

The Weibull PDF has a general form as follows:

$$f(x,\alpha,\beta)=\frac{\alpha}{\beta^{\alpha}}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \tag{1}$$

where $x$, $\alpha$ and $\beta$ are respectively the independent variable, the shape and the scale parameters. $\alpha$ determines the breadth of the distribution, while $\beta$ controls the modal value of the distribution (McGee et al., 2013).

The Weibull CFT assumes that a polymodal GSD comprises a series of unimodal subpopulations (Ashley, 1978), which can be described with a set of Weibull PDFs. It is expressed mathematically as:

$$f=p_1f_1+p_2f_2+\ldots+p_if_i+\ldots+p_nf_n \tag{2}$$

$$p_i\geq 0 \tag{3}$$

$$p_1+p_2+\ldots+p_i+\ldots+p_n=1 \tag{4}$$

where $f_i$ represents the *ith* unimodal subpopulation, and $p_i$ is the contribution of the *ith* subpopulation to the polymodal GSD $f$. There are n unimodal GSDs in Eq. (2) in total. The Weibull CFT thus can be considered as a problem of constrained nonlinear programming expressed as:

Minimize

$$abs\left(f-c_1\frac{\alpha_1}{\beta_1^{\alpha_1}}x^{\alpha_1-1}e^{-\left(\frac{x}{\beta_1}\right)^{\alpha_1}}+\ldots+c_i\frac{\alpha_i}{\beta_i^{\alpha_i}}x^{\alpha_i-1}e^{-\left(\frac{x}{\beta_i}\right)^{\alpha_i}}+\ldots+c_n\frac{\alpha_n}{\beta_n^{\alpha_n}}x^{\alpha_n-1}e^{-\left(\frac{x}{\beta_n}\right)^{\alpha_n}}\right) \tag{5}$$

when

$$c_i\geq 0 \tag{6}$$

$$c_1+\ldots+c_i+\ldots+c_n=1 \tag{7}$$

where $c_i$, $\alpha_i$, and $\beta_i$ are respectively the proportion, the shape and the scale parameters of the *ith* Weibull subpopulation the same as those in Eqs. (1) and (2), and $f$ is the raw sediment GSD data. The objective of the above nonlinear programming problem is to minimize Eq. (5) by finding a set of optimal values for $c_i$, $\alpha_i$, and $\beta_i$ under the constraints

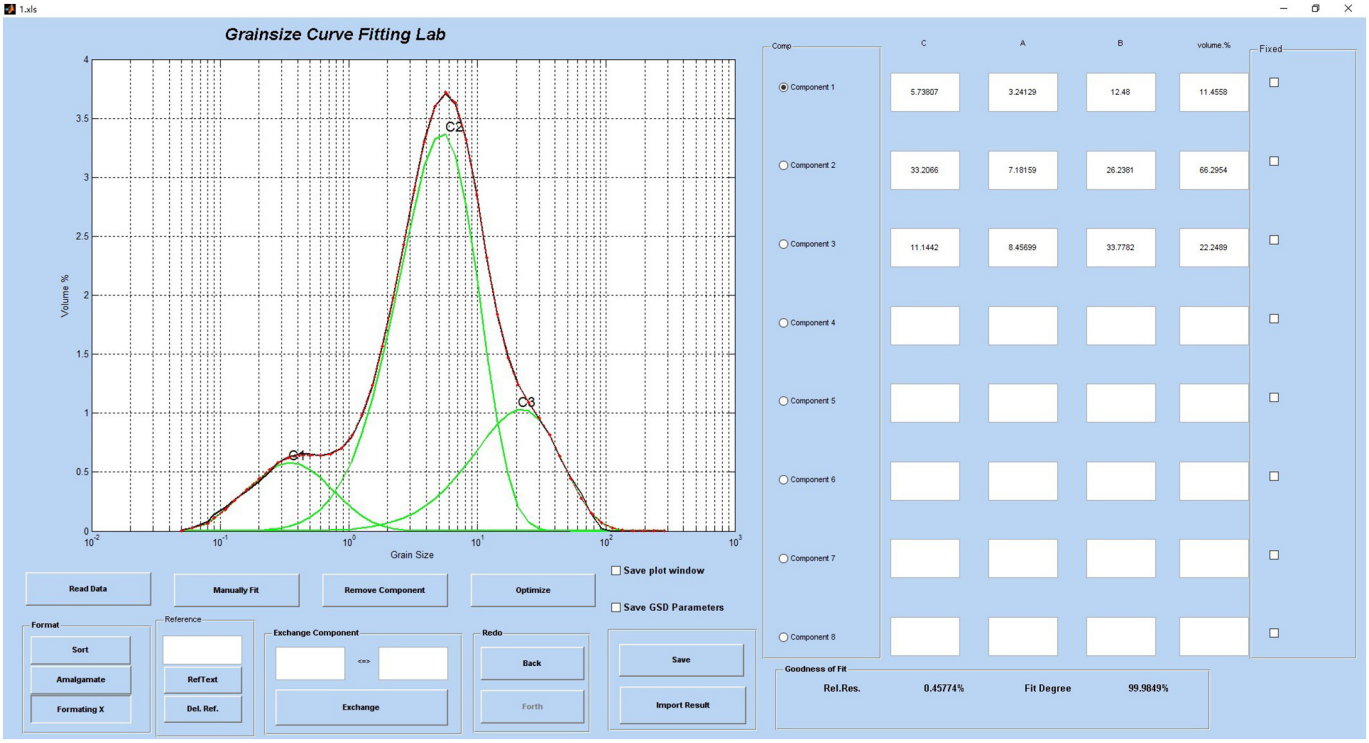


**Fig. 1.** The graphic interface of CFLab.

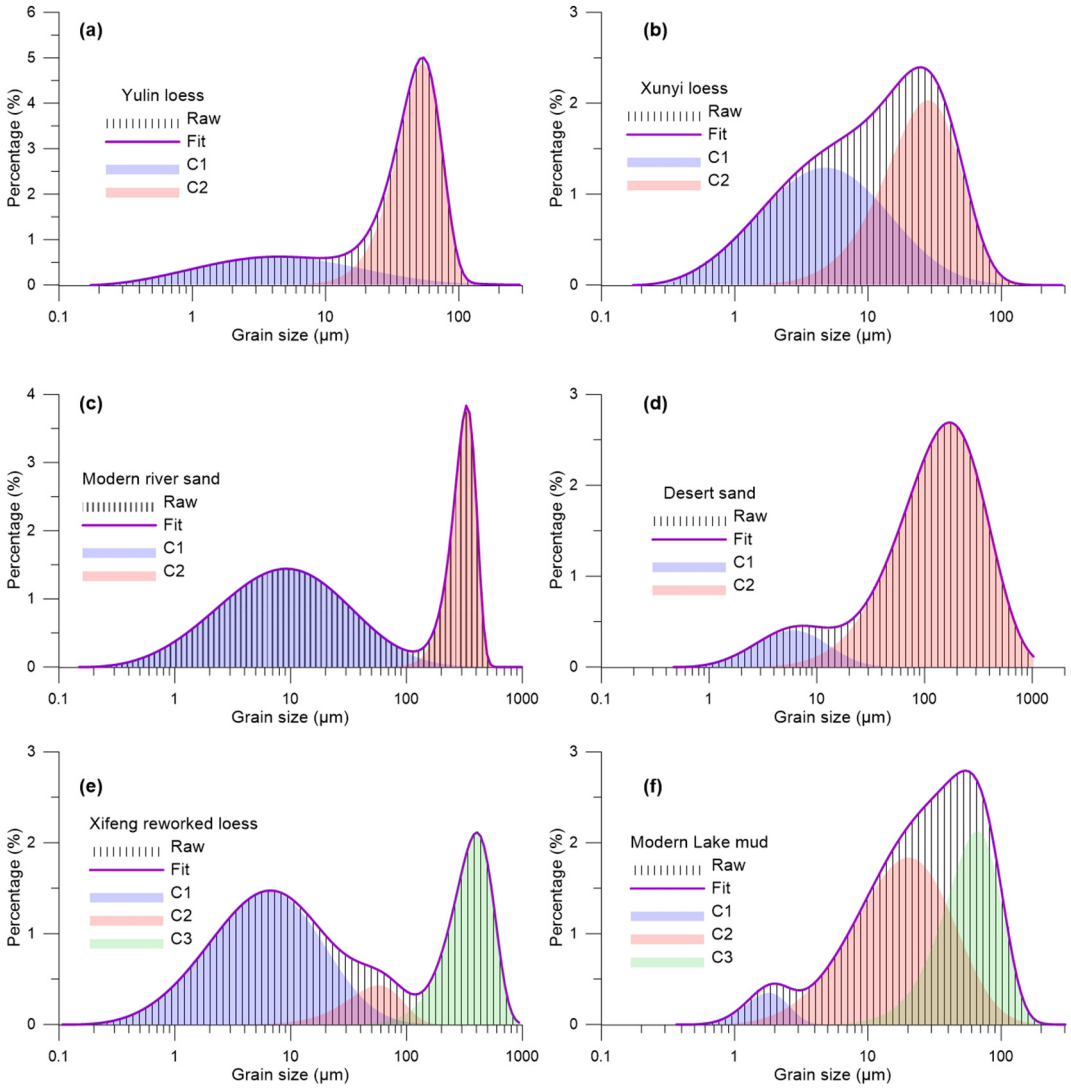


**Fig. 2.** Grain size distributions of typical eolian and hydraulic sediments and their subpopulations obtained using CFLab. (a) Yulin loess; (b) Xunyi loess; (c) Modern riverbed sand; (d) Desert sand from northwest China; (e) Xifeng reworked loess; (f) Modern lake mud from Taibei lake. The Grain size distribution datasets were from Sun et al. (2002, 2004). See Table 1 for data sources.

of Eqs. (6) and (7). Thus, the raw GSD *f* can be fitted with a set of subpopulations that can be described using Weibull PDFs.

## 3. CFLab

CFLab is a MATLAB-based GUI program designed specifically for practising CFT using Weibull PDFs (Fig. 1). It utilizes the trust-region-reflective algorithm (Coleman and Li, 1994, 1996) to solve the constrained nonlinear programming problem described in Eqs. (5)-(7). Fitting residue and fitting degree in CFLab are calculated using:

$$dF = \frac{1}{m}\sum_{i=1}^{m}(g(x_i)-G(x_i))^2 \times 100\% \tag{8}$$

$$R^2 = \left(1-\frac{\sum_{i=1}^{m}(g(x_i)-G(x_i))^2}{\sum_{i=1}^{m}(G(x_i)-G_{mean})^2}\right) \times 100\% \tag{9}$$

where $m$ is the number of grain size intervals, $g(x_i)$ is the fitted percentage of the *ith* grain size interval, $G(x_i)$ is the measured percentage of the *ith* grain size interval, and $G_{mean}$ is the mean percentage of all grain size intervals. A lower value of dF indicates a better fitting result, and if the fitted percentages agree well with the measured percentages, the $R^2$ value would be close to 100% (e.g., Zhang et al., 2015).

CFLab reads raw GSD data from an Excel file (*.xls), as this file type is versatile in data presentation and manipulation. Raw data read by CFLab contains two data columns. The first column is the grain size intervals. The second contains the volume percentage corresponding to each sediment grain size interval.

After reading the raw GSD data into CFLab, one can fit the GSD with the software flexibly and efficiently. The fitting work starts with a guess of an initial solution, i.e., initial estimate of values of $c_i$, $\alpha_i$, and $\beta_i$ in Eq. (5). Then, CFLab seeks for the optimized least-squares solution by iteration (Coleman and Li, 1994, 1996).

Usually, to solve the above nonlinear programming problem, an initial solution needs to be set by directly designating a set of numbers for the undetermined parameters in the problem. This traditional way of initial solution setting requires rich experience for dealing with Eq. (5), since the shape parameter ($\alpha$) and scale parameter ($\beta$) of a Weibull PDF do not visually correspond to statistic parameters of the subpopulation GSDs it represents (e.g., mean, median), resulting in difficulties in making an appropriate initial estimate of them.

In CFLab, the initial solution estimating is of great convenience. It can be either set by the traditional way of designating a set of values, or set

**Table 1**
Comparisons between the volume percentages of the subpopulations of sediment GSDs in Fig. 2 derived from this study and Sun et al. (2002, 2004). Fitting residuals and fitting degrees were given by this study.

| Sample | Volume percentage of subpopulations from this study (Sun et al.2002, 2004) | | | Fitting Residual (%) | Fitting Degree (%) |
|---|---|---|---|---|---|
| | C1 (%) | C2 (%) | C3 (%) | | |
| Yulin loess[a] | 30.3 (30) | 69.7 (70) | - | 0.046 | 99.92 |
| Xunyi loess[a] | 51.2 (51) | 48.8 (49) | - | 0.050 | 99.98 |
| Modern river sand[b] | 67.0 (67) | 33.0 (33) | - | 0.051 | 99.98 |
| Desert sand[b] | 8.2 (8.5) | 91.8 (91.5) | - | 0.041 | 99.97 |
| Xifeng reworked loess[b] | 59.6 (60) | 8.5 (8) | 31.9 (31) | 0.043 | 99.99 |
| Taibei lake mud[b] | 4.7 (4.6) | 56.8 (57.4) | 38.5 (37.9) | 0.052 | 99.98 |

[a] : data from Sun et al. (2004).
[b] : data from Sun et al. (2002).

interactively on its plot window. Namely, one can visually determine where a subpopulation should be on the raw sediment GSD plot, and then draw a sketch of the potential subpopulation as its initial estimate by clicking the mouse and selecting on the plot window of the CFLab software.

Once initial estimation for all potential subpopulations has been done, one needs only to press the 'Optimize' button, then the optimized solution will be calculated and show on the plot window of CFLab (Fig. 1). At any time, a subpopulation can be modified and/or deleted if it does not meet the fitting demand. In addition, the shape parameter and the scale parameter of the subpopulations can be fixed during a fitting experiment to allow only its proportion to be changed. This function is useful when a constant GSD of a subpopulation throughout the studied sample set is assumed and/or desired (like an invariable end-member in EMMA, Weltje, 1997). Moreover, every step of a fitting experiment will be stored and can be undone and recovered. These novel designs make CFLab flexible and efficient.

CFLab can deal with at most eight subpopulations for a single fitting run. This amount meets the need of most types of sediment GSDs suitable to be explored using Weibull CFT, because geologically meaningful fitting usually involves no more than five subpopulations (e.g., Sun et al., 2002, 2004; McGee et al., 2013; Zhang et al., 2015).

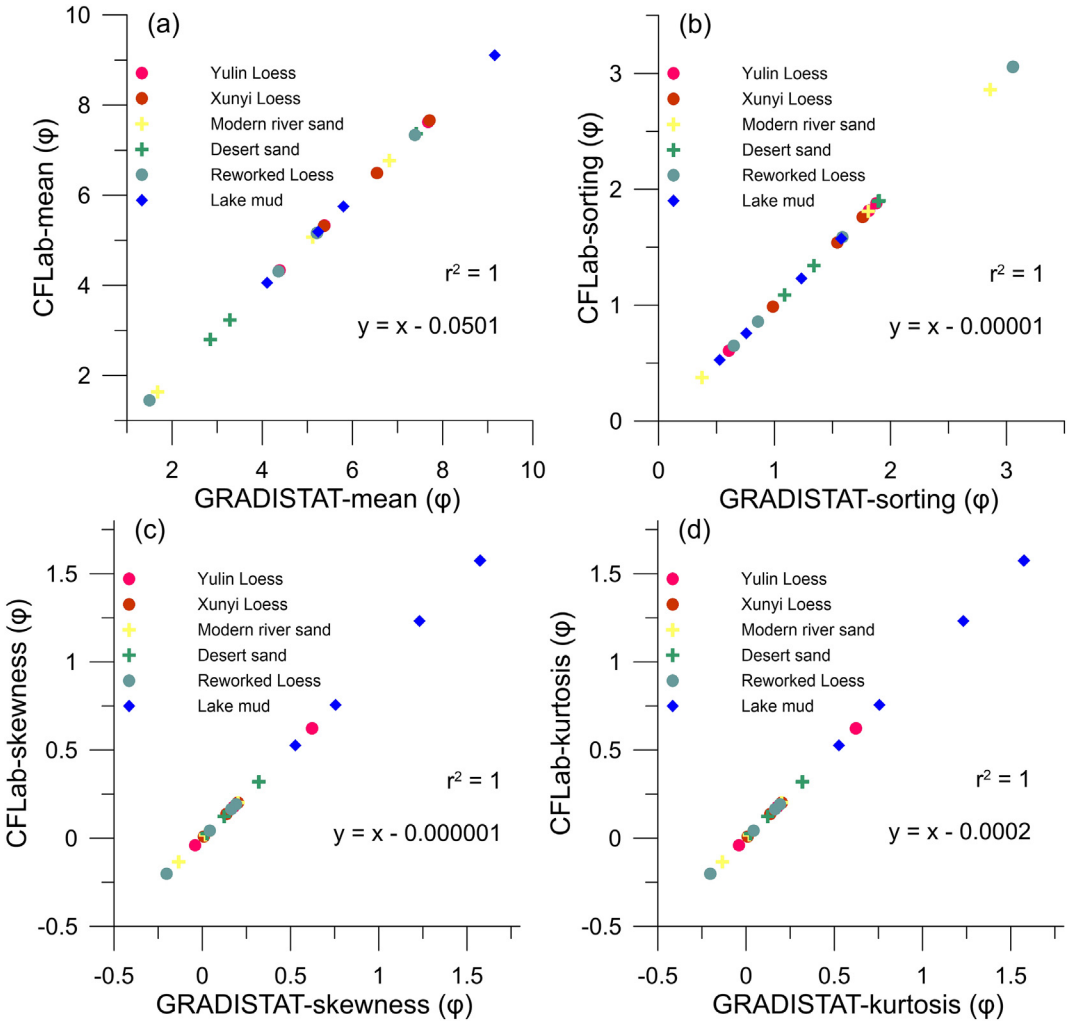


**Fig. 3.** Comparisons between CFLab- and GRADISTAT-derived statistic statistic parameters of the GSDs in Fig. 2, including the raw and the fitted GSDs and GSDs of all the subpopulations.

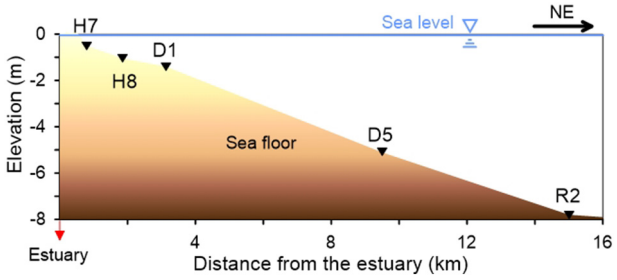


**Fig. 4.** Surface sample sites from the estuary of Laizhou Bay.

After a fitting is done, data of the percentages and GSDs of the subpopulations fitting the raw GSD will be obtained and can be saved as a worksheet entitled *'Components'* appending to the raw data file. A series of grain size statistic parameters including mean, median, sorting, skewness, kurtosis etc of the raw GSD, the fitted GSD, and all the generated subpopulations can be calculated and saved in another worksheet entitled *'Parameters'*. These statistic parameters are calculated by both the graphic method of Folk and Ward (1957) and the method of moments of McManus (1988). The plot window of CFLab can be saved as a '*.emf' file with the same name of the raw data file as well if needed. In addition, one can read the fitting data from the *'Components'* worksheet of the raw data file back to the CFLab software once it has been produced.

## 4. Performance of CFLab

To test the fitting performance of CFLab, GSDs of typical eolian and water-deposited sediments previously investigated with Weibull CFT by Sun et al. (2002, 2004) were analyzed using CFLab (Fig. 2). The results show that the fitting residuals generated by CFLab range from 0.041% to 0.052%, and fitting degrees range from 99.92% to 99.99% (Table 1). Both the two indices demonstrate the fitting is excellent. The resulted unimodal Weibull subpopulations and their contributions to the raw GSDs are basically the same as those reported by Sun et al. (2002, 2004), suggesting the fitting results of CFLab are reliable, whereas the CFLab software is more efficient, and the fitting results it generates are quite easy to be read, checked and re-made due to its graphic interface and interactive features.

On the other hand, to test the reliability of the statistic parameters generated by CFLab, We compare them with those generated by the GRADISTAT software (Blott and Pye, 2001) for the same data set (Fig. 3). The result shows that each of the parameters generated by CFLab is linearly correlated to its counterpart generated by GRADISTAT with a slop of unit, an intercept approaching to zero and a coefficient of determination ($r^2$) of unit, indicating that the two sets of parameters are excellently consistent with each other as well, albeit with some negligible discrepancy, which we ascribe to calculation error.

## 5. Application

CFLab is not only able to provide the volume percentages and GSDs of subpopulations that fit the raw GSD data, but also able to output many statistical parameters of the GSDs of the subpopulations. Investigation of the subpopulations with their statistics might provide important information on sediment GSD evolution in a specific environment. To demonstrate the application of this kind of investigation, we provide a case study by exploring the general evolution of sediment GSDs of an estuary using CFLab. The estuary environment is characterized by unique interactions between fluvial and tidal processes (Milliman and Syvitski, 1991). Characterizing sediment GSD from estuary environment allows for a better understanding of coastal evolution, and may assist in coastal management efforts (Qiao et al., 2010).

The sediment GSD dataset used in this case study was from Yi et al. (2012). A profile of the sampling sites is sketched in Fig. 4. Five surface sediment samples were collected off the estuary of Laizhou Bay (China) at variable water depths and distances. The GSD data were unmixed using CFLab. Three subpopulations (C1 to C3) for each GSD were obtained. The unmixed result is shown in Fig. 5. Parameters of the raw and fitted GSDs as well as resulted subpopulations were also calculated using the software and partly shown in Fig. 6. Among the three resulted

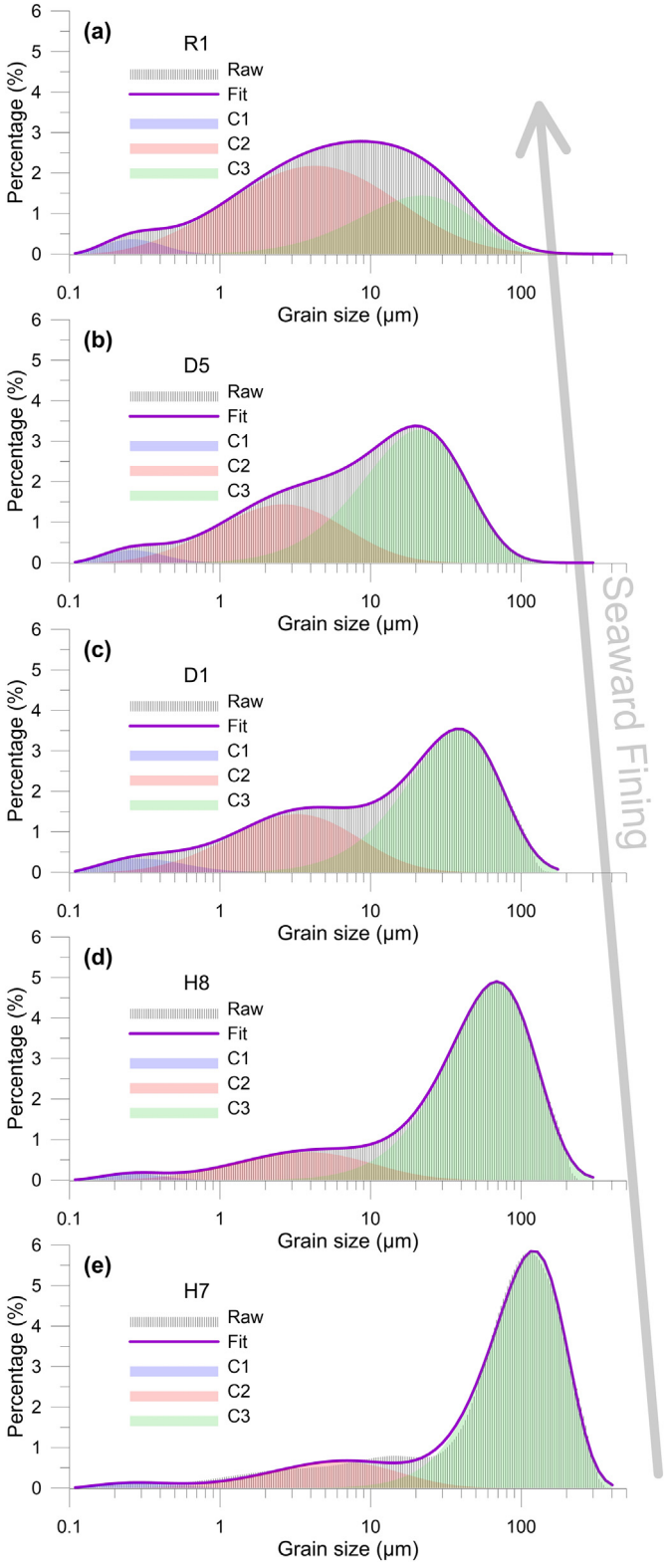


**Fig. 5.** Sediment GSDs showing a seaward fining trend along the transect in Fig. 4. They are decomposed into Weibull subpopulations using CFT.

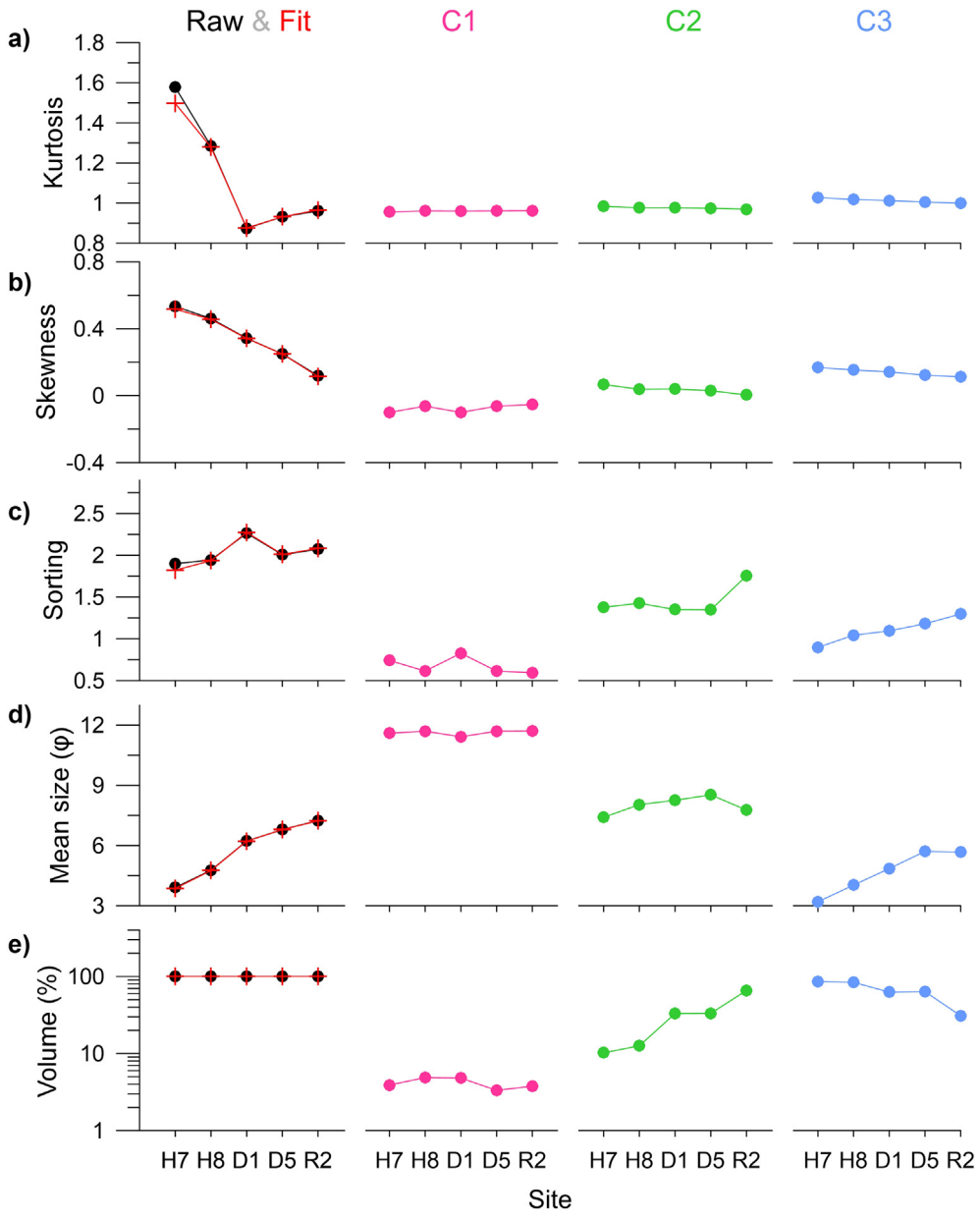


**Fig. 6.** Evolution of sediment grain size parameters along the transect in Fig. 4.

subpopulations, C1 has a stable mean grain size of ~0.4 μm in all samples, and its volume percentage never exceeds 5%. It is likely an artifact generated by sonification during sample preparation (Yi et al., 2010). Considering its constantly small volume percentage and stable GSD pattern, it was excluded from the analyses below.

After excluding C1, each GSD comprises two distinct major subpopulations, i.e., a well-sorted coarse subpopulation C1 with a mean grain size changing from 3.2-5.7 φ and a poorly-sorted fine subpopulation C2 with a mean grain size between 7.4-8.5 φ. C3 and C2 were transported by saltation and suspension, respectively according to Ashley (1978).

Site H7 was located on a shallow locus near the river mouth (Fig. 3). The sediment GSD at site H7 is characterized by a high saltation/suspension (C3/C2) ratio of 8.4 relative to that of the typical river bed sediment (1.5-2.5, Sun et al., 2002), indicating a lack of fines. This is possibly due to repeated washing of the sediment by waves and tides. As a result, fines were transported where hydraulic energy is low (Folk and Ward, 1957).

With increasing water depth and distance from the river mouth, the raw GSD progressively becomes finer (Fig. 5). This seaward fining trend is accompanied by decrease in the proportion, and mean grain size of C3 and poorer sorting of C3 (Fig. 6c-d), whereas the mean grain size and sorting parameter of C2 do not change significantly.

The grain size of C3 is mostly between 10-200 μm, incorporating the 'sortable silt' and the fine sand fractions. These grain size fractions are sensitive to the strength of bottom currents (McCave, 2008; Lamy et al., 2015). In particular, there is a strong negative linear correlation between the mean grain size (in φ scale) and the volume percentage of C3 ($r=-0.82$, $n=5$), robustly indicating the seaward fining of sediment grain size is due to a gradual weakening of bottom currents that might be associated with decrease in the energy of waves and tides with increasing water depth and distance from the bank (cf., Lamy et al., 2015; McCave and Andrews, 2019). The less variable C2, by contrast, is mostly < 10 μm. It is insensitive to changes in hydraulic dynamics, and even very weak flows can suspend it perennially in the water column once it is initiated (McCave, 2008).

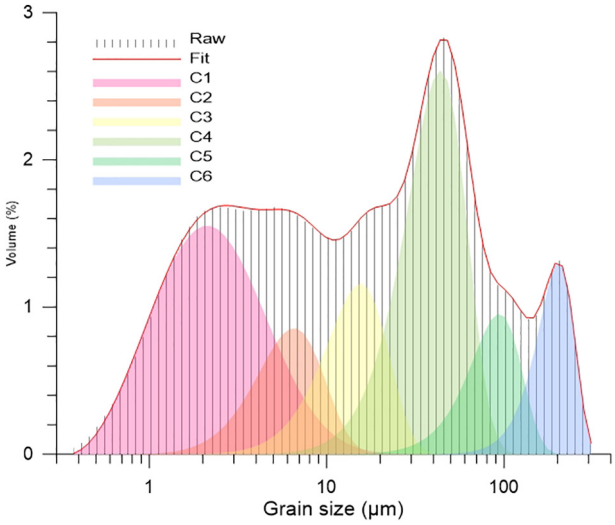


**Fig. 7.** A complex polymodal GSD of glaciomarine sediment from the Southern Ocean (Wu et al., 2018). It can be decomposed into six subpopulations (C1-C6) using CFLab, but with undetermined geological significance.

Generally, the sorting parameters of the raw GSDs are poorer than those of their subpopulations (Fig. 6c). Particularly, sorting of the raw GSD at site D1 is the poorest. This occurs primarily when C2 and C3 have comparable contributions to the raw GSD, indicating the mixing proportion of C2/C3 determines the sorting of the bulk GSD. However, the sorting of C2 and C3 themselves may also exert a secondary influence, such that the sorting of the raw GSDs become increasingly poor with progressive addition (reduction) of poorly (well) sorted C2 (C3) to (from) the sediment.

A similar but different scenario occurs for the kurtosis parameter (Fig. 6a). Kurtosis of the raw GSDs decreases from sites H7 to D1, and then it increases somewhat from sites D1 to R2. The lowest kurtosis of raw GSDs corresponds to a time when contributions of the two subpopulations to the raw GSDs are comparable as well. However, C2 and C3 have rather stable and similar kurtosis values independent of changes in water depth, resulting in their limited influence on the kurtosis of the raw GSD. The increasing trend of kurtosis both from sites D1 to H7 and from sites D1 to R2 thus essentially reflects progressive deviations of the contributions of C2 and C3 to the raw GSD from a comparable state. The gradual seaward decrease in the skewness of the raw GSDs (Fig. 6b), on the other hand, is also due to successive reduction (addition) of C3 (C2) from (to) the sediments, resulting in a less and less obvious tail on the left (the fine side) (Fig. 5).

## 6. Discussion

### 6.1. Scope of applicability of CFLab

In combination with CFT and calculations on statistics of GSDs of the subpopulations, the described case study shows that CFLab is powerful in exploring the evolution of sediment GSDs in a specific environment. To reasonably use CFLab, however, we emphasize that one must know the studied sedimentary environment well. In particular, it would be helpful if the number and the geological significance of the potential subpopulations contributing to the sediment are already known, because (1) theoretically, CFLab itself can not determine the number of end-members sufficient to represent a given GSD, and (2) in complex cases (e.g., Fig. 7), although a polymodal GSD can be decomposed into a series of subpopulations in terms of good fitting degree, caution must be paid when interpreting these subpopulations in the sense of geology/sedimentology as suggested by Weltje and Prins (2007).

Sun et al. (2002) summarized that a spectrum of eolian and water-deposited sediments are suitable to be unmixed using the Weibull CFT, including (reworked) loess, desert sand, fluvial and lacustrine sediments etc, and our re-unmixing work supports this point of view (Fig. 2). These sediment types are generally comprises two to three distinct subpopulations. The geological significance of these subpopulations has been well-documented (e.g., Ashley, 1978; Sun et al., 2002, 2004). With such background knowledge, CFLab thus can be confidently used.

### 6.2. Compare CFLab with AnalySize

To our knowledge, the AnalySize software made by Paterson and Heslop (2015) is also able to conduct curve fitting using four types of PDFs, including Lognormal, Weibull, General Weibull and Skewed Generalized Gaussian PDFs. There are several major differences between AnalySize and CFLab, as follows:

(1) Unlike AnalySize, the present version of CFLab can only practise the Weibull CFT, because we found Weibull PDF is the most popularly used PDF type in curving fitting studies (e.g., Sun et al., 2002; Weltje and Prins, 2007; Yi et al., 2012; McGee et al., 2013; Li et al., 2014; Serno et al., 2014; Paterson and Heslop, 2015; Wang et al., 2015). But it is easy to integrate other PDF types into CFLab if necessary in the future.
(2) Unmixing sediment GSDs using AnalySize starts with a non-parametric EMA (End Member Aanlysis) procedure to help estimate the number of subpopulations potentially contributing to the sediment GSDs and the initial parameters of these subpopulations. This design aims to take advantage of the genetic information present in a set of GSDs and to produce unimodal end members (subpopulations). However, it also means that AnalySize is unable to process a single GSD or a sparse GSD dataset (i.e., the number of GSDs in a dataset is < 10; Paterson and Heslop, 2015). By contrast, CFLab was designed to deal with a single GSD each time. The above limitation of AnalySize thus does not exist in CFLab. In particular, GSDs of subpopulations produced by AnalySize are always constant and apply to the whole sample dataset. This is sometimes just a low-order approximation of the 'real' GSD subpopulations (e.g., Paterson and Heslop, 2015), and cannot reflect the dynamic evolution of the sediment GSDs from the studied environment through space and/or time as CFLab does in our case study (cf., Fig. 5).
(3) CFLab is able to compute and output many statistics of subpopulations fitting the raw sediment GSD. Such information, however, cannot be produced directly by AnalySize.

## 7. Conclusions

The MATLAB® GUI program CFLab is a flexible and efficient tool that can be used to conduct sediment GSD curve fitting using Weibull PDFs. CFLab can be safely used to explore GSDs of a wide range of eolian and water-deposited sediments, including (reworked) loess, desert sand, fluvial, lacustrine and estuary sediments. It generates not only the optimized subpopulations that fit the GSD of a sample, but also outputs much useful statistic information on those subpopulations. Such outputs facilitate studies on the evolution of sediment GSDs in different sedimentary environments. To our knowledge, CFLab is the first open-source software that designed specifically for curve fitting on sediment GSD. It can be updated readily to integrate more functions that are relevant to GSD curve fitting in its future versions. The CFLab software as well as its manual can be downloaded via https://data.mendeley.com/datasets/7khd858vfm/draft?a=b501588c-4174-4a6db5f2-7d4ea8e00bfe.

## Acknowledgement

We are most grateful to Prof. Jasper Knight and another anonymous reviewer who kindly helped to improve the manuscript a lot. We also thank those authors who have provided us with the sediment grain size distribution data used in this work. This work is jointly supported by Chinese National Natural Science Foundation and China Postdoctoral Science Foundation (Grant No. 41806223, 2018M632158, 41572351 and 40776004).

## Data Availability

This work does not produce new data. The CFLab program and its manual can be found in the supplementary information, or can be obtained by asking the authors.

## References

Ashley, G.M., 1978. Interpretation of polymodal sediments. J. Geol. 86, 411–421.
Bagnold, R.A., Barndorff-Nielsen, O., 1980. The pattern of natural size distributions. Sedimentology 27, 199–207.
Blott, S.J., Pye, K., 2001. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. Earth Surf. Process. Landf. 26, 1237–1248.
Clark, M.W., 1976. Some methods for statistical analysis of multimodal distributions and their application to grain-size data. Math. Geol. 8, 267–282.
Coleman, T., Li, Y., 1994. On the Convergence of Reflective Newton Methods for Large-Scale Nonlinear Minimization Subject to Bounds. Math. Program. 67, 189–224.
Coleman, T.F., Li, Y., 1996. An Interior, Trust Region Approach for Nonlinear Minimization Subject to Bounds. SIAM J. Optim. 6, 418–445.
Curray, J.R., 1960. Tracing sediment masses by grain size modes. Report of the Twenty-First Session Norden. Int. Geol. Congr. 119–130.
Dietze, E., Hartmann, K., Diekmann, B., IJmker, J., Lehmkuhl, F., Opitz, S., Stauch, G., Wünnemann, B., Borchers, A., 2012. An end-member algorithm for decipheringmodern detrital processes from lake sediments of Lake Donggi Cona, NE Tibetan Plateau, China. Sediment. Geol. 243, 169–180.
Folk, R.L., Ward, W.C., 1957. Brazos River bar: a study in the significance of grain size parameters. J. Sediment. Res. 27, 3–31.
Lamy, F., Arz, H.W., Kilianc, R., Lange, C.B., Lembke-Jene, L., Wengler, M., Kaiser, J., Baeza-Urrea, O., Hall, I.R., Harada, N., Tiedemann, R., 2015. Glacial reduction and millennialscale variations in Drake Passage throughflow. Proc. Natl. Acad. Sci. U. S. A. 112, 13496–13501.
Li, Z., Sun, D., Chen, F., Wang, F., Zhang, Y., Guo, F., Wang, X., Li, B., 2014. Chronology and paleoenvironmental records of a drill core in the central Tengger Desert of China. Quat. Sci. Rev. 85, 85–98.
Li, B., Sun, D., Xu, W., Wang, F., Liang, B., Ma, Z., Wang, X., Li, Z., Chen, F., 2017. Paleomagnetic chronology and paleoenvironmental records from drill cores from the Hetao Basin and their implications for the formation of the Hobq Desert and the Yellow River. Quat. Sci. Rev. 156, 69–89.
McCave, I., 2008. Size sorting during transport and deposition of fine sediments: sortable silt and flow speed. Contourites. Dev. Sedimentol. 60, 121–142.
McCave, I., Andrews, J., 2019. Distinguishing current effects in sediments delivered to the ocean by ice. I. Principles, methods and examples. Quat. Sci. Rev. 212, 92–107.
McGee, D., deMenocal, P.B., Winckler, G., Stuut, J.B.W., Bradtmiller, L.I., 2013. The magnitude, timing and abruptness of changes in North African dust deposition over the last 20,000-yr. Earth Planet. Sci. Lett. 371–372, 163–176.
McManus, J., 1988. Grain size determination and interpretation. In: Tucker, M. (Ed.), Techniques in Sedimentology. Backwell, Oxford, pp. 63–85.
Menéndez-Aguado, J.M., Peña-Carpio, E., Sierra, C., 2015. Particle size distribution fitting of surface detrital sediment using the Swrebec function. J. Soils Sediments 15, 2004–2011.
Middleton, G.V., 1976. Hydraulic Interpretation of Sand Size Distributions. J. Geol. 84, 405–426.
Milliman, J., Syvitski, J., 1991. Geomorphic Tectonic Control of Sediment Discharge to Ocean–The Importance of Small Mountainous Rivers. J. Geol. 100, 525–544.
Osorio, A.M., Menendez-Aguado, J., Bustamante-Rúa, M., Restrepo, G., 2014. Fine grinding size distribution analysis using the Swrebec function. Powder Technol. 258, 206–208.
Paterson, G.A., Heslop, D., 2015. New methods for unmixing sediment grain size data. Geochem. Geophys. Geosyst. 16, 4494–4506.
Qiao, S., Shi, X., Zhu, A., Liu, Y., Bi, N., Fang, X., Yang, G., 2010. Distribution and transport of suspended sediments off the Yellow River (Huanghe) mouth and the nearby Bohai Sea. Estuar. Coast. Shelf Sci. 86, 337–344.
Seidel, M., Hlawitschka, M., 2015. An R-Based Function for Modeling of End Member Compositions. Math. Geosci. 47, 995–1007.
Serno, S., Winckler, G., Anderson, R.F., Hayes, C.T., McGee, D., Machalett, B., Ren, H., Straub, S.M., Gersonde, R., Haug, G.H., 2014. Eolian dust input to the Subarctic North Pacific. Earth Planet. Sci. Lett. 387, 252–263.
Sheridan, M.F., Wohletz, K.H., Dehn, J., 1987. Discrimination of grain-size subpopulations in pyroclastic deposits. Geology 15, 367–370.
Shyuer-Ming, S., Komar, P.D., 1994. Sediments, Beach Morphology and Sea Cliff Erosion within an Oregon Coast Littoral Cell. J. Coast. Res. 10, 144–157.
Sun, D., Bloemendal, J., Rea, D.K., An, Z., Vandenberghe, J., Lu, H., Su, R., Liu, T., 2004. Bimodal grain-size distribution of Chinese loess, and its palaeoclimatic implications. CATENA 55, 325–340.
Sun, D., Bloemendal, J., Rea, D., Vandenberghe, J., Jiang, F., An, Z., Su, R., 2002. Grain-size distribution function of polymodal sediments in hydraulic and aeolian environments, and numerical partitioning of the sedimentary components. Sediment. Geol. 152, 263–277.
Tanner, W.F., 1964. Modification of sediment distributions. J. Sediment. Petrol. 34, 156–164.
Wang, F., Sun, D., Chen, F., Bloemendal, J., Guo, F., Li, Z., Zhang, Y., Li, B., Wang, X., 2015. Formation and evolution of the Badain Jaran Desert, North China, as revealed by a drill core from the desert centre and by geological survey. Palaeogeogr. Palaeoclimatol. Palaeoecol. 426, 139–158.
Weibull, W., 1951. A Statistical Distribution Function of Wide Applicability. J. Appl. Mech. 18, 293–297.
Weltje, G.J., 1997. End-member modeling of compositional data: numerical-statistical algorithms for solving the explicit mixing problem. Math. Geol. 29, 503–549.
Weltje, G.J., Prins, M.A., 2007. Genetically meaningful decomposition of grain-size distributions. Sediment. Geol. 202, 409–424.
Wu, L., Wang, R., Xiao, W., Krijgsman, W., Li, Q., Ge, S., Ma, T., 2018. Late Quaternary Deep Stratification-Climate Coupling in the Southern Ocean: Implications for Changes in Abyssal Carbon Storage. Geochem. Geophys. Geosyst. 19. https://doi.org/10.1002/2017GC007250.
Xiao, J., Chang, Z., Si, B., Qin, X., Itoh, S., Lomtatidze, Z., 2009. Partitioning of the grain-size components of Dali Lake core sediments: evidence for lake-level changes during the Holocene. J. Paleolimnol. 42, 249–260.
Yi, L., Yu, H., Ortiz, J.D., Xu, X., Qiang, X., Huang, H., Shi, X., Deng, C., 2012. A reconstruction of late Pleistocene relative sea level in the south Bohai Sea, China, based on sediment grain-size analysis. Sediment. Geol. 281, 88–100.
Yi, L., Yu, H., Xu, X., Liu, Y., Yao, J., Zhao, N., 2010. Influences of carbonate contents on the grain-size measurements of Borehole sediments from southern shore of Laizhou Bay (In Chinese with English abstract). Adv. Mar. Sci. 28, 325–331.
Yu, S.-Y., Colman, S.M., Li, L., 2016. BEMMA: A Hierarchical Bayesian End-MemberModeling Analysis of Sediment Grain-Size Distributions. Math. Geosci. 48, 723–741.
Zhang, X., Li, Z., Li, P., Cheng, S., Zhang, Y., Tang, S., Wang, T., 2015. A model to study the grain size components of the sediment deposited in aeolian–fluvial interplay erosion watershed. Sediment. Geol. 330, 132–140.