Computers & Geosciences 35 (2009) 1552–1556

Contents lists available at ScienceDirect

Computers & Geosciences

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

Windows-based software for optimising entropy-based groupings of textural data[☆]

Lachlan K. Stewart[a,b,*], Vladmir E. Kostylev[c], Alan R. Orpin[c,d]

[a] CSIRO Land and Water, Davies Laboratory, PMB Aitkenvale, Townsville Qld 4814, Australia
[b] School of Earth and Environmental Sciences, James Cook University, Townsville, Qld 4811, Australia
[c] Geological Survey of Canada (Atlantic), P.O. Box 1006, Dartmouth NS, Canada B2Y 4A2
[d] NIWA, Private Bag 14 901 Kilbirnie, Wellington, New Zealand

ARTICLE INFO

Article history:
Received 22 August 2007
Received in revised form
27 October 2008
Accepted 2 December 2008

Keywords:
Entropy
Polymodal
Distribution
EntrpoyMax
Statistical

ABSTRACT

EntropyMax is a new 32-bit Windows-based software that groups large matrices of grain-size distribution data into a finite number of groups. The software utilises statistical algorithms that minimise the entropy within a group while maximising the entropy between groups. EntropyMax builds upon an existing DOS–BASIC program through the addition of a more robust computational routine, a graphical user interface, incorporation of an improved and more statistically meaningful method for identifying the optimal number of groupings, and better graphical presentation of results. These refinements add significantly to the functionality and statistical rigour of the program on modern computer platforms and large data sets.

1. Introduction

The advent of computerised laser-based grain-size analysis facilitates the generation of very large sets of grain-size distributions data that, through their sheer size, are almost impossible to group by inspection. Sediment grain size is a key descriptor for a multitude of disciplines but commonly only one distribution-dependent granulometric parameter, such as mean grain size or sorting, is used to characterise the texture or correlation against other physical parameters. However, both in the geological record and many active depositional environments today, sediments often have bimodal or polymodal size distributions, placing significant limitations on these traditional distribution-dependent statistics.

Cognisant of the need for an automated system for classifying such distribution data into statistically meaningful groups Woolfe and Michibayashi (1995) developed "ENTROPY", a QBASIC language, DOS-based software tool that could be used to partition numerical data, such as grain-size distributions, into self-similar groups. They built on the initiative of Forrest and Clark (1989) who recognised the requirement of a statistically robust technique that efficiently dealt with the total information contained within the distribution curve. ENTROPY is based upon the technique of information entropy analysis for grouping data (Shannon and Weaver, 1963; Semple and Gollege, 1970; Semple et al., 1972; Johnston, 1978; Thomas, 1981; Johnston and Semple, 1983) and represents a QBasic adaptation of the Fortran code provided in Johnston and Semple (1983). As ENTROPY is a DOS-based program the user interface and processing efficiency is poor by today's standards. In addition, there are significant limitations that impact upon its usability and practicality, such as a limit of 999 on the maximum number of samples that may be analysed.

Various geoscientists have sought a statistical basis of identifying the optimal classification of a sample set into groups. A multivariate extension to an entropy-based approach overcomes the problem of pre-selecting the optimal number of class intervals common to all samples in the data set (see Full et al., 1983). Here, the multivariate technique will group samples in terms of the totality of their distributions, irrespective of the optimal number of intervals for particular samples. Unlike other grouping techniques, entropy analysis minimises the amount of within-group variance by testing for all possible groupings of samples. For the multivariate case described by Semple et al. (1972), with N size intervals and K samples, the total inequality statistic is given by Eq. (2)

$$I(Y) = \sum_{j=1}^{J} Y_j \sum_{i=1}^{N} Y_i \log_2 NY_i \tag{1}$$

[☆] Code available from www.iamg.org/CGEditor/index.htm.

[*] Corresponding author. Current address: CSIRO Land and Water, Davies Laboratory, PMB Aitkenvale, Townsville Qld 4814, Australia.
Tel.: +61 7 4753 8659; fax: +61 7 4753 8650.
E-mail address: Lachlan.Stewart@csiro.au (L.K. Stewart).

0098-3004/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2008.12.002

where Y_j is the frequency value (volume of grains in this case) in size intervals (column) j; J is the number of intervals; N is the number of samples; Y_i is the frequency value (of grains) in size interval (column) j that are in sample (row) i, such that $Y_i = Y_{ij}/Y_j$, where Y_{ij} is the proportion of the total population (of all K samples) in row i, column j. Then

$$\sum_{j=1}^{J} Y_j = 1.0 \quad \text{and} \quad \sum_{i=1}^{K} Y_i = 1.0 \tag{2}$$

$I(Y)$ is thus the inequality in the distribution of size intervals across all samples, weighted by the frequency of samples in each size category. Given the number of groups (r), derivation of the weighted inter-row inequality value allows the calculation of between-group inequality $I_B(Y)$ from

$$I_B(Y) = \sum_{j=1}^{J} X_j \sum_{r=1}^{M} Y_{jr} \log_2 \frac{Y_{jr}}{N_r/N} \tag{3}$$

where M is the maximum number of groups considered. The procedure optimises the classification of the N samples into r classes (groupings of samples) through maximisation of between-class and minimisation of within-class entropy. Following Forrest and Clark (1989), a statistically optimum grouping is attained when the growth rate of the between-group entropy significantly decreases with the addition of yet further groups. Woolfe et al. (1998a, b, 2000) adaptation of Semple et al. (1972) multivariate extension introduces an addition term, the Rs statistic, which is the percentage of the data set "explained", which is defined numerically as

$$Rs = \frac{\text{between group inequality}}{\text{total inequality}} \times 100 \tag{4}$$

Inequality measures are based on Shannon and Weaver (1963) information (entropy) index. Graphically, in a plot of the number of groups vs. the percentage of total entropy explained, the optimal number of groups is approximated by the point from which a consistent and dramatic reduction in slope is noted (e.g., Woolfe et al., 1998a, b, 2000).

General agreement with field observations of sediment textural facies has also been used as an additional semi-quantitative measure of the appropriate number of meaningful textural subdivisions (e.g., Woolfe et al., 1998a, 2000; Orpin et al., 1999, 2004; Hamilton, 2007). However, Legendre et al. (2002) in acoustic classification of sedimentary facies and Orpin and Kostylev (2006a) in grain-size analysis showed that the C–H criterion (after Calinski and Harabasz, 1974) provides a more reliable indication of the optimal number of groups that exist within a complete sample population.

While Rs represents a ratio of between-groups entropy to total entropy of the set (Forrest and Clark, 1989), the C–H criterion, C, represents the pseudo F-statistic of multivariate analysis of variance and canonical analysis (Legendre et al., 2002) and is defined by the ratio of the mean square for the given grouping divided by the mean square of the residuals (Eq. (2))

$$C = \frac{[R^2/(K-1)]}{[(1-R^2)/(n-K)]} \tag{5}$$

where n is the number of samples, K is the number of groups into which samples were clustered, and

$$R^2 = \frac{(SST - SSE)}{SST} \tag{6}$$

where SST is the total sum of the squared distances of the sample components to the overall centroid (their respective component means). SST is similar to the total inequality or the between groups sum of the squares. SSE is the sum of the squared distances of the objects to the groups own centroids i.e., within group sum of the squares (e.g., Fig. 1). It then follows, that the grouping in which the C is at a maximum, is the optimal grouping solution in terms of a least-squares (similar to pseudo-F test).

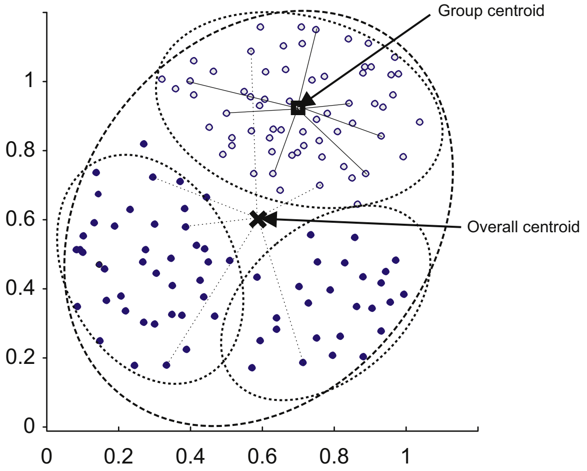


Fig. 1. Conceptual representation of overall and group centroids, points are for illustration only.

It is important to note that in the case of grain-size data each object (sample) is represented by the set of relative abundances of grain-size classes (components) describing the sample (i.e., the data of interest is compositional). Tolosana-Delgado et al. (2005) emphasise that care must be taken in the calculation of distances between samples (vectors) of compositional data because the formula for Euclidean distance is not appropriate for such data. Within the C–H methodology, the SST and SSE measures do not represent the sum of squared distances between samples (compositions) and the mean composition, but the cumulative, squared difference between sample components and their respective (overall or group) component means. The addition, subtraction and multiplication of vector components is treated identically in compositional-data space (simplex) as it is in Euclidean space (Tolosana-Delgado et al., 2005), therefore, such a measure is appropriate for use with compositional data.

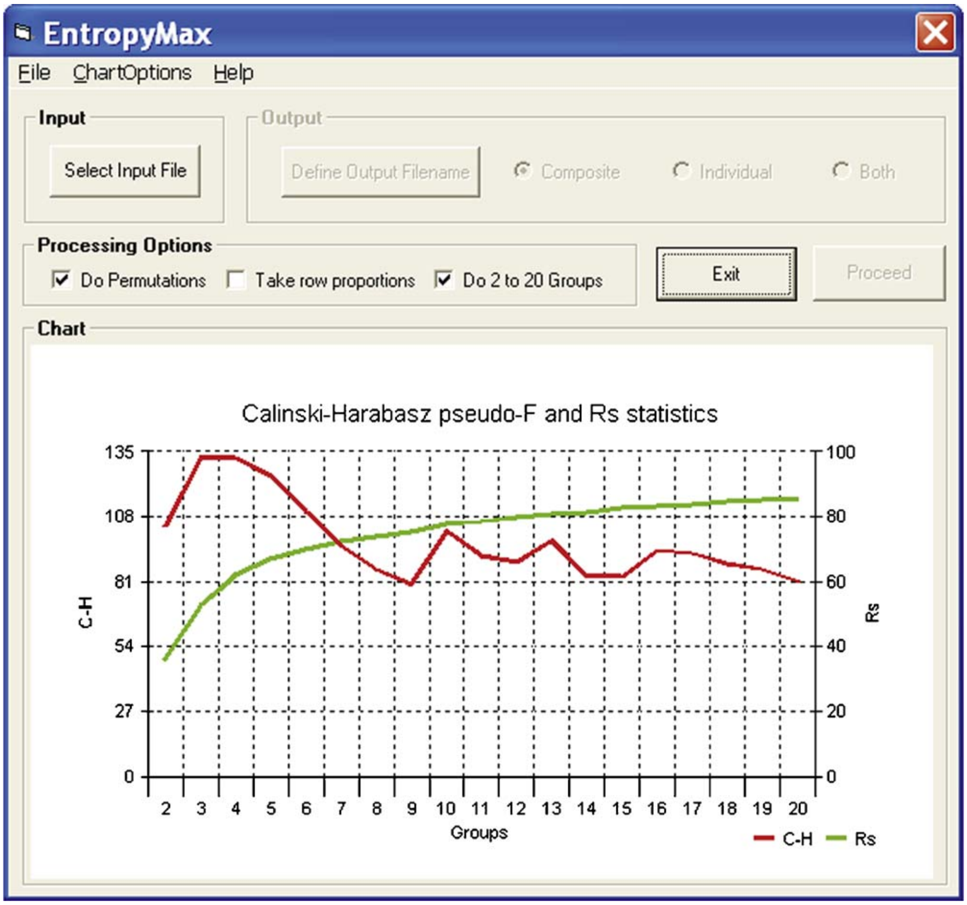
An advantage the C–H statistic has over the Rs statistic is that the optimal number of groups is more easily obtained via inspection of graphical results (e.g., Fig. 2). The optimum grouping indicated by the Rs statistic, which asymptotically grows until between-groups entropy equals the total, is not always immediately apparent.

No significance values are associated with the C–H statistics to and neither can they be produced in the partitioning procedure (Legendre and Legendre, 1998). Therefore, some debate remains as to the most appropriate and statistically valid criterion for determining the optimal number of groups into which the primary data set may be subdivided (e.g., Orpin and Kostylev, 2006a, b).

Herein, we describe the development of EntropyMax a Windows based, Visual Basic 6, software tool that represents an updated and improved implementation of ENTROPY, with the addition of a new tool, the C–H criterion and visual aids for assessing the optimal number of groups.

2. Graphical user interface (GUI) and code implementation

Core algorithms were implemented within EntropyMax through conversion of the original ENTROPY QBasic code into



● Composite ○ Individual ○ Both

☑ Do Permutations □ Take row proportions ☑ Do 2 to 20 Groups

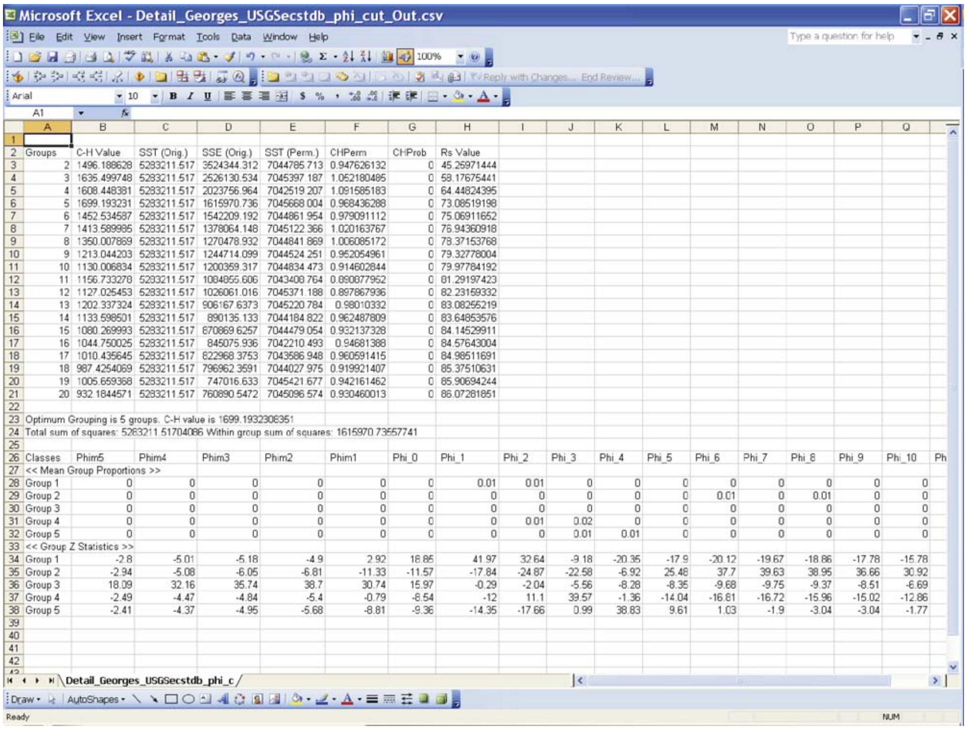
Fig. 2. EntropyMax GUI graphical display of values of *Rs* and C–H statistics for a set of 296, 32 class, seabed samples obtained adjacent to Herbert River delta, Queensland, Australia. Both traces suggest an optimal grouping of 3 or 4 groups (data from Woolfe et al., 2000).

Visual Basic 6. The implementation of the C–H criterion was achieved through the conversion of Orpin and Kostylev's (2006a) QBasic algorithms.

Critical to improving the functionality of this software was a modern graphical user interface (GUI) designed to provide ease of data input/output, the specification of the parameters of analysis, and the viewing of results (both on-the-fly and final result summary). Fig. 2 shows the primary form of the GUI that is designed for convenience and ready access to functionality options.

Within ENTROPY the "percentage explained" (*Rs*) statistic (Woolfe et al., 1998a, 2000) provides a means of identifying the optimal number of groups into which a sample population can be subdivided. In its original form the software did not collate or graphically present the *Rs* statistical data despite its importance to the grouping procedure. Therefore, collation and graphical presentation of this, and the C–H criterion, was implemented within EntropyMax. Data pertaining to both criteria are now available in a "Detail_*filename*.csv" output file, where *filename* is the name specified by the user for output files. Detail_*filename*.csv contains tabulated values of *Rs* and C–H values plus other statistical information (e.g., Fig. 3). In order to ensure that the groups are clustered based on the similarity of the shapes of grain-size frequency distributions rather than affinity of their average values, a permutation test was designed by Orpin and Kostylev (2006a) for evaluating the likelihood of the C–H statistic being greater than that expected from a random allocation of the existing data into grain-size classes. To create a null model, the average C–H value ("CHPerm") is calculated by randomly permutating values of grain-size classes within each sample, keeping samples in the originally distinguished clusters. To test the null model, the number of times the C–H statistic produced by each random permutation larger or equal to the original C–H value for that classification is calculated. The fraction "CHProb" of the number of permutated C–H values which are higher than the original C–H value is interpreted as the probability of the grain-size distribution shapes in each cluster being random. CHProb can be interpreted similarly to *p* values in testing the significance of statistical tests, for example if the CHProb is 1, then even the C–H maximum in this classification is not indicative of optimal grouping of the samples. One hundred permutations are performed in the calculation of CHProb, and grain-size classes are randomly interchanged within all samples with each permutation. As a result, with each permutation the position of the overall centroid and the corresponding *SST* will change. For completeness *SST* values from both the original and permutated data arrays are listed in the output table.

To enhance functionality of the software and expedite access to the graphical presentation of statistical outcomes, a significant enhancement in EntropyMax is a graphical display pane located at the bottom of the primary form. Within this pane values of the *Rs* and C–H criterion are plotted dynamically as analysis progresses for each successive increase in the number of groups. Upon completion of the analysis the user is able to immediately identify the optimal number of groups that exist within the sample population, identified as the peak in the value of C–H and the reduced rate of change in *Rs* with increasing group numbers (e.g., at 3 or 4 groups in Fig. 2).


File Edit View Insert Format Tools Data Window Help
Type a question for help
100%
Reply with Changes... End Review...
Arial
10
A1
Draw
AutoShapes
Ready
NUM

Microsoft Excel - Detail_Georges_USGSecstdb_phi_cut_Out.csv

Groups	C-H Value	SST (Orig.)	SSE (Orig.)	SST (Perm.)	CHPerm	CHProb	Rs Value
2	1496.188628	5283211.517	3524344.312	7044785.713	0.947626132	0	45.25971444
3	1635.499748	5283211.517	2526130.534	7045397.187	1.052180485	0	58.17675441
4	1608.448381	5283211.517	2023756.964	7042519.207	1.091585183	0	64.44824395
5	1699.193231	5283211.517	1615970.736	7045668.004	0.968436288	0	73.08519198
6	1452.534587	5283211.517	1542209.192	7044861.954	0.979091112	0	75.06911652
7	1413.589985	5283211.517	1378064.148	7045122.366	1.020163767	0	76.94360918
8	1350.007869	5283211.517	1270478.932	7044841.869	1.006085172	0	78.37153768
9	1213.044203	5283211.517	1244714.099	7044524.251	0.952054961	0	79.32778004
10	1130.006834	5283211.517	1200359.317	7044834.473	0.914602844	0	79.97784192
11	1156.733278	5283211.517	1084855.606	7043408.764	0.890877952	0	81.29197423
12	1127.025453	5283211.517	1026061.016	7045371.188	0.897867936	0	82.23159332
13	1202.337324	5283211.517	906167.6373	7045220.784	0.98010332	0	83.08255219
14	1133.598501	5283211.517	890135.133	7044184.822	0.962487809	0	83.64853576
15	1080.269993	5283211.517	870869.6257	7044479.054	0.932137328	0	84.14529911
16	1044.750025	5283211.517	845075.936	7042210.493	0.94681388	0	84.57643004
17	1010.435645	5283211.517	822968.3753	7043586.948	0.960591415	0	84.98511691
18	987.4254069	5283211.517	796962.3591	7044027.975	0.919921407	0	85.37510631
19	1005.659368	5283211.517	747016.633	7045421.677	0.942161462	0	85.90694244
20	932.1844571	5283211.517	760890.5472	7045096.574	0.930460013	0	86.07281851

Optimum Grouping is 5 groups. C-H value is 1699.1932308351

Total sum of squares: 5283211.51704086 Within group sum of squares: 1615970.73557741

Classes	Phim5	Phim4	Phim3	Phim2	Phim1	Phi_0	Phi_1	Phi_2	Phi_3	Phi_4	Phi_5	Phi_6	Phi_7	Phi_8	Phi_9	Phi_10	Ph
<< Mean Group Proportions >>																	
Group 1	0	0	0	0	0	0	0.01	0.01	0	0	0	0	0	0	0	0	
Group 2	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.01	0	0	
Group 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Group 4	0	0	0	0	0	0	0	0.01	0.02	0	0	0	0	0	0	0	
Group 5	0	0	0	0	0	0	0	0	0.01	0.01	0	0	0	0	0	0	
<< Group Z Statistics >>																	
Group 1	-2.8	-5.01	-5.18	-4.9	2.92	18.85	41.97	32.64	-9.18	-20.35	-17.9	-20.12	-19.67	-18.86	-17.78	-15.78	
Group 2	-2.94	-5.08	-6.05	-6.81	-11.33	-11.57	-17.84	-24.87	-22.58	-6.92	25.48	37.7	39.63	38.95	36.66	30.92	
Group 3	18.09	32.16	35.74	38.7	30.74	15.97	-0.29	-2.04	-5.56	-8.28	-8.35	-9.68	-9.75	-9.37	-8.51	-6.69	
Group 4	-2.49	-4.47	-4.84	-5.4	-0.79	-8.54	-12	11.1	39.57	-1.36	-14.04	-16.81	-16.72	-15.96	-15.02	-12.86	
Group 5	-2.41	-4.37	-4.95	-5.68	-8.81	-9.36	-14.35	-17.66	0.99	38.83	9.61	1.03	-1.9	-3.04	-3.04	-1.77	

Detail_Georges_USGSecstdb_phi_c

Fig. 3. Tabulated output file showing details of groupings and statistical analyses. Data are a 3000 sample subset of 17-class seabed grain sizes from Georges Bank and Gulf of Maine (Poppe et al., 2003).

This feature presents two alternatives to help identify the optimal statistical grouping of samples, which can now be compared graphically. Commonly the optima are the same, or very similar. Field-based research suggests that sedimentary facies relationships should also influence the selection of the most appropriate number of groups (e.g., Woolfe and Michibayashi, 1995; Woolfe et al., 1998a, b, 2000; Orpin et al., 1999, 2004), emphasising that discrete groupings of similar sediment type might be recognised by other criteria even with statistically limited sample populations in some cases. Nonetheless, the seamless incorporation of the *Rs* and C–H statistics now enables the user to more readily identify a statistically based optimal grouping.

Input and analysis-related options remain the same as for ENTROPY. However, output options have been extended to allow full control over the scope of file-based output. For example, as data are successively distilled into different numbers of groups, analyses are available as three outputs: (a) a composite file, "*filename*.csv", where data related to all groupings is represented; (b) individual files "#*filename*" ("#" represents the number of groups) where each file represents the data corresponding to a particular number of groups; or (c) both. C–H values associated with each grouping are also represented in the files.

Other code improvements include an increase in the allowable size of input datasets. Previously, datasets were limited to 999 samples. Under EntropyMax data sets of up to 32,767 rows may be processed. It should be noted that data sets of this size are time consuming (>day) to process.

3. Other features

Other features of the software include:

- The ability to save to the clipboard the graphical image displayed within the main form. This allows the graphics to be conveniently exported to a third party application.
- Help files describing the operation of the software.

4. Examples of application

The application of the earlier MS-DOS-based version of ENTROPY is now well established across a wide range of sedimentary environments. Examples of mixed carbonate–siliciclastic sediments from tropical Australia (e.g., Woolfe et al., 1998a, 2000; Orpin et al., 1999, 2004), periglacial environments from Antarctica (Woolfe et al., 1998b), lakes (Woolfe, 1995), and glaciated shelves (Orpin and Kostylev, 2006a). At these locations the ability to recognise discrete groups with similar granulometric properties assisted the delineation of distinct sedimentary facies and, by inference, distinct depositional environments. This is particularly true for environments with poorly sorted, multimodal sediments. The example shown in Fig. 4 from the Herbert River delta yielded four mappable sediment facies on the inner shelf, which can be correlated to local hydrodynamics and sediment supply (Woolfe et al., 2000).

Provisional analysis of grain-size data from other shallow shelf environments suggest that the amplitude of the C–H statistic has a positive relationship to the number of samples and a negative relationship with the number of classes within each sample (Orpin and Kostylev, 2006a). In essence, data order increases upon the addition of more samples and a reduction in resolution (number degrees of freedom). An outstanding question is how much textural information is required to obtain a meaningful result and how best to contend with natural variability, which can be continuous with no clear discontinuities and identifiable as groups or clusters in the data (e.g., Orpin and Kostylev, 2006b). The goal of clustering is to discover group structure in the data and whether discovered groups are meaningful (Zhao and Karypis,

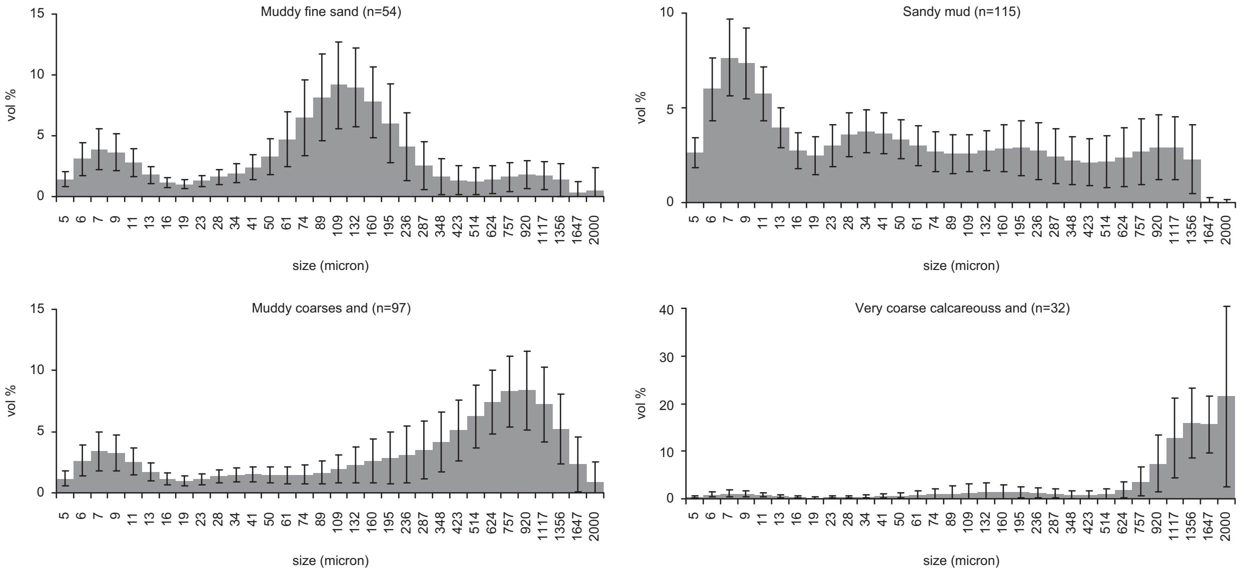


Fig. 4. (1) Example of ENTROPY-derived textural facies adjacent to Herbert River delta, Queensland, Australia (modified after Woolfe et al., 2000). Within-class variability is indicated by error-bars of one standard deviation from each class mean, and *n* denotes sample population within each group. Four sediment facies have been delineated: (1) a trimodal very sandy mud or muddy sand, typically containing a moderately sorted medium- to fine-grained sand mode (mostly quartz), combined with subordinate amounts of mixed terrigenous and calcareous, coarse-grained sand and terrigenous mud; (2) a very poorly sorted sandy mud. Sediments belonging to this group typically contain a primary mode finer than 20 μm, with a subordinate fine sand or silt mode and a calcareous coarse-sand mode; (3) muddy, medium- to coarse-grained terrigenous and/or calcareous sand; and, (4) a clean, calcareous coarse sand and/or fine gravel. Facies composition and distribution of these continental shelf sediments has been shown related to hydrodynamics and proximity to riverine sediment supply (Woolfe et al., 2000).

2005). In this regard the addition of the C–H criterion adds usefully to the utility of ENTROPY.

Acknowledgements

The authors gratefully acknowledge Dr James Lally for providing access to the original Entropy QBasic code. The reviewers are thanked for their helpful suggestions.

Appendix A. Supporting data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.cageo.2008.12.002.

References

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics 3, 1–27.

Forrest, J., Clark, N.R., 1989. Characterizing grain size distributions: evaluation of a new approach using a multivariate extension of entropy analysis. Sedimentology 36, 711–722.

Full, W.E., Ehrlich, R., Kennedy, S.K., 1983. Optimal definition of class intervals from frequency tables. Particulate Science and Technology 1, 281–293.

Hamilton, L.J., 2007. Clustering of cumulative grainsize distribution curves for shallow-marine samples with software program CLARA. Australian Journal of Earth Sciences 54, 503–519.

Johnston, R.J., 1978. Multivariate Statistical Analysis in Geography: A Primer on the General Linear Model. Longmans, London, 280pp.

Johnston, R.J., Semple, R.K., 1983. Classification Using Information Statistics: Concepts and Techniques in Modern Geography No. 37. Geobooks, Norwich, 46pp.

Legendre, P., Ellingsen, K.E., Bjørnbom, E., Casgrain, P., 2002. Acoustic seabed classification: improved statistical method. Canadian Journal of Fisheries and Aquatic Sciences 59, 1085–1089.

Legendre, P., Legendre, L., 1998. Numerical Ecology, second English ed. Elsevier, Amsterdam, 853pp.

Orpin, A.R., Brunskill, G.J., Zagorskis, I., Woolfe, K.J., 2004. Patterns of mixed siliciclastic-carbonate sedimentation adjacent to a large dry-tropics river on the central Great Barrier Reef shelf, Australia. Australian Journal of Earth Sciences 51, 665–683.

Orpin, A.R., Haig, D.W., Woolfe, K.J., 1999. Sedimentary and foraminiferal facies in Exmouth Gulf, arid tropical northwestern Australia. Australian Journal of Earth Sciences 46, 607–621.

Orpin, A.R., Kostylev, V.E., 2006a. Towards a statistically valid method of textural sea floor characterization of benthic habitats. Marine Geology 225, 209–222.

Orpin, A.R., Kostylev, V.E., 2006b. Reply to L.J. Hamilton's comment regarding Orpin and Kostylev, 2006 – towards a statistically valid method of textural sea floor characterization of benthic habitats. Marine Geology 232, 111–113.

Poppe, L.J., Paskevich, V.F., Williams, S.J., Hastings, M.E., Kelly, J.T., Belknap, D.F., Ward, L.G., FitzGerald, D.M., Larsen, P.F., 2003. Surficial sediment data from the Gulf of Maine, Georges Bank, and vicinity: a GIS compilation. US Geological Survey Open-File Report 03-001, CD-Rom. United States of America Department of the Interior.

Semple, R.K., Youngman, C.E., Zeller, R.E., 1972. Economic Regionalization and Information Theory: An Ohio Example. Discussion Paper 28. Department of Geography, Ohio State University, Columbus, USA, 64 pp.

Semple, R.K., Gollege, R.G., 1970. An analysis of entropy changes in a settlement pattern over time. Economic Geography 46, 157–160.

Shannon, C.E., Weaver, W. (Eds.), 1963. The Mathematical Theory of Communication. University of Illinois Press, Urbana, IL, 144 pp.

Thomas, R.W., 1981. Information statistics in geography. Concepts and Techniques in Modern Geography 31, 1–40.

Tolosana-Delgado, R., Otero, N., Pawlowsky-Glahn, V., 2005. Some basic concepts of compositional geometry. Mathematical Geology 37, 673–680.

Woolfe, K.J., 1995. Textural entropy groupings from a modern lake–lagoon system and its ancient analogue. New Zealand Journal of Geology and Geophysics 38, 259–262.

Woolfe, K.J., Fielding, C.R., Howe, J.A., Lavelle, M., Lally, J.H., 1998a. Laser-derived particle size characterisation of CRP-1, McMurdo Sound, Antarctica. Terra Antarctica 5, 383–391.

Woolfe, K.J., Larcombe, P., Orpin, A.R., Purdon, R.G., Michaelsen, P., McIntyre, C.M., Amjad, N., 1998b. Controls upon inner-shelf sedimentation, Cape York Peninsula, in the region of 12°S. Australian Journal of Earth Sciences 45, 611–621.

Woolfe, K.J., Larcombe, P., Stewart, L.K., 2000. Shelf sediments adjacent to the Herbert River delta, Great Barrier Reef, Australia. Australian Journal of Earth Sciences 47, 301–308.

Woolfe, K.J., Michibayashi, K., 1995. BASIC entropy grouping of laser-derived grain size data: an example from the Great Barrier Reef. Computers & Geosciences 21, 447–462.

Zhao, Y., Karypis, G., 2005. Data clustering in life sciences. Molecular Biotechnology 31, 55–80.