Sedimentology (2007) 1–24 doi: 10.1111/j.1365-3091.2007.00863.x

Turbidite bed thickness distributions: methods and pitfalls of analysis and modelling

ZOLTÁN SYLVESTER[1]
Department of Geological and Environmental Sciences, Stanford University, Stanford, CA 94305-2115, USA (E-mail: zoltan.sylvester@shell.com)

ABSTRACT

Turbidite bed thickness distributions are often interpreted in terms of power laws, even when there are significant departures from a single straight line on a log–log exceedence probability plot. Alternatively, these distributions have been described by a lognormal mixture model. Statistical methods used to analyse and distinguish the two models (power law and lognormal mixture) are presented here. In addition, the shortcomings of some frequently applied techniques are discussed, using a new data set from the Tarcău Sandstone of the East Carpathians, Romania, and published data from the Marnoso-Arenacea Formation of Italy. Log–log exceedence plots and least squares fitting by themselves are inappropriate tools for the analysis of bed thickness distributions; they must be accompanied by the assessment of other types of diagrams (cumulative probability, histogram of log-transformed values, q–q plots) and the use of a measure of goodness-of-fit other than R^2, such as the chi-square or the Kolmogorov–Smirnov statistics. When interpreting data that do not follow a single straight line on a log–log exceedence plot, it is important to take into account that 'segmented' power laws are not simple mixtures of power law populations with arbitrary parameters. Although a simple model of flow confinement does result in segmented plots at the centre of a basin, the segmented shape of the exceedence curve breaks down as the sampling location moves away from the basin centre. The lognormal mixture model is a sedimentologically intuitive alternative to the power law distribution. The expectation–maximization algorithm can be used to estimate the parameters and thus to model lognormal bed thickness mixtures. Taking into account these observations, the bed thickness data from the Tarcău Sandstone are best described by a lognormal mixture model with two components. Compared with the Marnoso-Arenacea Formation, in which bed thicknesses of thin beds have a larger variability than thicknesses of the thicker beds, the thinner-bedded population of the Tarcău Sandstone has a lower variability than the thicker-bedded population. Such differences might reflect contrasting depositional settings, such as the difference between channel levées and basin plains.

Keywords bed thickness, turbidites, power law distribution, lognormal distribution, log–log plots, lognormal mixture, Tarcău Sandstone, Marnoso-Arenacea Formation.

[1]Present address: Shell International Exploration and Production Inc., 3737 Bellaire Blvd., P.O. Box 481, Houston, TX 77001-0481, USA.

INTRODUCTION

One of the most striking features of turbidite sequences is the rhythmic alternation of sand and shale over great thicknesses. This is due to the fact that in deep-water settings coarse clastic sediment deposition is dominated by discrete sedimentation events and the background energy levels are low and, as a result, individual sedimentation events are often well preserved. The frequent interbedding of sand and shale strongly influences the overall heterogeneity of the resulting succession and the distribution of turbidite bed thicknesses along with the lateral thickness changes are important in modelling hydrocarbon reservoirs that were deposited in deep water (e.g. Flint & Bryant, 1993; Drinkwater & Pickering, 2001; Tye, 2004; Pyrcz *et al.*, 2005). A precise statistical description, along with appropriate graphical and computational tools for the analysis of this basic stratigraphic parameter, is essential for further advancements in the field. In addition, some parameters of the bed thickness distribution might prove to be useful in differentiating depositional settings (Malinverno, 1997; Carlson & Grotzinger, 2001; Mattern, 2002; Sinclair & Cowie, 2003; Clark & Steel, 2006), even when working with data of limited lateral extent, such as wells and smaller outcrops.

Four different types of statistical distributions have been proposed to describe sedimentary bed thickness data: truncated Gaussian, lognormal, exponential and power law (or fractal) distributions. According to an initial line of thought, the right-skewedness that seems to be ubiquitous in thickness data is the result of a truncated Gaussian distribution, reflecting a balance between deposition and erosion (Kolmogorov, 1951; Mizutani & Hattori, 1972; Muto, 1995). Muto (1995) has shown how Kolmogorov's truncated Gaussian model might lead to exponential distributions. Several researchers have suggested that the turbidite sequences are best characterized by lognormal distributions (e.g. McBride, 1962; Enos, 1969; Ricci Lucchi & Valmori, 1980; Murray *et al.*, 1996; Talling, 2001). Other studies have focused on whether *lognormal* or *exponential* bed thickness distributions dominate the geologic record. Drummond & Wilkinson (1996) suggested that the exponential distribution might describe sedimentary thicknesses over a wide range of scales, but the measurements are biased against very thin beds. Of late, in parallel with increasing interest in fractal phenomena and power law scaling in nature (e.g. Mandelbrot, 1983; Turcotte & Huang, 1995; Turcotte, 1997), the *power law* distribution has gained popularity in studies of turbidite sequences (Hiscott *et al.*, 1992, 1993; Rothman *et al.*, 1994; Malinverno, 1997; Pirmez *et al.*, 1997; Chen & Hiscott, 1999; Winkler & Gawenda, 1999; Awadallah *et al.*, 2001; Carlson & Grotzinger, 2001; Chakraborty *et al.*, 2002; Sinclair & Cowie, 2003).

Power law bed thickness distributions have been linked to earthquakes with comparably distributed magnitudes (Beattie & Dade, 1996; Awadallah *et al.*, 2001) and to self-organized criticality (Rothman *et al.*, 1994; Bak, 1996). The theory of self-organized criticality was introduced by Bak *et al.* (1988). This theory explains the widespread occurrence of fractal structures and $1/f$ noise by the tendency of large dissipative systems to develop a state of criticality and generate events of all sizes. The simplest model for such a system is a sand pile: once it reaches a critical slope, avalanches of all sizes will be generated. Obviously, turbidity current-initiating mechanisms are diverse and probably more complicated than sand piles (e.g. Normark & Piper, 1991). Not much is known about the size distribution of turbidity current initiating events, but the magnitudes of such events are likely to have a skewed distribution and a heavy-tailed distribution is not unreasonable. Whether this distribution is power law, exponential, lognormal or best fits some other model cannot be determined based on the available data. In addition, hundreds of kilometres might separate a slide on the continental slope, canyon wall or delta front and the final site of deposition and, as Talling (2001) has argued, 'it is unlikely that there is such a simple correspondence between the frequency distribution of sediment volumes failing on the continental slope, or the magnitude of seismic shaking, and the thickness of turbidites deposited much further downslope'. Well-defined power law distributions are relatively rare; most turbidite bed thickness data sets show significant departures from the ideal power law model. However, this has not prevented the power law model from being applied. Rather, significant departures from a single power law have been called 'segmented power law distributions' and interpreted as the results of modification of the initial power law distributed input volumes by erosion, amalgamation, confinement or a combination of these (Malinverno, 1997; Winkler & Gawenda, 1999; Carlson & Grotzinger, 2001; Sinclair & Cowie, 2003).

Talling (2001) has analysed bed thickness data from the Marnoso-Arenacea Formation in Italy

and from three other locations and has shown that turbidites with specific basal grain size classes have lognormal distributions. Therefore, sequences that consist only of thick-bedded turbidites or of thin-bedded turbidites tend to fit lognormal models. Distributions that plot as convex-up curves on log–log plots can result from the mixing of lognormal populations, and there are a number of sedimentological reasons why the lognormal mixture model is preferable to the segmented power law model. The results presented here build on and support these findings and the reader is referred to Talling (2001) for a detailed discussion of the origin and the arguments in favour of the lognormal mixture model. This paper focuses on the problems and potential solutions of deciding which distribution model best fits the data, by: (i) presenting the statistical methodology and the associated pitfalls in analysing bed thickness distributions; (ii) re-examining the idea of 'segmented power law distributions' from a statistical point of view; (iii) suggesting a methodology for estimating and modelling the two main components of a lognormal mixture; and (iv) applying these techniques to a turbidite succession and comparing the results with the Marnoso-Arenacea Formation. It is not the purpose of this paper to review all the possible distribution models or to cover a large number of turbidite bed thickness data sets. Rather, the focus is on the power law and lognormal models, using the bed thickness data from the Tarcău Sandstone as a case study.

GEOLOGIC SETTING OF THE TARCĂU SANDSTONE

The Tarcău Sandstone is one of the major sand-rich formations of the flysch belt of the Romanian East Carpathians. It represents a turbidite system deposited during the Palaeocene to Middle Eocene (Săndulescu & Săndulescu, 1973) between the active margin of a small continental block to the west (Fig. 1), termed the Tisza-Dacia Block (Csontos, 1995), and the passive margin of the East European Craton to the east. Palaeocurrent data suggest a longitudinal dispersal pattern (Contescu *et al.*, 1967; Jipa, 1967). A microfauna consisting mainly of agglutinant foraminifera characteristic of deep-water flysch assemblages (Săndulescu & Săndulescu, 1973) and an abundant ichnofauna typical of the Nereites ichnofacies (Buatois *et al.*, 2001) suggest deposition at bathyal water depths. Predominant lithologies are thick-bedded pebbly sandstones and medium-bedded to thin-bedded sandstones and shales (Figs 2 and 3), largely corresponding to the Bouma (1962) and Lowe (1982) turbidites. The presence of abundant amalgamation, erosion, debris flows and conglomerate beds within the thicker-bedded intervals suggests the presence of channels (Sylvester, 2002), and most of the thin-bedded intervals are likely to represent overbank deposits. The Tarcău Sandstone shows a number of important differences when compared with the well-studied Marnoso-Arenacea Formation: (i) it has numerous coarser-grained packages (not only pebbly sandstones, mainly, but also conglomerates); (ii) many of these coarse-grained deposits probably represent channel fills, whereas most of the Marnoso-Arenacea is interpreted as basin plain deposits; and (iii) the lateral extent of beds in the Tarcău Sandstone is limited and correlation from one outcrop to the next is impossible, whereas numerous beds in the Marnoso-Arenacea can be correlated across the whole basin (Ricci Lucchi & Valmori, 1980; Amy & Talling, 2006). Comparing bed thickness data from these two successions might give insights into the differences between basin plain deposits and channel levée systems.

FIELD METHODS

The Tarcău Sandstone is exposed in a series of roadcuts near the Siriu Dam in the Buzău Valley area of the East Carpathians (Fig. 1). More than 700 m of stratigraphic sections were measured (Fig. 2; Sylvester, 2002) and thicknesses of individual Bouma divisions and other characteristics, such as maximum grain size, presence of erosion, mud clasts and sedimentary structures were noted. The overall finer-grained and mainly thinner-bedded upper part of the Tarcău Sandstone is informally referred to below as the Upper Tarcău Sandstone and the bed thicknesses of this section have also been analysed separately (excluding the uppermost thick-bedded pebbly sandstone package; Fig. 2).

Precise measurement of bed thicknesses is only possible where abundant sandstone beds are present in the succession. Some relatively poorly exposed, predominantly shaly sections were therefore not included in the study. For the purpose of this paper, 'bed thickness' is equivalent to 'sedimentation unit thickness'. In parts of the succession where individual events consist of a sandstone–shale pair, the combined thickness of the two represents one value in the bed thickness

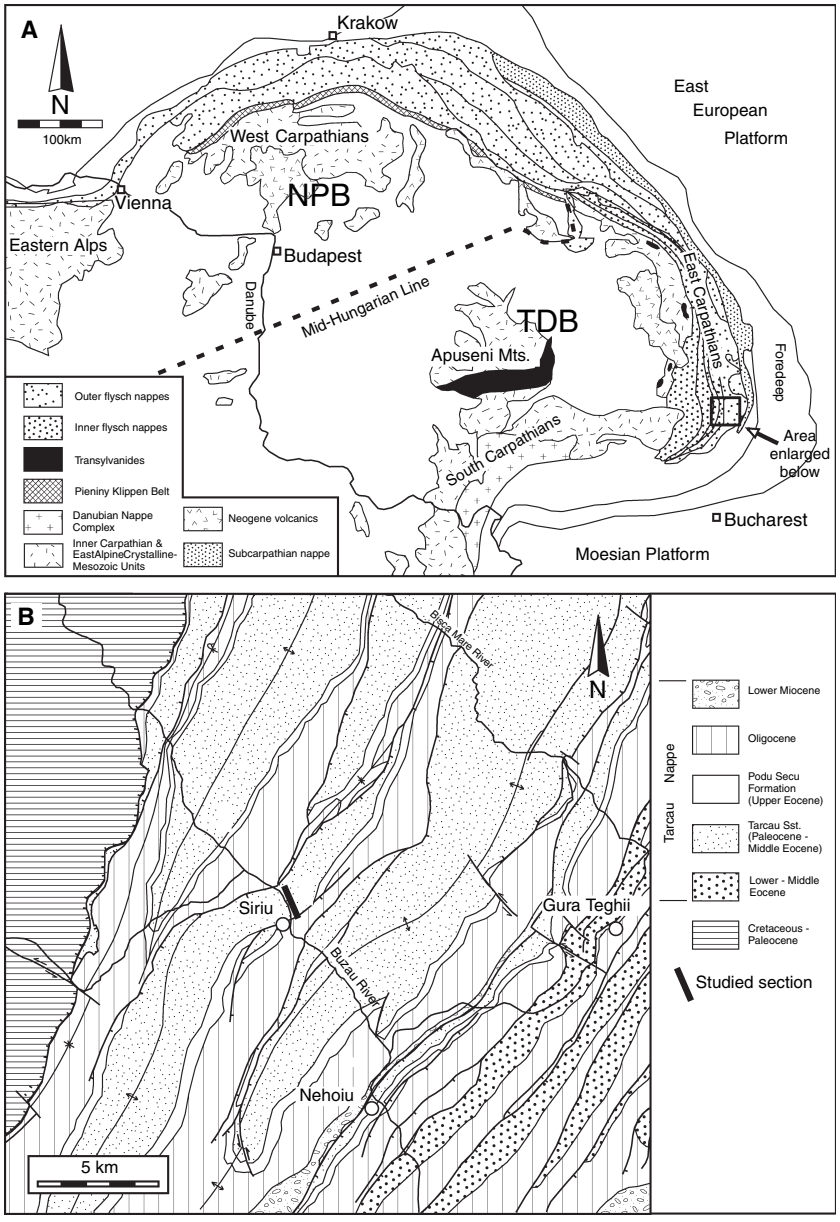

Fig. 1. (A) Geologic sketch of the Carpathians, based on Săndulescu (1988) and Zweigel *et al.* (1998) (TDB, Tisza-Dacia Block; NPB, North Pannonian Block). (B) Geologic map of the Tarcău Nappe in the Buzău Valley area, simplified from Murgeanu *et al.* (1968), with location of measured sections where bed thickness data were derived.

series (Fig. 4). Working with sedimentation unit thicknesses as opposed to treating sandstone and mudstone layers separately is reasonable if one of the underlying assumptions of the analysis is that bed thicknesses have a clear relationship with individual event bed volumes (e.g. Malinverno, 1997). In addition, if viewed separately, sandstone and mudstone thicknesses still have the signatures of mixtures (see below), suggesting that analysing event bed thicknesses, grouped as a function of their basal Bouma divisions or grain size (Talling, 2001), is a better approach.

In coarser-grained and thicker-bedded parts of the succession, where amalgamation is frequent, boundaries of individual event beds were carefully identified (Fig. 4). This was possible mainly by detecting subtle surfaces across which well-defined coarsening occurs. Breaking down amalgamated sections into individual event beds (see also Sinclair & Cowie, 2003) is in contrast with the lumping approach adopted by Carlson & Grotzinger (2001). Note that bed thickness distributions of the thick-bedded parts of the Tarcău Sandstone would be entirely different if the amalgamation surfaces were ignored. Working with event bed thicknesses makes it possible to study how bed thickness relates to bed volume and depositional processes, whereas the

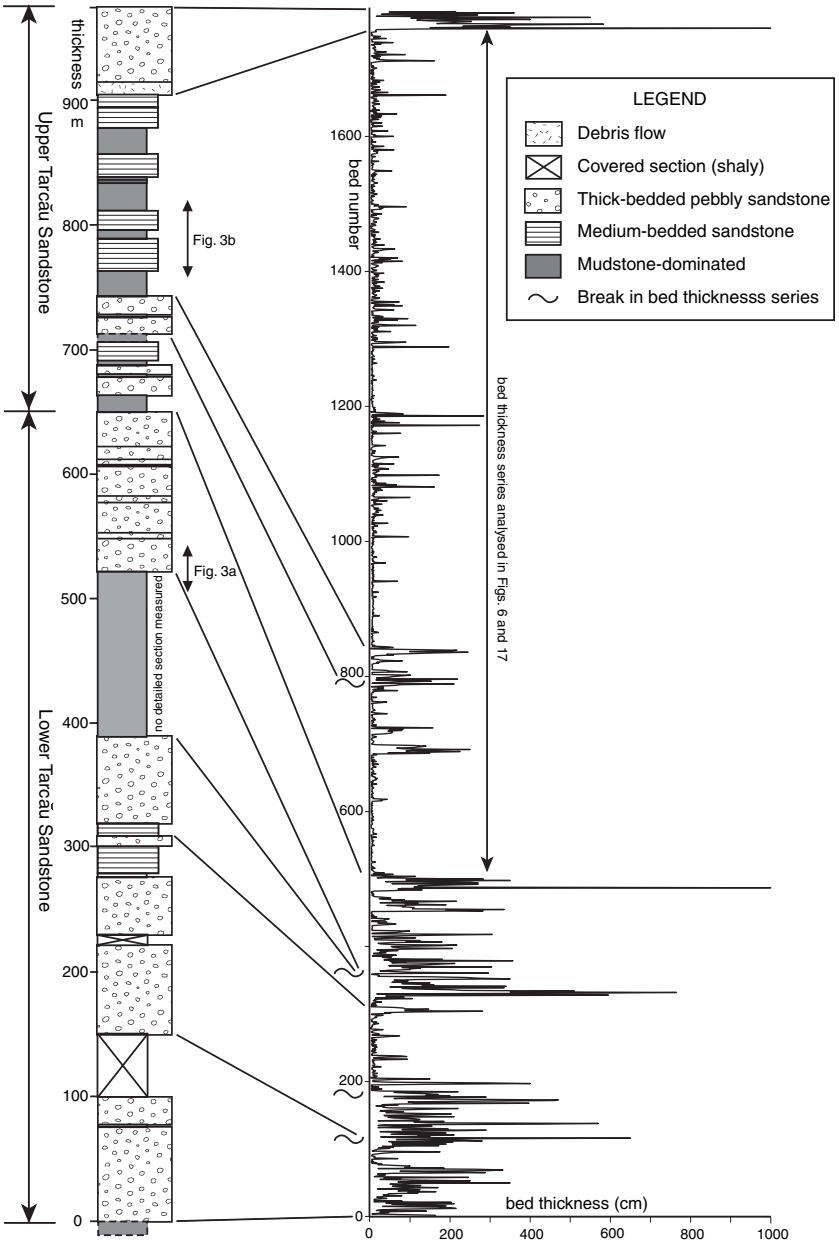

Fig. 2. Schematic stratigraphic column of the Tarcău Sandstone near Siriu Dam, Buzău Valley, with the measured bed thickness series shown on the right.

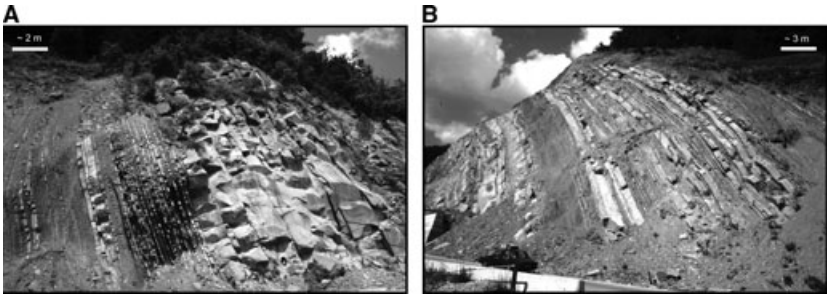

Fig. 3. (A) Thin-bedded, fine-grained turbidite package overlain by a thick-bedded, amalgamated pebbly sandstone unit. (B) Two medium-bedded, laminated sandstone packages separated by a thin-bedded, mudstone dominated unit. For the stratigraphic position of these sections, see Fig. 2.

thicknesses of amalgamated units convolve the effects of bed volume, depositional process *and* facies clustering.

No attempt has been made to separate turbidite mud from hemipelagic shale. However, true hemipelagic or pelagic deposits seem to be rare

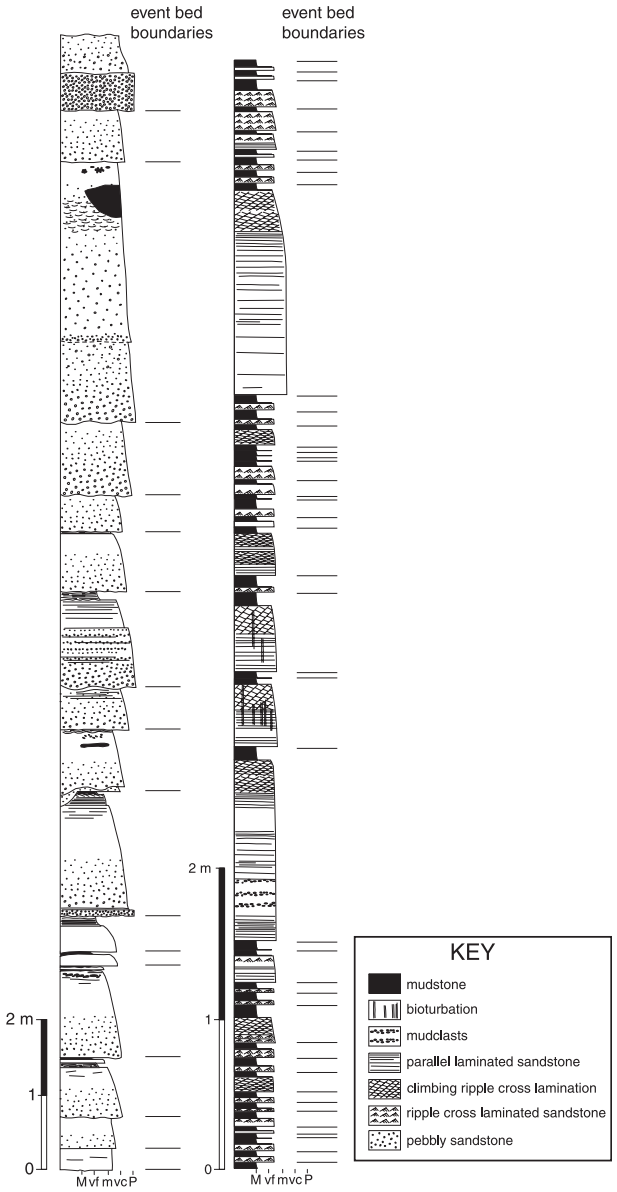

Fig. 4. Examples of measured sections from a thick-bedded, amalgamated pebbly sandstone unit and a medium-bedded, laminated sandstone unit. Event bed boundaries, used in most of the bed thickness calculations, are shown. Note different vertical scales.

in the succession and mainly present in the middle of shale-dominated sections. These deposits are unlikely to strongly affect the measured bed thickness distributions.

POWER LAW BED THICKNESS DISTRIBUTIONS

A power law distribution, called Pareto distribution by economists, is characterized by an exceedence probability (i.e. the probability that a value is greater than or equal to x) described by a power law:

$$P[X \geq x] = \left(\frac{m}{x}\right)^{\beta} \quad (1);$$

where m is the minimum value of x. The exceedence probability is often called complementary cumulative distribution function (CCDF). The cumulative distribution function (CDF) is

$$P[X < x] = 1 - \left(\frac{m}{x}\right)^{\beta} \quad (2);$$

and the probability density function (PDF) is:

$$P[X = x] = km^{\beta}x^{-(\beta+1)} \quad (3).$$

Note that the PDF of a Pareto distribution is also a power law, but the exponent is $-(k + 1)$, as opposed to $-k$ in the case of the exceedence probability (which equals 1 – CDF; Fig. 5). This means that, although both the PDF and the 1 – CDF plot as straight lines on log–log plots, the PDF line is steeper than the exceedence line (Fig. 5B). A power law distribution is defined by two parameters: the location parameter m and the shape parameter β. Both parameters must be positive. For comparison, the PDF, CDF and CCDF curves are illustrated for the lognormal distribution as well (Fig. 5).

The problems with least squares fitting

Most studies dealing with turbidite bed thickness distributions rely on the fact that power law distributions plot as straight lines on log–log exceedence probability diagrams. The best distribution model is chosen through visual inspection of these plots. The power law exponent is determined using least squares fitting on the log-transformed values of bed thickness and exceedence probability. In some cases, the coefficient of determination R^2, which is the ratio of the explained sum of squares to the total sum of squares, is used as a measure of goodness-of-fit either to support or to reject the power law model (Carlson & Grotzinger, 2001; Sinclair & Cowie, 2003).

Least squares fitting is a simple method for estimating the power law exponent. However, as noted by Goldstein *et al.* (2004), such estimates must be qualified by some assessment of their goodness-of-fit. R^2 is inappropriate for this purpose, especially when working with exceedence probability plots or cumulative plots in general. By definition, the exceedence

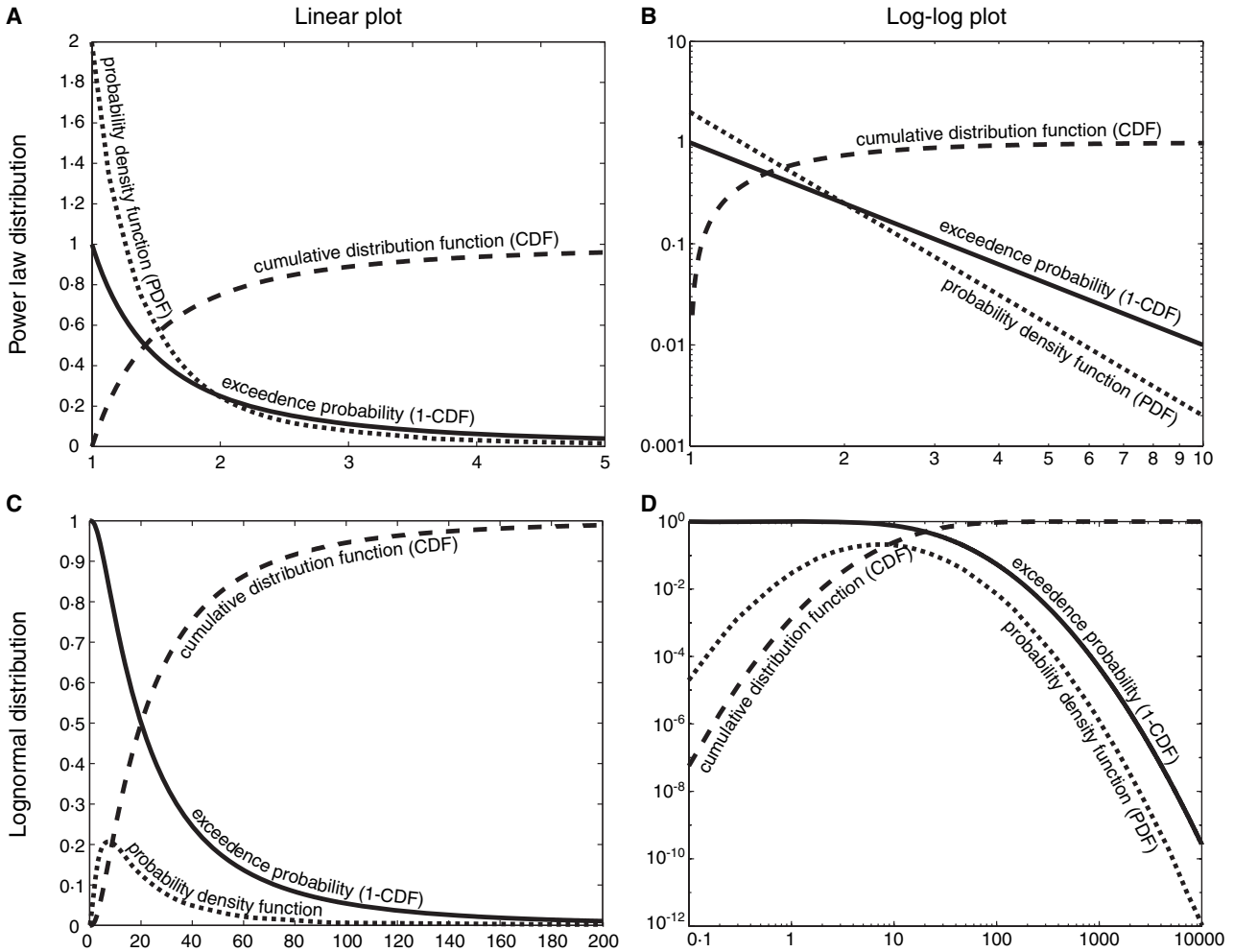

Fig. 5. Plots of the probability density function (PDF), cumulative distribution function (CDF), and exceedence probability (which equals 1 − CDF) of lognormal and power-law distributions on linear and log–log plots. Note that both the exceedence probability and the PDF of a power law distribution plot as straight lines on log–log diagrams (B). The exceedence probability of a lognormal distribution could be misinterpreted as consisting of two straight line segments on the log–log plot, that is, two power laws (D).

probability is decreasing as bed thickness increases (Fig. 5), and R^2 will be relatively large even when the curve is clearly non-linear. In addition, the significance of this statistic cannot be tested because some assumptions of classical regression are violated. The validity of the regression statistic depends on the distribution of the residuals (e.g. Swan & Sandilands, 1995): they must be homoscedastic, that is, there cannot be any trend in the distribution of variance; and they must be normally distributed. These assumptions are not satisfied by the residuals of a typical exceedence probability analysis. The Kolmogorov–Smirnov and the chi-square goodness-of-fit tests are more appropriate measures; they will be described briefly later in the paper.

The 'thin bed problem'

Another characteristic of the exceedence probability log–log plots is that they give emphasis to the upper tail of the distribution, that is, to the thick beds and tend to hide the shape of the distribution at the lower tail, where the much more numerous thin beds are located (see also Drummond & Coates, 2000). Departures from the power law model often occur at the thin tail of the distribution, but these departures are visually suppressed on these plots. For example, in the case of the Upper Tarcău Sandstone, the departure from a power law at the thin end of the distribution seems relatively insignificant on the exceedence plot (Fig. 6A). Based on this diagram, it looks like the data are relatively well described

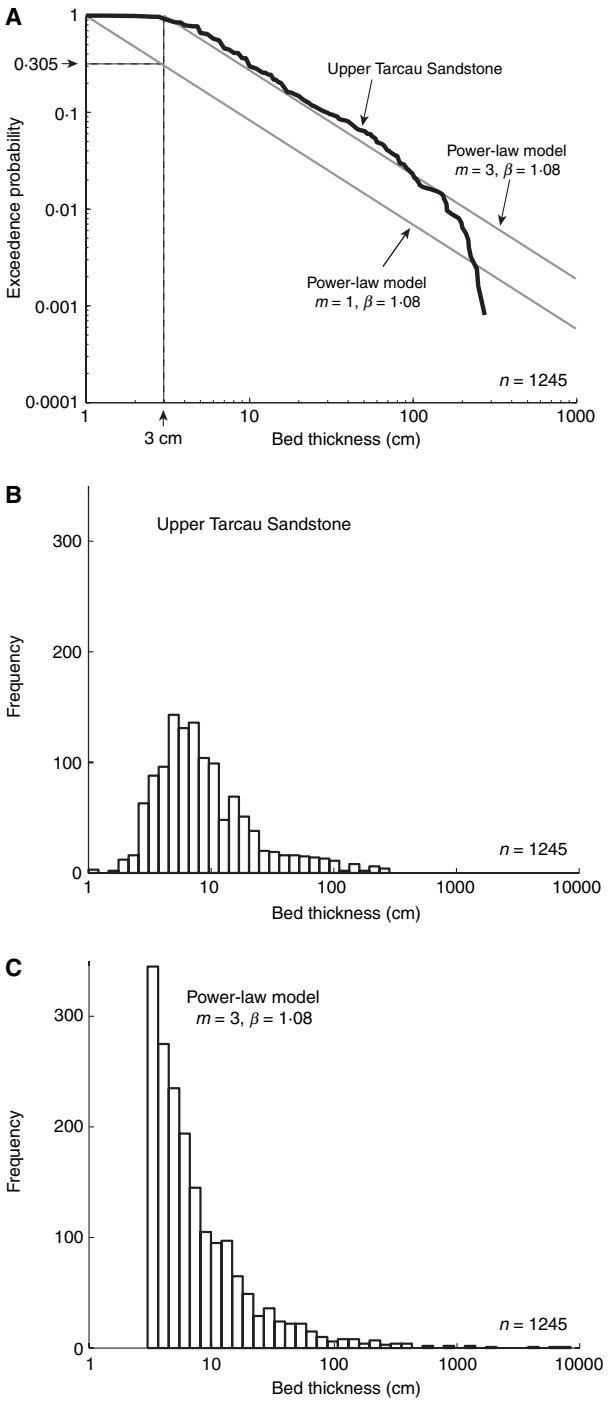

Fig. 6. (A) Log–log exceedence probability plot of bed thicknesses in the Upper Tarcău Sandstone (thick black line). Two power-law models with the same power-law exponent ($\beta = 1{\cdot}08$) but different location parameter m (or minimum bed thickness) are also shown. (B) Histogram of bed thickness data from the Upper Tarcău Sandstone. (C) Histogram of a synthetic bed thickness series that fits a power law with $\beta = 1{\cdot}08$ and m = 3. The histograms clearly indicate that the actual data set should have a lot more thin beds in order to fit a power law distribution, whereas the log–log plots tend to hide this disparity.

by a power law with a slope of about 1·08; R^2 is 0·964. However, there are several problems with such a model. Firstly, plotting the histograms of the log-transformed data for the real data set and a simulated distribution reveals that the departure from the power law at the thin beds is much more significant: the shapes of the two histograms are strongly dissimilar (Fig. 6). Secondly, a location parameter has to be chosen, that is, a minimum bed thickness for the power law distribution. A better visual fit can be achieved if this is set at 3 cm, the point where the departure starts. Then again, there are clearly a number of beds in the Tarcău data that are thinner than 3 cm. If the location parameter is set to 1 cm, the thinnest bed measured, the model exceedence probabilities drop significantly below the Tarcău data, again suggesting the mismatch in the number of thin beds. It may be calculated what proportion of beds should be thinner than 3 cm (Fig. 6A) in order to satisfy this model: almost 70% of the beds should be less than 3 cm thick. This means either that almost 3000 thin beds were not measured during the field work, or that the power law model is not appropriate in this case. Although it is true that the thin beds probably are under-represented in the data set, mainly because of the less-than-perfect exposure of some of the shaliest sections, such a degree of mismatch at least raises questions about the practicality of the power law model. On the one hand, it is only in the rarest circumstances that it is possible to measure correctly all bed thicknesses throughout a whole stratigraphic unit (e.g. Talling, 2001); on the other hand, the requirement to measure all the thin beds focuses the attention on the facies that commonly forms the non-reservoir parts of hydrocarbon accumulations. Compared with the 'thin bed problem', departures from a single straight line in the mid-range of the distribution probably represent a more significant challenge to the general applicability of the power law model.

Segmented power law distributions

In many cases, turbidite bed thickness distributions show significant departures from a straight line on the log–log plot and these departures are described frequently in terms of 'segmented power laws' (Rothman & Grotzinger, 1995; Pirmez *et al.*, 1997; Chen & Hiscott, 1999; Awadallah *et al.*, 2001; Sinclair & Cowie, 2003). Of 18 plots of bed thickness data in the literature, 14 are interpreted in terms of segmented power laws, two as impossible to fit with straight lines, and

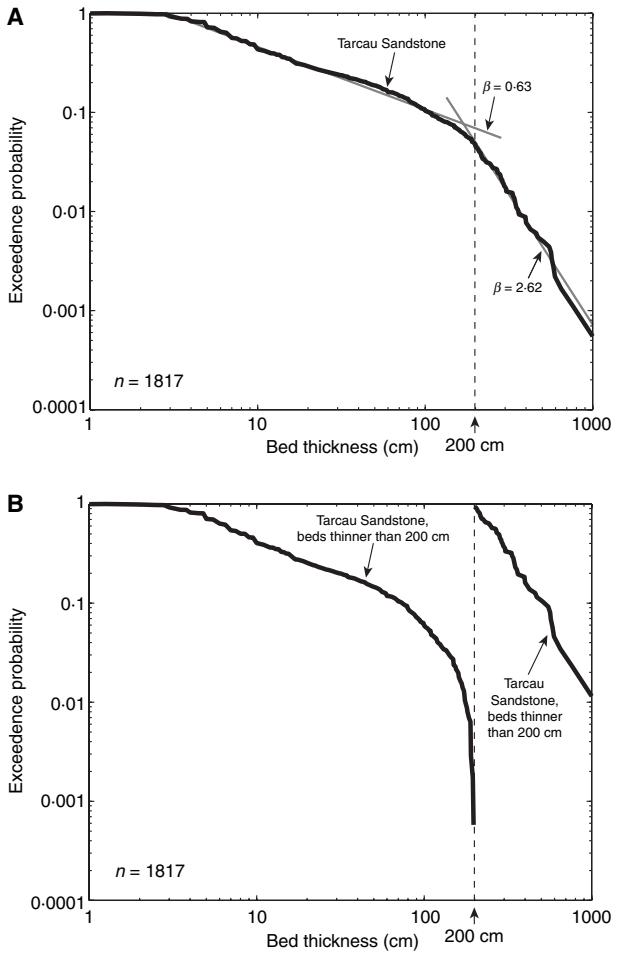

Fig. 7. (A) Bed thickness data from the Tarcău Sandstone interpreted as a 'segmented power law distribution'. The two straight lines were fitted to the data using least squares fitting and a threshold thickness of 200 cm. R^2 values are 0·988 for the thin-bedded part of the data and 0·987 for beds thicker than 200 cm. (B) Beds thinner and thicker than 200 cm plotted separately; the thinner-bedded set does not plot as a straight line (because it is truncated at 200 cm).

there are only three that seem to fit single power laws. The two segments with different power law exponents are treated as two different populations, purportedly well characterized by the exponent estimated from the exceedence plot. The implicit assumption is that the data are best described as a mixture of two power law distributions. A similar interpretation of the Tarcău Sandstone bed thickness data is shown in Fig. 7. Beds with thicknesses <200 cm seem to have a different exponent ($\beta \approx 0{\cdot}63$) compared with beds thicker than 200 cm ($\beta \approx 2{\cdot}62$). These values are usually found using linear regression, and sometimes large values of R^2 are treated as evidence for a good fit (Drummond & Coates, 2000; Sinclair & Cowie, 2003). Indeed, in the case of the Tarcău Sandstone data, R^2 is 0·9885 for the thin bedded part of the data, and it is 0·9871 for beds thicker than 200 cm.

However, if the two groups are plotted separately, it becomes obvious that the significance of this bed thickness threshold is unclear. Although the thicker bedded part of the data might be relatively well described as a power law distribution with a minimum bed thickness of 200 cm, the beds thinner than 200 cm do not plot as a straight line and, if trying to force them to conform with some power law models, in the best case they can be thought of as a truncated power law distribution.

A series of theoretical exceedence probability plots have been generated to see: (i) what is the behaviour of mixtures of two power law distributions?; and (ii) how can a segmented power law be generated on such plots (Fig. 8)?. In general, mixtures of two power law-distributed populations result in exceedence plots that are quite different from those observed in plots of real data. If two populations of roughly equal size with different exponents but with the same location parameter are mixed, the exceedence plot will hardly deviate from a single straight line and a conventional exceedence analysis would conclude that there was a single population (Fig. 8A). If the location parameters are different as well, a curve with two concave-up segments results that do not resemble any of the plots of actual data (Fig. 8B). Note that it would be difficult to derive meaningful exponents from this curve. Slopes on exceedence plots are not simply a function of the distribution of the values in the interval where the slope is roughly constant but also depend on how many data points are outside this interval. To generate a mixture that has a clear segmented character, the thinner-bedded component must be truncated at the minimum bed thickness for the second population (Fig. 8C). In addition, the number of beds in the thicker-bedded population, n_2, must satisfy the relationship:

$$\frac{n_2}{n_1 + n_2} = \left(\frac{m_1}{m_2}\right)^{\beta 1} \qquad (4),$$

where $n_1 + n_2$ is the total number of beds, m_1 and m_2 are the location parameters, and $\beta 1$ is the exponent of the thinner-bedded population. This also means that the second population must be a fraction of the thinner-bedded population, of the order of a maximum of a few percent (Fig. 8C).

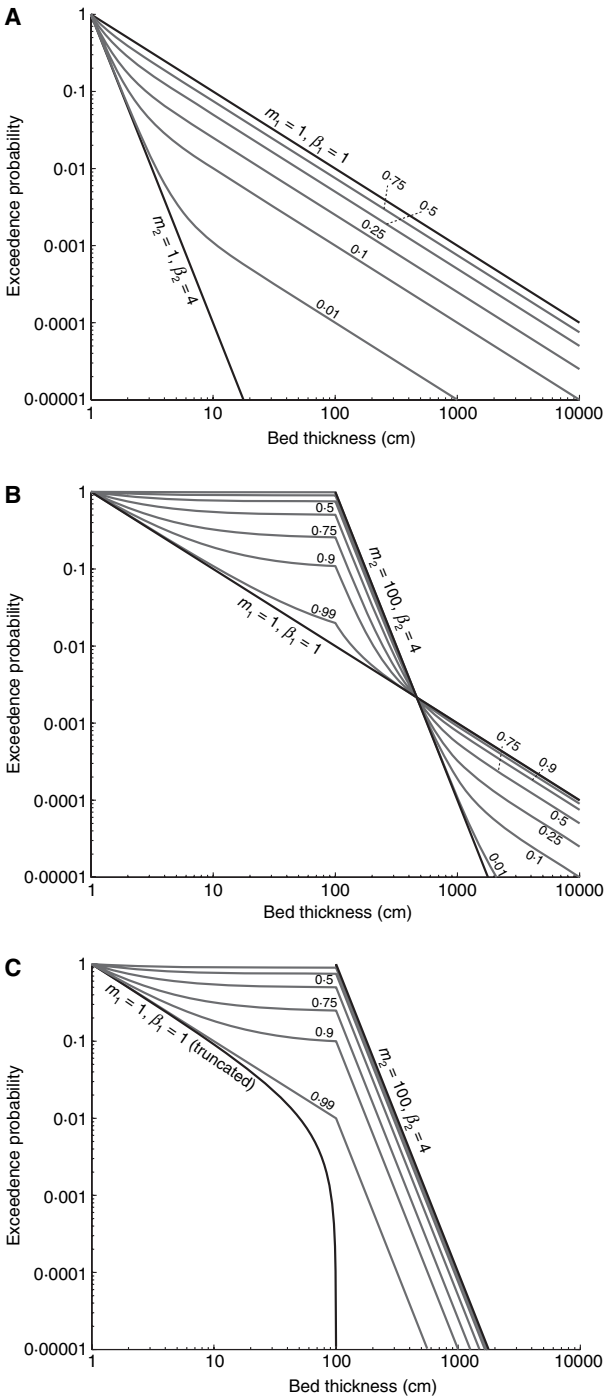

Fig. 8. Log–log exceedence probability curves of theoretical mixtures of two power law distributions. Of all these possible curve shapes, only one specific curve, which belongs to case (C), results in a segmented power law. Numbers near the curves indicate the proportion of the lower-exponent ($\beta_1 = 1$) population in the mixture. (A) Neither of the distributions is truncated; both have a minimum bed thickness of 1 cm, but they have different exponents. (B) The thicker-bedded population is truncated at its lower end: the minimum bed thickness is $m_2 = 100$ cm. (C) The thinner bedded population is truncated as well, at its upper end.

It is hard to think of a relatively simple process resulting in a power law mixture that satisfies these criteria for generating a segmented power law on a log–log plot. Yet there is a mathematical model of confinement of turbidite beds that can indeed result in segmented power laws if a series of assumptions are valid. This model will be the subject of the next section.

The Malinverno (1997) model for segmented power laws

Assuming a power law distribution of bed volumes, Malinverno (1997) derived a simple relationship among the scaling between bed length and bed thickness γ, the exponent of the bed volume distribution c, and the spatial distribution of the bed depocentres relative to the vertical sampling line, characterized by the exponent d:

$$\beta = c + \gamma(2c - d) \qquad (5).$$

Malinverno (1997) also developed a model of basinal confinement of turbidites and showed how the confinement can alter the bed thickness distribution at the basin centre. Beds of cylindrical shape are placed randomly into a circular basin and the bed thicknesses are measured at a vertical sampling line at the basin centre. Malinverno (1997) interpreted the resulting change in the shape of the exceedence probability curve as a transition to a segmented power law, and derived relationships for the exponents (or slopes) of the two segments or populations. Beds with a radius much less than the basin diameter will be distributed throughout the basin so that $d \approx 2$, and the original relationship becomes:

$$\beta_{\text{small}} = c(1 + 2\gamma) - 2\gamma \qquad (6).$$

Beds with diameters larger than the radius of the basin will be intersected by a sampling line at the basin centre, therefore $d \approx 1$, and:

$$B_{\text{large}} = c(1 + 2\gamma) \qquad (7).$$

Some beds from both these categories are in contact with the basin margin; they can be thought of as deposits of flows that locally interact with the basin margin (e.g. they are deflected or reflected). In contrast, 'fully ponded' beds should cover the whole basin floor and they form a third category that can be distinguished on a log–log exceedence plot if modelled appropriately.

In order to assess the general applicability of this 'segmentation through confinement' model, a series of numerical experiments comparable with those of Malinverno (1997) have been run. Although Malinverno (1997) used a bed volume distribution with the largest bed diameter equal to the basin size, in the present experiments there is no interdependence between the largest bed volume and the size of the basin. The basin size is fixed and beds that would exceed the basin diameter using the thickness–length scaling relationship are considered 'ponded', their diameter is set equal to the basin diameter and their thicknesses are increased accordingly. This setup makes it possible to run millions of flows with a power law distribution of volumes without having to change the basin size. The resulting bed thickness series can contain several tens of thousands of data points, an important advantage that helps visualize the subtle variations in the shapes of the exceedence probability curves.

Figure 9 shows the results of two experiments run with similar parameters, the only difference being the value of the bed volume distribution exponent *c*. Based on these results, the following observations can be made.

1 At the basin centre, the Malinverno model does generate exceedence probability curves with straight segments, although the transition between the first two segments is smoothly curved and does not show a well-defined breakpoint.

2 The value of β_{small} is overpredicted by Eq. 6 (Fig. 9) and is better approximated by Eq. 4. The reason for this is that the original bed thickness exponent (which, for $d = 2$, is the same as β_{small}) decreases as the oversampling of thick beds pushes up the first segment of the curve. For all bed thicknesses other than the thickest beds, the exceedence probability increases compared with the curve unaffected by confinement because more thick beds are intercepted by the sampling line. If Eq. 6 holds, for given c and γ values, the smaller-than-predicted β_{small} would require a d value larger than 2 which is impossible by definition. Because only a few beds exceed the thicknesses characterized by β_{large}, the value of β_{large} is affected to a much smaller degree and the second straight segment of the numerical experiment has a slope similar to the predicted value.

3 As suggested by Malinverno (1997), extremely large bed volumes result in a third segment with a slope equal to the bed volume exponent *c*. These volumes would be the equivalent of flows that entirely cover the basin floor and are fully ponded.

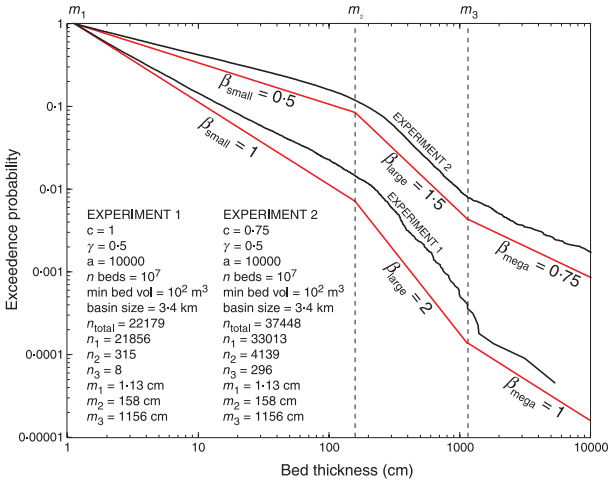

Fig. 9. Results of two numerical experiments that use the assumptions of the confined basin model of Malinverno (1997). The black lines represent bed thicknesses measured at the centre of the circular basin; the red lines correspond to the curves predicted by the equations of Malinverno (1997). The only difference between the setup of experiment 1 and experiment 2 is the power law exponent of the bed-volume distributions *c*; m_1 is the minimum bed thickness; m_2 is the threshold bed thickness that separates small-volume beds from beds with a diameter larger than the radius of the basin; and m_3 is the threshold bed thickness above which beds are 'fully ponded' (beds with a diameter equal to the basin diameter). The other parameters are: c = exponent of bed volume distribution; γ = scaling exponent between bed length and bed thickness; a = scaling constant between bed length and bed thickness ($l = ah^{\gamma}$, where l is the bed diameter and h the bed thickness); n beds = total number of beds deposited in the basin; min bed vol = minimum bed volume; basin size = basin diameter; n_{total} = number of beds intercepted by sampling line; n_1 = number of beds in the first population; n_2 = number of beds in the second population; n_3 = number of beds in 'fully ponded' population.

This is a relatively simple model that can result in straight segments on log–log exceedence probability plots of bed thickness data. However, before interpreting most departures on log–log plots from a single straight line as due to confinement, it is important to consider the limitations of the Malinverno (1997) model:

1 The input bed volumes must have a power law distribution. Although recent estimates from the Marnoso-Arenacea Formation suggest that bed volumes have a lognormal distribution (Talling *et al.*, 2007), it is possible that the power law assumption is valid in other systems. However, even if the power law input volume distribution is accepted, this initial range will be strongly

modified if there are well-developed channel levées. In such systems, most of the flows are guided by channels, and often deposit a significant part of their sediment in the overbank area and, as a result, go through an efficient modification of volumes before reaching the less channelized site of deposition.

2 If confinement was the cause of most deviations from a single power law of bed thickness data, at least in some cases the third segment of the plot should be seen as well, that is, a few 'megabeds' that cover the whole basin or container. No such outliers are reported in the literature. In a high-quality data set, even a small number of such data points would plot along a different trend than the second population characterized by β_{large} (Fig. 9). Many of the beds in the Marnoso-Arenacea Formation extend across the entire basin (Ricci Lucchi & Valmori, 1980; Amy & Talling, 2006), yet no third segment is observed on the log–log exceedence plots.

3 In channel levée systems, beds deposited in the overbank areas are likely to be characterized by different length–thickness relationships and different degrees of confinement than beds deposited in the channels. In addition, the geometric configurations of these settings are obviously more complicated than those of a circular basin.

4 The individual segments corresponding to β_{large} and β_{small} can easily be identified with large data sets collected exactly at the basin centre, but this task becomes increasingly difficult as the sampling location moves away from the centre and as the number of data points is reduced. Five experiments have been run with the same parameters but different sampling locations (Fig. 10). Although it is possible to derive reasonable estimates for β_{large} and β_{small} in the case of sampling location 1 and possibly location 2 as well, the clear segmented character of the exceedence plot disappears and determining β_{large} and β_{small} becomes impossible at location 3. The dashed circle in the centre of the basin shows the approximate area characterized by more or less segmented plots. This is <20% of the whole area of the basin. Sampling locations falling within the remaining 80% of the basin area would show no clear evidence of confinement.

In summary, a distribution described by a segmented power law on an exceedence plot is not a simple mixture of two power law distributions of arbitrary parameters. Although the Malinverno (1997) model of confinement produces

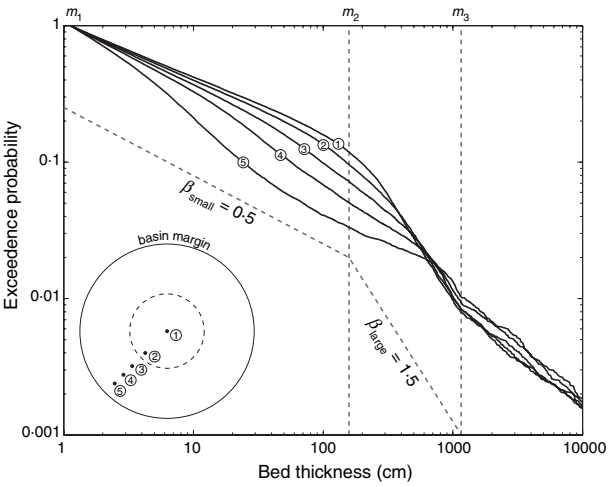

Fig. 10. Results of five numerical experiments that use the assumptions of the confined basin model of Malinverno (1997), but place the vertical sampling line progressively away from the basin centre. The individual segments corresponding to the three populations shown in Fig. 6 become increasingly difficult to recognize as the sampling line moves away from the immediate neighbourhood of the basin centre. Dashed lines show the segment slopes predicted by the Malinverno (1997) equations.

bed thickness distributions that do result in segmented power laws at the basin centre, a number of assumptions are necessary that restrict the applicability of this model. Identifying two or more relatively straight segments in exceedence probability plots can be highly subjective, needs high-quality data sets and should be used in the analysis of turbidite bed thickness data only if the assumptions of the Malinverno (1997) model are likely to be satisfied. Even in such cases, statistical models other than the power law distribution should be considered.

MIXTURES OF LOGNORMAL DISTRIBUTIONS

Because the use of the exponential distribution for describing bed thickness data also faces the 'thin bed problem' (see Drummond & Wilkinson, 1996 for a discussion), the lognormal model seems a more practical alternative. Lognormal distributions are frequent in nature; they result from the multiplicative effects of forces independent of each other, as opposed to the normal distribution that results from the addition of the effects (e.g. Aitchison & Brown, 1957; Limpert *et al.*, 2001). Talling (2001) suggested that the

lognormal distribution might be the most appropriate for characterizing turbidite bed thicknesses from a number of successions. Talling (2001) also drew attention to the bimodality of many turbidite bed thickness data sets, and pointed out that two lognormal populations may result in 'segmented power laws' on exceedence probability plots. Although the two populations seem to correspond to lithologic and facies differences, it would be useful to identify the two components of the lognormal mixture in a quantitative way. The questions addressed here are the following: (i) how can the parameters of the two (or more) lognormal components be extracted from a lognormal mixture?; and (ii) what is the appropriate statistical methodology for comparing the goodness-of-fit of the lognormal mixture model with the power law model?

Finding the lognormal components of a mixture

Separating components of a Gaussian mixture is an important subject in statistics that has been discussed in one of the early statistical papers (Pearson, 1894). Since the 1980s, in parallel with the increase in computational power, the number of publications on mixture modelling has amplified considerably (e.g. McLachlan & Peel, 2000, and references therein). Before the widespread availability of computers, identifying and separating the components of a grain-size mixture was not a simple exercise (e.g. Spencer, 1963; Tanner, 1964). However, today there are several algorithms available to address this problem, and the expectation–maximization algorithm is one of the best candidates for separating components both in grain size and in bed thickness data. Figure 11 illustrates the shapes of the cumulative probability plots for two types of lognormal mixtures: in the first type, the thinner-bedded population has a smaller standard deviation than the thicker bedded population; in the second type, the thicker-bedded group has a smaller variance.

The expectation–maximization algorithm (Dempster *et al.*, 1977) is a popular algorithm in machine learning and statistical computing that is often applied in parameter estimation problems, for example, in estimating parameters of Gaussian mixture models (e.g. Kung *et al.*, 2004). K-means clustering is an alternative approach, but the expectation–maximization algorithm has the advantage that the parameters of the two distributions are derived using maximum likelihood estimation during the iterations and the classification results are probabilistic. In the case of bed thicknesses, this means that the *probabilities* of each bed belonging to a thick-bedded or the thin-bedded population may be found, whereas class memberships in K-means clustering are either 0 or 1.

Using a Matlab® implementation of the EM algorithm (Nabney, 2001), the bed thickness data from the Tarcău Sandstone and the Marnoso-Arenacea Formation have been analysed (Marnoso-Arenacea Formation data from Talling, 2001) and the two most probable lognormal components separated. The results are shown in Figs 12 and 13. In addition to the log–log exceedence probability plot, other diagrams are also presented: the histogram of the log-transformed values, the PDFs of the model distributions and the cumulative probability plot of the log-transformed values. Using a range of diagram types guarantees that all potential departures from the statistical model are appropriately highlighted.

In both cases, the visual fit between the lognormal mixture model and the actual data is better than in the case of single or segmented power law models. Segmented power law interpretations applied to the log–log exceedence plots would ignore obvious non-linear variations in the curve shape that are remarkably well described by the lognormal mixtures (Figs 12D and 13D). In addition, there is a relatively good, although not perfect match, between the thin-bedded and thick-bedded lognormal components and beds starting with Bouma sequences T_c or T_d and T_a or T_b respectively (Fig. 14). This relationship could be used to estimate proportions of different turbidite facies where no data other than bed thicknesses are available (e.g. when bed thickness is estimated from high-resolution image logs).

If plotted separately, thicknesses of Bouma divisions also seem to have lognormal distributions (Fig. 15). In fact, T_a and T_b intervals match the lognormal model better than the thick-bedded population identified as the beds starting with T_a or T_b divisions (Fig. 14A). It is not clear how these divisions tend to combine to form the thick-bedded population, but it does seem that the basal Bouma division is one of the defining parameters of the thickness of the whole event bed. In fact, the slight divergence from the lognormal model, when beds starting with T_a or T_b divisions are grouped together, is reduced when beds starting with T_a and beds starting with T_b are plotted separately. Thus, it is possible that the Tarcău bed thickness data comprise more than two lognormal populations. However, the

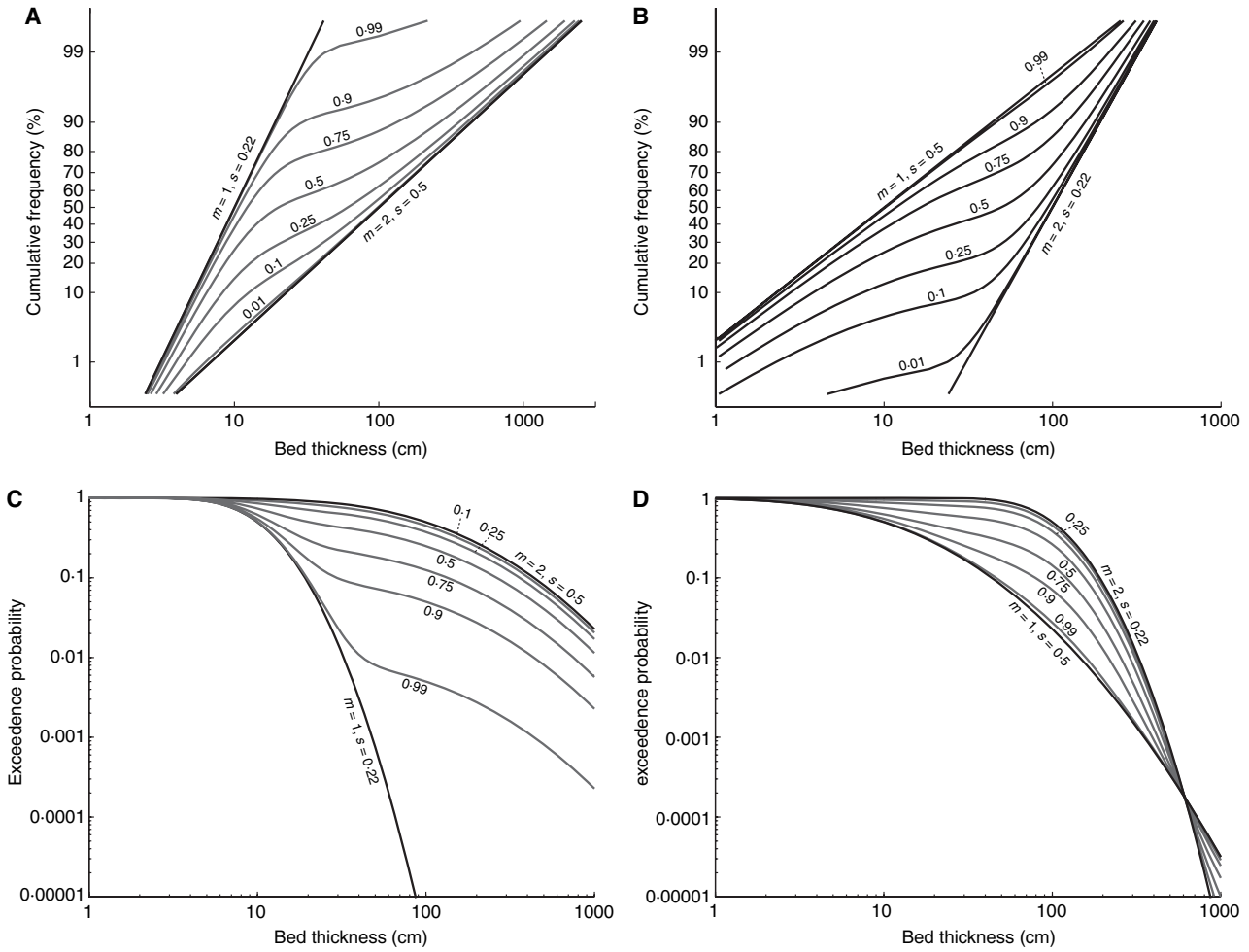

Fig. 11. Cumulative probability and log–log exceedence probability curves of theoretical mixtures of two lognormal distributions. Note that the exceedence curves shown in (D) could be misinterpreted as segmented power laws. Numbers near the curves indicate the proportion of the thinner bedded population in the mixture. (A) Cumulative probability curves of mixtures of two lognormal populations, the thinner bedded having a lower variance than the thicker bedded one. (B) Cumulative probability curves of mixtures of two lognormal populations, the thinner bedded having a higher variance than the thicker bedded one. (C) Exceedence probability curves of mixtures of two lognormal populations, the thinner bedded having a lower variance than the thicker bedded one. (D) Exceedence probability curves of mixtures of two lognormal populations, the thinner-bedded population having a higher variance than the thicker-bedded population. m and s are the mean and the standard deviation of the 10-base logarithm of the bed thicknesses.

two-component model gives a good match to the data (Fig. 12) and adding more populations would significantly – and probably unnecessarily – complicate the model. Sandstone bed thickness is also plotted on Fig. 15; this curve is clearly not a single lognormal population, but a mixture of several different groups.

Figure 16 shows the results of the classification of the Tarcău Sandstone bed thickness series into thin-bedded and thick-bedded groups. In the case of the Tarcău Sandstone, the limit between the thin-bedded and thick-bedded populations is *ca* 30 cm, whereas the segmented power law model would predict a limit of *ca* 200 cm. The 30 cm limit is notably similar to that in the Marnoso-Arenacea Formation (Talling, 2001). The classification results of the bed thickness series using the two different approaches show that the lognormal mixture model gives a more intuitive outcome: the lognormal thick-bedded population largely corresponds to the thick-bedded clusters in the succession, whereas the 200 cm limit only distinguishes a few very thick beds that are not clustered spatially.

An important difference between the Tarcău Sandstone and the Marnoso-Arenacea Formation is as follows. In the Tarcău Sandstone, the log-transformed thin-bedded population has a

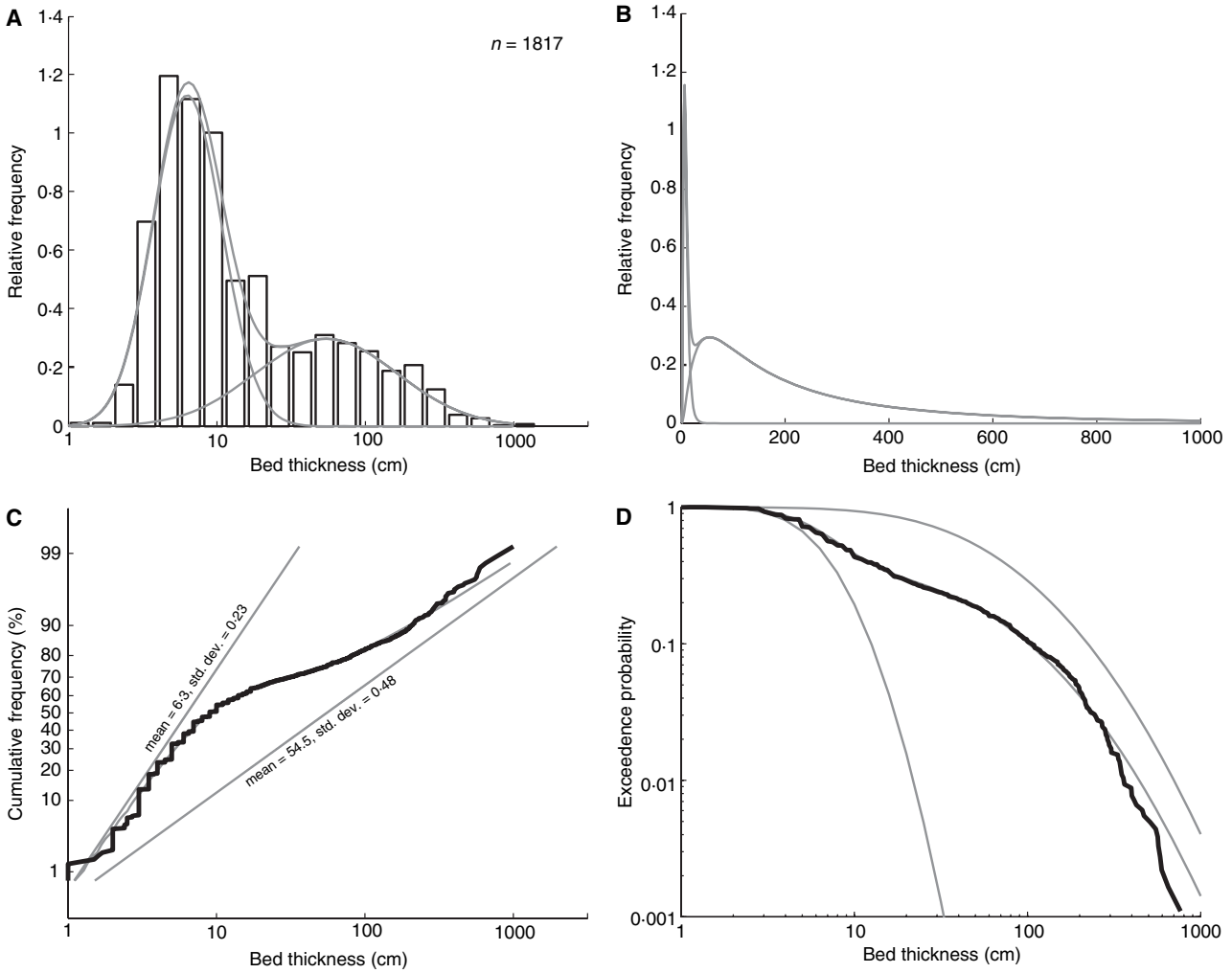

Fig. 12. Interpretation of the bed thickness data from the Tarcău Sandstone as a mixture of two lognormal populations. (A) Histogram of the log-transformed bed thicknesses, with the two model distribution PDFs shown. (B) Model PDFs and their sum shown on a linear x-axis. (C) Cumulative probability plot of the data and the model curves of the two lognormal distributions. Mean is the geometric mean, std. dev. is the standard deviation of the 10-base logarithm of the bed thickness series. (D) Log–log exceedence probability plot of the data and the model curves.

smaller standard deviation than the thick-bedded population; in the Marnoso-Arenacea Formation, the thicker-bedded cluster has less variability than the thinner-bedded cluster. These two different mixture types produce cumulative curves of different shapes on both probability 'paper' and on log–log exceedence plots (Figs 11 and 14). Note that 'Marnoso-Arenacea type' mixtures, in which the thicker-bedded population has a smaller standard deviation, tend to generate log–log exceedence plots that could easily be misinterpreted as segmented power laws. Unless strongly dominated by the thinner-bedded population, 'Tarcău-type' mixtures result in smoothly curving log–log exceedence plots that would be more difficult to interpret in terms of segmented power laws (but not impossible; see Fig. 7A).

Choosing between the power law and lognormal models using statistical testing

The Kolmogorov–Smirnov and the chi-square test statistics were used for testing both the single power law and the lognormal mixture models in the case of the Upper Tarcău Sandstone.

The most widely used goodness-of-fit tests for non-normal distributions are the chi-square test and the Kolmogorov–Smirnov test (e.g. Swan & Sandilands, 1995; Davis, 2002). The chi-square test compares the observed frequencies of values with the expected frequencies that correspond to the distribution model. The main disadvantage of the chi-square test is that the data must be grouped into bins and the result is influenced by the number of bins used. In contrast, the

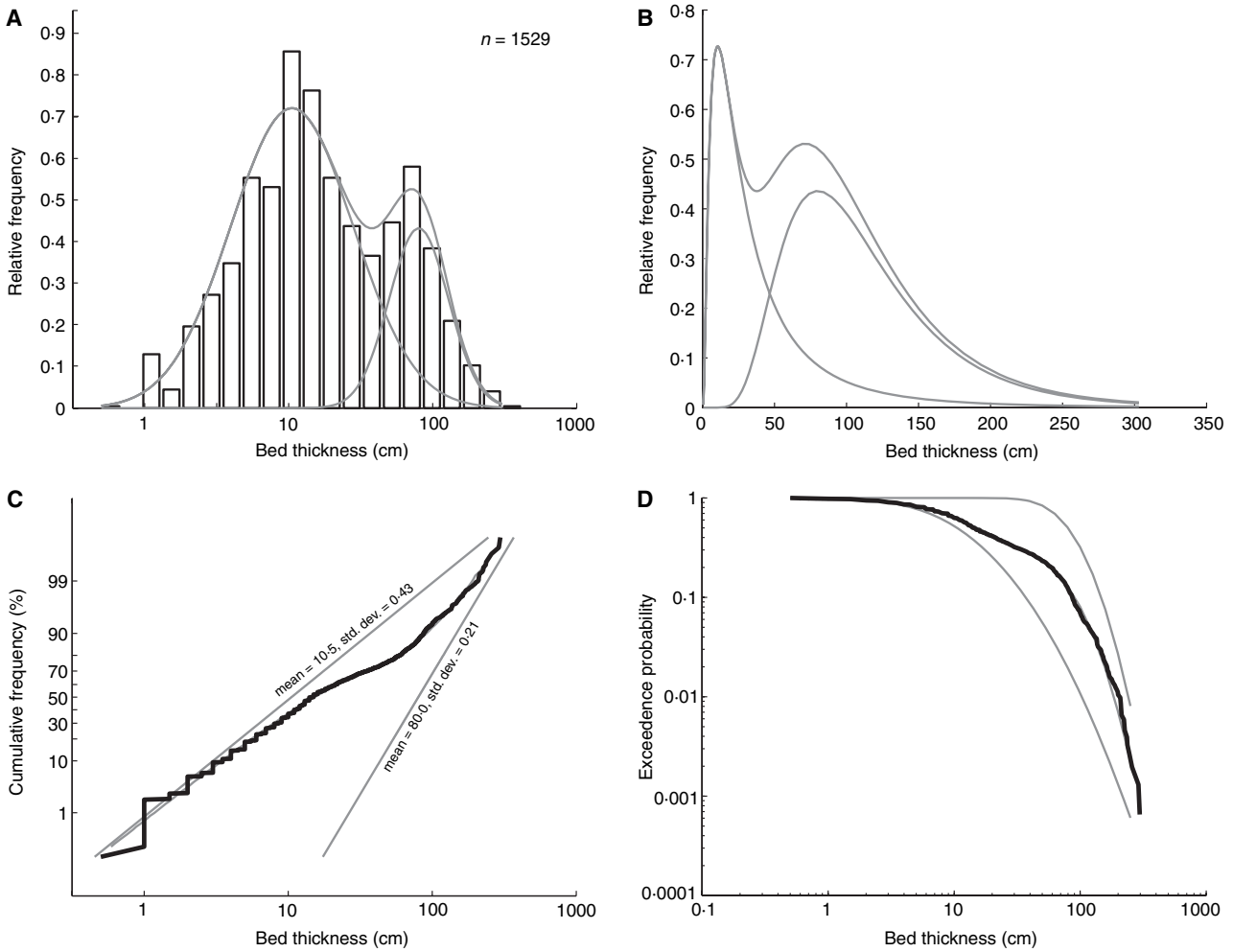

Fig. 13. Interpretation of the bed thickness data from the Marnoso-Arenacea Formation as a mixture of two lognormal populations. Bed thickness data from Talling (2001). (A) Histogram of the log-transformed bed thicknesses, with the two model distribution PDFs shown. (B) Model PDFs and their sum shown on a linear *x*-axis. (C) Cumulative probability plot of the data and the model curves of the two lognormal distributions. Mean is the geometric mean, std. dev. is the standard deviation of the 10-base logarithm of the bed thickness series. (D) Log–log exceedence probability plot of the data and the model curves.

Kolmogorov–Smirnov test compares the cumulative relative frequencies of the data with the cumulative distribution function (CDF) of the assumed underlying distribution model. The result does not depend on the way the data are divided into groups. The Kolmogorov–Smirnov test statistic corresponds to the largest difference between the cumulative relative frequency curve and the CDF of the model distribution.

Tables for critical values of the conventional Kolmogorov–Smirnov test statistic are valid only if the distribution parameters are not estimated from the data (Davis, 2002; Goldstein *et al.*, 2004). As in this study, both the location parameter and the shape parameter of the power law distribution are estimated from the data, the critical values must be recalculated. Largely based on the suggestions of Goldstein *et al.* (2004), the following methodology has been adopted and implemented in Matlab®: (i) using maximum likelihood estimation, the power law exponent of the data set is assessed; (ii) based on the CDF of a distribution with this exponent, the Kolmogorov–Smirnov statistic is found; (iii) a large number of artificial populations with the same number of values and the same parameters as the dataset are generated and steps (i) and (ii) are repeated for each to create a distribution of KS statistics; and (iv) critical values can be read from this distribution for different levels of significance. The null hypothesis states that the data come from a distribution with the estimated shape and location parameters. If the significance level is chosen as 5% and the Kolmogorov–Smirnov statistic of

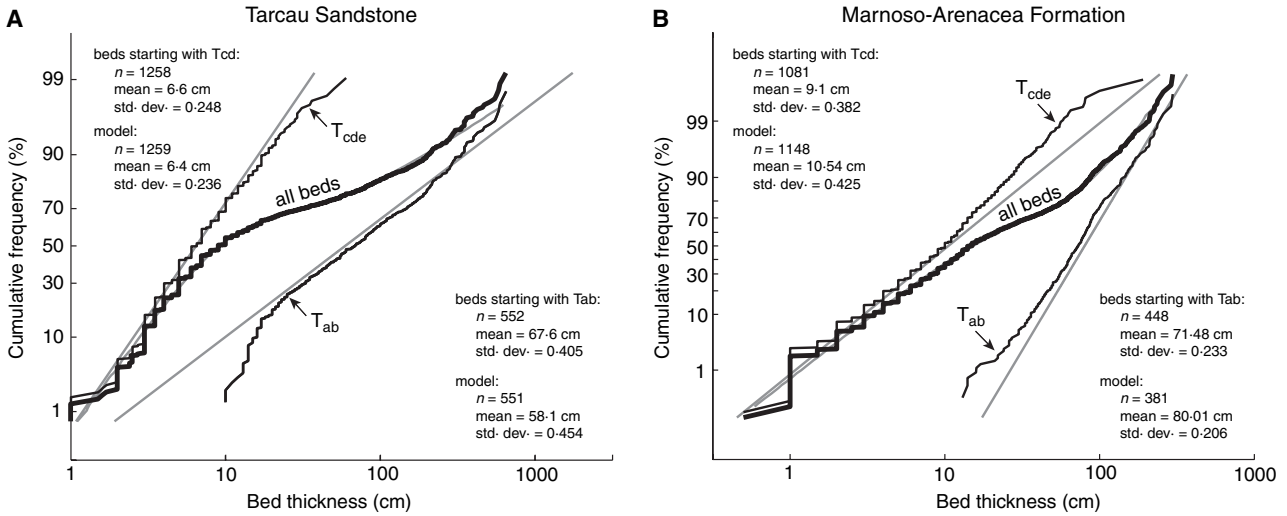

Fig. 14. (A) Cumulative probability plot of the data and the model curves for the Tarcău Sandstone, with plots of thicknesses of beds starting with Bouma divisions T_a or T_b and T_c, T_d or T_e. (B) Cumulative probability plot of the data and the model curves for the Marnoso-Arenacea Formation, with plots of thicknesses of beds starting with Bouma divisions T_a or T_b and T_c, T_d or T_e. Mean is the geometric mean, std. dev. is the standard deviation of the 10-base logarithm of the bed thickness series.

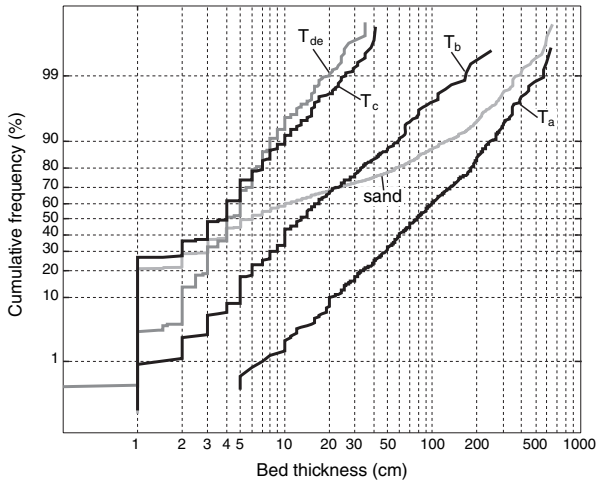

Fig. 15. Cumulative probability plots of Bouma division thicknesses. Individual divisions tend to have lognormal distributions.

the data set is larger than the 95th percentile of the values derived through randomization, the null hypothesis can be rejected. Alternatively, the difference between the actual and the expected Kolmogorov–Smirnov statistic, measured in standard deviations, can be used as a quantitative measure of goodness-of-fit to the distribution.

In a manner comparable with the procedure used to apply the Kolmogorov–Smirnov test to a power law model, the following steps were used to calculate test statistics for the lognormal mixture model: (i) using the expectation–maximization algorithm, the parameters of the two lognormal populations are estimated from the data; (ii) the test statistic (Kolmogorov–Smirnov or chi-square) of the original data is calculated; (iii) a large number of artificial populations with the same number of values and the same parameters as the data set are generated, the parameters are estimated and then the test statistics are calculated for each to create a distribution of test statistics; and (iv) critical values can be read from this distribution for different levels of significance.

The Upper Tarcău Sandstone bed thickness data plot close to a straight line on an exceedence log–log diagram, and have been previously interpreted as a good example of a power law distribution (Sylvester, 2002). The results of the Kolmogorov–Smirnov and chi-square tests are shown in Table 1. As a result of the large number of measurements, the null hypothesis can be rejected in the Kolmogorov–Smirnov test for both the power law and the lognormal mixture model. Using the chi-square statistic, the lognormal mixture model cannot be rejected at the 5% significance level, whereas the power law alternative fails the test. More importantly, the deviations from the expected test statistics are almost five times as large in the case of the power law model as in the case of the lognormal mixture model. This result strongly suggests that the latter provides a better description of this data set.

In addition to goodness-of-fit tests, quantile–quantile (q–q) plots are also useful in choosing the distribution model that provides a better fit. On a q–q plot, the quantiles of the data are plotted

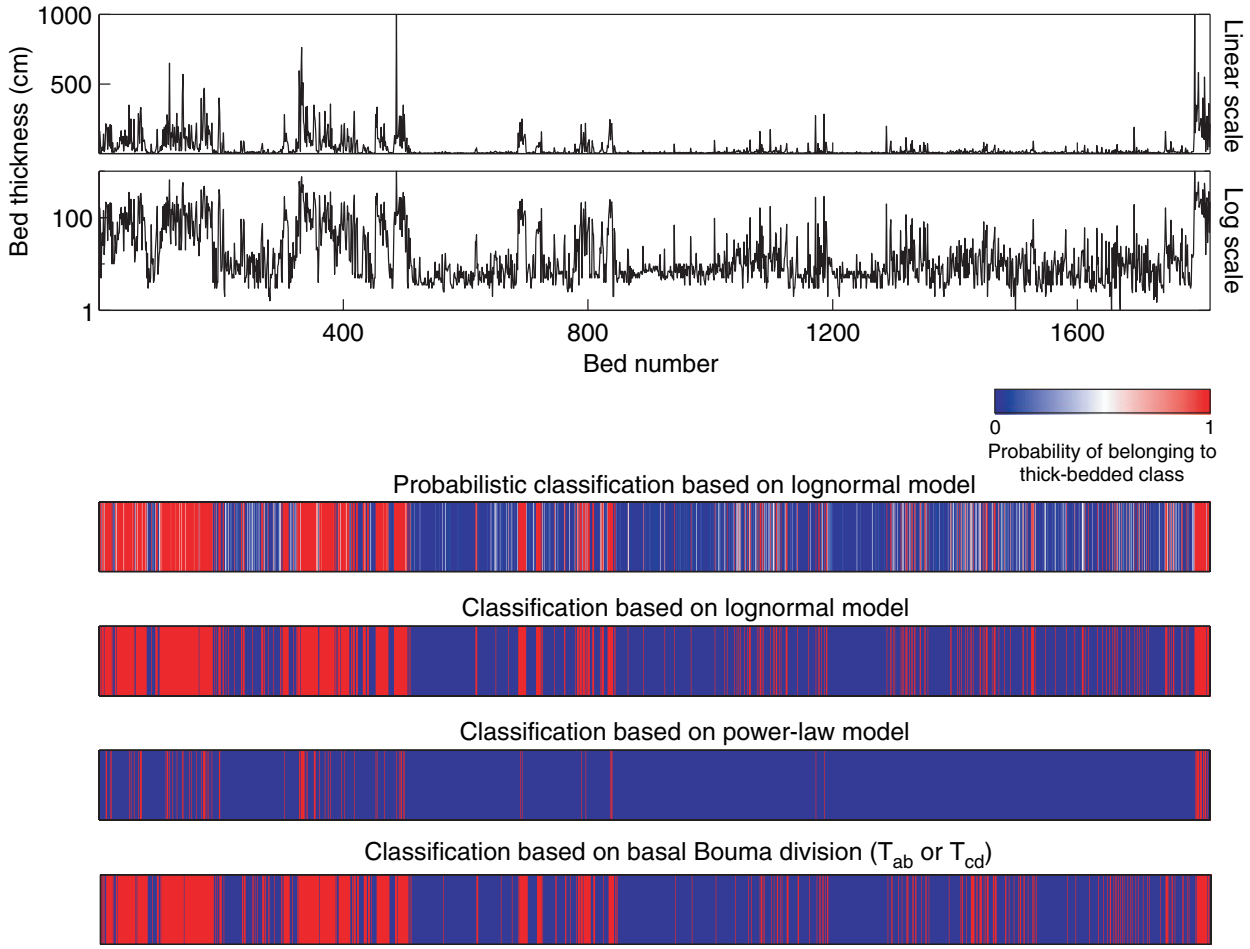

Fig. 16. Bed thickness vs. bed number plots for the Tarcău Sandstone and results of different classification methods. The classifications based on the lognormal mixture model give a better match to the sedimentological classification than the analysis based on the segmented power law model.

against the quantiles of the model distribution (Fig. 17); the better the fit between the model and the data, the closer the plot comes to the $x = y$ straight line on the diagram. The q–q plots for the Upper Tarcău Sandstone support the results of the goodness-of-fit testing: compared with the lognormal mixture model, the departures from the power law model are significantly larger at both the thick-bedded and thin-bedded ends.

Undeniably, the lognormal mixture model has five parameters (two mean, two standard deviation values and the proportion of the two populations), whereas the power law model has only two (the exponent and the minimum bed thickness); therefore, the better fit of the lognormal mixture does not necessarily qualify it as the 'correct' model. More parameters mean larger flexibility and easier fit to any data set. However, the lognormal mixture model is preferred here because it is sedimentologically meaningful and probably more useful in reservoir characterization, as explained below.

DISCUSSION

Power law distributions: the rule or the exception?

As noted by Mitzenmacher (2004), the question whether a particular empirical data set fits better the power law or the lognormal distribution has generated discussion in a number of different fields such as economics, biology, chemistry, information theory and computer science. The two types of distributions can result from similar basic generative models; and a significant portion of a lognormal distribution with a large variance behaves approximately as a straight line on a log–log exceedence plot (Mitzenmacher, 2004). Turbi-

Table 1. Results of statistical testing, using the Kolmogorov–Smirnov (KS) and chi-square statistics, for the Upper Tarcău Sandstone and the Tarcău Sandstone.

Formation	Model	Number of beds	Test	Statistic	Expected average statistic	*P*-value	SD of statistic	Number of SD from expected value	Reject H0?	Number of simulations	Number of bins
Upper Tarcău Sst.	Lognormal mixture	1245	KS	0·0792	0·0182	<0·002	0·0042	14·52	Yes	500	N/A
Upper Tarcău Sst.	Power law	1245	KS	0·3693	0·0234	<0·002	0·0061	56·70	Yes	500	N/A
Upper Tarcău Sst.	Lognormal mixture	1245	Chi-square	0·3422	0·1531	0·122	0·2666	0·71	No	500	40
Upper Tarcău Sst.	Power law	1245	Chi-square	0·1722	0·0115	<0·002	0·024	6·70	Yes	500	40
Tarcău Sandstone	Lognormal mixture	1817	KS	0·0595	0·0145	<0·002	0·0034	13·24	Yes	500	N/A
Tarcău Sandstone	Lognormal mixture	1817	Chi-square	0·3559	0·1755	0·14	0·3151	0·57	No	500	40

SD, standard deviation.

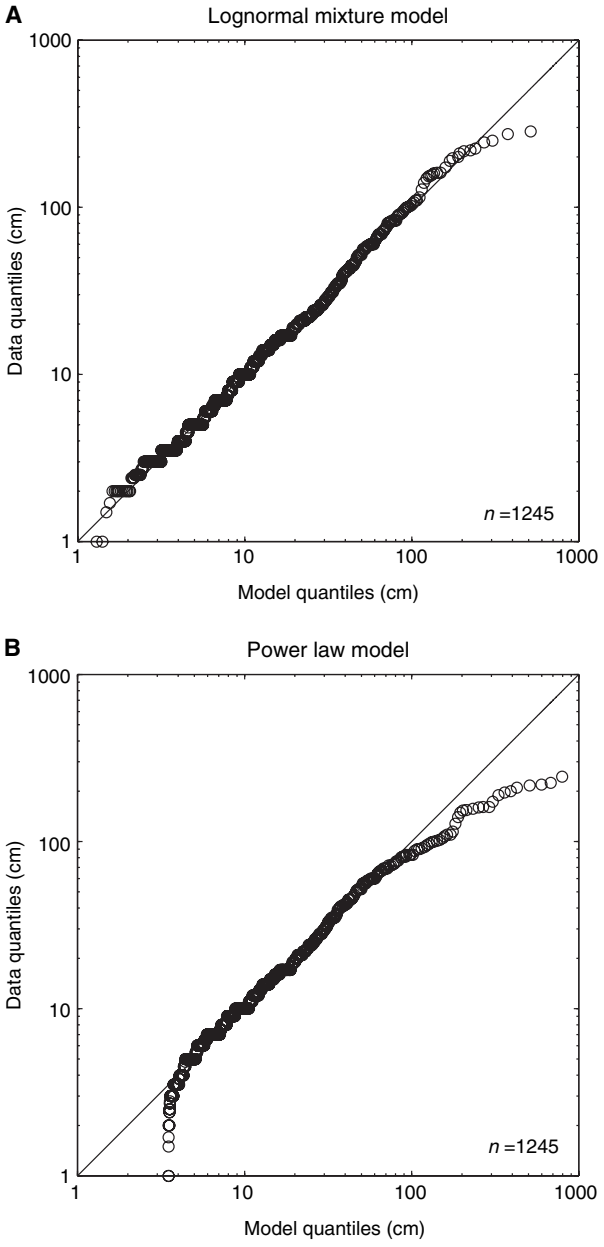

Fig. 17. (A) Quantile–quantile (qq) plot of the Upper Tarcău Sandstone bed thickness data, assuming a lognormal mixture model. (B) Quantile–quantile (qq) plot of the Upper Tarcău Sandstone bed thickness data, assuming a power law model.

dite bed thickness series that allow precise measurement of all bed thicknesses and indeed appear to be linear on such a plot remain interesting subjects of study. However, published turbidite bed thickness diagrams suggest that the power law distribution is the exception rather than the rule and explaining clearly non-power law distributions through various processes that modify an initial power law input remains speculative.

Bimodality of turbidite bed thicknesses

As pointed out by Talling (2001), the commonly observed bimodality of turbidite bed thickness distributions probably reflects the difference between deposition from more dilute and more dense suspensions, or the low-density and high-density turbidity currents of Lowe (1982). The thickness of a bed at any location depends on two immediate factors: duration of deposition and rate of sedimentation. It is probable that it is the rate of sedimentation that causes the bimodality: once it starts, deposition from high-concentration suspensions is likely to be rapid, as reflected by the common suppression of lamination or stratification and the relatively poor sorting in thick-bedded turbidites (Lowe, 1982; Hiscott, 1994; Sylvester & Lowe, 2004). From a slightly different perspective, the bimodality of bed thickness reflects the difference between competence-driven and capacity-driven sedimentation (Hiscott, 1994; Kneller & McCaffrey, 2003). Whether thick beds are usually related to larger originating events or strictly reflect local depositional conditions is a different and more difficult question. In any case, it is probable that bed thickness is controlled only in part by initial discharge (Pirmez & Imran, 2003) and flows depositing thick sand beds in channels can also deposit thin beds on levées or in more distal locations. Therefore, it is unlikely that bed thickness distributions of turbidite systems with well-differentiated morphologic and stratigraphic elements can be related directly to originating events like earthquakes; rather, they probably reflect the local depositional conditions (Talling, 2001). Taking into account the degree of complexity of many turbidite systems that can be observed on seafloor images and in high-quality seismic data (e.g. Deptuck *et al.*, 2003), it is also unlikely that a few numbers or curves derived from bed thickness data alone can give general guidance about depositional setting, degrees of confinement, erosion, bypass and other important characteristics of the turbidite system.

The two different types of deposition (from high-density and low-density suspensions) can occur probably from the same flow. Recent studies in the Marnoso-Arenacea Formation suggest that the bimodality in bed thicknesses can be related to a characteristic bed shape: beds that are relatively thick in proximal settings undergo rapid thinning towards more distal areas (Talling *et al.*, 2007). Any random vertical sampling line is likely to capture fewer transitional thicknesses compared with the laterally more extensive proximal thick and distal thin zones.

An important implication of this study is that segmented power laws and the corresponding characteristic β exponents on exceedence probability plots cannot be explained by different depositional processes leading to two or more different populations. For example, it is tempting to think that, in some cases, the two segments of the exceedence plot could be the result of deposition from turbidity currents vs. debris flows. If this was the case, one would expect to see a range of curve types representing power law mixtures, as shown in Fig. 8. There is no reason why different bed types could not mix in different proportions. However, there is no indication of 'mixed' curve shapes in the existing data sets; all the segmented plots can only correspond to one specific curve type of the family of curves illustrated in Fig. 8. At the moment, the only reasonable explanation for this curve type is the Malinverno (1997) model for confinement and, therefore, any claims made about the presence of segmented power law distributions should consider the applicability of the assumptions and implications of this mathematical representation to the data under investigation. In contrast, the lognormal mixture model is compatible with the two populations originating from different transport and/or depositional processes: the curve types seen on the cumulative plots of real data (Figs 12–14) correspond well to the various theoretical mixtures of two lognormal populations (Fig. 11). This model requires no relationship between the population parameters and the proportion between the two groups.

How important is erosion?

Pebbly sandstone packages with abundant amalgamation surfaces are common in the Tarcău Sandstone. If the erosional flows removed significant parts of the underlying beds, the bed thickness distribution was affected as well. A comparison of the distribution of truncated and non-truncated beds that start with T_a (Fig. 18) shows a discernible shift of the thick beds affected by erosion compared with non-eroded beds. The geometric mean values of the two distributions are 100 cm (134 non-truncated beds) and 75 cm (260 truncated beds). This suggests that, on average, most of the bed-scale erosion seen in the Tarcău Sandstone is restricted to *ca* 20–30% of the total bed thickness and does

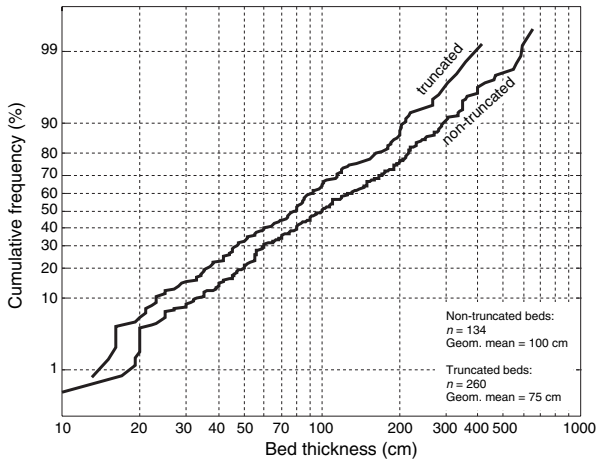

Fig. 18. Cumulative probability plot of thicknesses of beds starting with the T_a divisions, separated as a function of whether they were affected by erosion at the top (truncated) or not (non-truncated). Such a comparison gives an estimate of the magnitude of the bed-scale erosion.

not alter the lognormal nature of the original event bed thickness distribution.

Significance of thickness variability

It has been shown here that the Tarcău Sandstone bed thickness data represent a lognormal mixture different from the one that describes the Marnoso-Arenacea Formation. The log-transformed thinner-bedded population has a relatively small standard deviation, whereas the thick-bedded part shows more variability (Fig. 14). A possible cause for this is that the thin beds in the Tarcău Sandstone probably were deposited as levées and turbidites on levees of modern submarine fans have relatively uniform bed thicknesses (e.g. Pirmez & Imran, 2003). In contrast, the thin beds of the Marnoso-Arenacea Formation are *not* levée deposits; many thin beds have been correlated over most of the outcrop area, and they were deposited on a basin plain (Amy & Talling, 2006). Further studies, preferably of modern systems where the depositional settings are clearly known, are necessary to see whether the variances of the log-transformed bed thickness data, in conjunction with other facies characteristics, have some general implications for the interpretation of the deposits.

Practical advantages of the lognormal model

From a reservoir characterization perspective, it is worth noting that in most areas of research where power law distributions are of great interest (e.g. the study of networks, economics, natural hazard prediction), it is the 'heavy tail' that receives most of the attention, because it is important to predict rare but huge events. However, in reservoir characterization, the very large 'events', if any, are often already known from wells and seismic data. It is the subseismic scale, in areas unpenetrated by wells, that is largely unknown. On the other hand, very thin beds that usually cause the deficiencies in the bed thickness measurements commonly are not reservoirs and, from this practical point of view, the difficulties of measuring correctly the thinnest beds can probably be ignored. So a range of bed thicknesses, from a few centimetres to a few metres, are of most interest in reservoir characterization, and this range seems to be best described by the lognormal mixture model.

Regardless of what the theoretical implications and interpretations might be, the lognormal mixture model proposed by Talling (2001) is a relatively simple and intuitive model that offers a better description of the data analysed here than the power law models. Certainly, the statistical analysis of bed thicknesses can be refined further and other distribution models could be investigated as well. However, it is important to find the balance between the precision of the statistical models and their applicability: more complicated models are likely to give better fits, but they may not be practical for description and modelling of bed thicknesses. The lognormal mixture model may often fail to pass the goodness-of-fit statistical tests, especially with large data sets, but this should not preclude its use unless a better-fitting and relatively simple, sedimentologically insightful model is found.

CONCLUSIONS

1 Log–log plots and least squares fitting by themselves are inappropriate tools for the analysis of bed thickness distributions. Least squares fitting gives a reasonable estimate of the power law exponent, but says little about how well the distribution fits the power law model. The associated R^2 is unsuitable as a measure of goodness-of-fit. Log–log plots and least squares fitting should be accompanied by the assessment of other types of diagrams (cumulative probability, histogram of log-transformed values, q–q plots) and the use of a measure of goodness-of-fit other than R^2, such as the chi-squared and the Kolmogorov–Smirnov statistics.

2 Segmented power laws on exceedence plots are not simple mixtures of power law distributions with arbitrary parameters. Although a model of confinement does result in segmented plots at the centre of the basin (Malinverno, 1997), the segmented shape of the exceedence curve breaks down as the sampling location moves away from the basin centre. Identifying two or more relatively straight segments in exceedence probability plots is subjective, and should be used in the analysis of turbidite bed thickness data only if it is probable that the assumptions of the Malinverno (1997) model are satisfied.

3 The lognormal mixture model for turbidite bed thicknesses (Talling, 2001), mainly reflecting the differences in sedimentation rates between high-density and low-density turbidity currents, is a logical alternative to the power-law model. Identifying the parameters of the component distributions is difficult unless a detailed sedimentologic description of the internal structuring is available. It is suggested here that the expectation–maximization algorithm can be used to estimate the parameters and thus to model such bed thickness mixtures.

4 The bed thickness data from the Tarcău Sandstone of the East Carpathians (Romania) are best described by a lognormal mixture model with two components. In contrast with the Marnoso-Arenacea Formation, the thinner-bedded population has a lower variability than the thicker-bedded population; this could be the result of deposition of these beds on levees of submarine channels.

ACKNOWLEDGEMENTS

Part of this study was carried out with funding from the Stanford Project on Deepwater Depositional Systems (SPODDS), a consortium of sponsors that included Anadarko, BP, Chevron, Conoco, ENI-AGIP, ExxonMobil, JNOC, Marathon, Occidental, Perez Companc, Petrobras, Rohöl-Aufsuchungs AG, Shell, Texaco and Unocal. The publicly available notes of Cosma Shalizi on log–log plots have been influential in some of the discussions and conclusions of this paper. I am thankful for discussions with Donald Lowe, Carlos Pirmez, Stephan Graham, William Normark, Maarten Felix, Anikó Sándor; and for reviews and comments by Peter Haughton and Peter Talling.

REFERENCES

Aitchison, J. and **Brown, J.A.C.** (1957) *The Log-Normal Distribution: With Special Reference to its Uses in Economics*. Cambridge University Press, Cambridge, 176 pp.

Amy, L.A. and **Talling, P.J.** (2006) Anatomy of turbidites and linked debrites based on long distance (120 × 30 km) bed correlation, Marnoso Arenacea Formation, Northern Apennines, Italy. *Sedimentology*, **53,** 161–212.

Awadallah, S.A.M., Hiscott, R.N., Bidgood, M. and **Crowther, T.** (2001) Turbidite facies and bed-thickness characteristics inferred from microresistivity images of Lower to Upper Pliocene rift-basin deposits, Woodlark Basin, offshore Papua New Guinea (Eds P. Huchon, B. Taylor and A. Klaus), *Proc. ODP, Sci. Results*, **180,** 1–30.

Bak, P. (1996) *How Nature Works: The Science of Self-organized Criticality*. Springer Verlag, New York, 212 pp.

Bak, P., Tang, C. and **Wiesenfeld, K.** (1988) Self-organized criticality. *Phys. Rev. A*, **38,** 364–374.

Beattie, P. and **Dade, W.B.** (1996) Is scaling in turbidite deposition consistent with forcing by earthquakes? *J. Sed. Res.*, **66,** 909–915.

Bouma, A.H. (1962) *Sedimentology of Some Flysch Deposits; A Graphic Approach to Facies Interpretation*. Elsevier, Amsterdam, New York, 168 pp.

Buatois, L.A., Mangano, M.G. and **Sylvester, Z.** (2001) A diverse deep-marine ichnofauna from the Eocene Tarcău Sandstone of the Eastern Carpathians, Romania. *Ichnos*, **8,** 23–62.

Carlson, J. and **Grotzinger, J.P.** (2001) Submarine fan environment inferred from turbidite thickness distributions. *Sedimentology*, **48,** 1331–1351.

Chakraborty, P.P., Mukhopadhyay, B., Pal, T. and **Gupta, T.D.** (2002) Statistical appraisal of bed thickness patterns in turbidite successions, Andaman Flysch Group, Andaman Islands, India. *J. Asian Earth Sci.*, **21,** 189–196.

Chen, C. and **Hiscott, R.N.** (1999) Statistical analysis of facies clustering in submarine-fan turbidite successions. *J. Sed. Res.*, **69,** 505–517.

Clark, B.E. and **Steel, R.J.** (2006) Eocene turbidite – population statistics from shelf edge to basin floor, Spitsbergen, Svalbard. *J. Sed. Res.*, **76,** 903–918.

Contescu, L., Jipa, D., Mihailescu, N. and **Panin, N.** (1967) The internal Paleogene flysch of the eastern Carpathians; paleocurrents, source areas and facies significance. *Sedimentology*, **7,** 307–321.

Csontos, L. (1995) Tertiary tectonic evolution of the Intra-Carpathian area; a review. *Acta Vulcanol.*, **7,** 1–13.

Davis, J.C. (2002) *Statistics and Data Analysis in Geology*, 3rd edn. John Wiley and Sons, New York, 640 pp.

Dempster, A.P., Laird, N.M. and **Rubin, D.B.** (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B*, **39,** 1–22.

Deptuck, M.E., Steffens, G.S., Barton, M. and **Pirmez, C.** (2003) Architecture and evolution of upper fan channel-belts on the Niger Delta slope and in the Arabian Sea. *Mar. Petrol. Geol.*, **20,** 649–676.

Drinkwater, N.J. and **Pickering, K.T.** (2001) Architectural elements in a high-continuity sand-prone turbidite system, Late Precambrian Kongsfjord Formation, Northern Norway: application to hydrocarbon reservoir characterization. *AAPG Bull.*, **85,** 1731–1757.

Drummond, C. and **Coates, J.** (2000) Exploring the statistics of sedimentary bed thicknesses – two case studies. *J. Geosci. Educ.*, **48,** 487–499.

Drummond, C.N. and **Wilkinson, B.H.** (1996) Stratal thickness frequencies and the prevalence of orderedness in stratigraphic sections. *J. Geol.*, **104,** 1–18.

Enos, P. (1969) Anatomy of a flysch. *J. Sed. Petrol.*, **39,** 680–723.

Flint, S. and **Bryant, I.D.** (1993) The geological modelling of hydrocarbon reservoirs and outcrop analogues. *Int. Assoc. Sedimentol. Spec. Publ.*, **15,** 269 pp.

Goldstein, M.L., Morris, S.A. and **Yen, G.G.** (2004) Problems with fitting to the power law distribution. *Eur. Phys. J. B*, **41,** 255–258.

Hiscott, R.N. (1994) Loss of capacity, not competence, as the fundamental process governing deposition from turbidity currents. *J. Sed. Res.*, **A64,** 209–214.

Hiscott, R.N., Colella, A., Pezard, P., Lovell, M.A. and **Malinverno, A.** (1992) Sedimentology of deep-water volcaniclastics, Oligocene Izu-Bonin intraoceanic forearc basin, based on formation microscanner images (Eds B. Taylor and K. Fujioka), *Proc. ODP Sci. Results*, **126,** 75–96.

Hiscott, R.N., Colella, A., Pezard, P., Lovell, M.A. and **Malinverno, A.** (1993) Basin plain turbidite succession of the Oligocene Izu-Bonin intraoceanic forearc basin. *Mar. Petrol. Geol.*, **10,** 450–466.

Jipa, D. (1967) Relationship between longitudinal and transversal currents in the Paleogene of the Tarcău valley (eastern Carpathians). *Sedimentology*, **7,** 299–305.

Kneller, B.C. and **McCaffrey, W.D.** (2003) The interpretation of vertical sequences in turbidite beds; the influence of longitudinal flow structure. *J. Sed. Res.*, **73,** 706–713.

Kolmogorov, A.N. (1951) Solution of a problem in probability theory connected with the problem of the mechanics of stratification. *Ann. Maths Soc. Trans.*, **53,** 171–177.

Kung, S.Y., Mak, M.W. and **Lin, S.H.** (2004) *Biometric Authentication: A Machine Learning Approach*. Prentice Hall, Englewood Cliffs, 496 pp.

Limpert, E., Stahel, W.A. and **Abbt, M.** (2001) Lognormal distributions across the sciences: keys and clues. *Bioscience*, **51,** 341–352.

Lowe, D.R. (1982) Sediment gravity flows; II, Depositional models with special reference to the deposits of high-density turbidity currents. *J. Sed. Petrol.*, **52,** 279–297.

Malinverno, A. (1997) On the power law size distribution of turbidite beds. *Basin Res.*, **9,** 263–274.

Mandelbrot, B.B. (1983) *The Fractal Geometry of Nature*. Freeman, New York, 468 pp.

Mattern, F. (2002) Amalgamation surfaces, bed thicknesses, and dish structures in sand-rich submarine fans: numeric differences in channelized and unchannelized deposits and their diagnostic value. *Sed. Geol.*, **150,** 203–228.

McBride, E.F. (1962) Flysch and associated beds of the Martinsburg Formation (Ordovician), Central Appalachians. *J. Sed. Petrol.*, **32,** 39–91.

McLachlan, G. and **Peel, D.** (2000) *Finite Mixture Models*. Wiley Series in Probability and Statistics, Wiley-Interscience, New York, 456 pp.

Mitzenmacher, M. (2004) A brief history of generative models for power law and lognormal distributions. *Internet Math.*, **1,** 226–251.

Mizutani, S. and **Hattori, I.** (1972) Stochastic analysis of bed-thickness distribution of sediments. *Math. Geol.*, **4,** 123–146.

Murgeanu, G., Dumitrescu, I., Săndulescu, M., Bandrabur, I. and **Săndulescu, J.** (1968) *Geological map of Romania, 1:200,000, sheet 29 (Covasna)*. Romanian Geological Institute, Bucharest.

Murray, C.J., Lowe, D.R., Graham, S.A., Martinez, P.A., Zeng, J., Carroll, A., Cox, R., Hendrix, M., Heubeck, C., Miller, D., Moxon, I.W., Sobel, E., Wendebourg, J. and **Williams, T.** (1996) Statistical analysis of bed-thickness patterns in a turbidite section from the Great Valley Sequence, Cache Creek California. *J. Sed. Petrol.*, **66,** 900–908.

Muto, T. (1995) The Kolmogorov model of bed-thickness distribution: an assessment based on numerical simulation and field data analysis. *Terra Nova*, **7,** 417–423.

Nabney, I.T. (2001) *NETLAB: Algorithms for Pattern Recognition*. Springer, London, 420 pp.

Normark, W.R. and **Piper, D.J.W.** (1991) Initiation processes and flow evolution of turbidity currents; implications for the depositional record. *SEPM Spec. Publ.*, **46,** 207–230.

Pearson, K. (1894) Contributions to the mathematical theory of evolution. *Phil. Trans. Roy. Soc. London A*, **185,** 71–110.

Pirmez, C. and **Imran, J.** (2003) Reconstruction of turbidity currents in Amazon Channel. *Mar. Petrol. Geol.*, **20,** 823–849.

Pirmez, C., Hiscott, R.N. and **Kronen, J.D. Jr**, (1997) Sandy turbidite successions at the base of channel levee systems of the Amazon Fan revealed by FMS logs and cores: unraveling the facies architecture of large submarine fans. *Proc. ODP Sci. Results*, **155,** 7–33.

Pyrcz, M.J., Cătuneanu, O. and **Deutsch, C.V.** (2005) Stochastic surface-based modeling of turbidite lobes. *AAPG Bull.*, **89,** 177–191.

Ricci Lucchi, F. and **Valmori, E.** (1980) Basin-wide turbidites in a Miocene, over-supplied deep-sea plain: a geometric analysis. *Sedimentology*, **27,** 241–270.

Rothman, D.H. and **Grotzinger, J.P.** (1995) Scaling properties of gravity-driven sediments. *Nonlinear Processes Geophys.*, **2,** 178–185.

Rothman, D.H., Grotzinger, J.P. and **Flemings, P.** (1994) Scaling in turbidite deposition. *J. Sed. Res.*, **0A64,** 59–67.

Săndulescu, M. (1988) Cenozoic tectonic history of the Carpathians. In: *The Pannonian Basin – A Study in Basin Evolution* (Eds L.H. Royden and F. Horvath), *AAPG Mem.*, **45,** 17–25.

Săndulescu, M. and **Săndulescu, J.** (1973) La stratigraphie du facies du gres du Tarcău et sa position dans le schema stratigraphique du flysch Paleogene des Carpathes Orientales et septentrionales. *Assoc. Geol. Carpatho-Balkanique, Congr. Bull.*, **1,** 189–200.

Sinclair, H.D. and **Cowie, P.A.** (2003) Basin-Floor Topography and the Scaling of Turbidites. *J. Geol.*, **111,** 277–299.

Spencer, D.W. (1963) The interpretation of grain size distribution curves of clastic sediments. *J. Sed. Petrol.*, **33,** 180–190.

Swan, A.R.H. and **Sandilands, M.** (1995) *Introduction to Geological Data Analysis*. Blackwell Science, Oxford, 446 pp.

Sylvester, Z. (2002) *Facies, architecture, and bed-thickness structure of turbidite systems: Examples from the East Carpathian Flysch, Romania, and the Great Valley Group, California*, Ph.D. Thesis. Stanford University, Stanford, CA, 294 pp.

Sylvester, Z. and **Lowe, D.R.** (2004) Textural trends in turbidites and slurry beds from the Oligocene flysch of the East Carpathians, Romania. *Sedimentology*, **51,** 945–972.

Talling, P.J. (2001) On the frequency distribution of turbidite thickness. *Sedimentology*, **48,** 1297–1331.

Talling, P.J., Amy, L.A. and **Wynn, R.B.** (2007) New insight into the evolution of large-volume turbidity currents: comparision of turbidite shape and previous modelling results. *Sedimentology*, doi: 10.1111/j.1365-3091.2007.00858.x.

Tanner, W.F. (1964) Modification of sediment size distributions. *J. Sed. Petrol.*, **34,** 156–164.

Turcotte, D.L. (1997) *Fractals and Chaos in Geology and Geophysics*. Cambridge University Press, New York, 370 pp.

Turcotte, D.L. and **Huang, J.** (1995) Fractal distributions in geology, scale invariance, and deterministic chaos. In: *Fractals in the Earth Sciences* (Eds P.R. La Pointe and C.C. Barton), pp. 1–40. Plenum Press, New York.

Tye, R.S. (2004) Geomorphology: an approach to determining subsurface reservoir dimensions. *AAPG Bull.*, **88,** 1123–1147.

Winkler, W. and **Gawenda, P.** (1999) Distinguishing climatic and tectonic forcing of turbidite sedimentation, and the bearing on turbidite bed scaling: Palaeocene-Eocene of northern Spain. *J. Geol. Soc. London*, **156,** 791–800.

Zweigel, P., Ratschbacher, L. and **Frisch, W.** (1998) Kinematics of an arcuate fold-thrust belt: thesouthern Eastern Carpathians (Romania). *Tectonophysics*, **297,** 177–207.

Manuscript received 19 October 2005; revision accepted 26 January 2007