# Chapter 26
# Statistical Modeling of Regional and Worldwide Size-Frequency Distributions of Metal Deposits


**Frits Agterberg**

**Abstract** Publicly available large metal deposit size data bases allow new kinds of statistical modeling of regional and worldwide metal resources. The two models most frequently used are lognormal size-grade and Pareto upper tail modeling. These two approaches can be combined with one another in applications of the Pareto-lognormal size-frequency distribution model. The six metals considered in this chapter are copper, zinc, lead, nickel, molybdenum and silver. The worldwide metal size-frequency distributions for these metals are similar indicating that a central, basic lognormal distribution is flanked by two Pareto distributions from which it is separated by upper and lower tail bridge functions. The lower tail Pareto distribution shows an excess of small deposits which are not economically important. Number frequencies of the upper tail Pareto are mostly less than those of the basic lognormal. Parameters of regional metal size-frequency distributions are probably less than those of the worldwide distributions. Uranium differs from other metals in that its worldwide size-frequency distribution is approximately lognormal. This may indicate that the lognormal model remains valid as a standard model of size-frequency distribution not only for uranium but also for the metals considered in this chapter, which are predominantly mined from hydrothermal and porphyry-type orebodies. A new version of the model of de Wijs may provide a framework for explaining differences between regional and worldwide distributions. The Pareto tails may reflect history of mining methods with bulk mining taking over from earlier methods in the 20th century. A new method of estimating the Pareto coefficients of the economically important upper tails of the metal size-frequency distributions is presented. A non-parametric method for long-term projection of future metal resource on the basis of past discovery trend is illustrated for copper.

**Keywords** Pareto-lognormal distribution · Size-frequency distributions Worldwide metal resources · Future metal supply · Model of de Wijs

F. Agterberg (✉)
Geological Survey of Canada, 601 Booth Street, Ottawa, ON K1A 0E8, Canada
e-mail: frits.agterberg@canada.ca; frits@rogers.com



B. S. Daya Sagar et al. (eds.), *Handbook of Mathematical Geosciences*,
https://doi.org/10.1007/978-3-319-78999-6_26

## 26.1 Introduction

Most models for regional or worldwide mineral or hydrocarbon resource appraisal assume either a lognormal or a Pareto model for the size-frequency distribution of the deposits considered. It can also be assumed that both models apply with the lognormal distribution providing a good fit to all sizes except for the smallest and largest deposits that satisfy fractal/multifractal Pareto distributions. The largest deposits obviously are rare and may be too few in number for adequate modeling in regional studies. However, recently, very large data bases have become available for metal deposits (Patiño Douce 2016a, b, c, 2017). In a newly proposed Pareto-lognormal model for worldwide metal deposit size-frequency distributions (Agterberg 2017a, b, in press), a basic lognormal distribution is flanked by two Pareto distributions. In this chapter this model is applied to copper, zinc, lead, nickel, molybdenum and silver. The upper and lower tail Pareto's are separated from the central lognormal by bridge functions to ensure continuity. An improved version of the Pareto-lognormal model will be applied to the upper tails of the size-frequency distributions for the six metals considered.

Previously, this approach was also applied to gold and uranium (Agterberg 2017b). For gold, the Pareto-lognormal model is not fully satisfied in that there is a shortage of gold deposits with sizes in the vicinity of the median of the worldwide gold size-frequency distribution. For uranium (size measured in tons of $U_3O_8$), a lognormal size-frequency distribution without Pareto tails provides a good fit. In the earlier publications (Agterberg 2017a, b, in press) comparisons were made between regional and worldwide size-frequency distributions for copper and gold. Logarithmic variances of worldwide size-frequency distributions exceed those of regional distributions and worldwide separate mineral deposit-type distributions. This observation also applies to the upper tail Pareto size-frequency distributions. A new variant of the model of de Wijs, to be discussed in more detail later in this chapter, can provide a partial explanation of the fact that the worldwide basic lognormal can be regarded as a mixture of regional lognormal distributions with parameters less than those of the worldwide basic lognormals and Pareto's. For example, within the Abitibi volcanic belt on the Precambrian Canadian Shield, the largest deposits for copper and gold satisfy Pareto size-frequency distributions with Pareto parameters ($\alpha_{Cu} = 0.45$; $\alpha_{Au} = 0.88$) that are less than those of their worldwide distributions ($\alpha_{Cu} = 1.21$; $\alpha_{Au} = 1.16$) illustrating that upper tail size parameter estimates for individual metal deposits are not stochastically independent data but subject to spatial correlation.

It should be pointed out that worldwide size-frequency distributions for some metals including copper (2541 deposits) are sufficiently large so that original data (without use of parametric statistical models) can be employed for long-term projections into the future at specific cut-off metal sizes (Agterberg 2017b; also see later in this chapter). Main emphasis in this chapter will be on size-frequency distribution modeling of the upper tail Pareto distribution and its transition into the basic lognormal. This is because total amount of metal in the lower tail of each Pareto-lognormal distribution is negligibly small. For example, 1340 copper

deposits with greater than median size contain 99.7% of all copper in the complete data set of 2541 deposits so that information provided by the approximately 50% smaller deposits can be neglected (cf. Patiño Douce 2016c).

Patiño Douce (2016a, b, c, 2017) has published four important papers that are helpful in planning future metal supply; showing, for example, that for copper there would be a deficit of about $2.39 \times 10^9$ t (tonnes) by the end of this century if recent discovery rates are maintained. For comparison, according to the USGS Mineral Commodity Summaries (2015), proven copper reserves currently are $0.68 \times 10^9$ t. According to Patiño Douce (2017), current copper resources including the estimated reserves are $2.32 \times 10^9$ t whereas new demand by 2100 will be $4.70 \times 10^9$ t. Consequently, estimated future copper deficit is approximately equal to currently known copper resources. Using a non-parametric statistical method, this forecast was confirmed by Agterberg (2017b) who estimated copper resources to be discovered by the end of this century at $2.77 \times 10^9$ t with 95% confidence interval of $\pm 0.994 \times 10^9$ that contains Patiño Douce's estimate (also see Sect. 26.5).

Patiño Douce (2016b) is accompanied by a supplementary database with sizes and grades for 20 metals. For example, his data on 2541 copper deposits were compiled from as many as 49 different sources. Patiño Douce (2016b) initially fitted lognormal distributions to the metal deposit size-frequency distributions in this data base pointing out that the logarithmic (base $e$) standard deviation ranges from about 2 to 3 for different metals, although average metal deposit sizes are greatly different. Both Patiño Douce (2016c) and Agterberg (2017a) showed that the largest deposits for different metals can be described by means of Pareto distributions. In the Pareto-lognormal metal size-frequency distribution model of Agterberg (2017a, b) the lognormal has a Pareto upper tail separated from the central lognormal by a bridge zone. This model recognizes both (1) lognormality of metal content of ore deposits from within smaller regions and those belonging to different mineral deposit types (see, e.g. Singer 2013), and (2) Pareto size-frequency distribution of the largest deposits but also for the economically unimportant smallest metal deposits that exhibit Pareto size-frequency distributions as well.

The Pareto-lognormal model for metal deposits provides an alternative to other size-frequency distribution models, which until about 30 years ago almost exclusively were based on the lognormal model. Mandelbrot (1983, p. 263) stated that oil and other natural resources have Pareto distributions and "this finding disagrees with the dominant opinion, that the quantities in question are lognormally distributed. This difference is extremely significant, the reserves being much higher under the hyperbolic than under the lognormal law." It will be seen in this chapter that size frequencies in the upper Pareto tails of the worldwide metal deposits taken for example are less than those of the basic lognormals when these are projected to the largest sizes. In this sense, the metal size frequency distributions are not "heavy-tailed". It can, however, be assumed that the Pareto represents a stable limiting form for the largest as well as the smallest metal deposits. Pareto size-frequency distribution modeling of the largest deposits has during the past 35 years been used by many authors including Drew et al. (1982) and Crovelli (1995) for oil and gas fields, and Cargill (1981), Cargill et al. (1980, 1981) and

Turcotte (1997) for metal deposits. The latter author has developed a modification of the model of de Wijs (1951) that results in a Pareto instead of a lognormal distribution. Turcotte (1997) based this model on original publications by Cargill et al. (1980, 1981) and Cargill (1981) who had assumed power-law instead of lognormal models for U.S. mercury, lode gold and copper production. Like the lognormal, the Pareto-lognormal distribution is not universally applicable to all elements, which show bimodal or multimodal size-frequency distributions when all the many different rock bodies within the Earth's crust would be considered.

The fact that uranium has lognormal distribution without Pareto tails suggests that a multiplicative form of the central limit theorem is applicable for this metal and possibly for other metals in different kinds of mineral deposits as well. A new variant of the model of de Wijs described in the next section provides a partial explanation of the fact that the basic lognormal probably can be regarded as a mixture of regional lognormals with parameters that are less than those of the worldwide basic lognormal.

## 26.2 Modified Version of the Model of de Wijs Applied to Worldwide Metal Deposits

In the original model of de Wijs (1951) for metal concentration values in blocks of rock, any block with metal concentration model $\zeta$ is repeatedly divided into halves with concentration values $(1 + d) \cdot \zeta$ and $(1 - d) \cdot \zeta$ where $d$ is the coefficient of dispersion which us assumed to be independent of block size. The frequency distribution for metal concentration values in increasingly smaller blocks then satisfies the so-called logbinomial distribution that rapidly approaches lognormal form. If there are $p$ subdivisions, the logbinomial distribution of the $\binom{p}{K}$ concentration values of the resulting $n = 2^p$ blocks is

$$X(p,K) = \zeta \cdot (1+d)^{p-K}(1-d)^{K}$$

where $K$ satisfies the binomial distribution with $\mu(K) = p/2$ and variance $\sigma^2(K) = p/4$ (cf. Agterberg 1974, p. 322). This logbinomial has $\mu(X) = \zeta$ and variance $\sigma^2(X)$ approaching to:

$$\sigma^2(X) = \frac{p}{4} \cdot \left[\ln\frac{1-d}{1+d}\right]^2$$

Various modifications of the original model of de Wijs (1951) were developed by Matheron (1962), Turcotte (1997) and Agterberg (2007). These modifications were primarily concerned with randomizing the model of the Wijs (e.g. in the

random-cut model), spatial realizations to account for spatial autocorrelation, maximizing $p$ (three-parameter model of de Wijs) and producing a Pareto tail (or other types of tail) on the logbinomial (e.g., as in the accelerated dispersion model, Agterberg (2007)). As discussed by Mandelbrot (1983), the model of de Wijs was the earliest example of a multifractal cascade. Lovejoy and Schertzer (2007) have pointed out that this original cascade is micro-canonical in that average metal concentration value is preserved locally at every cut. In universal multifractal theory these authors have generalized the cascade-type approach by preserving regionalized instead of strictly local averages. Their approach can result in a cascade that is largely lognormal but generates tails which are exactly Pareto-type. Here, another modified version of the original model of de Wijs (1951) is introduced as follows.

Suppose that the sizes of all deposits are combined with one another into a single very large block which is assigned to an arbitrary point in the upper part of the Earth's crust that contains metal deposits that have been or can be discovered. Suppose further that this block is divided into halves and the two smaller blocks are assigned to two points randomly located within halves of the upper part of the Earth's crust. This process can be repeated $2^p$ times. At each step, the two resulting half-blocks of metal are further divided into halves that, after every cut, are randomly assigned to successively smaller segments of the upper Earth's crust. If there are $n$ known deposits the cascade process is repeated until $n \leq 2^p$. For example, in relatively well-known parts of the Earth's crust there occur 2541 copper deposits. Suppose that $p = 12$ so that total number of subdivisions would be 4096. The 2541 copper deposits then can be regarded as a random subset of this larger population, so that the overall mean copper content value $\zeta$ and the coefficient of dispersion $d$ can be estimated. From the parameters of the straight line representing the basic copper lognormal distribution (Fig. 26.2a, see later) it follows that the logarithmic (base $e$) mean and standard deviation are $\mu = 10.445$ and $\sigma = 3.1062$. Consequently, $\zeta = \exp(\mu + \sigma^2/2) = 4.277 \times 10^9$. It then follows that $d = 0.7276$.

"Observed" frequencies satisfying the log-binomial model are shown in Fig. 26.1. The best-fitting straight line ($y = 0.755x - 3.8123$) in this diagram has coefficients corresponding to mean $\mu' = 11.627$ and standard deviation $\sigma' = 3.050$ which are relatively poor estimates in comparison to the values to derived later for the basic lognormal for copper in Fig. 26.2a. Main reason for this minor discrepancy is relatively strong influence on the best-fitting regression line of logbinomial frequencies represented by first and last points which are for single blocks only. Positions of these two points illustrate that the logbinomial produces slightly weaker upper and lower tails in comparison with the lognormal. On the whole, the logbinomial very closely approximates the lognormal in this application.

The preceding model would allow for spatial autocorrelation of metal deposit size observations, which is known to exist. For example, the largest copper deposits are porphyry type and largely clustered in the Andes mountain chain of South America. On the other hand, the largest copper deposits in the Abitibi volcanic belt on the Canadian Shield are volcanogenic massive sulphide deposits which are smaller than the South American porphyry coppers. Because of the close resemblance of the

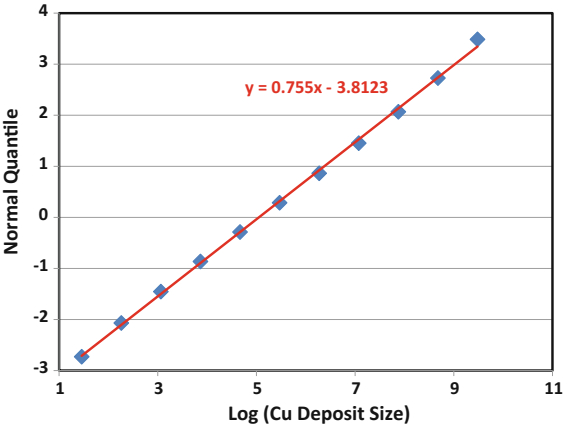


**Fig. 26.1** Model of de Wijs applied to worldwide copper deposit size-frequency distribution. Overall mean set equal to $\zeta = 4.277$ Mt copper; dispersion index $d = 0.7276$; number of subdivisions $p = 12$. "Observed" frequencies satisfy log-binomial model. Best-fitting straight line represents lognormal distribution. Logbinomial frequencies represented by first and last point are for single blocks only (*Source* Agterberg, in press)

logbinomial to the lognormal, preceding results also can be represented as follows. The characteristic function of a random variable $X$ is:

$$g(u) = E(e^{iux}) = \int_{-\infty}^{\infty} e^{-iux} f(x)dx$$

where $f(x)$ is the probability density function of $X$. Characteristic functions are discussed in statistical textbooks including Billingsley (1986) and Bickel and Doksum (2001). For a normal distribution:

$$g(u) = e^{i\mu u - \sigma^2 u^2/2}$$

If $Z$, with mean $\mu_z$ and variance $\sigma_z^2$, represents the sum of two random variables $X$ and $Y$, then the respective three characteristic functions satisfy:

$$g_z(u) = g_x(u) \cdot g_y(u)$$

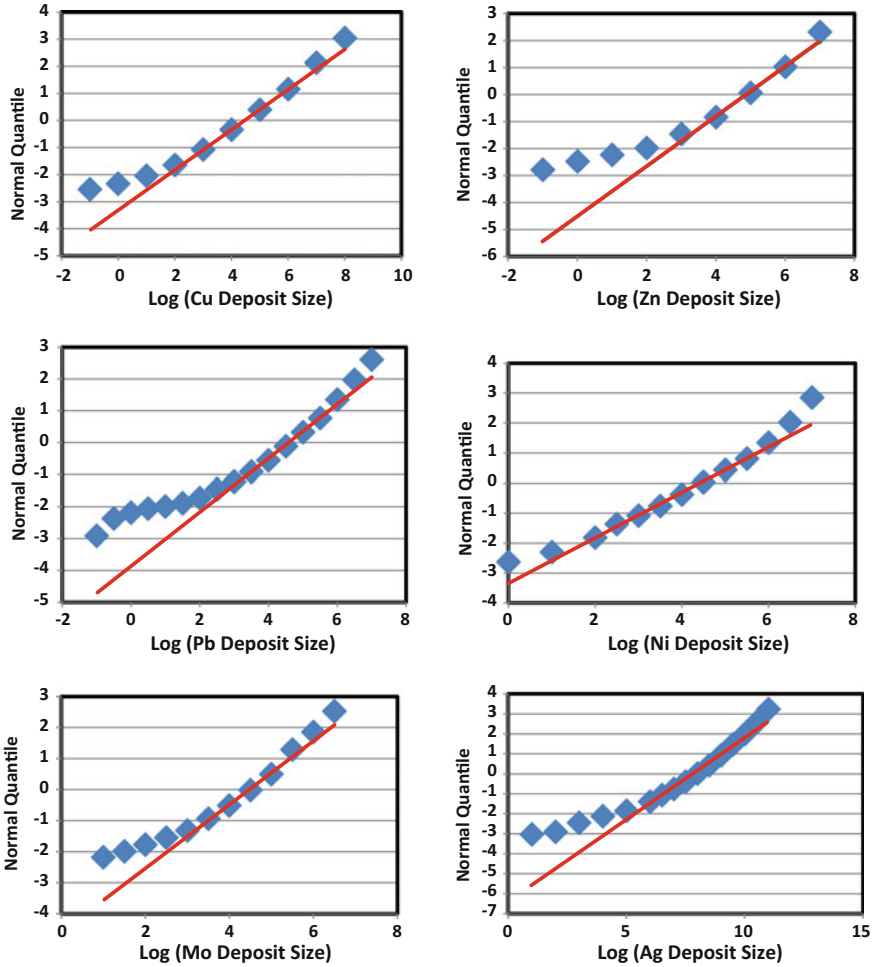


**Fig. 26.2** Lognormal *Q-Q* plots for six metals (Cu, Zn, Pb, Ni, Mo and Ag). Coefficients of straight lines representing truncated lognormal distributions are shown in Table 26.1. Sample sizes are shown in Table 26.2. In each case, frequencies for the largest and smallest deposits deviate from the straight-line pattern indicating lower and higher number frequencies than expected on the basis of the lognormal size frequency distribution models represented by the straight lines

If $X$ is normal with zero mean and variance $\sigma_x^2$, and $Y$ is normal as well with mean $\mu_y$ and variance $\sigma_y^2$, then $Z$ is normal with:

$$g_z(u) = e^{\left[i(\mu_x + \mu_y) \cdot u - (\sigma_x^2 + \sigma_y^2) \cdot u^2/2\right]}$$

Consequently, the probability density function of $Z$ is:

$$f(z) = \frac{1}{\sqrt{\sigma_x^2 + \sigma_y^2} \cdot \sqrt{2\pi}} e^{-\left\{z-(\mu_x+\mu_y)\right\}^2 \cdot \left\{2 \cdot (\sigma_x^2+\sigma_y^2)\right\}^{-1}}$$

Interpretation of this result in the context of worldwide metal deposits can be as follows. Suppose that log $Z$ represents the basic lognormal metal deposit size-frequency distribution with logarithmic mean $\mu_z = \mu_x + \mu_y$ and logarithmic variance $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$. Then log $Z$ can be regarded as a composite of many regional lognormal distributions with different means and lesser logarithmic variances, much like as in the previous version of the model of de Wijs the overall logbinomial would consist of regional logbinomials with different parameters.

## 26.3 Theory and Applications of the Pareto-Lognormal Model

The cumulative frequency distribution for the Pareto-lognormal distribution $F(x) = F(\log x)$ can be written as

$$F(\log x) \approx \Phi\left(\frac{\log x - \mu}{\sigma}\right) + H(\log x - \mu) \cdot B_1(\log x) \cdot (\log x - \mu)^{-\alpha} + H(\mu - \log x) \cdot B_2(\log x) \cdot (\mu - \log x)^{-\kappa}$$

where $\Phi\left(\frac{\log x - \mu}{\mu}\right)$ represents the basic lognormal (logs base 10). $H$ (…) is the Heaviside function that applies to two filtered Pareto distributions, for positive and negative values of (log $x$ - $\mu$), respectively; it signifies that values at the other side of $\mu$ are set equal to zero when the equation is applied to either the upper tail or the lower tail of the Pareto-lognormal distribution.. The bridge functions $B_1(\log x)$ and $B_2(\log x)$ span relatively short intervals between the basic lognormal and the Pareto distributions for the largest and smallest values, respectively. They satisfy $\lim_{x\to\infty} B_1(\log x) = \lim_{x\to 0} B_2(\log x) = 1$ and $\lim_{x\to 0} B_1(\log x) = \lim_{x\to\infty} B_2(\log x) = 0$.

The Pareto-lognormal probability density function $f(\log x)$ corresponding to $F(\log x)$ can be written as

$$f(\log x) \approx \varphi\left(\frac{\log x - \mu}{\sigma}\right) + H(\log x - \mu) \cdot B'_1(\log x) \cdot (\log x - \mu)^{-\alpha - 1} + H(\mu - \log x) \cdot B'_2(\log x) \cdot (\mu - \log x)^{-\kappa - 1}$$

It may be useful for prediction of resources to be discovered in the future. The exponents in $(\log x - \mu)^{-\alpha-1}$ and $(\mu - \log x)^{-\kappa-1}$ reflect the fact that the Pareto probability density function remains linear on a plot with logarithmic scales for both frequency and deposit size, but has a steeper dip.

The lognormal *QQ*-plot (logarithmic probability paper) provides a useful first step in fitting the Pareto-lognormal distribution. Figure 26.2 contains results for the six metals. Original data were taken from Patiño Douce (2016b). Each graph shows a straight-line pattern with departures from lognormality in the upper and lower frequency distribution tail. Relatively, there are too many smallest deposits and too few largest deposits. In the Pareto-lognormal model both the upper and lower tail distributions have transitions to the central lognormal that are gradual and described by the two bridge functions. For projections into the future (or for global downward projections into the Earth's crust) only the upper tails of the size-frequency distributions are of economic interest. In the next section, a new, relatively simple method will be described for fitting the upper tail Pareto distributions. The upper tail bridge function will be fitted empirically by connecting this Pareto to the central lognormal distribution. Copper can be used for illustrating details of the methods used. The straight line $y = bx + a$ in Fig. 26.2a for copper represents the basic lognormal with coefficients $a = -3.314$ and $b = 0.741$ derived from the logarithmic mean $\mu = -a/b = 4.469$ and standard deviation $\sigma = 1/b = 1.349$ of a truncated lognormal for which 10% (or 254 values) in both upper and lower tail were excluded from the sample of 2541 original copper deposit size values. The mean $\mu$ of this truncated distribution is only slightly different from 4.403 representing the logarithmic mean of all values. The basic lognormal standard deviation $\sigma = 1.349$ is slightly less than 1.423 representing the standard deviation based on all values because there are relatively many copper deposit size values in the lower tail. It was obtained by dividing 0.893 representing the standard deviation of the truncated copper data set by 0.662, representing a value taken from Johnson and Kotz (1970, Table 10, p. 84). Other published truncation correction factors were used for metals with wider upper or lower tails. Coefficients for all six straight lines shown in Fig. 26.2 are given in Table 26.1. The basic statistics estimated for all six metals shown in Table 26.2 were taken from Agterberg (2017a, b and in press) except for the upper tail Pareto coefficients with slightly different values newly derived by the method to be described in the next section.

**Table 26.1** Constants $a$ and $b$ in equations $y = a + bx$ for straight lines shown in Fig. 26.2 representing truncated lognormals for six metals

| Metal | $a$ | $b$ |
|---|---|---|
| Cu | −3.314 | 0.741 |
| Zn | −4.538 | 0.924 |
| Pb | −3.894 | 0.847 |
| Ni | −3.400 | 0.758 |
| Mo | −4.616 | 1.030 |
| Ag | −6.501 | 0.819 |

**Table 26.2** Comparison of basic statistics for eight metals including the six metals represented in Table 26.1 and Figs. 26.2, 26.3 and 26.4. N—number of deposits; Mt—million tons, t metric tons; LM, LS—logarithmic mean and standard deviation; $\mu$, $\sigma$—ditto for truncated lognormal; $\alpha$, $\kappa$—upper and lower tail Pareto coefficients

| Metal | N | Total metal | Mean | LM | LS | $\mu$ | $\sigma$ | $\alpha$ | $\kappa$ |
|---|---|---|---|---|---|---|---|---|---|
| Cu | 2541 | 2319.11 Mt | 0.9127 Mt | 4.403 | 1.423 | 4.469 | 1.349 | 1.206 | 0.332 |
| Zn | 1476 | 1111.51 Mt | 0.7531 Mt | 4.821 | 1.215 | 4.910 | 1.082 | 1.162 | 0.318 |
| Pb | 1102 | 481.43 Mt | 0.4367 Mt | 4.479 | 1.337 | 4.596 | 1.180 | 1.654 | 0.340 |
| Ni | 464 | 171.05 Mt | 0.3686 Mt | 4.384 | 1.261 | 4.484 | 1.319 | 1.352 | 0.515 |
| Mo | 343 | 59.76 Mt | 0.1742 Mt | 4.371 | 1.131 | 4.480 | 0.971 | 1.093 | 0.358 |
| Ag | 1644 | 1899.43 t | 1.1554 t | 7.832 | 1.342 | 7.936 | 1.221 | 1.382 | 0.361 |
| Au | 2106 | 284.33 t | 0.1350 t | 6.551 | 1.168 | 6.629 | 0.994 | 1.164 | 0.297 |
| U | 172 | 59.76 Mt | 0.3474 Mt | 2.979 | 1.177 | 2.979 | 1.177 | 0.000 | 0.000 |

## 26.4 Upper Tail Pareto Distribution and Its Connection to the Basic Lognormal Distribution

The cumulative Pareto distribution function satisfies

$$F(x) = 1 - \left(\frac{k}{x}\right)^{\alpha}$$

where $\alpha > 0$ and $k > 0$ are its two parameters. The following maximum likelihood estimator of the Pareto coefficient $\alpha$ has been used in several publications (Clauset et al. 2009; Patiño Douce 2016c; Agterberg 2017b) in various ways:

$$\alpha = \frac{n}{\sum_{i=1}^{n} \ln \frac{x_i}{k}}$$

where $n$ represents number of metal deposits selected in an ordered sequence of values $x_i$ ($i = 1, 2, \ldots, n$), and $k$ is the critical size parameter representing the truncation point at which the maximum value of the Pareto probability—density drops to zero. In the original algorithm of Clauset et al. (2009), which was used by Patiño Douce (2016c), all possible values of $k$ are tested for sizes $x_1 < x_2 < x_3 \ldots < x_n$. Minimum size ($x_1$) was set at median size and $x_n$ at maximum size. Each sample of $n$ sizes provides a different estimate of $k$ and $\alpha$. The Kolmogorov-Smirnov test was used to find the Pareto distribution that provides the best fit.

In Agterberg (2017b)'s application, $x_1 > x_2 > x_3 \ldots > x_n$, was used instead. This reversal of order was based on the following three premises: (1) worldwide metal deposit size sample sizes are very large ensuring that cumulative frequencies become increasingly precise when $n$ is increased, regardless of whether or not the Pareto distribution model is satisfied; (2) starting with the largest deposits and increasing sample size by including progressively more deposits improves results if

the Pareto distribution model would indeed be satisfied; and (3) for increasingly large values of $n$, observed frequencies become increasingly less than expected Pareto distribution model frequencies because the upper tail Pareto gradually passes into the lower frequency density basic lognormal via the upper tail bridge function. Theoretically, if $\alpha$ is known, $k$ could be derived from $\alpha$ by using the preceding equation for the maximum likelihood estimator. In Agterberg (2017a, b, in press), $\alpha$ was pre-determined by visual inspection for 7 metals that all show approximately linear patterns in log rank—log size plots for their largest deposits.

Initially, for small values of $n$, the resulting patterns for copper and other metals show large random fluctuations. For larger values of $n$ the plots develop multi-peak patterns for $\alpha$ that are superimposed on a gradational decrease. In Agterberg (2017b) a straight line was fitted by least squares for copper and gold avoiding the large small-sample fluctuations at the largest size values end capture the downward bend of log rank values toward lower log size values. This procedure produced estimates of $k_{Cu} = 6.98$ and $k_{Au} = 8.98$. Both estimates were confirmed by more detailed analysis of cumulative frequencies for largest deposits yielding $k_{Cu} \approx 7.0$ and $k_{Au} \approx 9.2$.

However, the preceding method does not work very well for some metals with fewer data than copper and gold. The following relatively simple method gave good results for six metals as shown in Figs. 26.3 and 26.4. The value of $n$ was set equal to 20 in each application for a window that was slid along the series of ordered metal deposit size values from the largest deposit downward. Initial random fluctuations connected to the largest values were avoided and so were windows on the upper bridge function transition zone toward the basic lognormal size-frequency distribution. For copper this procedure gives $\alpha = -1.2059$ for $k = 6.996$. The straight line with slope $\alpha$ passing through the point with average log size and average log rank for the 20 pairs of copper deposit size values used is shown in Fig. 26.3a. Similar results for the other five metals are shown in Figs. 26.3 and 26.4. According to the Pareto-lognormal model, a decrease of estimated values of $\alpha$ at the point where the upper tail Pareto ceases to be applicable is indeed expected. However, it is not clear why there is an equally strong decrease of estimated values of $\alpha$ in the patterns of Fig. 26.3 from the peak outward toward increasing values of log (deposit size). Very large random fluctuations are known to exist for the largest deposits. However, the upper tail downward trends in Fig. 26.3 could mean larger sizes than expected for the largest deposits although there are no indications of this in Fig. 26.4. Neither are there obvious deviations from linearity in log rank—log size plots that include the largest deposits for various metals (Agterberg 2017a, b). Residuals from the straight lines representing the Pareto distributions show relatively strong autocorrelation. Because of this uncertainty, it remains important to look for alternative upper tail models like the lognormals proposed by Patiño Douce (2016c, 2017) and shown for copper and gold in Agterberg (2017b). These alternate lognormals differ from the basic lognormals primarily in that they have much large mean deposit size values.

In order to fully represent the upper tail cumulative size-frequency distribution, the Pareto's have to be connected to the basic lognormals. Taking copper for

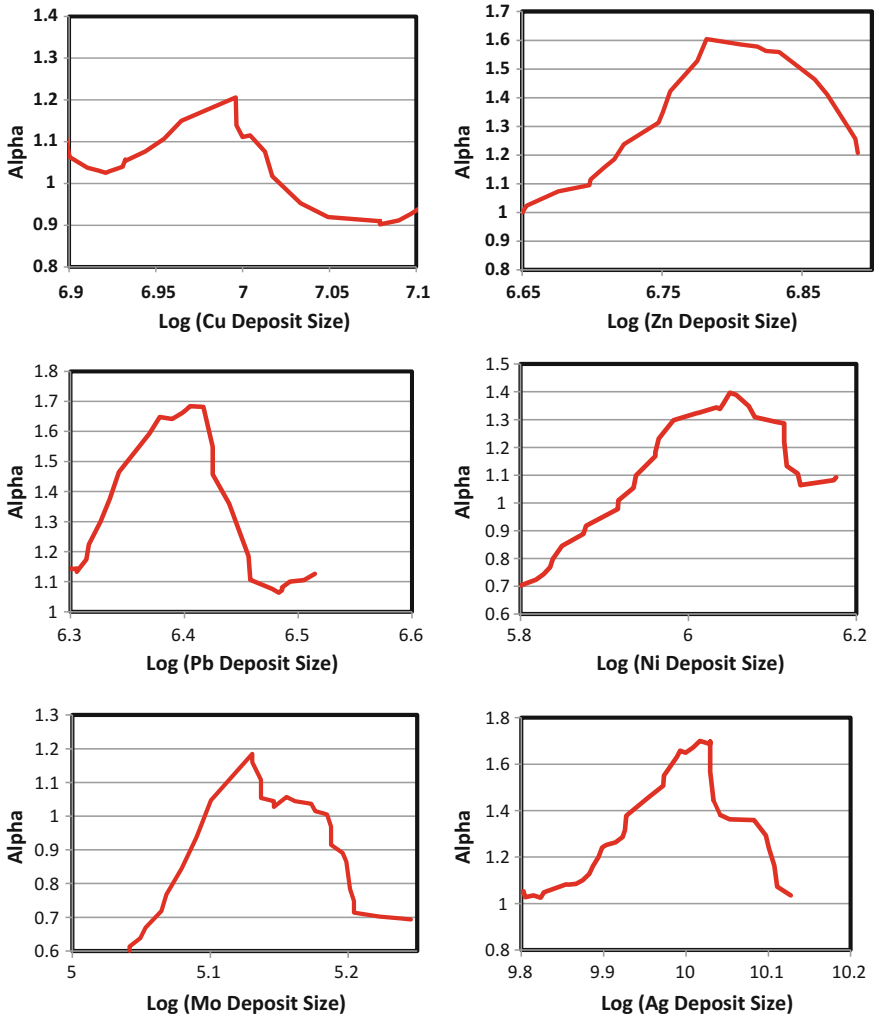


**Fig. 26.3** Pareto coefficient ($\alpha$) for log of metal deposit size as obtained in the text, setting $n$ equal to 20 for overlapping data sets moving from larger to smaller log (deposit size) values. Maximum $\alpha$ will be taken as optimum value with data sets, on which it is based, for the six metals shown in Fig. 26.4

example again, it can be seen in Fig. 26.2a that observed frequencies deviate from the best-filling straight line for log Cu deposit size values greater than 6. In total 42 deposits have log Cu deposit size values greater than 7 and their observed cumulative frequency of 42/2524 can be used as an anchor point to connect the upper tail Pareto to the upper bridge function which represents the transition zone between the basic lognormal (for values $< 6$) and the Pareto (for values $\geq 7$). Table 26.3 shows anchor points used for all six metals considered. Figure 26.5 shows best-fitting

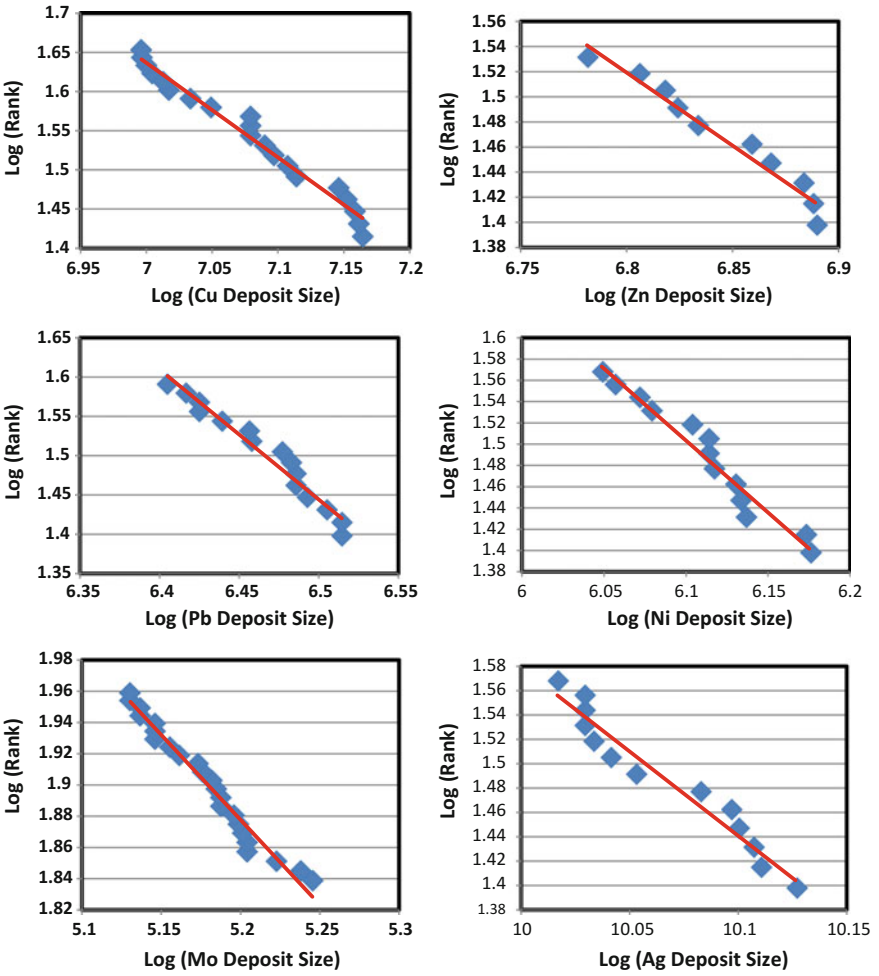


**Fig. 26.4** Sets of twenty log (metal deposit size) values corresponding to maximum value of $\alpha$ in Fig. 26.3 for the six metals. Corresponding Pareto distribution functions are shown as straight lines on these log rank—log size plots

**Table 26.3** Pareto coefficients $\alpha$ and $k$ corresponding to peaks for six metals in Fig. 26.3. Anchor point is log metal deposit size on upper tail Pareto distribution with relative frequency (Rel Freq) and observed $y$-value—$\log_{10}$ (1—cumulative frequency)

| Metal | α | $k$ | Anchor pnt | Rel Freq | Obs $y$-value |
|---|---|---|---|---|---|
| Cu | 1.2059 | 6.996 | 7 | 42 | −1.7818 |
| Zn | 1.1615 | 6.782 | 7 | 15 | −1.9930 |
| Pb | 1.6535 | 6.405 | 6.5 | 27 | −1.6108 |
| Ni | 1.3523 | 6.049 | 6 | 42 | −1.0433 |
| Mo | 1.0926 | 5.130 | 5 | 106 | −0.5286 |
| Ag | 1.3820 | 10.017 | 10 | 38 | −1.6361 |

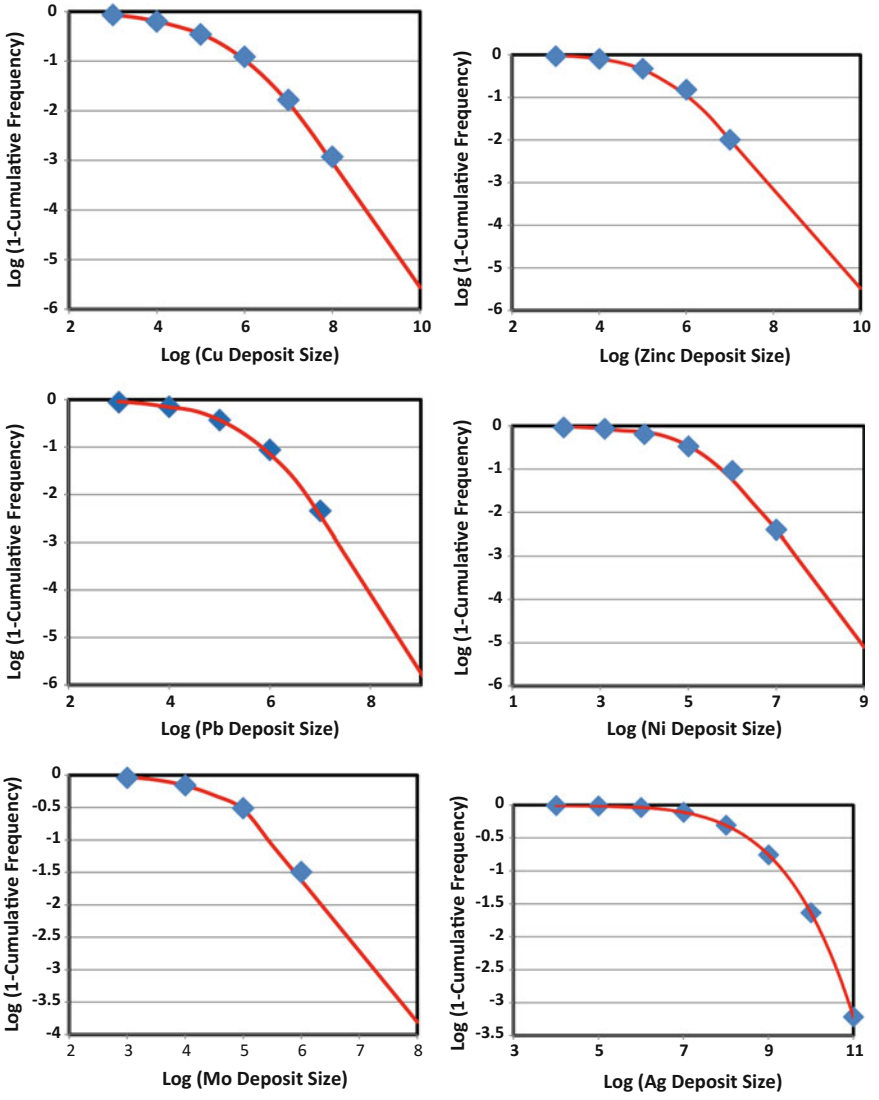


**Fig. 26.5** Upper tails of Pareto-lognormal size-frequency distributions for the six metals constructed by using the method explained in Table 26.4. Upper tails bridge functions are smooth curves that satisfy quartic polynomials fitted by least squares to log size values satisfying basic lognormal on the left and upper tail Pareto on the right side. For copper the result does not differ significantly from sextic polynomial previously shown in Agterberg (2017b, Fig. 14). For molybdenum no bridge function was fitted. Points with log (Mo deposit size) $\leq 5$ satisfy basic lognormal shown as straight line on Fig. 26.2e; points with log (Mo deposit size) $\geq 5$ belong to the upper tail Pareto distribution

frequency distribution curves that are Pareto-type for log deposit size values exceeding the anchor points. Some anchor points slightly exceed the estimated values of the truncation parameters $k$ without significantly changing the results. Quartic polynomials were used to approximate the smooth shapes of each frequency distribution within the upper tail bridge function that connects the Pareto with the basic lognormal. Table 26.4 shows results of this interpolation procedure for copper. The curve in Fig. 26.5a resembles the curve previously shown in Agterberg (2017a) where it was a best-fitting sextic polynomial. Contrary to the fitting of sextic polynomials to other metals, the method using a quartic explained in Table 26.4 gave good results for the other metals considered with the exception of molybdenum that does not seem to need a bridge function to pass from the Pareto into the basic lognormal. It is the only metal for which the upper tail Pareto and the central lognormal almost continuously pass into one another. Molybdenum, therefore, almost exactly satisfies the model proposed by Patiño Douce (2016b, Appendix 1) in which the probability density function of the lognormal as well as its first derivative pass continuously into the density function of the Pareto. The value at log (Mo deposit size) = 5 predicted by the basic lognormal is equal to the value of the Pareto at this point. Figure 26.5e, however, shows that there is a slight change of dip of the curve for log (1—cumulative frequency) at this point. All frequency distribution curves in Fig. 26.5 are close to their observed cumulative frequencies also shown in these diagrams.

## 26.5 Prediction of Future Copper Resources

As previously pointed out in Agterberg (2017b; in press), one of the purposes of developing statistical models of the size-frequency distributions of worldwide metal deposits is to use these models for prediction purposes either spatially (e.g., from relatively well-explored regions to unexplored regions, or deeper down from the Earth's surface), or in time. For multifractal modeling of the spatial distribution of mineral deposits, see Cheng (1994) or Cheng and Agterberg (1995). Use of parametric models is discussed by many authors including Agterberg (1974), Patiño Douce (2017) and Agterberg (2017b). The following non-parametric approach was first presented in the latter paper.

Suppose that $X$ is a continuous random variable denoting mineral deposit size and that $K$ is a discrete random variable for number of deposits per unit of area, volume or time; then the continuous random variable $Y$ representing the sum of the sizes of the $K$ deposits satisfies:

$$Y = X_1 + X_2 + \cdots + X_K$$

**Table 26.4** Curve connecting smoothed $y$-values for $\log_{10}$ (1—cumulative frequency) in Fig. 26.5a in comparison with observed $y$-values for copper. Commencing as lognormal, the curve passes gradually into the straight line for its upper tail Pareto. Smoothed $y$-values for the intermediate bridge function satisfy a quartic polynomial equation fitted by least squares to lognormal values for $x \leq 5$ and $x \geq 7$. Smoothed values include quartic polynomial for $x = 5$ and $x = 7$

| $x$ | Lognormal | Pareto | Quartic | Smoothed | Observed |
|---|---|---|---|---|---|
| 4 | −0.196 | | −0.199 | −0.196 | −0.198 |
| 4.5 | −0.309 | | −0.298 | −0.309 | |
| 5 | −0.460 | | −0.449 | −0.449 | −0.460 |
| 5.5 | −0.653 | | −0.671 | −0.671 | |
| 6 | −0.892 | | −0.974 | −0.974 | −0.908 |
| 6.5 | −1.180 | | −1.364 | −1.364 | |
| 7 | −1.518 | −1.782 | −1.838 | −1.838 | −1.782 |
| 7.5 | −1.909 | −2.411 | −2.388 | −2.411 | |
| 8 | −2.354 | −3.041 | −2.999 | −3.041 | −2.928 |
| 8.5 | −2.853 | −3.670 | −3.650 | −3.670 | |
| 9 | −3.408 | −4.299 | −4.313 | −4.299 | |

The mean $E(Y)$ and variance $\sigma^2(Y)$ satisfy:

$$E(Y) = E(K) \cdot E(X); \quad \sigma^2(Y) = E(K) \cdot \sigma^2(X) + \sigma^2(K) \cdot E^2(X)$$

These equations were previously used in Agterberg (1974, Eq. 7.72) who had adopted them from Feller (1968, Chap. 12) where they are derived for $K$ and $X$ both representing integral-valued random variables. The approach also is applicable when $X$ is a continuous random variable. The variance equation can be found in an online article on compound distributions (Lin 2014, Eq. (4)) with many additional references. Specific distribution models can be assumed to hold true for $K$ and $X$. However, as shown earlier in this chapter, significant uncertainties remain in modeling the upper tail of worldwide metal size-frequency distributions that contain most metal. Fortunately, samples now available for statistical modeling are so large that the following non-parametric approach can be used.

Patiño Douce (2017) contains tables with statistics based on number of 1950–2007 copper deposit discoveries originally derived from a plot by Schodde (2010) for copper deposits with size $> 3 \times 10^5$ t Cu. Mean and variance of yearly number of discoveries are 8.621 and 14.304, respectively. Extrapolation of these two parameters over 85 years, toward the end of this century, would yield an expected number of 732.8 discoveries with variance of $12.158 \times 10^3$. Patiño Douce (2016b)'s original data base contains 591 copper deposits with sizes $> 3 \times 10^5$ t Cu resulting in estimated values of $E(X) = 3.784 \times 10^6$ t and $\sigma^2(X) = 1.135 \times 10^{14}$. Because of the large sample size, the 95% confidence limits on the estimated mean value are $3.784 \times 10^6 \pm 0.859 \times 10^6$ t with the large sample ensuring approximate normality of the frequency distribution of this mean. Consequently, this estimate is rather

precise. Using the preceding equations for mean $E(Y)$ and variance $\sigma^2(Y)$, it follows that estimated total tonnage copper value amounts to $732.8 \times 3.784 \times 10^6 = 2.773 \times 10^9$ t. The corresponding variance amounts to $25.726 \times 10^{16}$, from which it follows that the 95% confidence limits on the estimated mean value are $2.773 \times 10^9 \pm 0.994 \times 10^9$ t. This mean value is approximately normally distributed as well. Although the method for deriving this result differs significantly from the computer simulation method used by Patiño Douce (2017), the end result is only $0.654 \times 10^9$ t greater and the difference between the two estimates is not statistically significant. These results confirm Patiño Douce (2017)'s conclusion that there would be a significant shortage of copper if current rates of discovery will be maintained. The problem would become even worse if future rates would decrease.

## 26.6 Concluding Remarks

In this chapter it was argued that publicly available large metal deposit size data bases (especially Patiño Douce 2016b) allow new kinds of statistical modeling of regional and worldwide metal resources. The two models most frequently used in the past are lognormal size-grade and Pareto upper tail modeling. Both approaches are probably valid for several metals including copper, zinc, lead, nickel, molybdenum and silver taken for example because the upper tails of their mostly lognormal size frequency distributions satisfy the Pareto distribution model. The worldwide metal size-frequency distributions for these metals are similar indicating that a central, basic lognormal distribution is flanked by two Pareto distributions from which it is separated by upper and lower tail bridge functions. The lower tail Pareto distribution shows an excess of small deposits which are not economically important. Number frequencies of the upper Pareto are mostly less than those of the basic lognormal. A new method for fitting the upper tail Pareto was introduced and produces good results for the six metals taken for example. Parameters of regional metal size-frequency distributions as well as those of mineral deposit type distributions are less than those of the worldwide distributions. Uranium differs from other metals in that its worldwide size-frequency distribution is approximately lognormal. This may indicate that the lognormal model remains a standard model of size-frequency distributions of metals predominantly mined from hydrothermal and porphyry-type orebodies. A new version of the model of de Wijs may provide a framework for explaining the differences between regional and worldwide distributions. Further research on this topic remains to be carried out. The Pareto tails may reflect historical mining methods with bulk mining becoming prevalent in the 20th century. A new method of estimating the Pareto coefficients of the economically important upper tails of the size-frequency distributions was presented, and a non-parametric method for long-term projection of future metal resource on the basis of past discovery trend was illustrated for copper.

## References

Agterberg FP (1974) Geomathematics. Elsevier, Amsterdam
Agterberg FP (2007) Mixtures of multiplicative cascade models in geochemistry. Nonlinear Process Geophys 14:201–209
Agterberg FP (2017a) Pareto-lognormal modeling of known and unknown metal resources. Nat Resour Res 26:3–20 (with erratum on p 21)
Agterberg FP (2017b) Pareto-lognormal modeling of known and unknown metal resources. II. Method refinement and further applications. Nat Resour 26:265–283
Bickel PJ, Doksum KA (2001) Mathematical statistics, 2nd edn. Prentice-Hall, Upper Saddle River, New Jersey
Billingsley P (1986) Probability and measure, 2nd edn. Wiley, New York
Cargill SM (1981) United States gold resource profile. Econ Geol 76:937–943
Cargill SM, Root DH, Bailey EH (1980) Resources estimation from historical data: mercury, a test case. Math Geol 12:489–522
Cargill SM, Root DH, Bailey EH (1981) Estimating unstable resources from historical industrial data. Econ Geol 76:1081–1095
Cheng Q (1994) Multifractal modeling and spatial analysis with GIS: gold mineral potential estimation in the Mitchell-Sulphurets area, northwestern British Columbia: unpublished doctoral dissertation. University of Ottawa, Canada
Cheng Q, Agterberg FP (1995) Multifractal modeling and spatial statistics. Math Geol 28(1):1–16
Clauset A, Shalizi CR, Newman MEJ (2009) Power-law distributions in empirical data. SIAM Rev 51:661–703
Crovelli RA (1995) The generalized 20/80 law using probabilistic fractals applied to petroleum field size. Nonrenewable Resour 4(3):223–241
De Wijs HJ (1951) Statistics of ore distribution, I. Geol Mijnbouw 30:365–375
Drew LJ, Schuenemeyer JH, Bawlee WJ (1982) Estimation of the future rates of oil and gas discoveries in the western Gulf of Mexico, US Geological Survey Professional Paper 1252. GPO, Washington, DC, US
Feller W (1968) An introduction to probability theory and its applications, vol. I, 3rd edn. Wiley, New York
Johnson NL, Kotz S (1970) Continuous univariate distributions-1. Wiley, New York
Lin XS (2014) Compound distributions. Wiley StatsRef: Statistics Reference Online http://dx.doi.org/10.1002/9791118445112.stat04411
Lovejoy S, Schertzer D (2007) Scaling and multifractal fields in the solid Earth and topography. Nonlinear Process Geophys 14:465–502
Mandelbrot BB (1983) The fractal geometry of nature. Freeman, San Francisco
Matheron G (1962) Traité de Géologique Appliquée. Mémoire BRGM 14. Paris, Éditions Technip
Patiño Douce AE (2016a) Metallic mineral resources in the twenty first century. I. Historical extraction trends and expected demand. Nat Resour Res 25:71–90
Patiño Douce AE (2016b) Metallic mineral resources in the twenty first century. II. Constraints on future supply. Nat Resour Res 25:97–124
Patiño Douce AE (2016c) Statistical distribution laws for metallic mineral deposit Sizes. Nat Resour Res 25:365–387
Patiño Douce AE (2017) Loss distribution model for metal discovery probabilities. Nat Resour Res 26:241–263
Singer DA (2013) The lognormal distribution of metal resources in mineral deposits. Ore Geol Rev 55:80–86

Singer DA, Menzie DW (2004) Quantitative mineral resource assessments. Oxford University Press

Schodde R (2010) The declining discovery rate—What is the real story? https://www.minexconsulting.com/publications/mar2010a.html

Turcotte DL (1997) Fractals and chaos in geology and geophysics, 2nd edn. Cambridge University Press

USGS Mineral Commodity Summaries (2015). http://minerals.usgs.gov/minerals/pubs/mcs2015.pdf