Multifractal interpolation method for spatial data with singularities

by Q. Cheng*

Synopsis

This paper introduces the multifractal interpolation method (MIM) developed for handling singularities in data analysis and for data interpolation. The MIM is a new moving average model for spatial mapping and interpolation. The model decomposes the raw data into two components: singular and nonsingular components. The former can be characterized by a localized singularity index that quantifies the scaling invariance property of measures from a multifractal point of view. The latter is a smooth component that can be estimated using ordinary kriging or other moving average models. The local singularity index characterizes the concave/convex properties of the neighbourhood values. The paper utilizes a binomial multiplicative cascade model to demonstrate the generation of one- and two-dimensional data with multi-scale singularities which can be modelled by asymmetrical multifractal distribution. It then introduces a generalized moving average mathematical model for analysing and interpolating data with singularities. Finally, it is demonstrated by a one-dimensional case study of de Wijs’ data from a profile in a zinc mine, that incorporation of spatial association and singularity can improve the interpolation result, especially for observed values with significant singularities.

Keywords: data analysis, spatial mapping, moving average models, multifractal interpolation.

Introduction

Since the term ‘kriging’ was coined by Georges Matheron in the early 1960s on the basis of Krige’s master’s thesis dealing with interpolation of point samples, geostatistics has been rapidly developed as a branch of science and relevant techniques have been commonly applied in many fields of science for mapping, estimation, simulation, and prediction (Journel and Huijbregts, 1978; Goovaerts, 1997). The International Association for Mathematical Geosciences (IAMG) is proud of the invention and further development of the subject by our members. Kriging and other geostatistical techniques have been widely applied outside of geosciences, where users unaware of its origins and mathematical evolution refer to it simply as a type of spatial analysis. The semivariogram, a function of distance between locations, can measure the spatial autocorrelation between values at locations separated by a distance. Models empirically fitted to semivariograms are used for assigning weights to linear equations whose solutions provide weighted averages for kriging data with stationarity (Goovaerts, 1997; Deutsch and Journel, 2008). Traditionally, kriging is for interpolating data in the neighbourhood and estimating values at locations where no data is available. Interpolation algorithms have been developed for a variety of simple, indicator, and higher-order kriging as well as kriging with transformed and compositional data. Algorithms for interpolation of data with anisotropic spatial association (e.g. Chiles and Delfiner, 1999), mixed categorical and/or continuous data (Journel and Huijbregts, 1978; Goovaerts, 1997), and compositional data (Pawlowsky-Glahn and Olea, 2004), have been created.

Case studies comparing these methods are available in the literature (e.g. Park and Jang, 2014). Application of kriging depends heavily on stationarity of the mean and second-order moments involving the variogram and standard deviation of a regionalized random variable. Simple kriging (SK) may be applied if the mean of the data has a known but constant trend, whereas ordinary kriging (OK) may be applied if the trend is constant but unknown. If the trend is unknown but follows some polynomial model, other types of kriging accounting for trends can be used (Hansen et al., 2010). In most cases stationarity of second order moments is also assumed. However, the real data, especially exploratory data involved in characterizing mineralization and hazardous events, often does not meet stationarity requirements because of singularities.

* Department of Earth and Space Science and Engineering, Department of Geography, York University, Toronto.
© The Southern African Institute of Mining and Metallurgy, 2015. ISSN 2225-6253.

The Journal of The Southern African Institute of Mining and Metallurgy

VOLUME 115 MARCH 2015 235
Multifractal interpolation method for spatial data with singularities

A new approach is the multiple point geostatistics which is a new field of spatial and temporal analysis (Mariethoz and Caers, 2014). Multiple point geostatistics techniques train images to quantify and to include structural information into stochastic simulation (Guardiano and Srivastava, 1993; Strebe, 2002).

The multifractal interpolation method (MIM) based on multifractal theory (Cheng, 1999a, 2000) has been developed for the analysis and interpolation of data with singularities. Multifractal theory integrates both spatial association and local singularities and can enhance and retain the local structure properties (Cheng, 2006a,b). This paper introduces a generalized binomial multiplicative cascade process to demonstrate the generation of one- and two-dimensional data with multi-scale singularities which can be modelled by asymmetrical multifractal distribution. It will then introduce a generalized moving average mathematical model for analyzing and interpolating between data with singularities. Finally the method will be demonstrated by a one-dimensional example.

Multifractal processes and multifractal distributions

Singular physical, chemical, and biological processes can result in anomalous energy release, mass accumulation, or matter concentration, all of which are generally confined to narrow intervals in space or time (Cheng, 2007a). The end products of these nonlinear processes can be modelled as fractals or multifractals. Singularity is a property of nonlinear natural processes, examples of which include, but are not limited to, cloud formation (Schertzer and Lovejoy 1987), rainfall (Veneziani 2002), hurricanes (Sornette, 2004), flooding (Malamud et al., 1996; Cheng 2000; Cheng et al., 2009), landslides (Malamud et al., 2004), forest fires (Malamud et al.,1996), earthquakes (Turrotte 1997), mineral deposits and related geochemical anomalies (Agterberg, 1995; Cheng et al., 1994; Xie and Bao, 2004; Cheng and Agterberg, 2009), solar wind turbulence (Macek, 2007), DNA series (Rosas et al., 2002), heat flow at the mid oceanic ridges (to be published by the author elsewhere) and landscape textures (Plotnick et al., 1993). Multifractal modeling involves quantification of multi-scale singularities and various types of properties associated with spatial distribution of the singularities (Halsey et al., 1986; Feder, 1988; Mandelbrot, 1989; Evertsz and Mandelbrot, 1992).

This section introduces an asymmetrical cascade process that generates results with singularities characterized by asymmetrical multifractal models. There are several formalisms for describing multifractal distributions, one of which is the multifractal model based on the partition function (Halsey et al., 1986). This model involves three functions: the mass exponent function or Renyi exponent \(\tau(q)\), the coarse Hölder exponent \(\alpha(q)\), and the fractal spectrum function \(f(\alpha)\). In order to demonstrate the singularities and their distributions in one- or two-dimensional data, I introduce the theories and concepts of multiplicative cascade processes (MCPs), which play a fundamental role in quantifying turbulent intermittency and other nonlinear processes (Schertzer and Lovejoy 1985, Schertzer et al. 1997). MCPs have been extensively discussed in the literature (e.g. Gupta and Waymire 1993, Over and Gupta 1996, Menabde and Sivapalan 2000; Serinaldi 2010). The model of de Wijs is a simple two-dimensional multiplicative cascade model (de Wijs 1951, Agterberg 2001, 2007a) described in terms of multiplicative canonical cascade processes (Lovejoy and Schertzer, 2007). Other modifications exist, e.g. a cascade model with functional redistribution rate (Agterberg 2007b); a two-dimensional cascade model with anisotropic partition (Cheng 2005); a generalized two-parameter binominal multiplicative model proposed by Koscielny-Bunde et al. (2006) and applied for describing multifractal spectra of runoff time series; a three-parameter binominal multifractal model proposed by Macek (2007) and applied to characterize solar wind turbulence data based on a generalized two-scale weighted Cantor set for characterizing asymmetrical multifractal distribution; a two-dimensional cascade model with variable and conditional dependence partition (Cheng, 2012); and a five-parameter binominal multiplicative cascade model has been recently proposed by the author (Cheng, 2014) for describing fundamental multifractal indices characterizing asymmetrical multifractal distribution of real-world data.

The five-parameter generalized multiplicative cascade processes involve the partitioning of with a length \(L\) (e.g. in unit of meter) into \(h\) equal segments of which \(m_1\), receive \(d_1\) \((>0)\); and \(m_2\), receive \(d_2\) \((>0)\) proportion of mass in the previous segment, respectively, where \(m_1 + m_2 = h\). For a closed system with preservation of unit measure (e.g. with total mass \(M\)), \(d_1 + d_2 = 1\). Otherwise, \(d_1 + d_2 < 1\) or \(d_1 + d_2 > 1\), corresponding to a loss or gain of mass during the cascade process, respectively. At the nth partition, the segment length will be \(e_n = L(1/h)^n\); the segments are subject to \(k\) times segments with measure \(d_1 m_1\) and \(n-k\) times segments with measure \(d_2 m_2\). Thus the measures of these segments are  
\[\mu_k = M d_k m_k (d_2/m_2)^k (d_1/m_1)^{n-k}\]. Therefore, the numbers of segments with \(k\) will be \(N_k = m_k (d_2/m_2)^k (d_1/m_1)^{n-k}\). Letting \(n \to \infty\), we can find series of \(n\) and \(k\) with \(k = \lfloor n \rfloor\), \(0 \leq \xi \leq 1\), where \(\xi\) is independent of \(n\) or \(k\). We then obtain the following relationships

\[\mu_k = M \left( \frac{d_k}{m_k} \right)^{\xi_k} \left( \frac{d_1}{m_1} \right)^{1-\xi_k} = M \left( \frac{L}{L(1/h)^n} \right)^{\alpha_k \xi_k} \left( \frac{L}{L(1/h)^n} \right)^{1-\alpha_k \xi_k} \]

where \(\alpha_k\) is the coarse Hölder exponent which quantifies the singularity of the distribution of

\[\alpha_k = -\frac{\xi_k \log(d_k/m_k) + (1-\xi_k) \log(d_1/m_1)}{\log h}\]

and the subset of segments with singularity \(\alpha_k\) is an intertwined set which is a fractal with fractal dimension \(f(\alpha_k)\). The number of segments in each of the intertwined fractals can be expressed as

\[N_k(\alpha_k) = \mathcal{E}_s^{-\xi_k} \]

where \(f(\alpha_k)\) is the fractal dimension spectral function, which can be expressed as

\[f(\alpha) = -\frac{\xi_k \log(d_k/m_k) + (1-\xi_k) \log(d_1/m_1)}{\log h}\]
Multifractal interpolation method for spatial data with singularities

It can be seen from Equations [1] to [4] that both the measure \( \varepsilon \) and \( n_0 \) follow power-law relationships at scale \( \varepsilon_n \). Since the value of \( \varepsilon \) falls within the range \([0, 1]\), the value of the singularity index takes any value between \( \varepsilon_{\min} \) and \( \varepsilon_{\max} \) following the linear relationship between \( \alpha(\varepsilon) \) and \( \varepsilon \):

\[
\alpha(\varepsilon) = \frac{\varepsilon \alpha_{\min} + (1 - \varepsilon) \alpha_{\max}}{\varepsilon_{\min}} = \frac{\varepsilon \alpha_{\min} - \alpha_{\max}}{\varepsilon_{\min}} + \alpha_{\max} \tag{[5]}
\]

where the maximum and minimum values of singularity \( \alpha \) are

\[
\alpha_{\min} = -\frac{\log(\varepsilon / \varepsilon_{\min})}{\log h}, \quad \alpha_{\max} = \frac{\log(\varepsilon / \varepsilon_{\min})}{\log h} \tag{[6]}
\]

assuming \( d_1/m_1 > d_2/m_2 \); otherwise, the two extremes will be reversed. Accordingly, the corresponding fractal dimensions with special singularities are shown to be

\[
f'(\alpha_{\varepsilon}) = \frac{\log m_\alpha}{\log h}, \quad f''(\alpha_{\varepsilon}) = \frac{\log m_{\alpha}}{\log h} \tag{[7]}
\]

The multifractality and symmetry of the multifractal distribution can be characterized by the asymmetry and multifractality indices

\[
\Delta\alpha_{m\rightarrow m} = \frac{\log(\varepsilon / \varepsilon_{\min})}{\log h}, \quad R = \frac{\Delta \alpha_{\varepsilon}}{\Delta \alpha_{\varepsilon}} = \frac{n_m}{m_\alpha} \tag{[8]}
\]

The asymmetry index corresponds to the ratio of the values \( m_\alpha \) and \( m_{\varepsilon} \); whereas the multifractality is proportional to the ratio of average measures, \( (d_1/m_1)/(d_2/m_2) \).

Singularities and nonstationarity

The singularity in the multifractal model introduced in the previous section characterizes how the statistical behaviour varies as the measuring scale changes. For example, at some locations the mean value calculated from the neighbourhood values might be independent of the size of the vicinity within which the values are averaged. In other cases, the mean value might proportionally depend on the size of the vicinity. The former case represents nonsingular location but the latter is for singular location. Singularity property has been commonly observed in geochemical and geophysical quantities (Cheng, 1994). Generally speaking, as the size of segment \( \varepsilon \) tends to zero and the number of segments tends to infinity, in order to explain the singularity of the measure at locations with singularity \( \varepsilon \), the former component is a non-singular component whereas the latter is singular component if the singularity index \( \alpha \neq 0 \). The inconvenience property of the measure with the following singularity properties implies nonstationarity of the measure or the density (Cheng, 1999a):

1. \( \alpha = 1 \) or \( \alpha = 0 \), if \( p(\varepsilon) \) is constant, independent of scale size \( \varepsilon \)
2. \( \alpha < 1 \), \( \alpha > 0 \), if \( p(\varepsilon) \) is a decreasing function of \( \varepsilon \), which normally implies the ‘convex’ property of \( \phi(\varepsilon) \) at the location with \( \alpha \)
3. \( \alpha > 1 \), \( \alpha < 0 \), if \( p(\varepsilon) \) is an increasing function of \( \varepsilon \), which normally implies the ‘concave’ property of \( \phi(\varepsilon) \) at the location with \( \alpha \).

Therefore, the \( \alpha \)-values as the fractal dimension of the fractal density \( \rho \) – value as the co-dimension can be used to characterize the nonlinear structural property of the measure \( \phi(\varepsilon) \). This approach has already been used for texture analysis of remote sensing Landsat TM images (Cheng, 1997b, 1999c), in multifractal interpolation of geochemical concentration values for mineral exploration (Cheng, 1999a, 2000a, 2000b), in well log curve reconstruction (Li and Cheng, 2001), flood event modelling (Cheng et al., 2009), in hyperspectral image analysis (Neta et al., 2010), faults and geochemistry (Wang et al., 2013), and in geochemical element concentration mapping (Chen et al., 2007). Zuo et al., 2009).

In order to introduce how singularities can be included in data interpolation, we here introduce the following scale-invariant property of the measure, \( \phi(\varepsilon) \) and density, \( p(\varepsilon) \). Due to the properties of power-law type functions we can associate the densities at two different scales \( (\varepsilon_n \text{ and } \varepsilon_m) \) as follows

\[
\rho(\varepsilon_n) \rho(\varepsilon_m) = \left(\varepsilon_n/\varepsilon_m\right)^{\Delta}\rho(\varepsilon_n)\rho(\varepsilon_m) \tag{[11]}
\]

The MIM to be introduced in the next section is developed according to the scale invariance property (Equation [11]).

MIM incorporating spatial association and singularity

Statistical properties derived at one scale may be used to estimate properties at another scale on the basis of the scaling property (Cheng, 1999a, 2000). The main purpose of data interpolation, including kriging, is to estimate values at unknown locations based on the neighbourhood values and their spatial association with the value being estimated. Spatial association represents a type of statistical dependency of values at separate locations. If the value at a location \( x \) is considered as the realization of a regionalized random variable \( Z(x) \), the spatial association or variability can be measured by means of the variogram

\[
2\gamma(x, h) = E[(Z(x) - Z(x + h))^2] \tag{[12]}
\]

where the semivariogram \( \gamma(x, h) \) is a function of vector distance \( h \) separating locations \( x \) and \( x + h \), measuring the symmetrical variability between \( Z(x) \) and \( Z(x + h) \). Under an assumption of second-order stationarity, the semivariogram (Equation [12]) becomes a function of \( h \) that is independent of location \( x \). This strong assumption on the nature of the regionalized random variable is generally required in kriging. Equation [12] has been commonly used for structural...
Multifractal interpolation method for spatial data with singularities

analysis and interpolation in geostatistics (Jouzel and Huijbregts, 1978). It has also been applied for texture analysis in image processing (Atkinson and Lewis, 2000; Herzfeld and Higginson, 1996).

To incorporate both spatial association and singularity in the supporting interpolation model based on Equation (11), the following average density within a small vicinity \((x_0, \lambda)\) around location \(x_0\) with linear size \(\lambda\) was defined by Cheng (2006)

\[
\rho(x, x_0) = \frac{E}{\lambda} \left( \frac{\lambda}{\max} \right)^{-\alpha} \sum_{x \in (x_0, \lambda)} \rho(x) \tag{13}
\]

where \(\rho(x)\) is the density of the sample at location \(x\) and \(E\) is the number of samples within the vicinity \((x_0, \lambda)\). The parameter \(\alpha\) controls the singularity of the data, with \(\alpha > 0\) indicating convexity and \(\alpha < 0\) concavity.

Assume that Equation (11) holds true within a range of window sizes, \(\epsilon < \lambda < \lambda_{\text{max}}\). Then the average density \(\rho(x, x_0)\) within the window \(\{|x_0, \lambda\}\) where it may not contain samples with observed values, can be estimated by kriging as follows:

\[
\overline{\rho}(x, x_0) = \frac{E}{\lambda_{\text{max}}} \left( \frac{\lambda}{\lambda_{\text{max}}} \right)^{-\alpha} \sum_{x \in (x_0, \lambda_{\text{max}})} \rho(x) \tag{14}
\]

Equation (14) is a general weighted average model that can be used to estimate the value at the centre of \(\rho(x, x_0)\) from the neutron values within \(\{|x_0, \lambda_{\text{max}}\}\) (Cheng, 2006). Since the above discussions are valid for all dimensions and here we will use \(E\) to present the dimension of problem, \(E=1, 2, 3\) stand for 1D, 2D and 3D problems. It has the following properties:

1. If it does not show singularity, \(\alpha = E\) or \(\alpha = 0\), then Equation (14) reduces to the ordinary moving average function that has been used commonly in kriging and other data interpolation methods.
2. If all locations show the same singularity strength with \(\alpha = \text{constant}\) and \(\alpha = \text{constant}\), then Equation (14) becomes the same as the ordinary moving average function used in kriging and other methods.
3. If the singularity varies from location to location, \(\alpha = \text{constant}\) or \(\alpha = \text{constant}\), then Equation (14) is equivalent to the ordinary moving average function multiplied by a scale ratio factor, \(\left(\lambda / \lambda_{\text{max}}\right)^{-\alpha}\), with three possible situations given \(\epsilon < \lambda_{\text{max}}\):

\[
(\epsilon / \lambda_{\text{max}})^{-\alpha} = \begin{cases} > 1, & \text{if } \Delta \alpha > 0 \\ 1, & \text{if } \Delta \alpha = 0 \\ < 1, & \text{if } \Delta \alpha < 0 
\end{cases} \tag{15}
\]

These properties indicate that if the data used for interpolation satisfies a multifractal distribution, then Equation (14) must be used as an extended form of the ordinary weighted averaging model. In this case, the scale ratio factor modifies the ordinary average in such a way that if there is positive singularity with \(\alpha > 0\), then the new result is to be increased by a factor (Equation [15]), whereas if \(\alpha < 0\), then the new result is reduced by a factor (Equation [15]). This modification is reasonable because \(\alpha > 0\) and \(\alpha < 0\) correspond to convex and concave properties of the surface \(\rho(x, x_0)\) around the location \(x_0\), respectively.

The new model (Equation (14)) not only describes the spatial association reflected in the calculation of the weight \(\lambda\), but also incorporates the singularity characterized by the singularity index \(\alpha\). The new model therefore has two obvious advantages: it not only improves the accuracy of the interpolated results but also retains the local structure of the interpolation map. The latter property is essential for geochemical and geophysical data processing and for pattern recognition. This will be demonstrated using the assay values from the Pulacayo sphalerite-quartz vein in Bolivia studied by De Wijs (1951).

Analysis of de Wijs's Bolivia sphalerite data

De Wijs (1951) studied assay values from the Pulacayo sphalerite-bearing quartz vein in Bolivia. Along a drift 118 channel samples had been collected at 2.00 m intervals (Figure 1A). These values have been analysed by multifractal modeling and spatial analysis (Cheng and Agterberg, 1996; Cheng, 1997b) and they can be approximated by five-parameter binomial multiplicative cascade models (Cheng, 1999c, 2014). The fractal dimension spectra of the distribution of de Wijs's zinc values are estimated by the gliding window method with order of moment ranging from -20 to 20 and cell size \(\epsilon\) ranging from 2 to 50 m (Cheng, 1999c) (Figure 2) and fitted by the five-parameter binomial multiplicative cascade model (Cheng, 2014) seen in Figure 2.

The estimated values of \(f(\alpha)\) are as follows:

\[f(\alpha) = 1.32, f(\alpha_{\text{min}}) = 0.75, f(\alpha_{\text{max}}) = 1.00\]

Therefore,

\[\Delta \alpha_{\text{max}} = \alpha_{\text{max}} - \alpha_{\text{min}} = 0.75\] and \(R = 1.1\).

Figure 1—Multifractal interpolation of de Wijs’s zinc values (de Wijs, 1951). (A) Observed values %Zn; (B) singularity \(\alpha\)-values estimated by multifractal interpolation method; (C) correlation coefficients associated with the estimation of the \(\alpha\)-values; and (D) interpolated results for zinc values. Blue dots represent the observed values; red and yellow lines present the results obtained by moving averaging method and the multifractal interpolation method, respectively.
Multifractal interpolation method for spatial data with singularities

The results indicate that the distribution follows a multifractal model that is nearly symmetrical ($\Delta_{\text{max}} - \Delta_{\text{min}} > 0$).

In order to show the distribution of singularity and the data interpolation results, Figure 1B shows the resulting distribution of singularity values calculated for the data and the correlation coefficients associated with the linear model fitted after double log-transformation of measure and scale. It shows that the estimated values of $\alpha$ are within a range from 0.6 to 1.4 with correlation coefficients greater than 0.975 (Figure 1C). Figure 2D illustrates the interpolated and reconstructed results obtained by means of MIM and moving averaging. The yellow line represents the results obtained using MIM with window size 20 m (10 point values) and the thicker red line represents the results obtained using the averaging technique with window size 6 m (two to three point values). The blue dots represent the observed data. Comparing the results obtained using MIM and moving averaging shows that MIM provides better results not only with smaller fitting errors for the observed data, but also that localized multifractality of the data is preserved.

Discussion and conclusions

It has been demonstrated that the multifractal distribution generated by binomial multiplicative cascade processes has multiple singularities that can be quantified by singularity index and fractal dimension spectrum. According to MIM, the singularity of multifractally distributed data can be used in data interpolation for mapping purposes with the localized structural properties (multifractality) preserved. The model used in MIM can be considered as an extended form of the ordinary moving average or weighted average used in various data interpolation methods, including inverse distance weighting and kriging. For most quantities in exploration geochemistry showing singularities, in order to retain the localized structural property, the multifractal interpolation method can be used to extend the ordinary moving average techniques, including ordinary kriging.

Since the singularity can be estimated using various methods, for example, integration of multiple patterns by weights of evidence method (Cheng, 2012) and other

The Journal of The Southern African Institute of Mining and Metallurgy

Multifractal interpolation method for spatial data with singularities

Figure 2 – Results showing fractal dimension functions (dots) calculated for de Wijs’s zinc values by means of gliding box method (Cheng, 1999) and binomial multiplicative cascade model with $m_1 = 1$, $m_2 = 1.1$, $h = 2.1$, $d_1 = 0.38$ and $d_2 = 0.63$. $\Delta_{\text{max}} - \Delta_{\text{min}} = 0.75$, $\alpha = 0.033$ and $R = 1.1$

Acknowledgements

The author wishes to thank Professor Richard Minnitt for accepting this paper for the Danie Krige Commemorative Volume. Thanks are due to Dr Frits Agterberg and Dr Zhijun Chen for their critical review of the paper and constructive comments. This research has been jointly supported by a research project on ‘Quantitative models for prediction of strategic mineral resources in China’ (201211022) by China Geological Survey, and by NSERC Discovery Research ‘Research and development of multifractal methods and GIS technology for mineral exploration and environmental assessments’ (ERC-OGP018599).

References

Multifractal interpolation method for spatial data with singularities


