

Spatial Analysis of Cell Composition Data

F.P. Agterberg¹

¹Geological Survey of Canada, Ottawa, Canada; agterber@NRCan.gc.ca

Abstract

A version of Matheron's discrete Gaussian model is applied to cell composition data. The examples are for map patterns of felsic metavolcanics in two different areas. Q-Q plots of the model for cell values representing proportion of 10 km x 10 km cell area underlain by this rock type are approximately linear, and the line of best fit can be used to estimate the parameters of the model. It is also shown that felsic metavolcanics in the Abitibi area of the Canadian Shield can be modeled as a fractal.

Key words: Discrete Gaussian model, Q-Q plots, felsic metavolcanics, fractals

1 Introduction

In an application of the discrete Gaussian model, Matheron (1974, p. 57-61) considered the following cell-value distribution model for binary patterns:

$$X = \Phi \left[\frac{\rho Z - \beta}{\sqrt{1 - \rho^2}} \right]$$

In this equation, X is a random variable for cell values generated from a binary variable X_0 , Z is a standard normal random variable, while β and ρ are constants (to be explained later); Φ represents quantile of the normal distribution function.

The cell values are compositional data in that they represent proportions of cell area underlain by the binary pattern (e.g. for a rock type on a geological map) that is being studied. The cell can be square, rectangular or have another shape. One property of this model is that the complement ($1 - X$) also is the quantile of a normal random variable with different mean but the same variance. This prompted the following statement (Agterberg, 1981, p. 20): "The fact that percentage values for a rock type and its complement can both have a frequency distribution which satisfies (this) equation may be of interest in the study of closed number systems. If the values for a set of random variables sum to one at all observation points, these variables cannot have the same type of distribution if one or more of them has a normal, lognormal, or gamma distribution. On the other hand, each variable in a set of random variables summing to one can have the distribution function given by (this) equation."

The following section contains a brief review of this approach. If the model is valid, observed cell values should have a frequency distribution that plots as a straightline on "prob-prob" paper with standard normal quantile scales along both x - and y -axis. Then the parameters β and ρ can be estimated from the intercept and slope of this straightline. In this respect the approach seems to form a natural extension of two other straightline fitting techniques that are widely applied: use of log-log paper in fractal-multifractal analysis, and log-prob paper in fitting a lognormal frequency distribution to data.

In the second part of this paper, two previously used data sets (felsic metavolcanics in the Bathurst and Abitibi areas) will be re-analyzed to illustrate the approach. The multifractal method of moments also will be applied to the larger of these two data sets (Abitibi felsic metavolcanics).

2 Review of underlying theory

Matheron (1974, 1976) proposed the "discrete Gaussian model" that can be applied to cell values (cf. Agterberg, 1981, 1984) as follows: Let X_1 with density function $f(x_1)$ represent a random variable for average concentration values of small cells with size S_1 and X_2 , with $f(x_2)$, that of large cells with size S_2 . Suppose that X_1 can be transformed into Z_1 by $X_1 = \psi_1(Z_1)$ and X_2 into Z_2 by $X_2 = \psi_2(Z_2)$ so that the

random variable Z_1 , with marginal density function $\varphi(z_1)$, and Z_2 with $\varphi(z_2)$, together satisfy the bivariate standard normal density function

$$\varphi(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{-(z_1^2 - 2\rho z_1 z_2 + z_2^2) / 2(1-\rho^2)\} \quad (1)$$

where ρ represents the product-moment correlation coefficient of Z_1 and Z_2 . In general, if the regression of X_1 on X_2 satisfies

$$E(X_1|X_2) = X_2 \quad (2)$$

$f(x_2)$ can be derived from $f(x_1)$. The interpretation of Equation (2) for a binary pattern is as follows. Suppose that X_2 represents the average value for amount of pattern in a large cell superimposed on the map, while X_1 is this value for a small cell sampled at random from within the large cell. Then Equation (2) is satisfied exactly. If the small cell is made infinitely small, X_1 becomes a binary random variable that can be written as X_0 for presence or absence of the pattern at a point.

From Equation (1) it follows that

$$\varphi(z_2|z_1) = \frac{\varphi(z_1, z_2)}{\varphi(z_1)} = \varphi\left[\frac{z_2 - \rho z_1}{\sqrt{1-\rho^2}}\right] \quad (3)$$

A bivariate standard normal distribution also can be expressed as the product of its marginal distributions by means of the Mehler identity:

$$\varphi(z_1, z_2) = \varphi(z_1) \cdot \varphi(z_2) \left[1 + \sum_{j=1}^{\infty} \rho^j H_j(z_1) H_j(z_2) / j! \right] \quad (4)$$

where $H_j(z_1)$ and $H_j(z_2)$ are Hermite polynomials with as a property:

$$\int_{-\infty}^{z_2} H_j(z_2) \varphi(z_2) dz_2 = H_{j-1}(z_2) \varphi(z_2)$$

Use of this property during integration of Equation (3) with respect to z_2 , replacement of z_1 by the standard normal random variable Z , and replacing z_2 by the constant β yields the random variable

$$X = \Phi\left[\frac{\rho Z - \beta}{\sqrt{1-\rho^2}}\right] = 1 - \Phi(\beta) - \varphi(\beta) \sum_{j=1}^{\infty} \rho^j H_j(Z) H_{j-1}(\beta) / j! \quad (5)$$

This is equivalent to assuming that the "probit" of X has normal distribution. The binary random variable X_0 for points random located within cells with values assumed by X can be defined by letting ρ tend to one. Then ρ can be interpreted as the correlation coefficient between X_0 and X . The mean of X satisfies

$$\mu = 1 - \Phi(\beta) \quad (6)$$

It can be shown (Agterberg, 1984) that its variance is

$$\sigma^2 = \varphi^2(\beta) \sum_{j=1}^{\infty} \rho^{2j} [H_{j-1}(\beta)]^2 (j!)^{-1} \quad (7)$$

By means of Equations (6) and (7) it is possible to obtain estimates of β and ρ . However, it is also possible to estimate these parameters by fitting a straightline to a quantile-quantile (Q-Q) plot of cell value percentage values and their observed frequencies using prob-prob paper.

3 Examples of application

The first example is for a relatively small area measuring 80 km in the east-west and north-south directions in the vicinity of Bathurst, New Brunswick, Canada (cf. Agterberg, 1981, p. 19-20). This square study area was subdivided into 64 square 10 km x 10 km cells and amount of felsic metavolcanics was measured for each cell. The frequency distribution of the cell values is shown as a Q-Q plot on prob-prob paper in Figure 1.

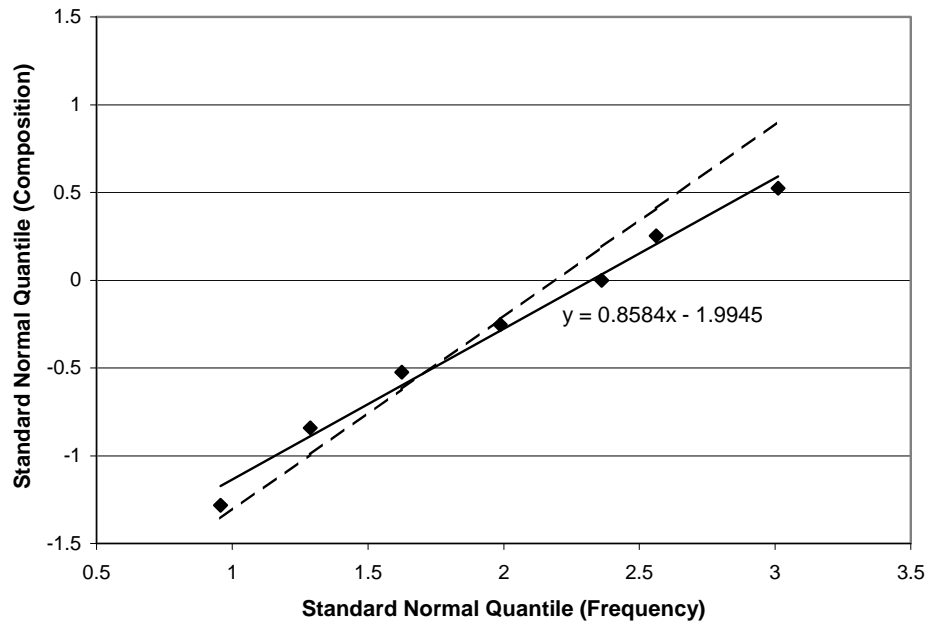


Figure 1: Content of felsic metavolcanics in 64 square cells (Bathurst area, New Brunswick) measuring 10 km on a side; Q-Q plot of cell values for model of Equation (5) with line of best fit. Broken line represents solution based on Equations (6) and (7). See text for further explanation.

The broken line in Figure 1 was derived previously (Agterberg, 1981, Figure 5) by estimating β and ρ from estimated values of the mean and variance of the 64 cell values using Equations (6) and (7). The variance was estimated by standard geostatistical methods from an empirically derived spatial covariance function (for details, see Agterberg and Fabbri, 1978). The resulting estimates of β and ρ were 0.85 and 0.91, respectively. Figure 1 also shows a best-fitting straightline obtained by linear regression of the quantile for frequency (y) on the quantile for composition (x). The intercept and slope of this best-fitting line yield estimates of 0.82 and 0.86 for β and ρ , respectively. The difference between the two straightlines in Figure 1 is probably due to the fact that on the Q-Q plot more weight is given to points with larger relative frequencies.

The second example is for a larger area measuring 480 km in the east-west direction and 160 km in the north-south direction located within the Abitibi Subprovince on the Canadian Shield (cf. Agterberg, 1984). This rectangular study area was subdivided into 768 square 10 km x 10 km cells and amount of felsic metavolcanics was measured for each cell. The resulting cell value frequency distribution is shown as a Q-Q plot on prob-prob paper in Figure 2.

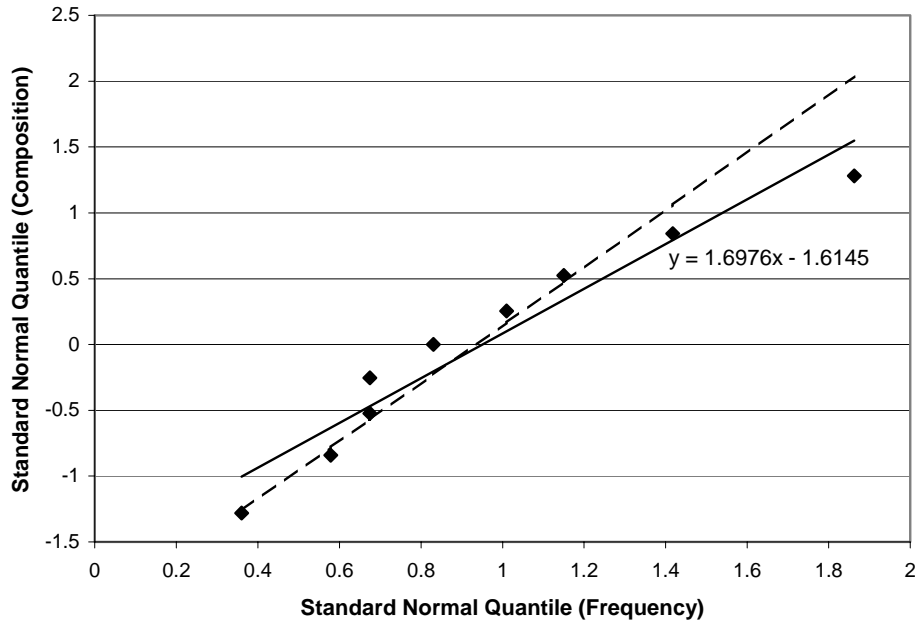


Figure 2: Content of felsic metavolcanics in 768 square cells (Abitibi area, Canadian Shield) measuring 10 km on a side; Q-Q plot of cell values for model of Equation (5) with line of best fit. Broken line represents solution based on Equations (6) and (7). See text for further explanation.

As in the example of Figure 1, the broken line in Figure 2 was derived previously (Agterberg, 1984, Figure 9) by estimating β and ρ from the mean and variance using Equations (6) and (7). In this earlier application, the variance was estimated from the sample of 768 cell values. The resulting estimates of β and ρ were 1.62 and 0.74, respectively; versus 1.51 and 0.65 as derived from the best-fitting line. Again, the difference between the two straightlines is probably due to differences in weighting of the observed frequencies.

In both examples of application, the straightline fitted to the Q-Q plot is probably slightly better than the straightline derived from estimates of the mean and variance. This would be because the smallest percentage values in both examples had greater relative errors than the larger percentage values.

4 Fractal/multifractal analysis of Abitibi felsic metavolcanics

For comparison, a simplified fractal/multifractal analysis was performed on the larger Abitibi data set using the method of power moment sums (cf. Falconer, 2003) also known as “method of moments”. Multifractal analysis is appropriate if the pattern being studied is self-similar or scale-independent. Equation (2) would be satisfied for self-similar multifractals but the assumption of self-similarity is stronger than that of Equation (2). A (mono-) fractal can arise as a special case of a self-similar multifractal.

Suppose that a grid of cells with length of cell side ϵ is superimposed on a pattern, and that amount of pattern in a cell is called the cell’s “measure” μ_ϵ . A self-similar multifractal can be characterized by its multifractal spectrum in which the fractal dimension $f(\alpha)$ is plotted against the singularity (or Hölder exponent) α . The measure satisfies $\mu_\epsilon \sim \epsilon^\alpha$ where \sim denotes proportionality, and the fractal dimension $f(\alpha)$ satisfies $N_\epsilon \sim \epsilon^{-f(\alpha)}$ where N_ϵ represents number of cells with singularity approximately equal to α . In practice, N_ϵ is defined for singularities from a narrow interval around α .

The method of moments to determine the multifractal spectrum consists of three consecutive steps. Initially, power moment sums are calculated for different cell sizes. In our application, square cells with sizes of 10 km x 10 km, 20 km x 20 km and 40 km x 40 km were used. As unit for the measure (amount of felsic metavolcanics per cell), decimal fraction per 10 km x 10 km cell was used. The power moment sums are plotted against length of cell side using log-log paper. Figure 3 shows results for power moments q between -1 and 5. Logarithms (base 10) were used to plot the power moment sums, and the three cell sizes are labelled 1, 2, and 3 in Figure 3.

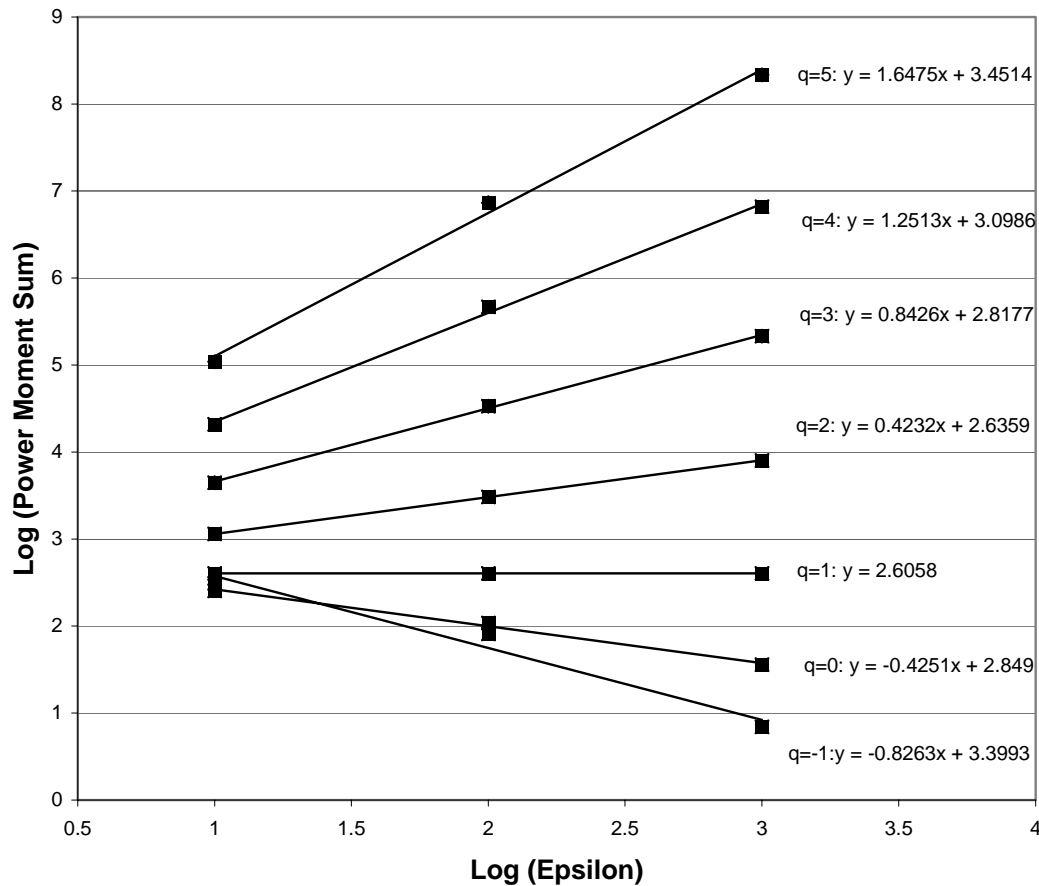


Figure 3: First step of multifractal analysis of felsic metavolcanics in 768 square cells (Abitibi area, Canadian Shield) measuring 10 km on a side; power moment sums (logarithms base 10) versus length of cell side. See text for further explanation.

The underlying pattern would be multifractal if the power moment sums all exhibit straightline patterns on log-log paper. This condition seems to be satisfied in Figure 3. The second step in the method of moments consists of assuming that the straightline slopes represent multifractal mass exponents (τ or “tau”) of the pattern for the power moments q . The first derivative of $\tau(q)$ with respect to q then yields an estimate of $\alpha(q)$ representing the singularity.

Figure 4 is the plot of $\tau(q)$ versus q using the slope estimates of Figure 3. The pattern of Figure 4 is closely approximated by the straightline $\tau(q) = 0.4149 \cdot q - 0.4136$. This would imply $\alpha(q) = 0.41$ representing a (mono-) fractal with constant singularity of 0.41 rather than a multifractal with different values of α and $f(\alpha)$. The final step in the method of moments consists of constructing the multifractal spectrum that is a plot of $f(\alpha)$ versus α using the relation: $f(\alpha) = q \cdot \alpha(q) - \tau(q)$. In our application to Abitibi felsic metavolcanics, the multifractal spectrum is reduced to a single spike representing a fractal. The intercept of the straightline, which also is 0.41, can be regarded as the (constant) fractal dimension $f(\alpha)$ of the pattern.

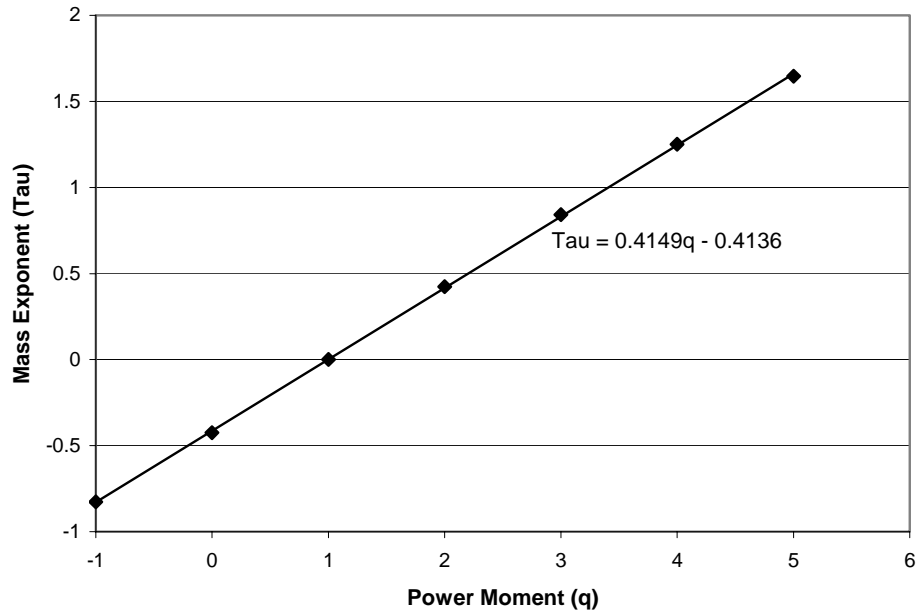


Figure 4: Second step of multifractal analysis of felsic metavolcanics in 768 square cells (Abitibi area, Canadian Shield) measuring 10 km on a side; mass exponent (representing slopes of straight lines in Fig. 3) versus power moment. Result is a fractal with multifractal spectrum (not shown) reduced to a single spike with $f(\alpha) = \alpha(q) = 0.41$. See text for further explanation.

References

- Agterberg, F.P. (1981). Cell-value distribution models in spatial pattern analysis. In R.G. Craig and M.L. Labovitz (Eds.), *Future Trends in Geomathematics*, pp. 5-28. London: Pion.
- Agterberg, F.P. (1984). Use of spatial analysis in mineral resource evaluation. *Mathematical Geology* 16(6), 565-589.
- Agterberg, F.P. and Fabbri, A.G. (1978). Spatial correlation of stratigraphic unit quantified from geological maps. *Computers & Geosciences* 4, 285-294.
- Falconer, K. (2003). *Fractal geometry*, Second Edition. Chichester: Wiley
- Matheron, G. (1974). Les fonctions de transfert des petites panneaux. *Note Géostatistique 127*, Centre de Morphologie Mathématique, Fontainebleau, France.
- Matheron, G. (1976). Forecasting block grade distributions: the transfer functions. In M. Guarascio, C.J. Huijbregts and M.L.R. David (Eds.), *Advanced Geostatistics in the Mining Industry*, pp. 221-236. Dordrecht: Reidel.