

# AN OBJECT-BASED MULTIPLE-POINT CONNECTIVITY MEASURE

ANDREW RICHMOND

Golder Associates

## ABSTRACT

*A new object-based method of measuring multiple-point connectivity is based on fitting power law (fractal) models to cluster size distributions. The multiple-point connectivity (implicit to the cluster size distribution) is fully described by the scaling factor and fractal dimension of the power model fitted to experimental data points. A vector interpretation of the power models indicate that they form the upper probability bound of traditional multiple-point connectivity measures. This new multiple-point measure is demonstrated for: (1) characterising the spatial continuity of mineral deposits; (2) selecting conditional simulation algorithms; and (3) post-processing stochastic images.*

## INTRODUCTION

Specific multiple-point ( $n$ -point) connectivity measures were introduced by Journel and Alabert (1989) as the expected value of the product of  $n$  indicator variables, *i.e.* the  $n$ -point non-centered indicator covariance is:

$$\delta(h; z_k) = E \left\{ \prod_{j=1}^n I(x + (j-1)h; z_k) \right\} \quad (1)$$

where  $\Pi$  represents a product, the  $n$  points are separated spatially by a fixed vector  $h \in A$ , and  $I(x; z_k) = 1$  if the RF  $Z(x) \geq z_k$  and 0 otherwise. Unlike bivariate measures of spatial continuity,  $\delta(h; z_k)$  cannot be modelled in two or three dimensions. Consequently, its value is limited to the explicit definition of  $h$ . The connectivity measure in (1) was expanded by Deutsch (1992) to include  $n$ -point configurations with variable separation vectors and threshold values as:

$$\psi(h_1, \dots, h_n; z_1, \dots, z_n) = E \left\{ \prod_{j=1}^n I(x + h_j; z_j) \right\} \quad (2)$$

A. RICHMOND

where  $h_1, \dots, h_n$  are  $n$  separation vectors, with by convention  $h_1=0$ , and  $z_1, \dots, z_n$  are  $n$  threshold values.

(1) and (2) can be interpreted as the probability that the  $n$  values are jointly above  $z_k$ , provided that  $z_j=z_k$  for all  $j$  in (2) (assumed for the remainder of the paper). To compare directly the connectivity at different threshold grades (1) and (2) can be standardised by dividing by the expectation of the corresponding indicator random variable:

$$\delta_R(h; z_k) = \delta(h; z_k) / E\{I(x; z_k)\} \in [0,1] \quad (3)$$

$$\psi_R(h_1, \dots, h_n; z_k) = \psi(h_1, \dots, h_n; z_k) / E\{I(x; z_j)\} \in [0,1] \quad (4)$$

(3) and (4) are then interpreted as the probability that the  $n$  values are jointly  $\geq z_k$  given that  $x$  is  $\geq z_k$ . As  $E\{I(x; z_k)\}$  may vary ergodically between simulations, (3) and (4) also enables the direct comparison of connectivity between multiple simulations for specific  $z_k$ .

A significant practical problem with the use of connectivity measures shown in (3) and (4) is that if the continuous variable  $z$  is discretised into  $K$  classes, there are  $K$  and  $K^n$  possible measures of  $\delta_R$  and  $\psi_R$  respectively for each (implicit) set of separation vectors. Consequently, (3) and (4) are rarely used in geostatistics except for the posterior calculation of multiple-point statistics of stochastic images. The exception being in objective functions of combinatorial optimisation algorithms (COA), where the  $K^n$  product tends to impose practical limitations due to limited computer memory (eg Deutsch, 1992; Deutsch and Gringarten, 2000). When considering multiple sets of separation vectors these practical limitations are further deteriorated. Hence, a connectivity measure that is vector-independent may be of significant practical importance.

## A NEW OBJECT-BASED APPROACH

Consider the local indicator values plotted in Figure 1 for the Walker Lake dataset. We denote a contiguous set of locations (or geo-object or cluster) for which  $i(x; z_k)=1$  as  $g(z_k)=m$ , where  $m$  is the number of adjoining locations. For example, in Figure 1 the largest geo-object is grey. Figure 1 shows that any number  $G < N$  (the total number of locations) of such geo-objects exist. Algorithms for determining the clustering of 3D objects is outside the scope of this paper, see Mehlhorn (1984) and Deutsch (1998). If we consider the  $n$ -point configurations in (3) and (4) as the vector identification of geo-objects, we can interpret  $\delta_R$  and  $\psi_R$  as the probability of belonging to a geo-object of size  $m \geq n$ , provided that the separation vectors  $h_1, \dots, h_n$  define adjoining locations. Hence, geo-objects contain implicitly some multiple-point connectivity information.

Figure 2 shows that geo-object size distributions may closely follow power-law (fractal) models, defined as:

$$\log(G_m) = \alpha - \beta \log(m) \quad (5)$$

where  $G_m$  is the cumulative number of geo-objects  $\geq m$ ;  $\alpha$  is a scaling factor function of the geo-objects in the study area  $A$ , and  $\beta$  is the fractal dimension of the geo-object population that defines the relative number of large and small geo-objects. In (5), when  $\alpha$  is high the number of geo-objects is high, and when  $\beta$  is high the number of small geo-objects is high relative to the number of large geo-objects. In Figure 2, note that  $\alpha$  and  $\beta$  vary with the indicator threshold value  $z_k$ , and the experimental data points close to the  $x$ -axis tend to deviate somewhat from the fitted models. The variability of  $\alpha$  and  $\beta$  characterise the multiple-point connectivity of nodes at the various threshold values. The bottom end deviation is known as a truncation effect, and is due to limitations on the study area  $A$ .

The information in Figure 1 can be used to calculate the probability  $p(m; z_k)$  that a location  $x$  belongs to a geo-object of at least size  $m$ , given that  $z(x) \geq z_k$ , where

$$p(m; z_k) = \begin{cases} G_m / G_1 & \text{for the experimental points} \\ [\alpha - \beta \log(m)]^{10} / \alpha^{10} & \text{for the fitted models} \end{cases} \quad (6)$$

which calculates the probability that a location is jointly connected to at least  $n$  other points  $\geq z_k$ . If the geo-objects are all one-dimensional with orientation  $\theta$ , then  $p(m; z_k) = \delta_R(h; z_k)$  if  $h = \theta$ . Hence, the probabilities  $p(m; z_k)$ ,  $m = 1, \dots, M$  also form the upper bound of the vector-based measures  $\delta_R$  and  $\psi_R$ , i.e.  $\delta_R(h; z_k)$  and  $\psi_R(h_1, \dots, h_n; z_k) \in [0, p(m; z_k)]$ ,  $n = m$ . Thus, (5) could be considered analogous to an omni-directional indicator bivariate measure of spatial continuity. Furthermore, just as the bivariate spatial continuity is fully defined by the parameter values of the variogram model, the multiple-point connectivity (implicit to the cluster size distribution) is fully described by the scaling factor  $\alpha$  and the fractal dimension  $\beta$  of the fitted power model.

(5) provides an easy and valuable tool for measuring and validating the multiple-point statistics of pixel-based simulation algorithms. In addition, the practical limitations of vector-based multiple-point continuity measurements in COA are removed, allowing rapid post processing of stochastic images.

## EXAMPLES

The connectivity of data extremes in the simulated images is of particular interest when the realisations represent metal values in an ore body that is being mined with a high cut-off grade. Geo-object distributions at 0.87% Cu ( $\pm 90^{\text{th}}$  cumulative percentile) for 3 benches in a copper mine are shown in Figure 3. The geo-objects were determined from grade control data adjusted to a regular grid. Note that,  $\beta$  was consistent at around 1.05 for these 3 benches selected at 50 m vertical intervals. This has two important implications for their imaging by

conditional simulation (CS). Firstly, for deeper parts of the mine (or analogous mineral deposits) the CS algorithm chosen should result in stochastic images with  $\beta \approx 1.05$  for 0.87% Cu ( $\pm 90^{\text{th}}$  cumulative percentile). Secondly, if the adopted CS algorithm results in an unsuitable  $\beta$ , then the stochastic images should be post-processed to more accurately reflect the adopted power law model.

Figure 4 shows the experimental power law models for three sequential simulation realisations generated from a subset of an exhaustive image using Gaussian (SGS), Mosaic (SMS) and Indicator (SIS) RFs. Note that, the high entropy Gaussian RF results in the highest  $\beta$  value, the Indicator RF produces the lowest  $\beta$  value. The Indicator RF image most closely represents the geo-object distribution of the original image. Figure 5 indicates that SGS is a poor CS algorithm choice for the Walker Lake dataset if reproducing the connectivity of high valued samples is important.

If images from alternate CS algorithms cannot reproduce reasonably the adopted power model then post-processing by COAs is an option. The optimisation constraint used could be the reproduction of either the experimental or modeled (7) geo-object size distributions, *i.e.* minimisation of an objective function (OF)

$$OF = w_{\alpha} (\alpha_k^{ref} - \alpha_k^*)^2 + w_{\beta} (\beta_k^{ref} - \beta_k^*)^2 + w_t \sum_{j=m}^M t_j \quad (7)$$

where  $M$  is the number of geo-object sizes to consider, *ref* and \* represent the reference and current images,  $t_j$  is the tortuosity of geo-object  $j$ , and  $w_{\alpha}$ ,  $w_{\beta}$ , and  $w_t$  are the OF component weights. The last component of the OF was key in controlling the “texture” of the geo-objects during computational experiments. Low values of  $w_t$  result in highly anastomosing or “sieve-like” geo-objects. Conversely, geo-objects with compact textures result from high values of  $w_t$ .

## CONCLUSIONS

A vector-independent multiple-point measure was proposed that demonstrates significant promise in characterising stochastic images in a computationally tractable manner. This new object-based multiple-point measure requires the fitting of a power law model to geo-object distributions calculated from complete stochastic images. In practice, these parameters could be calculated from a training image or adopted from an adjacent area or similar deposit. The power law models provide support for selecting CS algorithms and can be incorporated in objective functions when using COAs

## REFERENCES

- Deusch, C.V. (1992) *Annealing techniques applied to reservoir modeling and the integration of geological and engineering (well test) data*, PhD thesis, Stanford University.

- Deutsch, C.V. (1998) *Fortran programs for calculating connectivity of three-dimensional numerical models and for ranking multiple simulations*, Computers & Geosciences, vol. 24, no. 1, pp. 69-76.
- Deutsch, C.V. and Gringarten, E. (2000) *Accounting for multiple-point continuity in geostatistical modelling*, in: Proceedings of the 6<sup>th</sup> International Geostatistics Congress, pp. 156-165.
- Journel, A.G. and Alabert, F. (1989) *Non-gaussian data expansion in the earth sciences*, Terra Nova, vol. 1, pp. 123-134.
- Mehlhorn, K. (1984) *Data structures and algorithms*, Volume 3: multi-dimensional searching and computational geometry, Springer-Verlag.

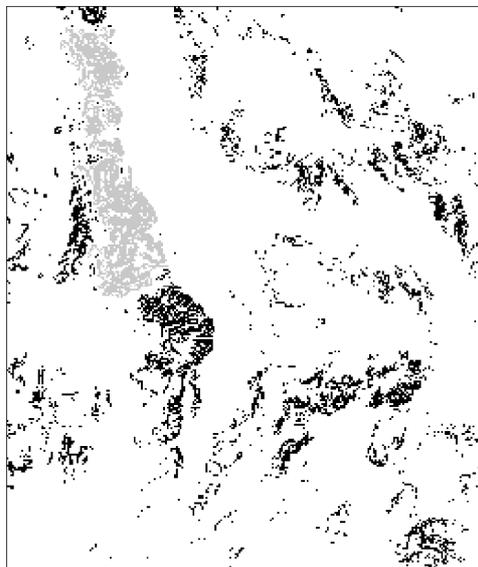


Figure 1: geo-objects for 90<sup>th</sup> cumulative % (gray/black  $\geq 784$ ; white  $< 784$ ).

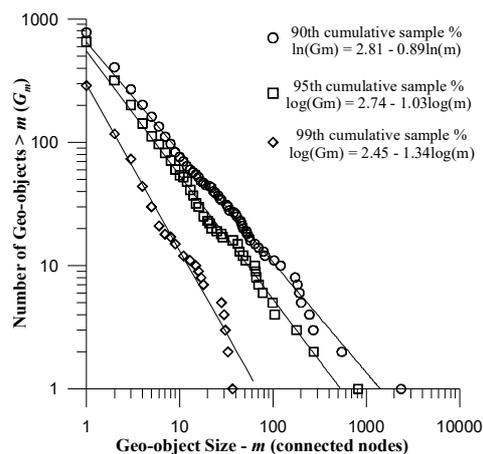


Figure 2: geo-object distributions and fitted power models.

A. RICHMOND

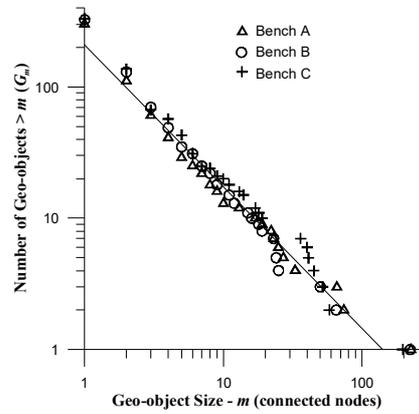


Figure 3: Geo-object distributions for 3 benches in a copper deposit.

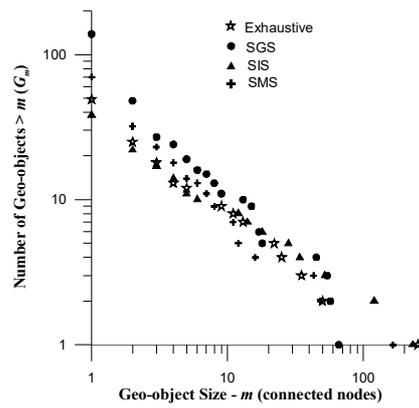


Figure 4: Geo-object distributions for alternate RF models.

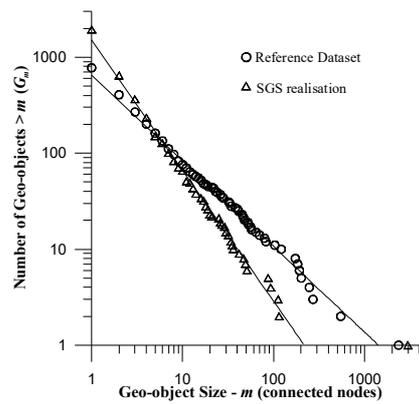


Figure 5: Geo-object distributions for SGS realisation and reference Walker Lake dataset.