Experimental Variogram Calculation using Bernstein Polynomials

John Manchuk and Oy Leuangthong

Inference of the variogram commonly relies on the method-of-moments approach to obtain an experimental variogram, on which variogram modeling is based. Conventional sampling practice leads to a set of irregularly spaced data that is available for variogram calculation; this necessarily requires some binning of the paired data to obtain reliable estimates of the variogram. For some software, this grouping of data pairs depends on the specified lag and angle tolerances, and bandwidths. We can consider this as one form of smoothing of the h-scatterplot for reliable inference. This paper proposes an alternative smoothing approach for improved variogram estimation; rather than binning data, we propose to interpolate the variogram cloud directly. Bernstein polynomials can be used to infer a multivariate density function from a set of sample data. We can apply this approach to the variogram cloud, which consists of the experimental variogram and lag vectors $\mathbf{h}$*, and for any lag of interest* $\mathbf{h}_0$*, we use the multivariate densities to calculate the covariance and determine the corresponding variogram value. This permits inference of the experimental variogram in any direction and at any lag not beyond the extents of the data.*

Introduction

Acceptability of estimation via Kriging is strongly dependent on the variogram, which describes the spatial correlation of a random variable. Acquiring the best estimate of the true variogram is an important component of geostatistics. When a substantial amount of sample data is available and those samples adequately partition the domain of interest, calculation of the experimental variogram directly is a good estimate of the correct variogram. However, when sample data are sparse, little information is available in any given direction for which the experimental variogram is to be calculated. Data and directions must be mixed to obtain any insight in regards to directional variograms – in the worst case it may only be possible to obtain an omnidirectional isotropic variogram. Practitioners may be required to use analogs or other similar phenomena to infer a variogram.

For sparse data, mixing of directions to infer the variogram is needed to obtain an adequate number of sample pairs and lag separations for a particular direction. A common technique is used in the *GSLIB* program for variogram calculation of irregularly spaced data, *gamv* (Deutsch and Journel, 1998). For a particular lag, data are binned based on tolerances of a spherical coordinate system. In cases of sparse data, tolerances must be set large to obtain a reasonable estimate, which approaches the omnidirectional case.

A more robust method of mixing directional information is developed. It involves modeling the distribution that is defined by the experimental variogram. The distribution is modeled using Bernstein polynomials: a type of non-parametric function interpolation (Bernstein, 1912; Lorentz, 1986). Once the model is obtained, it is possible to calculate an experimental variogram in any direction and for any lag distance within the sample data domain. After developing some theory on Bernstein polynomials for distribution modeling, the distribution defined by the variogram will be described. This will be followed by two examples involving sparse data in two dimensional space and a third applying the method to variogram map generation.

Bernstein Polynomials

Bernstein polynomials are a form of non-parametric function interpolator. They have several amenable characteristics for modeling distribution functions including non-negativity, free of boundary bias, differentiable and integrable, and monotonic. Bernstein polynomials have been applied to modeling univariate distribution functions (Kakizawa, 2004; Babu, Canty, and Chaubey, 2002; Petrone, 1999) as well as multivariate distributions (Sancetta and Satchell, 2004; Sancetta, 2007; Kolev, Anjos, and Mendes, 2006).

Any continuous function defined on the interval [0,1] can be approximated by a sequence of Bernstein polynomials. Various transformations exist for functions that do not meet this condition, for example (1), (2), and (3). An approximation to a function $f(x)$ by Bernstein polynomials of degree n, $B_n(f\,;x)$, with x in

[0,1] is given by (4). B_n also interpolates the endpoints of the interval being approximated: $B_n(f\,;\,0) = f(0)$ and $B_n(f\,;\,1) = f(1)$. Each $B_{i,n}(x)$ in (4) are referred to as basis functions.

$$g(y) \in [a,b] \rightarrow f(x) \in [0,1] : x = (y-a)/(b-a) \tag{1}$$

$$g(y) \in [0,\infty] \rightarrow f(x) \in [0,1] : x = y/(y+1) \tag{2}$$

$$g(y) \in [-\infty,\infty] \rightarrow f(x) \in [0,1] : x = 1/2 + (1/\pi)\tan^{-1}(y) \tag{3}$$

$$B_n(f;x) = \sum_{i=1}^{n} f\left(\tfrac{i}{n}\right)\binom{n}{i} x^i (1-x)^{n-i} = \sum_{i=1}^{n} f\left(\tfrac{i}{n}\right) B_{i,n}(x) \quad \text{where } \binom{n}{i} = \frac{n!}{i!(n-i)!} \tag{4}$$

The underlying function f must be known at $x = (i\,/\,n)$, $i = 0,\ldots,n$. For purposes of density function estimation, the densities for each $x = (i\,/\,n)$ could be calculated with a histogram. The density for any other x would be approximated by $B_n(f\,;\,x)$. For a multivariate function of d dimensions defined on the $[0,1]^d$ hypercube, the approximation is defined by (5). By estimating f using a multivariate histogram, $B_n(f\,;\,x_1,\ldots,x_d)$ approximates the density for any vector $\mathbf{x} = (x_1\ x_2\ \ldots\ x_d)$.

$$B_n(f;\mathbf{x}) = \sum_{i_1=0}^{n_1} \cdots \sum_{i_d=0}^{n_d} f\left(\tfrac{i_1}{n_1},\ldots,\tfrac{i_d}{n_d}\right) \prod_{k=1}^{d} B_{i_k,n_k}(x_k) \tag{5}$$

Bernstein polynomials can be used to model cumulative densities as well. Assuming the cdf is differentiable, Bernstein polynomials can be used to model the integral of the density function defined by f. The previously mentioned properties of Bernstein polynomials are important:

Non-negative: if f is positive on [0,1], $B_n(f\,;\,x)$ is also positive: B_n is a monotone operator:

$$0 \le f \le 1 \rightarrow 0 \le B_n(f;x) \le 1 \tag{6}$$

Differentiable and integrable: if f is positive on [0,1], its integral is monotonically increasing on [0,1] and the integral of $B_n(f\,;\,x)$ is also monotonically increasing. For a proof see Phillips (2003). This is stated in terms of derivatives:

$$f' \ge 0 \rightarrow B_n'(f;x) \ge 0 \tag{7}$$

The integral of (5) is required for a multivariate cumulative distribution model. Because each basis function depends only on one x_k, integration can be done independently for each variable by (8) where $\beta_{i,n}(x)$ represents the integrated basis functions. The result is (9) where B has been replaced by β in the product and the sums start at 1 rather than 0. An additional advantage of the multivariate Bernstein polynomial model is that conditional distributions are accessible. The conditional distribution in terms of a particular variable x_j is accessible via Bayes law (10). Note that $\mathbf{x}_{(j)}$ denotes $\mathbf{x}$ excluding element j.

$$\begin{aligned} B_n(f;x) &= \sum_{i=1}^{n} f\left(\tfrac{i}{n}\right) n \binom{n-1}{i-1} \int_0^x t^{i-1}(1-t)^{n-i+1}\,dt \\ &= \sum_{i=1}^{n} f\left(\tfrac{i}{n}\right) \beta_{i,n}(x) \end{aligned} \tag{8}$$

$$B_n(f;\mathbf{x}) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_d=1}^{n_d} f\left(\tfrac{i_1}{n_1},\ldots,\tfrac{i_d}{n_d}\right) \prod_{k=1}^{d} \beta_{i_k,n_k}(x_k) \tag{9}$$

$$B_n(f;\mathbf{x}_{(j)} \mid x_j) = \frac{B_n(f;\mathbf{x})}{B_n(f;x_j)} \tag{10}$$

Constructing a multivariate distribution using Bernstein polynomials amounts to choosing the number of known points per dimension, n_k, $k = 1,\ldots,d$, and calculating the density at each point $\mathbf{x} = (i_1\,/\,n_1,\ i_2\,/\,n_2,\ldots,\ i_d\,/\,n_d)$, $i_j = 1,\ldots,n_j$, $j = 1,\ldots d$. Densities can be calculated using any technique such as histograms or kernel density estimation. Multivariate histograms have the advantage of sparsity, which is ideal for problems involving several variables and a large number of samples. For example, a problem consisting of 5 variables and 100 points per dimension involves 100^5 or 10 billion floating point numbers. However, most points would be characterized by zero density. For any histogram and a sample size of N, there will be at most N non-zero densities.

Experimental Variogram Distribution

Calculating an experimental variogram, $\hat{\gamma}(\mathbf{h})$, is accomplished by (11) where $\mathbf{h} = (r, \theta, \varphi)$ is the vector in spherical coordinates separating two samples, $z(\mathbf{u}_i)$ and $z(\mathbf{u}_j)$, $\mathbf{u} = (u_x, u_y, u_z)$, and $N(\mathbf{h})$ is the total number of samples separated by $\mathbf{h}$. Spherical and Cartesian coordinate systems are defined by (12). For irregularly spaced data, the condition $\mathbf{u}_i - \mathbf{u}_j = \mathbf{h}$ must be relaxed: tolerances, $\boldsymbol{\varepsilon}$, are placed on $\mathbf{h}$ in (13) such that an adequate number of pairs $N(\mathbf{h})$ can be obtained for the calculation. Specification of $\boldsymbol{\varepsilon}$ imparts tolerances on the direction and magnitude of $\mathbf{h}$.

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 \cdot N(\mathbf{h})} \sum_{\mathbf{u}_i - \mathbf{u}_j = \mathbf{h}}^{N(\mathbf{h})} \left[z(\mathbf{u}_i) - z(\mathbf{u}_j) \right]^2 \tag{11}$$

$$\left.\begin{aligned} \Delta x &= (\mathbf{u}_i - \mathbf{u}_j)_x = r\cos\theta\sin\varphi \\ \Delta y &= (\mathbf{u}_i - \mathbf{u}_j)_y = r\sin\theta\sin\varphi \\ \Delta z &= (\mathbf{u}_i - \mathbf{u}_j)_z = r\cos\varphi \end{aligned}\right\} \leftrightarrow \left\{\begin{aligned} r &= \left\|\mathbf{u}_i - \mathbf{u}_j\right\| \\ \theta &= \tan^{-1}\left(\Delta y / \Delta x\right) \\ \varphi &= \cos^{-1}\left(\Delta z / r\right) \end{aligned}\right. \tag{12}$$

$$\mathbf{h} - \boldsymbol{\varepsilon} \le (\mathbf{u}_i - \mathbf{u}_j) \le \mathbf{h} + \boldsymbol{\varepsilon} \tag{13}$$

For sparse data, $\boldsymbol{\varepsilon}$ may have to be set large to adequately characterize the experimental variogram leading to substantial mixing of directions and distances. For a particular $\mathbf{h}$, the set $S: \mathbf{u}_i - \mathbf{u}_j \in \mathbf{h} \pm \boldsymbol{\varepsilon}$ characterizes a bivariate distribution conditional to $\mathbf{h}$ with variables $Z[\mathbf{u}]$ and $Z[\mathbf{u} + (\mathbf{h} \pm \boldsymbol{\varepsilon})]$ from which the variogram is calculated. Without considering a specific lag and tolerance, the full set, S: $\mathbf{u}_i - \mathbf{u}_j$, specifies a multivariate distribution with variables $\mathbf{h}$, $Z[\mathbf{u}]$, and $Z[\mathbf{u} + \mathbf{h}]$. In three dimensional space the distribution would be five dimensional: r, θ, φ, $Z[\mathbf{u}]$, and $Z[\mathbf{u} + \mathbf{h}]$. This concept can be illustrated with a one dimensional example where data may be available from along an exploration drillhole or well. Ten synthetic data are defined along the x coordinate direction with positions and values specified by Table 1. The multivariate distribution is defined by $h = |x_i - x_j|$, $Z(x)$, and $Z(x + h)$. Absolute values of h are considered in this example because the experimental variogram is an even function. In higher dimensions, this symmetry is handled as $\mathbf{h}_{ij} = -\mathbf{h}_{ji}$: $\mathbf{h}_{ij} = \mathbf{u}_i - \mathbf{u}_j$. In this example, there are a total of 100 pairs characterizing the multivariate distribution. To calculate the experimental variogram as in (11) for a particular $h \pm \varepsilon$, all pairs within the domain defined by $\{h \pm \varepsilon, Z[x], Z[x + (h \pm \varepsilon)]\}$ are identified and used in (11). This is done with $h \pm \varepsilon = [0,5)$ and $h \pm \varepsilon = [5,10)$ for the example in Figure 1. The associated standardized experimental variogram values are 0.29 and 1.21 respectively.

Table 1: Synthetic data for one dimensional example

x	1	4	6	11	15	16	18	21	25	29
$Z(x)$	2	2	4	6	11	14	13	1	3	0

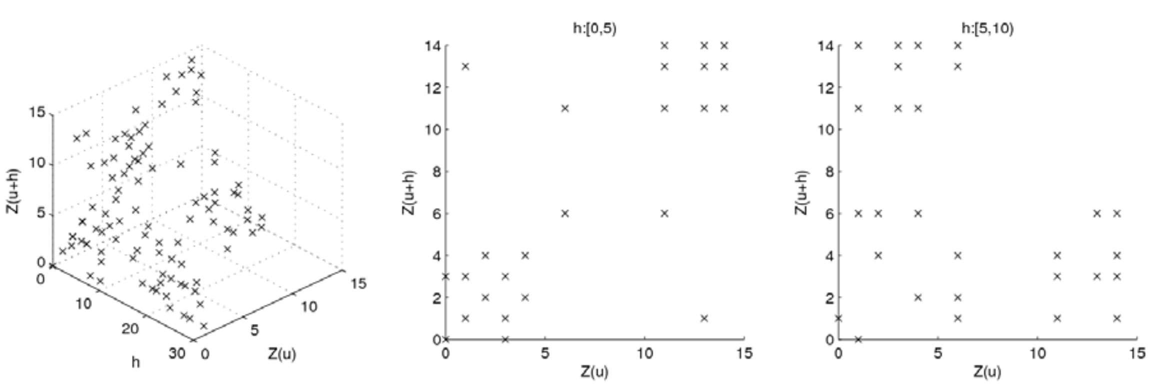

Figure 1: Variogram cloud (left) and experimental bivariate plots

Rather than specifying tolerance parameters to acquire an acceptable experimental variogram, the multivariate distribution can be modeled with Bernstein polynomials. The cloud defined by **h**, Z[**u**], and Z[**u** + **h**] must first be transformed to the $[0,1]^d$ hypercube, where $d = 5$ for $Z(\mathbf{u}) \in \mathbb{R}^3$. Z[**u**] is transformed according to (1) where the minimum and maximum of Z are determined from the sample data (14). Transformation of the spherical coordinates is dependent on the direction, (θ_0,φ_0), being explored. Bernstein polynomial interpolation was not designed for periodic coordinate systems; however, the space can be adjusted to accommodate this. The system is first rotated so that θ_0 is zero degrees. Angles for each $\mathbf{u}_i - \mathbf{u}_j$ pair are calculated in reference to θ_0 and made to range from $-\pi/2$ to $\pi/2$. Angles in quadrant 1 remain unchanged, quadrants 2 and 3 are reflected into quadrants 4 and 1 respectively, and quadrant 4 angles are reduced by 2π. Reflections are permitted because the variogram cloud is an even function. A similar transformation is used with φ_0 as the reference. Instead of being set to zero, φ_0 is set to $\pi/2$ and calculated φ range from 0 to π. These transformations are supported by Figure 2. Afterwards, θ and φ are changed to [0,1] according to (15) and the radius r according to (16). Transformed variables will be denoted with a tilde overbar.

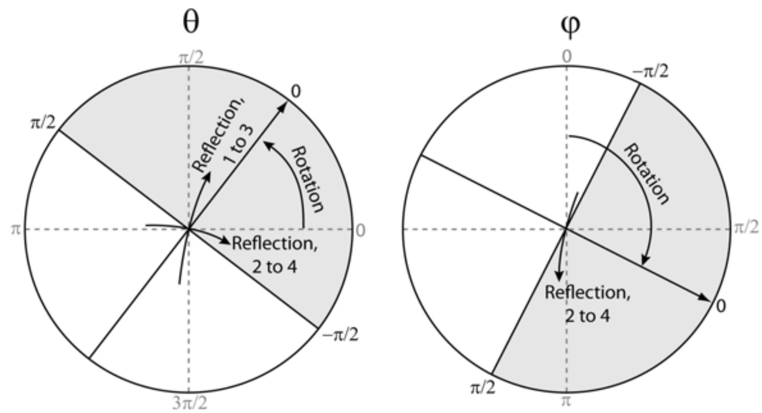

Figure 2: Spherical coordinate transformations

$$\tilde{z}(\mathbf{u}) = \left(z(\mathbf{u}) - \min\{Z\}\right) / \left(\max\{Z\} - \min\{Z\}\right) \tag{14}$$

$$\left\{\tilde{\theta},\tilde{\varphi}\right\} = \{\theta,\varphi\} / \pi + 1/2 \tag{15}$$

$$\tilde{r} = r / \max\{r\} \tag{16}$$

A multivariate histogram is calculated for the cloud defined by $\tilde{\mathbf{h}}$, $\tilde{Z}[\tilde{\mathbf{u}}]$, and $\tilde{Z}[\tilde{\mathbf{u}}+\tilde{\mathbf{h}}]$. Frequencies correspond to the values of $f(\cdot)$ in (5) or (9). Parameters required from the user are the number of bins to use per dimension for the multivariate histogram. For three dimensional data there are four parameters to choose: the number of bins to partition θ, which is analogous to an azimuth tolerance; the number of bins to partition φ, which is analogous to a dip tolerance; the number of bins to partition the radius, similar to a lag tolerance; and the number of bins to partition Z. The first three create bins in a spherical coordinate system and the last creates two dimensional Cartesian bins in the Z[**u**]Z[**u** + **h**] plane (Figure 3). Within this plane and for a particular direction and lag, Bernstein polynomials interpolate a bivariate density function from which a covariance, thus a variogram, is calculated.

Calculating the experimental variogram for a particular direction first requires transformation of the vector into the same space as the distribution via (15) and (16). The $[0,1]^2$ plane that represents $\tilde{Z}$ is then sampled using a regular grid of points, where each point represents a pair $(\tilde{z}_u, \tilde{z}_{u+h})$. Letting $\mathbf{x} = [\tilde{\theta}_0, \tilde{\varphi}_0, \tilde{r}, \tilde{z}_u, \tilde{z}_{u+h}]$ and using (5) along with the multivariate histogram frequencies, f, calculated previously, an interpolated density is determined. If the set of densities for the full grid of points is denoted $w_{ij}(\mathbf{x})$ and the inverse of (14) is used to map $(\tilde{z}_u, \tilde{z}_{u+h}) \to (z_u, z_{u+h})$, then the variogram is calculated using (17), where $\mathbf{h}_0=[\theta_0,\varphi_0,r]$ and the covariance, $C(\mathbf{h}_0)$, is the usual weighted covariance calculated for the grid of w and z values.

$$\hat{\gamma}\left(\mathbf{h}_0\right) = 1 - \mathrm{Cov}\left\{z_u, z_{u+h}\right\} / \left(\sigma_u \cdot \sigma_{u+h}\right) = 1 - C(\mathbf{h}_0) / \sigma^2(\mathbf{h}_0) \tag{17}$$

Using this methodology the distribution for the previous example was fit and the experimental variogram calculated for *h*=2.5 and *h*=7.5. Experimental values were 0.51 and 1.20 respectively using a multivariate histogram with 20 radius bins, 100 z bins, and a 51 by 51 grid to sample w_{ij} and z_{ij} for (16). Visualization of the three dimensional distribution is done with volume rendering (Figure 4) and the density functions at the two prescribed *h* values are shown in Figure 5.

Measuring Information Content

The importance of experimental variogram points from *gamv* are commonly measured using the number of $Z[\mathbf{u}]Z[\mathbf{u}+\mathbf{h}]$ pairs involved. A similar measure can be constructed for experimental variograms calculated using Bernstein polynomials. For a particular direction, each experimental point is calculated with the same resampling grid for *Z*, so the number of pairs may not be used. However, each resampled point is interpolated from all the available data and receives weights calculated from the basis functions in (5). Summing the weights that all data have received for a resample point is a measure of its information content. The sum of all information content values for an entire resampling grid provides a measure of significance for the associated experimental variogram point.

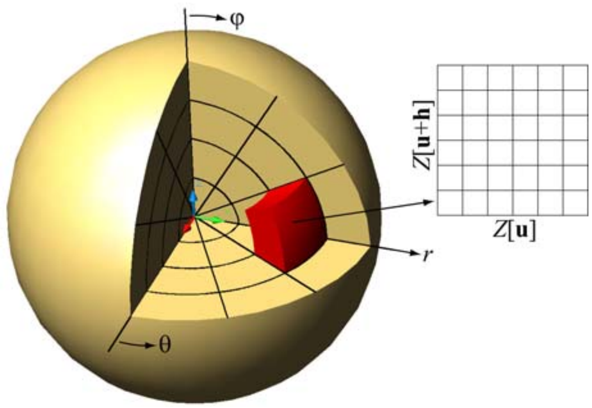

Figure 3: Multivariate histogram bin geometry

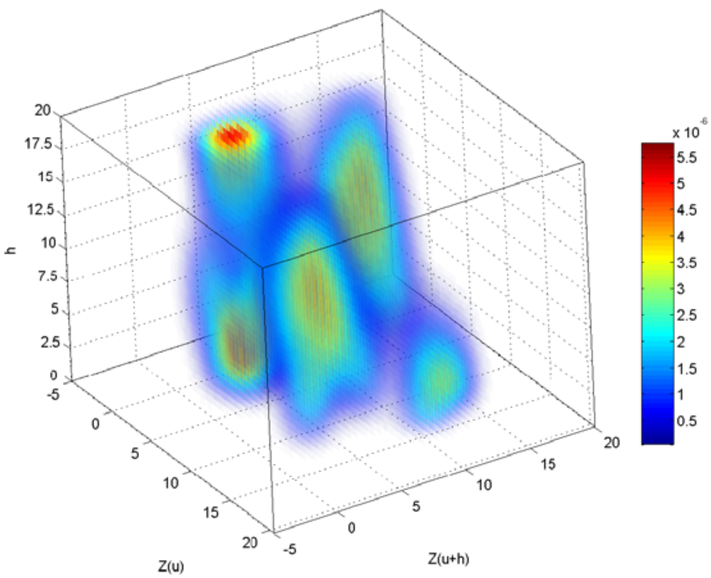

Figure 4: Three dimensional density function

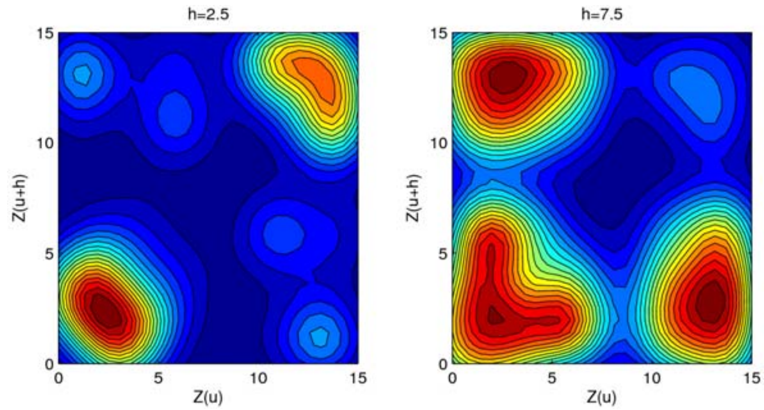

Figure 5: Experimental bivariate plots

Examples

To illustrate the usefulness of interpolating the distribution defined by the full variogram cloud when data is sparse, two examples will be covered and compared to existing methods for determining experimental variograms that utilize (11) and a third example applying the method to variogram map generation. Examples are two dimensional and consist of data that is sparse relative to the field they represent. In the first example, a model was synthetically created with a known spherical variogram (Table 2) and a set of 76 random samples was extracted from the model (Figure 6). The total field size is 15000 square and samples have a minimum spacing of roughly 1500. Very few pairs exist below this spacing. Inferring the short range structure of variograms is important, where the short range in this example is considered < 1500.

Table 2: Variogram for first synthetic example

Structure	Type	Variance	Azimuth	Major Range	Minor Range
1	Nugget	0.0			
2	Spherical	0.37	45	3000	1000
3	Spherical	0.63	45	3000	3000

Experimental variograms were calculated using *gamv* for the sample set. Directions explored were 45 and 135 degrees. For *gamv*, the lag size was 600 and tolerance 300. Referring to Figure 6, the second and third points for the experimental variogram at 45 degrees are characterized by only one pair. Variograms calculated with the Bernstein polynomial method (*gamest*) used the bin parameters in Table 3 and 20 lags of size 300. This model infers short range information by mixing what is known for all available **h** values in the data set not including pairs at **h** = 0.

Table 3: Example 1 bernstein polynomial fit parameters

Direction	Radius bins	Azimuth bins	Z bins	Resampling grid
45	20	70	200	20×20
135	20	20	200	20×20

The resulting experimental variogram using the Bernstein polynomial method is smooth and allows calculation of $\gamma(\mathbf{h})$ for any **h**. Although the results are not identical to the true underlying variogram, which should be considered unavailable, several components of a variogram can be interpreted from this first example: a nugget effect; range values; the possibility of a significant trend where the variogram extends above the sill; and the variogram shape. One might interpret a variogram with a nugget effect of 0, two spherical nested structures, and ranges of 2000 and 3000 at 45 and 135 degrees respectively as well as a slight trend in the 45 degree direction. This information would be difficult to ascertain from the *gamv* experimental variograms, especially if the points characterized with only one pair were disregarded.

The second example uses sample data from Deutsch and Journel (1998). In one of the examples on pages 257-259, the first 97 samples of cluster.dat (97data.dat) are used to calculate an experimental variogram, which is shown on page 258, and the exhaustive 50 by 50 grid of data (true.dat) were used to produce the variograms on page 259. Parameters for *gamv* and *gam* were set identical to those in the text, the only difference being that standardized variograms are used here. Data and results are shown in Figure 7. A comparison of the Bernstein polynomial variograms with the exhaustive data are shown in Figure 8.

Generating the Bernstein polynomial fit was done using bins parameters in Table 4. In Figure 7 and Figure 8, the nugget effect and ranges selected using the Bernstein polynomial fit would be more accurate than those selected using the experimental variograms when compared to the exhaustive results. This is especially true for the nugget effect. The zonal anisotropy that is shown in Figure 8 is not captured by the sample set, so is not represented by the results of either the *gamv* or the Bernstein methods.

This last example applies the Bernstein polynomial method to generating a variogram map using the data from the previous example. Because the directional information is interpolated in spherical coordinates, it is appropriate to create a variogram map with the same coordinate system. A structured grid of r, θ, φ values is used to evaluate the experimental variogram (Figure 9). Each value is calculated exactly as in the previous two examples with a resampling grid for $Z[\mathbf{u}]$ and $Z[\mathbf{u}+\mathbf{h}]$. Calculating a reasonable variogram at this resolution would be difficult for the traditional method of moments.

Table 4: Example 2 bernstein polynomial fit parameters

Direction	Radius bins	Azimuth bins	Z bins	Resampling grid
0	50	170	500	20×20
90	50	170	500	20×20

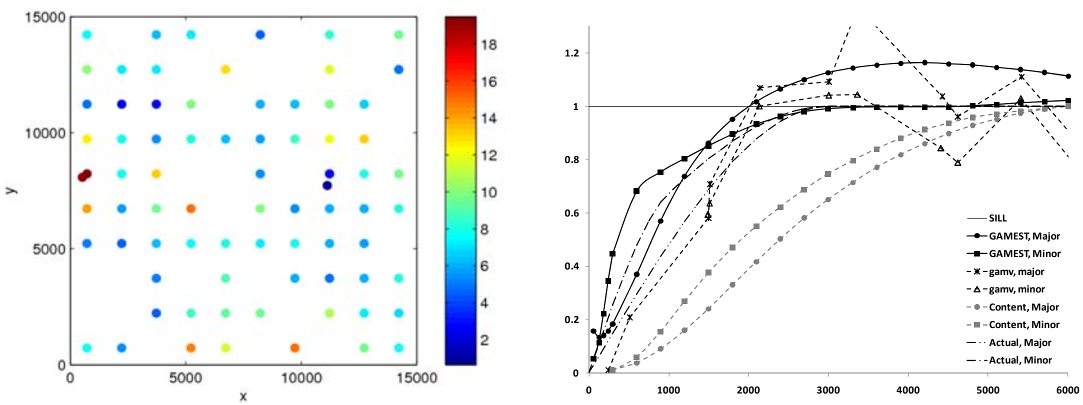

Figure 6: Example 1 samples (left) and variograms

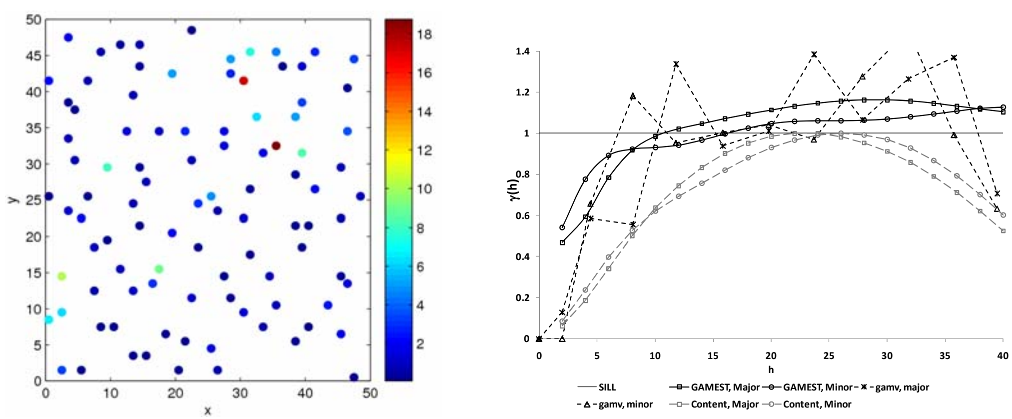

Figure 7: Example 2 sample data (left) and variograms

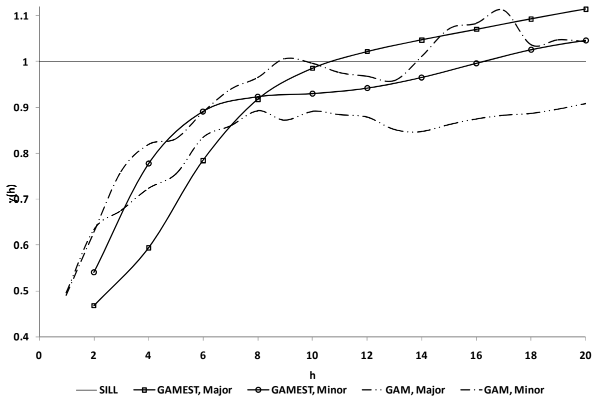

Figure 8: Exhaustive experimental variograms and Bernstein models

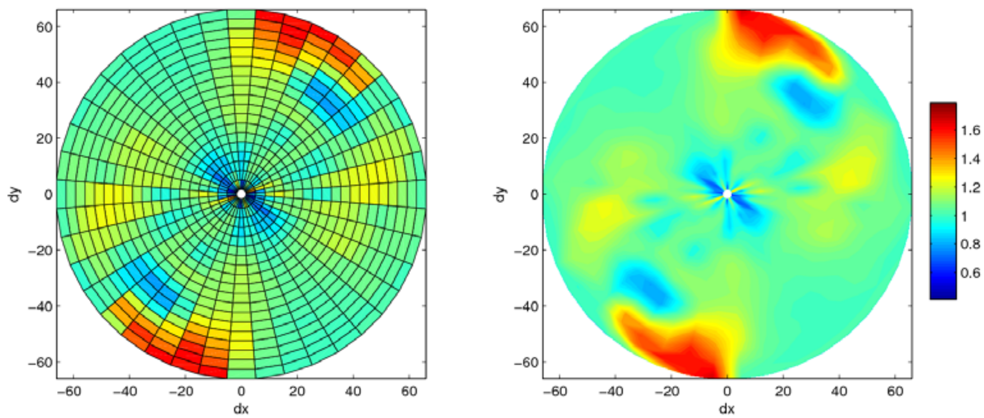

Figure 9: Variogram map as calculated on a polar grid (left) and the contoured equivalent

Conclusions

Inferring the correct variogram in a geostatistical study with sparse data can be difficult. In some cases, existing methods are unsatisfactory. This paper introduced a new method of mixing available information to obtain more interpretable experimental variograms. A non parametric density estimation method that uses Bernstein polynomials was used to fit the full variogram cloud of a sample data set. Then the variogram value can be calculated for any direction and any lag within the extents of the data. Two examples show this method can be used to identify structural elements such as the nugget effect, range, trends, and anisotropy and in a more interpretable way. A third example applied the method to produce a higher resolution variogram map than would be obtainable using the traditional method of moments approach. Non parametric methods are promising for variogram interpretation and can result in more appropriate variogram models.

References

Babu, J., Canty, A. J., Chaubey, Y. P., 2002, Applications of Bernstein polynomials for smooth estimation of a distribution and density function: Journal of Statistical Planning and Inference, No. 105, p. 377–392

Bernstein, S., 1912, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilities: Comm. Soc. Math. Kharkov 13, 1-2

Deutsch, C. V., and Journal, A. G., 1998, GSLIB: Geostatistical software library and user's guide: Oxford, 384 p.

Kakizawa, Y., 2004, Bernstein polynomial probability density estimation: Journal of Nonparametric Statistics, Vol. 16, No. 5, p. 709–729

Kolev, N., Anjos, U., and Mendes, B. V., 2006, Copulas: A review and recent developments: Stochastic Models, Vol. 22, No. 4, p. 617 – 660

Lorentz, G. G., 1986, Bernstein polynomials: American Mathematical Society, 134 p.

Petrone, S., 1999, Bayesian density estimation using Bernstein polynomials: The Canadian Journal of Statistics, Vol. 27, No. 1, p. 105–126

Sancetta, A., 2007, Nonparametric estimation of distributions with given marginals via Bernstein–Kantorovich polynomials: L1 and pointwise convergence theory: Journal of Multivariate Analysis, No. 98, p. 1376 – 1390

Sancetta, A., and Satchell, S., 2004, The Bernstein copula and its applications to modeling and approximations of multivariate distributions: Economic Theory, No. 20, p. 535–562