

J-Bessel

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \frac{2^{\theta_d} \Gamma(\theta_d + 1)}{(\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r)^{\theta_d}} J_{\theta_d} (\Omega_{\theta_d} \|\mathbf{h}\| / \theta_r) \right] \text{ for all } \mathbf{h}$$

where $\theta_s \geq 0$, $\theta_r \geq 0$, $\theta_d \geq 0$, Ω_{θ_d} must satisfy,

$$\min_{B > 0, \gamma(B) = \theta_s, \gamma'(B) < 0} B = \theta_r$$

$\Gamma(\theta_k)$ is the gamma function,

$$\Gamma(y) = \int_0^{\infty} x^{y-1} \exp(-x) dx$$

and $J_{\theta_d}(\bullet)$ is the J-Bessel function (Abramowitz and Stegun, 1965, p. 358).

Stable

The semivariogram model is

$$\gamma(\mathbf{h}; \boldsymbol{\theta}) = \theta_s \left[1 - \exp \left(-3 \left(\frac{\|\mathbf{h}\|}{\theta_r} \right)^{\theta_e} \right) \right] \text{ for all } \mathbf{h},$$

where $\theta_s \geq 0$ and $0 \leq \theta_e \leq 2$. Because this model has unstable behavior without nugget, by default the Geostatistical Analyst adds a small nugget to the model, equal to 1/1000 of the sample variance computed from the data.

Crosscovariance models

Crosscovariance models in the Geostatistical Analyst use “coregionalization” models, which means that they are of the same families as the covariance forms of the semivariogram models listed above. Crossvariograms are not used in the Geostatistical Analyst. The traditional crossvariogram (Matheron, 1965) can only be used under certain conditions (Journel and Huijbregts, 1978, p. 236; Cressie, 1993, p. 67; Ver Hoef and Cressie, 1993), and it is not optimal otherwise. Crosscovariances allow models that can have spatial shifts (Journel and Huijbregts, 1978, p. 41; Ver Hoef and Cressie, 1993), and the empirical crosscovariance surface allows the user to visually inspect for such shifts.

Coregionalization models for crosscovariance (sometimes given as crossvariograms, but the ideas and models are readily adapted to crosscovariances) are described in Journel and Huijbregts (1978, p. 40), Isaaks and Srivastava (1989, p. 390), Goovaerts, (1997, p. 107), and Chiles and Delfiner (1999, p. 339). The Geostatistical Analyst adapts these models by allowing a spatial shift between any two variables (Ver Hoef and Cressie, 1993). This adds two parameters to the model to describe the shift in the x-coordinate and y-coordinate.

Fitting semivariogram and covariance models

The fitting algorithm begins by obtaining a preliminary estimate for the range of the data called stage 1. Use $Z_j^k(\mathbf{s}_i)$ to denote the j th measurement of variable type k at the i th spatial location \mathbf{s}_i .

Stage 1

Geostatistical Analyst first scales each dataset, $\tilde{Z}_j^k(\mathbf{s}_i) = Z_j^k(\mathbf{s}_i) / s_k$ where s_k is the sample standard deviation. Stage 1 begins by assuming an isotropic model, and it computes the empirical semivariogram (or covariance) on the scaled data $\tilde{Z}_j^k(\mathbf{s}_i)$ using the sector method (as defined earlier in the section ‘Binning the variogram and covariance estimates into lag classes’) over a large range of lag classes that progress in a geometric series. The lag classes are formed from intervals $[d^{k-1/2}, d^{k+1/2})$, where $d = 1.25$ and k ranges from the smallest machine integer to the largest. The center of each lag class is taken to be $d^k \cosh(\frac{1}{2} \log d^k)$. Obviously, many lag classes are empty, and the Geostatistical Analyst only uses those that have data in them. Call this empirical (cross)covariance $\hat{C}_{ij}(h_k)$, where i indicates the i th variable type, j indicates the j th variable type, and k indicates the k th lag class. The first iteration of parameter estimates is obtained by minimizing,

$$\sum_{i=1}^T \sum_{j=1}^T \sum_{k=1}^{n_{ij}} w_{ij}(\mathbf{h}_k) \left(\tilde{C}_{ij}(\mathbf{h}_k; \boldsymbol{\theta}_{ij}) - \hat{C}_{ij}(\mathbf{h}_k) \right)^2 \quad (1)$$

for θ , where θ_{ij} is the vector of parameters for the i, j th covariance function and θ contains all covariance parameters, where

$$w_{ij}(\mathbf{h}_k) = N_{ij}(\mathbf{h}_k) / \sum_{m=1}^{n_{ij}} N_{ij}(\mathbf{h}_m) \quad (2)$$

and $N_{ij}(\mathbf{h}_k)$ is the number of pairs in the empirical (cross) covariance function for variables i and j in lag class k . Call this

estimate $\theta^{(1)}$. In the next iteration, the Geostatistical Analyst uses a Cressie’s (1985) weighted least squares by minimizing (1) again, only this time let,

$$\bar{w}_{ij}(\mathbf{h}_k; \boldsymbol{\theta}_{ij}^{(1)}) = \frac{N_{ij}(\mathbf{h}_k)}{\tilde{C}_{ii}(\mathbf{0}; \boldsymbol{\theta}_{ii}^{(1)}) \tilde{C}_{jj}(\mathbf{0}; \boldsymbol{\theta}_{jj}^{(1)}) + \tilde{C}_{ij}^2(\mathbf{h}_k; \boldsymbol{\theta}_{ij}^{(1)})} \quad (3)$$

and then these weights are normalized so that each (cross)covariance gets equal weight,

$$w_{ij}(\mathbf{h}_k) = \bar{w}_{ij}(\mathbf{h}_k; \boldsymbol{\theta}_{ij}^{(1)}) / \sum_{m=1}^{n_{ij}} \bar{w}_{ij}(\mathbf{h}_m; \boldsymbol{\theta}_{ij}^{(1)}) \quad (4)$$

Call this estimate $\theta^{(2)}$. Notice that if we use variograms rather than covariances $\boldsymbol{\theta}_{ii}^{(2)}$ is,

$$\arg \min_{\boldsymbol{\theta}_{ii}} \left[\sum_{k=1}^{n_{ij}} w_{ii}(\mathbf{h}_k) \left(\tilde{\gamma}_{ii}(\mathbf{h}_k; \boldsymbol{\theta}_{ii}) - \hat{\gamma}_{ii}(\mathbf{h}_k) \right)^2 \right] \quad (5)$$

where $w_{ii}(\mathbf{h}_k)$ is given by (2) and then $\boldsymbol{\theta}_{ii}^{(2)}$ is obtained from (5) with weights as in (4) but now,

$$\bar{w}_{ij}(\mathbf{h}_k; \boldsymbol{\theta}_{ii}^{(1)}) = \frac{N_{ij}(\mathbf{h}_k)}{\gamma_{ii}^2(\mathbf{h}; \boldsymbol{\theta}_{ii}^{(1)})}$$

The estimates $\theta^{(1)}$ and $\theta^{(2)}$ are two steps in an iteratively reweighted least-squares algorithm.

The estimate $\theta^{(2)}$ is only used to provide a range estimate for a default lag size for the grid method in estimating the empirical semivariogram or covariance. The default number of lags is 12, so the lag size for the grid method in the next section is taken to be $2 * \text{range} / 12$.

Stage 2

Stage 2 essentially repeats stage 1, but on an empirical semivariogram or (cross)covariance on the scaled data $\tilde{Z}_j^k(\mathbf{s}_i)$ that uses the grid method (as defined earlier in the section ‘Binning the semivariogram and covariance estimate’) where the default lag size is obtained from the range estimate in $\theta^{(2)}$ from Stage 1. It also allows for anisotropy and linear combinations of up to three (cross)covariance or semivariogram models in additions into lag classes to the nugget effect for each dataset,

$$\tilde{C}_{ij}(\mathbf{h}; \boldsymbol{\theta}) = \sum_{u=1}^S B_u(i, j) \rho_u(\mathbf{h}; \boldsymbol{\varphi}_u).$$

Here, $B_u(i, j)$ is a partial sill parameter and is the i, j th component of B_u , a $T \times T$ positive-definite matrix, where T is the number of *types* of variables, S is the number of different (cross)covariance models used in a linear combination, and the function $\rho_u(\mathbf{h}; \boldsymbol{\varphi}_u)$ is a normalized covariance model; $\rho_u(\mathbf{0}; \boldsymbol{\varphi}_u) = 1$, where $\boldsymbol{\varphi}_u$ are parameters that typically control the range (and/or shape) of the covariance model. As before, $\boldsymbol{\theta}$ contains all of the parameters. The third iteration of parameter estimates, $\theta^{(3)}$ is obtained by minimizing (1) with weights (2) on the empirical covariance using the grid method, and then $\theta^{(4)}$ is obtained by minimizing (1) with weights from (4) and (3) on the empirical covariance using the grid method. These formulas are modified in an obvious way if we are using semivariograms, as was shown for stage 1. Now, change back to the original scale. The final (cross)covariance models are

$$C_{ij}(\mathbf{h}) = s_i s_j \tilde{C}_{ij}(\mathbf{h}; \boldsymbol{\theta}_{ij}^{(4)}),$$

and for variograms they are

$$\gamma_{ii}(\mathbf{h}) = s_i^2 \tilde{\gamma}_{ii}(\mathbf{h}; \boldsymbol{\theta}_{ii}^{(4)}).$$

If the user changes any parameters, such as lag size, then estimates are recalculated beginning at stage 2.

Bivariate distribution dialog box

Disjunctive kriging requires that all pairs of data have a bivariate normal distribution. This assumption is difficult to test in practice. The Geostatistical Analyst gives a visual tool to help assess the bivariate normal assumption. A theoretical curve, as a function of lag, can be developed based on various threshold values for the cumulative distribution function (see Deutsch and Journel, 1992, p. 139 and Goovaerts, 1998, p. 265). This theoretical curve can be compared to an empirical semivariogram based on indicators. More generally, if QQPlots show a marginal normal distribution, and the data appears to have bivariate normal distributions, it is reasonable to assume a full multivariate normal distribution for the data. Thus, the check for the bivariate normal distribution can be used for simple kriging, allowing a user to assure that quantile and probability maps are based on reasonable assumptions.