We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Model Averaging
for Semivariogram Model Parameters

Asim Biswas and Bing Cheng Si

Additional information is available at the end of the chapter

1. Introduction

Soil varies considerably from location to location (Nielsen et al., 1973). Knowledge on soil spatial variability is important in ecological modelling (Burrough, 1983; Corwin et al., 2006), environmental prediction (Trangmar et al., 1985), precision agriculture (Goderya, 1998), soil quality assessment (Heuvelink and Pebesma, 1999; McBratney et al., 2000), and natural resources management. Adequate understanding of the variability in soil properties as a function of space and time is necessary for developing logical, empirical and physical models of soil and landscape processes (Burrough, 1993; Foussereau et al., 1993; Wilding et al., 1994). Geostatistics, a widely used approach, has been used to identify the spatial structure in the variability of soil attributes (Vieira, 2000; Carvalho et al., 2002; Vieira et al., 2002). Semivariance function characterizes the spatial continuity between points. When the semivariance is plotted against the lag distance or separation distance between points, the plot is called semivariogram (Fig. 1; McBratney and Webster, 1986; Isaaks and Srivastava, 1989). The structure of the semivariogram is explained by three properties; the nugget, the sill and the range (Fig. 1). These spatial structures of semivariogram help in identifying autocorrelation and replicating samples, revealing dominant pattern in data series, identifying major ongoing processes (Si et al., 2007), designing experiments (Fagroud and van Meirvenne, 2002) and monitoring networks (Prakash and Singh, 2000), selecting proper data analysis method and interpreting data (Lambert et al., 2004), and assessing simulation and uncertainty analysis in a better way (Papritz and Dubois, 1999). The semivariogram structures also help to quantify spatial dependence between observations. Modelling of observed semivariance values helps in predicting the spatial distribution of attribute values (Goovaerts, 1998). The spatial distribution of attribute values is very important in separating random noise in semivariance, and interpolation and mapping analysis such as kriging (Deutsch and Journel, 1998; Nielsen and Wendorth, 2003).
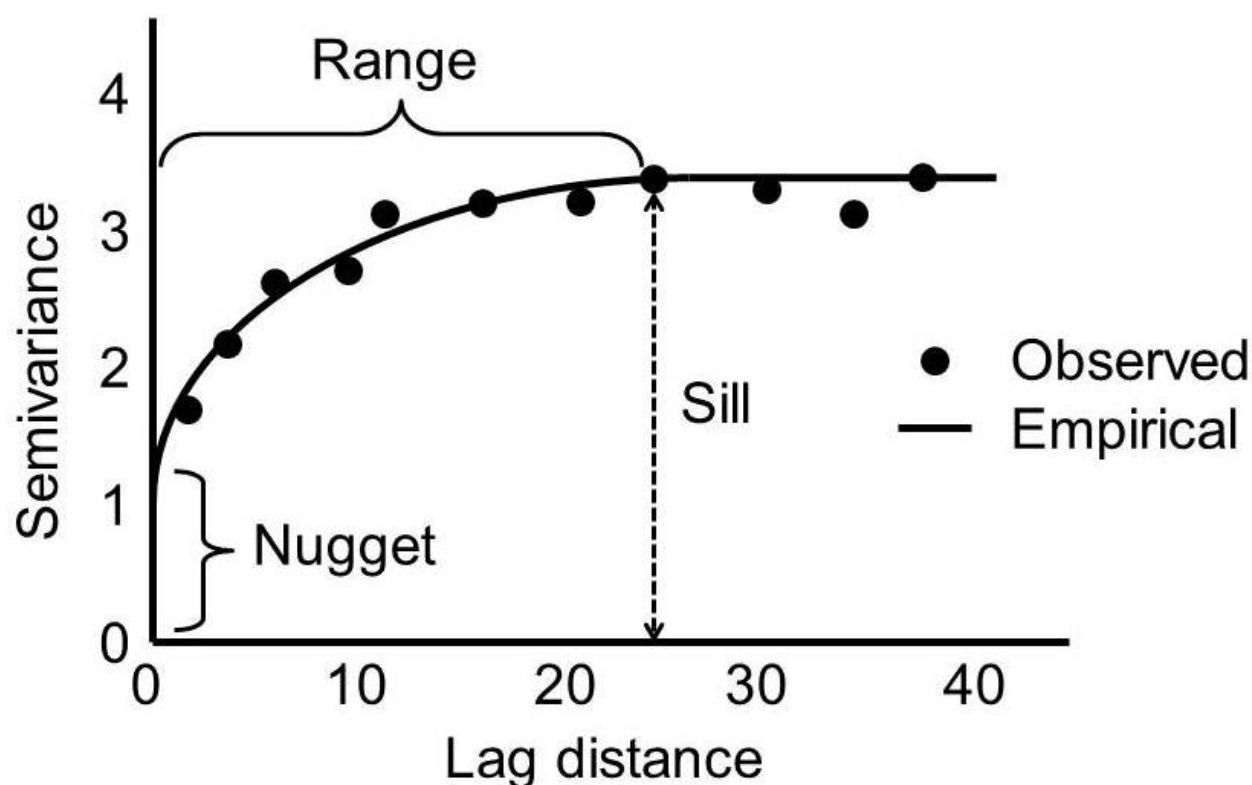
Figure 1. A typical example of semivariogram showing different components.

The choice of theoretical models and its fitting procedure is very important to get a better pre‐
diction of unsampled locations (McBratney and Webster, 1986). Spherical model (Burgess and
Webster, 1980; Vieira et al., 1981; Van Kuilenburg et al., 1982; Vauclin et al., 1983; Trangmar,
1985), exponential model (David, 1977), Gaussian model (McBratney and Webster, 1986) and
linear plateau model (Burgess and Webster, 1980; Hajrasuliha et al., 1980; Vauclin et al., 1983)
are important among the most commonly used semivariogram models in the field of soil sci‐
ence. The maximum likelihood method (Cressie, 1991) or least square regression (Vieira et al.,
1981; Yost et al., 1982; Trangmar et al., 1985) including the weighted least square methods
(Cressie, 1985) optimize the parameter value by minimizing the deviations of model predic‐
tion from the experimental semivariances. Small sum of deviations between the model and the
experimental values indicate superior performance of the models. Small and comparable sum
of the deviations indicates comparable performance of the models. Therefore, selection of
model is prerequisite for better prediction (Burnham and Anderson, 2002). In reality, there can
be several models to choose from. Selection of good models requires balancing of goodness of
fit and complexity, which is generally, determined using a likelihood ratio approach leading to
a chi-square test. The Akaike Information Criterion (AIC) values can be calculated from maxi‐
mum likelihood function (Akaike, 1973), which is used to evaluate the performance of the
models (Webster and McBratney, 1989). The smallest AIC value indicates the best model, but
the values of parameters such as nugget, sill and range can be different from model to model,
even though they have the same physical meaning (Trangmar, 1985). Different value of a pa‐
rameter for different models clearly indicates the uncertainty associated with the parameter,

which makes the interpretation difficult. Therefore, interpretation of a particular spatial process from the parameter values of a particular model will be biased. In addition the parameter value from individual models may have large uncertainty. This uncertainty is inherent to the model selection and is well beyond just the issue of determining best model(s). Model averaging is a technique designed to help in accounting the uncertainty associated with the models and their parameters.

The model averaging is not a new concept in statistics. People started combining and averaging models since 1970s for different purposes (Hoeting et al., 1999). During averaging models are assigned with weights based on their performance. The AIC values are used to calculate the weights for the models (Burnham and Anderson, 2002). Before the recent development of computational power, the averaging procedure ignored the uncertainty associated with models. Recently, the use of model averaging is increased to reduce the uncertainty associated with the model selection. Information on the use of model averaging procedure in soil science is scarce (Webster and McBratney, 1989) and there is no information of reducing uncertainty associated with the semivariogram model parameters from averaging of commonly used semivariogram model parameters. Therefore, the objective of this paper is to reduce the uncertainty associated with semivariogram model parameters through averaging. The weighted average of the parameter values of commonly used models can be a better way of describing the spatial processes.

Name of Model	Equations †
Spherical Model	$\gamma_s(h;a, b, c) = \begin{cases} a + (b-a)\left[1.5\dfrac{h}{c} - 0.5\left(\dfrac{h}{c}\right)^3\right] & if\ h \le c \\ b & otherwise \end{cases}$
Exponential Model	$\gamma_e(h;a, b, c) = a + (b-a)\left\{1 - \exp\left(\dfrac{-h}{c}\right)\right\}$
Gaussian Model	$\gamma_g(h;a, b, c) = a + (b-a)\left\{1 - \exp\left(\dfrac{-h^2}{c^2}\right)\right\}$
Linear Plateau Model	$\gamma_l(h;a, b, c) = \begin{cases} a + (b-a)\dfrac{h}{c} & if\ h \le c \\ b & otherwise \end{cases}$

† In all models; γ = 0 when h = 0

Table 1. Commonly used models for semivariogram fitting in soil science

2. Theory

Geostatistics can explain the spatial variability and the patterns of a variable in field from its autocorrelation. It describes the relationship between measurements at different locations

(or times) separated by certain distance (or time). For example, a soil property $Z(x_i)$ measured at location x_i, where $i = 1, 2, …, n$ and n is the number of samples along a transect. The continuity of this relationship can be investigated from h scatterplots, which can be done by plotting the measured values $Z(x_i)$ as horizontal axis and $Z(x_i+h)$ as vertical axis, where $i = 1, 2, …$. The h scatterplot will show a cloud of points, where every point represents one pair of sample locations $Z(x_i)$ and $Z(x_i+h)$ (Pannatier, 1996; Zawadzki and Fabijanczyk, 2007). For each h, half of the mean value of the squared difference $Z(x_i)$ - $Z(x_i+h)$ is defined as semivariance (Matheron, 1962).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z(x_i) - Z(x_i + h) \right]^2 \tag{1}$$

The variance (σ_i^2) of the squared difference $Z(x_i)$ - $Z(x_i+h)$ can be calculated using Eq. (2).

$$\sigma_i^2 = \frac{1}{N(h)-1} \sum_{i=1}^{N(h)} \left(a_i - \bar{a}_i \right)^2 \tag{2}$$

where, $a_i = [Z(x_i) - Z(x_i + h)]^2$ and $N(h)$ is the number of pairs in a cloud for a particular h.

The experimental semivariogram can be fitted to a mathematical model. Not all the mathematical functions that seem to fit the observed values can be considered as semivariogram model. One important criterion for the semivariogram models is to be "positive-definite" to ensure the nonnegative covariance values restricting the models to be permissible for fitting (Isaaks and Srivastava, 1989; Goovaerts, 1997; Deutsch and Journel, 1998). There are a number of permissible semivariogram models including the most commonly used spherical, exponential, Gaussian, and linear plateau (Table 1).

Among the fitting procedures, the weighted nonlinear least square method is the most robust and reliable method (Cressie, 1985). This procedure minimizes the residual sum of squared errors (RSS) between experimental semivariance data and the models by optimizing the model parameters: nugget, sill and range values. The RSS can be calculated using Eq. 3.

$$RSS = \sum_{i=1}^{m} w_i \left[\tilde{\gamma}_i - \gamma_i \right]^2 \tag{3}$$

where, m is the number of lags, $\tilde{\gamma}_i$ is the semivariance value for lag i, γ_i are the corresponding model predictions, and w_i are weighting factors. The weighing factor used in calculating RSS is related to the variance associated with the semivariance calculation (Jian et al., 1996).

$$w_i = \frac{1}{\sigma_i^2} \qquad (4)$$

where, σ_i^2 is the variance calculated in Eq. 2 for lag i. In this weighted least squares approximation the fit can always be improved by diminishing the residual sum of squared errors with addition of parameters to the models (Webster and McBratney, 1989). In case of the same number of parameters in the most commonly used semivariogram models, the small and comparable RSS values indicate the comparable performance of different models. A solution can be achieved by fitting each model using the least square approximation and then comparing the goodness of fit for each model (Webster and McBratney, 1989). The performance of a model can be evaluated and the fitting can be said as the best with the lowest Akaike Information Criterion (AIC) (Akaike, 1973). AIC is an estimate of the expected Kullback-Leiler information (a ruler to measure the similarity between the statistical model and the true distribution) lost by using a model to approximate the process that generated observed pattern (Burnham and Anderson, 2002; Johnson and Omland, 2004). It is also defined as Eq. 5:

$$AIC = -2\ln\left(\text{maximized likelihood}\right) + 2 \times \left(\text{number of parameters}\right) \qquad (5)$$

Or it can be estimated as in Eq. 6.

$$AIC = -2\ln\left[L\left(\hat{\theta}\middle|y\right)\right] + 2p \qquad (6)$$

where, p is the number of parameters and $\ln[L(\hat{\theta} \mid y]$ is the maximized likelihood and can be estimated from Eq. 7.

$$\ln\left[L\left(\hat{\theta}\middle|y\right)\right] = -\frac{n}{2}\ln\left(\frac{RSS}{n}\right) \qquad (7)$$

where, n is the number of data points. With close performance, different models with same number of parameters having the same physical meaning produces different optimized parameter values. A single process in field will be represented by different values based on the selected models. In this case the estimated parameter values will be associated with uncertainty, which can be reduced by making weighted average based on the performance of the selected model. If there are R numbers of models with parameters θ_j ($j = 1, 2, 3,...,p$), the estimated parameters can be calculated from the Eq. 8 (Burnham and Anderson, 2002; Johnson and Omland, 2004).

$$\hat{\bar{\theta}} = \sum_{j=1}^{R} W_j \hat{\theta}_j \tag{8}$$

where, W_j are the weights assigned to a particular model. The weights can be calculated from the AIC values for each model (Eq. 9).

$$W_j = \frac{\exp\left(-\Delta_j/2\right)}{\sum_{j=1}^{R} \exp\left(-\Delta_j/2\right)} \tag{9}$$

where, $\Delta_j = AIC_j - AIC_{min}$ and AIC_j and AIC_{min} are the AIC values for a particular model and the lowest AIC value model.

To explain the uncertainty associated with the parameter, we need to calculate the variance associated with each optimized parameter. If a model has p parameters, we can calculate $p \times p$ covariance matrix. From the calculation of Hessian Matrix (the second derivative matrix of χ^2 merit function) we can estimate the standard error in fitted parameters. In the calculation of Hessian Matrix (Press et al., 1992), let us assume we have to fit a model:

$$\gamma = \gamma\left(h; a, b, c\right) \tag{10}$$

The χ^2 merit function, that will be equal to RSS (Eq. 11),

$$\chi^2\left(a, b, c\right) = \sum_{i=1}^{m} \frac{1}{\sigma_i^2} \left[\tilde{\gamma}_i - \gamma\left(h; a, b, c\right)\right]^2 = RSS\left(a, b, c\right) \tag{11}$$

The χ^2 gradient for the parameters, z, will be zero when χ^2 is minimum. The components can be calculated as Eq. (12).

$$\frac{d\chi^2}{dz_k} = -2\sum_{i=1}^{N} \frac{\left[\tilde{\gamma}_i - \gamma\left(h; a, b, c\right)\right]}{\sigma_i^2} \frac{d\gamma\left(h; a, b, c\right)}{dz_k} \quad \text{where, } k = a, b, c. \tag{12}$$

An additional derivative will give,

$$\frac{d^2\chi^2}{dz_a dz_b} = 2\sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[\frac{d\gamma\left(h; a, b, c\right)}{dz_a} \frac{d\gamma\left(h; a, b, c\right)}{dz_b} - \left\{\tilde{\gamma}_i - \gamma\left(h; a, b, c\right)\right\} \frac{d^2\gamma\left(h; a, b, c\right)}{dz_a dz_b}\right] \tag{13}$$

It is conventional to remove the factors of 2 by defining

$$\alpha_{ab} \equiv \frac{1}{2} \frac{d^2 \chi^2}{dz_a dz_b} \tag{14}$$

For a linear equation, the second derivative term can be dismissed because it is zero or small enough when compared to the term involved in first order derivative. So the α_{ab} formula can be written as (Press et al., 1992);

$$\alpha_{ab} = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \left[\frac{d\gamma(h;a,b,c)}{dz_a} \frac{d\gamma(h;a,b,c)}{dz_b} \right] \tag{15}$$

So for a three parameter (a, b, c) semivariogram model we will get a 3×3 Hessian Matrix, (Eq. 16).

$$[\alpha] = \begin{bmatrix} \sum_{i=1}^{N} \frac{d\gamma}{da} \frac{d\gamma}{da} & \sum_{i=1}^{N} \frac{d\gamma}{da} \frac{d\gamma}{db} & \sum_{i=1}^{N} \frac{d\gamma}{da} \frac{d\gamma}{dc} \\ \sum_{i=1}^{N} \frac{d\gamma}{db} \frac{d\gamma}{da} & \sum_{i=1}^{N} \frac{d\gamma}{db} \frac{d\gamma}{db} & \sum_{i=1}^{N} \frac{d\gamma}{db} \frac{d\gamma}{dc} \\ \sum_{i=1}^{N} \frac{d\gamma}{dc} \frac{d\gamma}{da} & \sum_{i=1}^{N} \frac{d\gamma}{dc} \frac{d\gamma}{db} & \sum_{i=1}^{N} \frac{d\gamma}{dc} \frac{d\gamma}{dc} \end{bmatrix} \tag{16}$$

The approximation of the covariance matrix can be done by inversing the matrix α (Press et al., 1992) (Eq. 17).

$$[C] = [\alpha]^{-1} \tag{17}$$

This matrix component indicates the variance associated with each parameter for different models. At the same time we need to calculate the variance associated with the averaged parameter. The parameter variance for each model and the weight assigned to the model are used to calculate the variance of the averaged parameter (Eq. 18) (Burnham and Anderson, 2002)

$$\hat{var}\left(\hat{\bar{\theta}}\right) = \sum_{j=1}^{R} W_j \left[var\left(\hat{\theta} \mid g_j\right) + \left(\hat{\theta} - \hat{\bar{\theta}}\right)^2 \right] \tag{18}$$

where, W is the weights assigned to R number of models, g, $\hat{\theta}$ is the estimated model parameter values from fitting, and $\hat{\bar{\theta}}$ is the average of estimated model parameter values.

Characteristics	Sand (Zeleke and Si, 2005)	Copper (Atteia et al., 1994)
Minimum	52.50	3.55 (1.27)
Maximum	78.75	166.40 (5.11)
Mean	64.23	23.58 (2.87)
Median	64.41	17.20 (2.84)
Mode	58.75	16.40 (2.80)
Skewness	0.28	3.02 (0.42)
Kurtosis	-0.73	11.88 (0.11)
Standard Deviation	6.06	22.24 (0.72)
Variance	36.76	494.48 (0.52)

Table 2. Descriptive statistics of the data sets used for model fitting

3. Materials and methods

Demonstration of this average parameter estimation procedure was done using two soil pa‐
rameters, which were chosen from two different experimental datasets. One soil property,
sand content, was collected from a regularly spaced (3 m) 128 point transect (Zeleke and Si,
2005). The sampling site was located at Smeaton, Saskatchewan, Canada (53°40′N latitude
and 104°58′W longitude). The hydrometer method was used to determine the particle sizes
(Gee and Bauder, 1986). Another soil property, copper (Cu) content, was selected from a da‐
taset consisting of 359 topsoil samples collected by the Swiss Federal Institute of Technology
in the Swiss Jura (Atteia et al., 1994). The data points were selected according to a regular
configuration of a mesh of 250 m × 250 m with several clusters from an area of approximate‐
ly 1450 ha. The cluster samples were collected following nested sampling design with 100,
40, 15, and 6 m sampling interval. Copper content was measured using a direct current plas‐
ma spectroscopy (ARL, Spectran V) with other 5 heavy metals (Atteia et al., 1994).

4. Results and discussion

The exploratory information about the dataset is given in Table 2. The skewness of the sand
0.28, which indicates a near normal statistical distribution of the dataset. The copper content
is highly skewed (skewness = 3.02) and exhibits a log normal statistical distribution. There‐
fore, we decided to normalize the copper content using natural logarithm transformation.
The exploratory information of logarithm transformed values for copper content is present‐
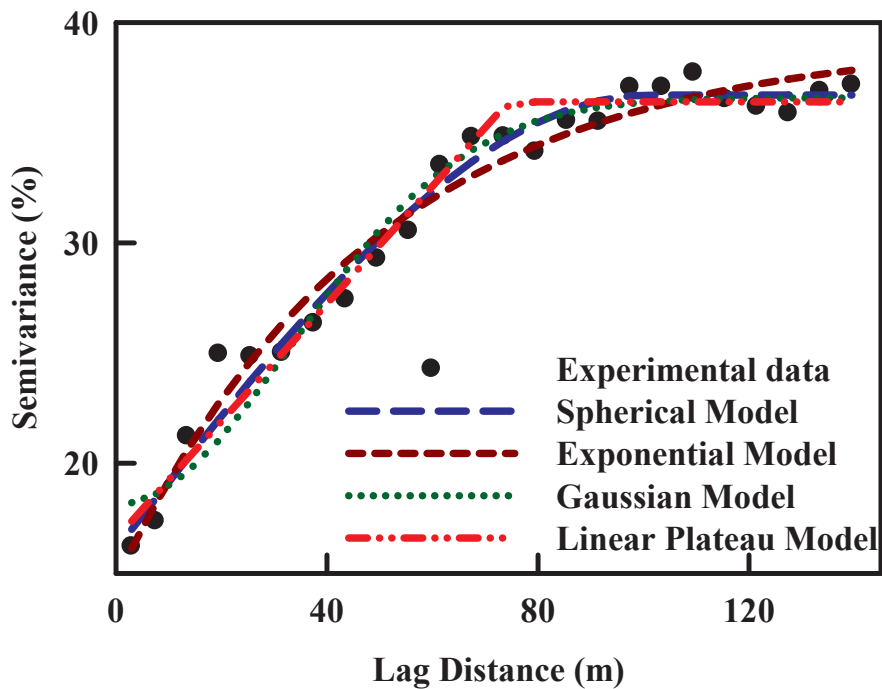ed in parentheses in Table 2.

Figure 2. Empirical semivariogram and fitted models for sand content data.
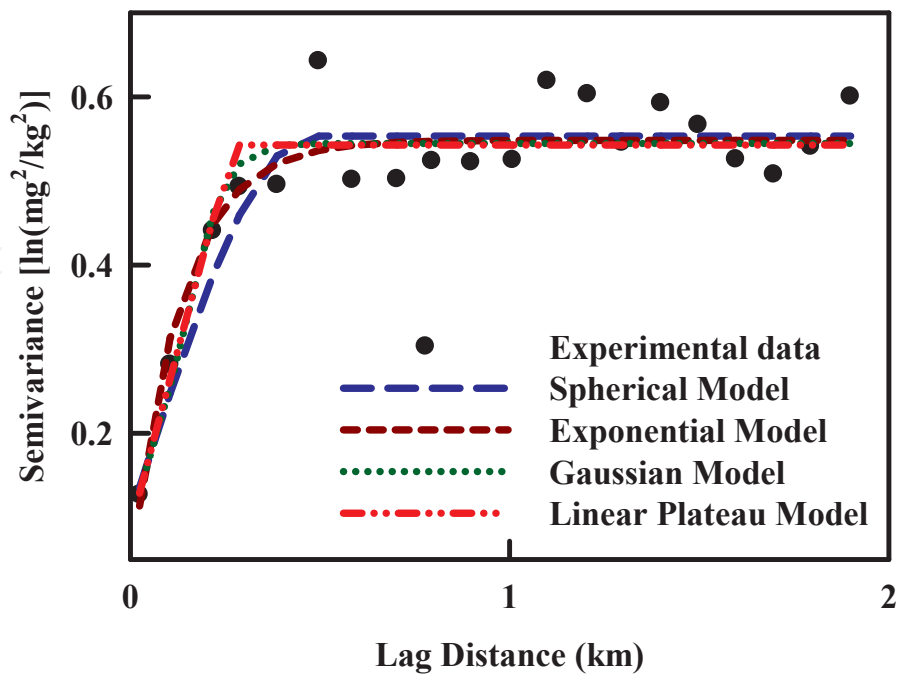

Figure 3. Empirical semivariogram and fitted models of logarithm-transformed copper content data.

Semivariance for sand content (Fig. 2) and logarithm transformed copper content (Fig. 3) are calculated and plotted as a function of lag interval. For the semivariance calculation of regularly spaced sand samples, the lag interval is 6 m, which is twice the minimum sampling interval. The maximum lag distance used for this calculation is 144 m and the minimum number of pairs used for the semivariogram calculation is 24. The semivariance of the copper content is calculated using minimum and maximum lag distance of 0.1 km and 2 km respectively, which leads to at least 20 pairs of data points. The experimental semivariograms are fitted with four most commonly used semivariogram models (Table 1) following weighted least square estimation. The fitted semivariogram models and the experimental semivariograms are shown in Fig. 2 and Fig. 3 for sand and copper, respectively. The optimized parameters (nugget, sill and range) values for each of the models are presented in Table 3. The nugget value optimized by different models for the sand content varies from 14.67 % (exponential model) to 18.13 % (Gaussian model). The value of sill varies from 36.40 % (linear plateau model) to 39.27 % (exponential model). The optimized range value that indicates the distance over which the processes are spatially dependent, for the sand content varies widely from 47.10 m (Gaussian model) to over 100 m (Spherical model). The variance (values in the paresis in Table 3) associated with each parameters are approximated from the calculation of Hessian Matrix (Press et al., 1992). The variance associated with the parameter estimate is very high as the semivariance is calculated from the semivariance cloud, which spreads over a range of values. The nugget, sill, and range have their own physical significance and different models should result in the same set of parameter values. However, due to model uncertainty, four models have different set of optimized parameter values for the sand content. The range of optimized parameter values (Table 3) from different models clearly indicate how uncertain our models and parameters are.

Similarly there are a wide range of optimized parameter values from different models for copper content. The optimized nugget values vary widely in different models (0.027 $\ln(mg^2/kg^2)$ for exponential model to 0.122 $\ln(mg^2/kg^2)$ for Gaussian model) (Table 3). Wide range of sill and range values clearly indicates the uncertainty associated with the parameters. Different parameter values for different models make the interpretation of the parameters and the semivariogram structures difficult and uncertain. The ratio of the nugget semivariance (V_N) to the sill semivariance (V_S) gives a measure of the strength or degree of spatial structure (Cambardella et al., 1994). Therefore, it is sometime difficult to explain the semivariogram structure of a particular property with different nugget and sill values for different models.

In this situation the goodness of fit for each model is calculated. The residual sum of squared error (RSS) between the experimental data and the model data is presented in Table 3. The comparable RSS values indicate that the four models are comparable. From the RSS value, the maximum likelihood is calculated for each model. Akaike Information Criterion (AIC) for different models is calculated from the maximum likelihood of the models and is presented in Table 4. The AIC values for the spherical (-167.269) and exponential models (-167.187) fitted for sand semivariogram are very close indicating a close performance of the models. The performance of the linear plateau and Gaussian models are also acceptable. The

AIC values for all four models are very similar for copper content (vary from -111.739 to -113.739). The similar AIC values for the models indicate that all four models are almost equivalent. Though the performance of the models is very comparable, the optimized parameter values from different models are not similar. In this situation, the selection of parameter value from a particular model can be associated with high uncertainty.

Parameter	Model	Nugget	Sill	Range	V_N/V_S	RSS
		---%---	---%---	---m---		
Sand	Spherical	16.08 (257.45)	36.71 (231.66)	100.94 (29610.00)	0.438	0.021
	Exponential	14.67 (423.95)	39.27 (1081.00)	49.15 (32550.00)	0.374	0.021
	Gaussian	18.13 (198.24)	36.57 (214.91)	47.10 (5233.00)	0.495	0.040
	Linear Plateau	16.57 (217.15)	36.40 (181.53)	74.37 (8379.00)	0.455	0.026
		$\ln(mg^2/kg^2)$	$\ln(mg^2/kg^2)$	--km--		
Copper	Spherical	0.101 (0.074)	0.553 (0.037)	0.480 (0.767)	0.183	0.067
	Exponential	0.027 (0.177)	0.548 (0.039)	0.132 (0.133)	0.049	0.061
	Gaussian	0.122 (0.062)	0.544 (0.034)	0.171 (0.097)	0.224	0.061
	Linear Plateau	0.089 (0.080)	0.543 (0.033)	0.271 (0.235)	0.164	0.063

Table 3. Optimized parameter values and their variance (in paresis) for different semivariogram models for sand content and copper

The model averaging is conducted to reduce parameter uncertainty. Weights are assigned to the models based on their performance and importance. The model with lowest AIC value indicates best model. Based on the smallest AIC, ∆AIC is calculated from the difference between AIC value of a particular model and the smallest AIC value out of four models. The ∆AIC is used to measure the likelihood of each model. The value of the likelihood of models helps in assigning weights for different models. The spherical and the exponential models for the sand semivariogram have the highest weight, which are almost equal. This indicates an equivalent performance of these two models. The Gaussian and linear plateau models have very small weights, indicating less importance of the models than that of the other two. Whereas for copper, the weights assigned to different models are close enough to explain the performance as equivalent. Based on the weights assigned to each model, the average

value of the parameter and the variance associated with each parameter is estimated. Sand has an average nugget, sill and range value of 15.42 %, 37.92 % and 75.53 m respectively. The variance associated with each averaged parameters are presented in the paresis in Table 4. The higher weight of the spherical and exponential model indicates that two models have large contribution to the average value of parameters. Similarly, the variance of the averaged parameters will have large contribution from spherical and exponential models. Whereas for copper, the contribution to the average value by each models is comparable. The average value of nugget, sill and range for copper are 0.081 $\ln(mg^2/kg^2)$, 0.546 $\ln(mg^2/kg^2)$, and 0.219 km respectively. The variance of each averaged parameter is also calculated and presented in Table 4 (in paresis). The averaged variance reduced the uncertainty associated with the parameters by taking the average.

Parameter	Model	AIC value	Weight	Average parameter value		
				Nugget	**Sill**	**Range**
Sand	Spherical	-167.269	0.4946	15.42 %	37.92 %	75.53 m
	Exponential	-167.187	0.4748	(335.79)	(635.06)	(31007.42)
	Gaussian	-151.279	0.0001			
	Linear Plateau	-161.687	0.0303			
Copper	Spherical	-111.739	0.1230	0.081 $\ln(mg^2/kg^2)$	0.546 $\ln(mg^2/kg^2)$	0.219 km
	Exponential	-113.703	0.3285			
	Gaussian	-113.658	0.3212	(0.106)	(0.036)	(0.235)
	Linear Plateau	-112.964	0.2270			

Table 4. Parameter values obtained through model averaging of commonly used models and their variance (in paresis).

When different models optimize parameters with a wide range of values, the weighted average of the estimated parameter values provide more concise information and better understanding about the parameters and thus the spatial structure. The variance of the optimised parameters clearly indicates the reduced uncertainty of the parameters. This weighted average value of the parameters provides a better prediction about the underlying processes by reducing the uncertainty associated with the parameters and the bias in their selection. Better understanding of the parameters provides guidance for sampling and insights on experimental design.

5. Conclusion

The performance of the commonly used semivariogram models, used for fitting the experimental semivariogram, is comparable. Different models optimize a particular parameter differently, which indicates the uncertainly associated with the parameter. The uncertain parameter value

makes the prediction of spatial processes difficult and uncertain. A model averaging procedure is used to obtain the parameter value and to reduce model parameter uncertainty.

Two soil properties (sand content and copper content) are taken as examples for demonstrating the average parameter estimation procedure. The sand content is measured at regular sampling interval along a transect, whereas the copper content is measured at variable sampling intervals in a two-dimensional field. The semivariogram is calculated for both properties and fitted with four most commonly used mathematical models (spherical, exponential, Gaussian and linear plateau). Weighted least square estimation is used for fitting these models to the experimental semivariogram. The goodness of fit for each model is calculated from their residual sum of squares. The parameter for each models are optimized during the fitting procedure. The likelihood of each model is calculated based on the Akaike Information Criterion (AIC) for each model. Different weights were assigned to each model based on their performance and importance from the AIC values. These weights are used for obtaining the weighted average of the optimized parameters. The weighted average of the estimated parameters reduced the uncertainty associated with the parameters and the bias in their selection. The average parameter values and reduced uncertainty provide more concise information about the spatial structure and consequently provide better guidance for sampling, experimental design, and interpolation and mapping.

Acknowledgements

Author details

Asim Biswas[1,2*] and Bing Cheng Si[2]

*Address all correspondence to: asim.biswas@csiro.au

1 Department of Natural Resource Sciences, McGill University, Canada

2 Department of Soil Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

References

[1] Akaike H (1973) Information theory as an extension of the maximum likelihood principle. In: Second International Symposium on Information Theory. Petrov, B.N., Csaki, F., Eds, Akademiai Kiado, Budapest, 267-281 p.

[2] Atteia O, Dubois JP, Webster R (1994) Geostatistical analysis of soil contamination in the Swiss Jura. Environ. poll. 86, 315-327.

[3] Burgess TM, Webster R (1980) Optimal interpolation and isarithmic mapping of soil properties. I. The semivariogram and punctual kriging. J. soil sci. 31, 315-331.

[4] Burnham KP, Anderson DR (2002) Model selection and multimodal inference: A Practical Information-Theoretic approach, Springer.

[5] Burrough PA (1983) Multiscale sources of spatial variation in soil. I. The application of fractal concepts of nested levels of soil variogram. J. soil. sci. 34: 577-597.

[6] Cambardella CA, Moorman TB, Novak JM, Parkin TB, Karlen DL, Turco RF, Konopka AE (1994) Field scale variability of soil properties in central Iowa soils. Soil sci. soc.am. j. 58, 1501-1511.

[7] Carvalho JRP, Silveira PM, Vieira SR (2002) Geostatistics in determining the spatial variability of soil chemical characteristics under different preparations. Pesq. agropec. bras. 37, 1151-1159.

[8] Corwin DL, Hopmans J, de Rooij GH (2006) From Field- to Landscape-Scale Vadose Zone Processes: Scale Issues, Modeling, and monitoring. Vadose zone j. 5, 129-139.

[9] Cressie N (1985) Fitting variogram models by weighted least squares. J. int. assoc. math. geol. 17, 563-586.

[10] Cressie NAC (1991) Statistics for spatial data. John Wiley & Sons, NY, 900 p.

[11] David M (1977) Geostatistical Ore reserve estimation. Elsevier, Amsterdam.

[12] Deutsch CV, Journel AG (1998) GSLIB Geostatistical software library and user's guide. 2nd eds. Oxford Uni. Press, NY.

[13] Fagroud M, Van Meirvenne M (2002) Accounting for soil spatial correlation in the design of experimental trails. Soil sci. soc.am. j. 66, 1134-1142.

[14] Foussereau X, Hornsby AG, Brown RB (1993) Accounting for variability within map units when linking a pesticide fate model to soil survey. Geoderma 60, 257-276.

[15] Gee GW, Bauder JW (1986) Particle size analyses. In: A. Klute (Eds.) Method of soil analyses. Part 1: Physical and Mineralogical Methods. ASA, Madison, Wisconsin, USA.

[16] Goderya FS (1998) Field scale variations in soil properties for spatially variable control: a review. J. soil contam. 7, 243-264.

[17] Goovaerts P (1997) Geostatistics for natural resources evaluation.Oxford Uni. Press., NY.

[18] Goovaerts P (1998) Geostatistical tools for characterizing the spatial variability of microbiological and physio-chemical soil properties. Biol. fertil.soils 27, 315-334.

[19] Hajrasuliha SW, Baniabassi N, Metthey J, Neilsen DR (1980) Spatial variability of soil sampling for salinity studies in South-West Iran. Irrig. sci. 1, 197-208.

[20] Heuvelink GBM, Pebesma EJ (1999) Spatial aggregation and soil process modelling. Geoderma 89, 47-65.

[21] Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. Statist. sci. 14, 382-417.

[22] Isaaks EH, Srivastava RM (1989) An introduction to applied geostatistics, Oxford University Press, Toronto, Canada.

[23] Jian X, Olea RA, Yu Y (1996) Semivariogram modelling by weighted least squares. Comp. Geosci. 22, 387-397.

[24] Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends ecol. evol. 19, 101-108.

[25] Lambert DM, Lowenberg-Debeor J, Bongiovanni R (2004) A comparison of four spatial regression models for yield monitor data: a case study from Argentina. Precis.agr. 5, 579-600.

[26] Matheron G (1962) Traité de Géostatistique Appliqué, Tome 1. Memoires du Bureau de RecherchesGéologiquesetMiniéres, Paris.

[27] McBratney AB, Bishop TFA, Teliatnikov IS (2000) Two soil profile reconstruction techniques. Geoderma 97, 209–221.

[28] McBratney AB, Webster R (1986) Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. J. soil sci. 37, 617-639.

[29] Nielsen DR, Biggar JW, Erh KT (1973) Spatial variation of field measured soil water properties. Hilgardia 42, 214-259

[30] Nielsen DR, Wendroth O (2003) Spatial and temporal statistics: sampling field soil and their vegetation. Catena Verlag, Reiskirchen, Germany.

[31] Pannatier Y (1996) VARIOWIN: Software for spatial data analysis in 2D. Springer-Verlag, NY.

[32] Papritz A, Dubois JP (1999) Mapping heavy metals in soil by (non-) linear kriging: an empirical validation. In: Gómez-Hernández, J. et al. (Eds.) geoENV II-Geostatistics for environmental applications. Kluwer Academic Publ. Dordrecht, The Netherlands. 429-440 p.

[33] Prakash MR, Singh VS (2000) Network for ground water monitoring-a case study. Environ. geol. 39, 628-632.

[34] Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipies in C. The art of scientific computing.Cambridge University press, NY.

[35] Si BC, Kachanoski RG, Reynolds WD (2007) Analysis of soil variability. In:E.G. Gregorich (Eds.) Soil sampling and methods of analysis. 1163-1191 p.

[36] Trangmar BB, Yost RS, Uehara G (1985) Application of geostatistics to spatial studies of soil properties.Adv. agro. 38, 45-94

[37] Van Kuilenburg J, De Gruijter JJ, Marsman BA, Bouma J (1982) Accuracy of spatial interpolation between point data on soil moisture supply capacity, compared with estimates from mapping units. Geoderma 27, 311-325.

[38] Vauclin M, Vieira SR, Vachaud G, Neilsen DR (1983) The use of cokriging with limited field soil observations. Soil sci. soc.am. j. 47, 175-184.

[39] Vieira SR (2000) Geostatistics in studies of spatial variability of soil. In: Novais, R.F. et al. (Eds.) Topics in soil science. Viçosa, Brazilian soc. soil sci. Vol 1, 1-54 p.

[40] Vieira SR, Millete J, Topp GC, Reynolds WD (2002) Handbook for geostatistical analysis of variability in soil and climate data. In: Alvarez, V. et al. (Eds.) Topics in soil science. Viçosa, Brazilian soc. soil sci. V.2.1-45 p.

[41] Webster R, McBratney AB (1989) On the Akaike Information Criterion for choosing models for variograms of soil properties. J. soil sci. 40, 493-496.

[42] Wilding LP, Bouma J, Goss D (1994) Impact of spatial variability on modeling. In: Bryant, R., Arnold, R.W., (Eds.) Quantitative modelling of soil forming processes. SSSA Special Publication #39. Soil Sci. Soc. Am. Inc. Madison, WI. 61-75 p.

[43] Yost RS, Uehara G, Fox RL (1982) Geostatistical analysis of soil chemical properties of large land areas. I. Semivariograms. Soil sci. soc.am. j. 46, 1028-1032.

[44] Zawadzaki J, Fabijanczyk P (2007) Use of variogramsfro field magnetory analysis in upper Silesia industrial area. Stud. geophys. geod. 51, 535-550.

[45] Zeleke TB, Si BC (2005) Scaling relationships between saturated hydraulic conductivity and soil physical properties. Soil sci. soc. am. j. 69, 1691-1702.