

New variogram modeling method using MGGP and SVR

Changik Han^{1,2} · Jiyang Wang³ · Mingguo Zheng³ · Ende Wang¹ · Jianming Xia¹ · GwangSu Li² · Sunchol Choe²

Received: 3 June 2015 / Accepted: 9 February 2016 / Published online: 16 February 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract A critical step for kriging in geostatistics is estimation of the variogram. Traditional variogram modeling comprise of the experimental variogram calculation, appropriate variogram model selection and model parameter determination. Selecting of the variogram model and fitting of model parameters is the most controversial aspect of geostatistics. Shapes of valid variogram models are finite, and sometimes, the optimal shape of the model can not be fitted, leading to reduced estimation accuracy. In this paper, a new method is presented to automatically construct a model shape and fit model parameters to experimental variograms using Support Vector Regression (SVR) and Multi-Gene Genetic Programming (MGGP). The proposed method does not require the selection of a variogram model and can directly provide the model shape and parameters of the optimal variogram. The validity of the proposed method is demonstrated in a number of cases.

Keywords Variogram modeling · Multi-gene genetic programming · Support vector regression · Geostatistics

Communicated by: H. A. Babaie

✉ Changik Han
han_6130@sina.com

¹ College of Resources & Civil, Northeastern University, Shenyang, China

² College of Geoporation Engineering, Kimchaek University of Technology, Pyongyang, Democratic People's Republic of Korea

³ College of Information Science & Engineering, Northeastern University, Shenyang, China

Introduction

Central to the application of most geostatistical studies is the precise modeling of the variogram, firstly termed by Matheron (1963), because it not only characterizes the spatial behavior of the interesting variables but also influences significantly on the kriging interpolation (Chilès and Delfiner 2012; Hilário and Manuela 2011; Desassis and Renard 2013). Despite many studies on the modeling of the variogram, it is still remained to the difficult problem in the application of geostatistics (Minasny and McBratney 2005; Li and Lu 2010; Oliver and Webster 2014).

Two usual variogram modeling approaches are the maximum likelihood (Kitanidis 1997) and least square methods (Jian et al. 1996).

The likelihood-based methods choose and fit the variogram models by maximizing of the joint likelihood function for the observed values, which gain ground among geostatisticians, especially to incorporate trend and external drift (Kitanidis 1983; Mardia and Marshall 1984; Zimmerman 1989; Kerry and Oliver 2007, 2010; Lark 2012; Lark and Webster 2006; Lark et al. 2006).

Compared with a least squares method, this method is more statistically efficient, whereas the latter has certain computational simplicity and availability to be widely used within geostatistical software packages (Zhang and Zimmerman 2007; Oliver and Webster 2014).

In the traditional least squares method, the estimation of the variogram includes two steps. At first, the experimental variogram is calculated directly from the observed values at specific lags. The experimental variogram is a finite set of discrete variances, whereas the underlying function should be continues among all lag distances. Therefore, the next step of the variogram estimation is to fit a smooth curve which ignores the point-to-point erratic fluctuation for the experimental values. The curve should

be expressed as a mathematical formula to describe the variances of the random processes with the changed lag and guarantee nonnegative variances in the spatial predictions. The modeling of the variogram is the process to replace the discrete experimental values with the closest negative definite function model. For this purpose, the valid variogram model, namely basic structure is selected, and then its model parameters are fitted.

The choice of model shape is particularly important, whereas the known valid variogram model shapes to satisfy above-mentioned conditions are limited to a few simple functions. The previous studies on the precision of variogram estimators (e.g. McBratney and Webster 1986; Pardo-Igúzquiza and Dowd 2001; Marchant and Lark 2007a) assumed that the variogram only conformed to some of commonly-used variogram model types, such as a spherical, exponential or Gaussian models. Those models, called basic structure depend on the limited set of parameters, such as the sill, the nugget effect, the range parameter, and the anisotropy ratios. Other basic structures need some shape parameters, for instance, the exponent in the power model. Furthermore, some practitioners still find valid models by eye, in which the final model is often selected by the specialist knowledge of the characteristics of the actual variable under consideration (Goovaerts 1997; Ricardo 2006; Oliver and Webster 2014). For several reasons including fluctuations in the experimental variogram, it was considered as a particularly hard problem for them to select the best model type by simple inspection. Hence, they adopt the manual method to carry out the selecting of valid variogram model with the naked eyes, or fitted all possible models to select the model with the best fit goodness. Although Pannatier (1996) proposed the improved method to combine the variogram modeling by eye with statistical evaluation, but this method consumed a lot of laborious and times with lacking of objectivity (Oliver and Webster 2014).

Once the set of basic structures, i.e. the shape of variogram model is defined, the optimal model parameters would be obtained by using standard minimization procedures. Several least squares methods have been proposed for a variogram modeling such as the ordinary least squares (OLS) (Journel and Huijbregts 1978; Clark 1979), weighted least squares (WLS) (Cressie 1985), and generalized least squares (GLS) (Genton 1998). The principle of least squares methods is to fit the model in order to minimize a cost function measuring the distance between the estimated model and the observed experimental variogram. Among of them, OLS method is the simplest and GLS method is recommended for modeling of variogram. Lahiri et al. (2002) proved that the GLS estimator was asymptotically efficient, but the GLS criterion was not

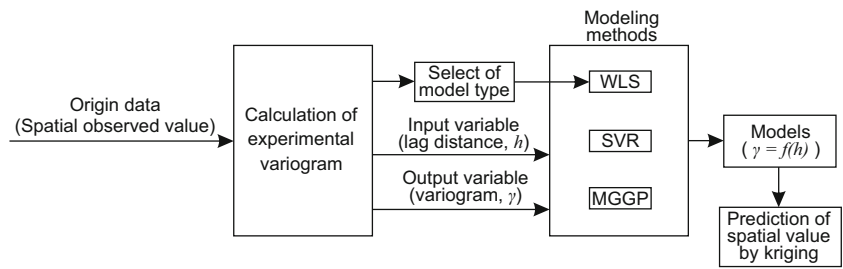
feasible since the exact expression for the covariance matrix of the variogram estimator was very difficult to obtain, even for Gaussian processes. This is the reason why the inversion of those covariance matrix and minimization by the GLS criterion often is computationally prohibitive (Lahiri et al. 2002; Hilário and Manuela 2011). Hilário and Manuela (2011) pointed that these difficulties of the GLS could be overcome by WLS. Recent studies showed that WLS method presented the most satisfactory results in fitting variogram model (Ricardo 2006; Emery 2010; Desassis and Renard 2013; Oliver and Webster 2014). In the several proposals for the weights, the approach proposed by Cressie (1985) are most commonly used.

As with any methods to fit the variogram model, all of them assumed the basic structure of model in advance and then found the optimal coefficients of predefined model structure. The known basic structure of variogram model are limited, it is very hard to pick the optimal variogram model even for the achieved good variogram fitting. Therefore, the estimation results lacked with the objectivity and the optimal type of the variogram model would not be found anytime to decrease the prediction accuracy. For this reason, the artificial intelligence (AI) methods could be recommended, which have the ability to pick up optimal model only based on input data.

Huang et al. (2012) and Chen et al. (2015) first introduced the SVR-based variogram fitting method. The SVR models based on the principle of structural risk minimization (SRM) were interested for their higher generalization ability and easy formulation only based on given data, which became popular among various researchers in the machine learning community (Kecman 2001; Garg et al. 2013b). Least square—support vector machines (LS-SVM) variant of SVR was adopted for predicting the performance of turning process (Çaydaş and Haşçalık 2008; Shi and Gindy 2007). However, it did not provide the explicit formulation between the input and output process parameters, and gave the output values in the crisp form (Garg and Tai 2014). In addition, the studies of Huang et al. (2012) and Chen et al. (2015) were limited to the local case studies, which did not prove the universality of their method in various variogram modeling. Therefore, we pay attention to genetic programming (GP), which possesses the ability to evolve the models structure and the coefficients automatically (Cevik and Guzelbey 2007; Cevik and Sonebi 2008; Gandomi et al. 2011). Most popularly used variant of GP is MGGP (Gandomi and Alavi 2011; Garg and Tai 2012, 2013; Garg et al. 2013a).

In this paper, we proposed the new variogram modeling method based on MGGP and demonstrated the practicalities of using the MGGP and SVR in the modeling of variogram through comparison analysis. The modeling problem formulation of the variogram is shown in Fig. 1.

Fig. 1 Formulation of modeling of variogram



Methodology

Standard variogram modeling

Consider a spatial stochastic process $\{Z(x): x \in D\}$, where the domain D is a subset of R^d , $d \geq 1$. Assume that $Z(x)$ satisfies the hypothesis of intrinsic stationarity:

$$\forall x, x + h \in D, \quad E[Z(x) - Z(x + h)] = 0, \tag{1}$$

$$\forall x, x + h \in D, \quad E\left[\{Z(x) - Z(x + h)\}^2\right] = Var[Z(x) - Z(x + h)] = 2\gamma(h). \tag{2}$$

The function $\gamma(h)$ is called the variogram, a function of lag distance h only, and it is defined as the variance of the increments $Z(x) - Z(x+h)$.

The standard variogram modeling consists of two stages.

The first stage is to calculate the experimental variogram. The classical estimator of the experimental variogram was proposed by Matheron (1962), and for a fixed $h \in R^d$, it is defined as

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i + h) - Z(x_i)]^2 \tag{3}$$

Where $Z(x_i+h)$ and $Z(x_i)$ are the actual values of Z at sampled locations x_i+h and x_i , and $N(h)$ is the number of pairs separated by the vector lag distance h defined as following:

$$N(h) = \{(x_i, x_j) : x_i - x_j = h; i, j = 1, \dots, n\}. \tag{4}$$

It is well known that this Matheron estimator has good properties, such as unbiasedness and consistency (Hilário and Manuela 2011).

Several factors affect the reliability of the experimental variogram. The most important factor is the sample size, which determines the reliability and accuracy of the experimental variogram. Generally, the more sample data would be helpful for the higher accuracy than ever. If the sampling interval is larger than the correlation range of the process under consideration, the experimental variogram would be flat:

‘pure nugget’ in the jargon. It is useless for prediction and tells us only that all variations occur within a shorter distance. The second factors are the lag interval and bin width. For data on a regular grid or at equal intervals on transects, the natural increment in the experimental variogram is one interval. Where the data are irregularly scattered, the data points must be grouped by the distance/direction. In irregular sampling schemes, for the purpose of grouping pairs, $Z(x_i+h)$ is regarded as a centroid of a distance class. Figure 2 illustrates the geometry of grouping, in which any measurement inside the shaded area (i.e. x_j) is considered for the calculation of $\hat{\gamma}(h)$, although it is not an exact distance h from x_i . The judgment is needed for choosing the lag interval and bin width. If the lag interval is short and bin width is narrow, there would be many estimates of $\gamma(h)$, each based on few sample points and subject to large error, and the variogram would appear ‘noisy’. Else if the lag interval is large and bin width is wide, there would be too few estimates of $\gamma(h)$ to reveal the form of the variogram. In practice, the lag interval h is typically taken close to the value of the average sampling distance and the bin width is sensibly determined by adjusting the lag tolerance dh , the lateral tolerance db and the angular tolerance δ .

The second stage in standard method is to select the authorized continuous variogram model and to fit it to the experimental variogram. The variogram characterizes spatial variability of the variable under consideration. Not any continuous function will serve, since the variogram models must satisfy the conditional negative definiteness property as follows:

$$\sum_{i=1}^m \sum_{j=1}^m a_i a_j \gamma(x_i - x_j) \leq 0 \tag{5}$$

for any $\{x_i \in D \subset R^d \mid 1 \leq i \leq m, m \in N\}$ and for any $\{a_i \in R \mid 1 \leq i \leq m\}$, such that $\sum_{i=1}^m a_i = 0$. This mathematical property ensures that the variogram is the licit measure of the lag distance and all resulting variances are non-negative for all possible configurations of conditioning data (Journel and Huijbregts 1978).

Although there is an infinite number of the negative definite function, the basic shape of the variogram rising from zero to reach the limiting value restricts to a few negative definite functions to be interested. Webster and

Oliver (2007) described those most commonly employed authorized functions.

For example, the spherical model

$$\gamma(h) = \begin{cases} c_0 + c \left(\frac{3h}{2a} - \frac{h^3}{2a^3} \right) & 0 \leq h \leq a, \\ c_0 + c & h > a \end{cases}, \quad (6)$$

the exponential model

$$\gamma(h) = c_0 + c \left(1 - e^{-\frac{h}{a}} \right) \quad (7)$$

and Gaussian model

$$\gamma(h) = c_0 + c \left(1 - e^{-\frac{h^2}{a^2}} \right) \quad (8)$$

Here, c_0 is the nugget effect, (c_0+c) the sill and a the range. They are the unknown parameters to be fitted. Nugget effect is used to show the jump at $h \rightarrow 0^+$, also means the white noise caused by observations or data. Nugget effect, c_0 , is partitioned into two sub-components: the error variance and the micro variance. Among them, the micro variance represents the uncorrelated variation at the scale of sampling while the error variance is the variation that remains unresolved including any measurement error. The sill shows the asymptotic value $\lim_{h \rightarrow \infty} \gamma(h)$ and range is the minimum distance where $\gamma(h)$ reaches sill. It means if the distance of two points is out of the range, the covariance of the two random variables becomes zero. In Eq. (6)–(8), we can also know that $c_0 \geq 0$, $(c_0+c) > 0$ and $a > 0$.

Wackernagel (2003) pointed the implausible results from the use of Gaussian model, and more recently Chilès and Delfiner (2012) and Oliver and Webster (2014) reported the ill-considered use of the Gaussian model, which was at the limit of the acceptability for the random process and led to bizarre predictions. Therefore, we only refer to the spherical and exponential models in this study.

If the basic variogram model is selected, the optimal model parameters would be determined by several least squares procedures. The most feasible approach is the weighted least squares. One of the advantages of this option is that it automatically gives more weight to early lags with the maximum number of the pairs and down the weight to lags with the small number of the pairs to produce the unbiased estimated minimum variance.

Suppose the vector of variogram parameters is denoted by $\alpha = (c_0, c, a)$ and the basic variogram model is denoted by $\gamma(h_i; \alpha)$, and the experimental variogram model is denoted by

$\hat{\gamma}(h_j)$. The objective is to find the value of α for minimizing the error variance as follows:

$$\sum_{i=1}^{n_b} w_i \left\{ \hat{\gamma}(h_i) - \gamma(h_i; \alpha) \right\}^2, \quad (9)$$

Where n_b is the number of bins and the $h_i, i = 1, \dots, n_b$ are the lag distances for which the experimental variogram is estimated. The weights w_i account for the varying reliability of each entry of the experimental variogram due to the number of the pairs used to calculate $\hat{\gamma}(h_i)$ and the inverse relation between the reliability of $\hat{\gamma}(h_i)$ and the actual value of $\gamma(h_i)$. The weights chosen in this paper are those given by Cressie (1985)

$$w_i = \frac{N(h_i)}{\hat{\gamma}(h_i)^2} \quad (10)$$

In above-mentioned traditional approach, the variogram modeling has relied on fitting the known conditional negative definite functions such as spherical, exponential, and Gaussian models. Any positive linear combinations of the variogram models were also considered as the valid functions (Deutsch and Journel 1998). While this provided the workable mechanism for modeling optimal variograms, there were some cases to do not well fit with this framework. Figure 3 shows the example commonly observed in the experimental variograms, which is not easy to fit with the conventional model shapes (Pyrz and Deutsch 2006).

The application of more flexible variogram modeling method is inhibited by the difficulty in ensuring the conditional negative definiteness. There is a largely unexplored suite of the conditional negative definite models to provide the additional flexibility. We find the corresponding candidate in the artificial intelligence (AI) methods such as SVR and MGGP.

In the next sections, we represent how to carry out more flexibly the modeling of variogram from the experiment variograms without assuming the basic structures of variogram based on the SVR and MGGP.

SVR-based method

The most popular and advanced technique in the field of artificial intelligence is the support vector machine (SVM). The structure of SVM is shown in Fig. 4.

The SVM model is comprised of the input process variables, support vectors, kernel function and output variable. SVM has been well applied to solve the classification problems and it is known as SVR for applying to regression problems (Gupta 2008; Hearst et al. 1998; Byvatov and Schneider 2003; Garg et al. 2013b). Unlike the regression analysis or other statistical models, SVR is not based on the statistical assumptions (model structure, error dependency, etc.), and does not require any assumption of the model structure and

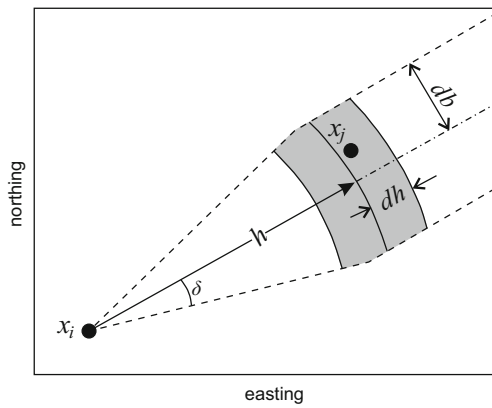


Fig. 2 Definition of the distance class in the estimation of the variogram

also easily formulates the models based only on the given data. The SVR models are known for their unique capability of imparting generalization ability to the model. Therefore, SVR has been extensively used in solving the symbolic regression problems (Kecman 2001; Hadi and Ahmed 2006; Al-Ahmari 2007; Basak et al. 2007).

SVR is based on the statistical learning theory of which the framework is formulated based on the structural risk minimization (SRM) principle. The SRM principle is the modified form of the empirical risk minimization principle. SRM minimizes the upper bound on the expected risk and therefore play a key role in the SVR algorithm formulation. The original input variables in the lower dimensional space are projected into the higher dimensional space so as to convert the regression problem with non-linearity to the linear regression problem. The conversion is carried out using several nonlinear hyperspace functions.

The training data $\{(x_i, y_i)\}_{i=1}^N \in R^m \times R$ are used to formulate the SVR model, where x_i and y_i are the input variable and the actual output value of the process, respectively. In the present work, there are one input (i.e. lag distance) and one output (i.e. variogram). The SVR model is given by:

$$y = f(x) = \sum_{i=1}^N w_i \partial_i(x) + b = w^T \partial(x) + b \tag{11}$$

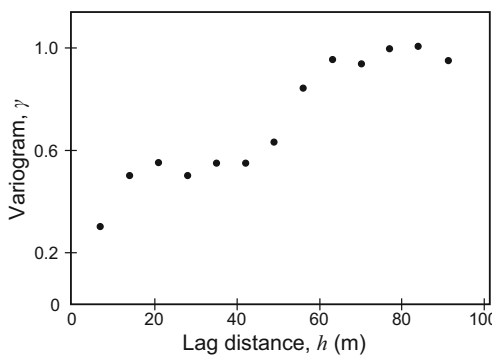


Fig. 3 Experimental variogram that is not well fit by nested sets of traditional variogram models

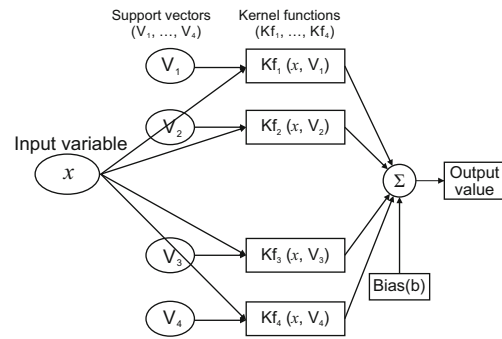


Fig. 4 Structure of SVM

where the function $\partial_i(x)$ is the feature space to be converted into the higher dimensional space, and $w = [w_1 w_2 \dots w_N]^T$ and $\partial = [\partial_1 \partial_2 \dots \partial_N]^T$.

Equation (11) represents the non-linear regression model as the hyper-surface projected from the input variable space into the higher dimensional space. The regression model given by $\partial(x)$ represents the converted linear form of the original non-linear model in the higher dimensional space. Based on the data obtained from the process, the chosen kernel function learns and minimizes the regularized risk function (L_r). By this optimizing risk function L_r , the parameters, namely, support vector weight (w) and bias (b), are evaluated.

$$L_r(w) = \frac{1}{2} w^T w + \lambda \sum_{i=1}^N |y_i - f(x)|_e \tag{12}$$

where

$$|y_i - f(x)|_e = \begin{cases} 0 & , \quad \text{if } |y_i - f(x)| < \varepsilon \\ |y_i - f(x)| - \varepsilon & , \quad \text{otherwise} \end{cases} \tag{13}$$

The regularization parameter (λ) regulates the trade-off between the approximation error and the weight vector norm ($\|w\| = \sqrt{w^T w}$). The approximation error will be decreased according to the increase λ or the weight vector norm, but this may not ensure the higher generalization ability of the model

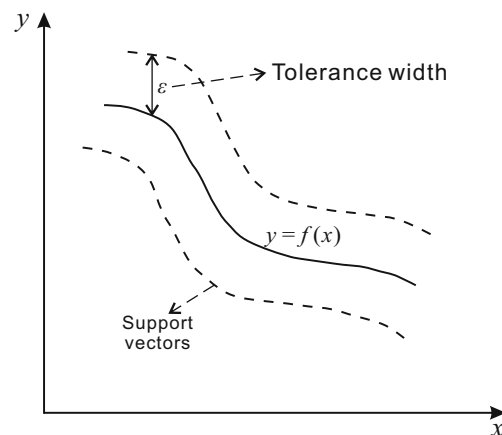


Fig. 5 SVR model and its support vectors and tolerance width

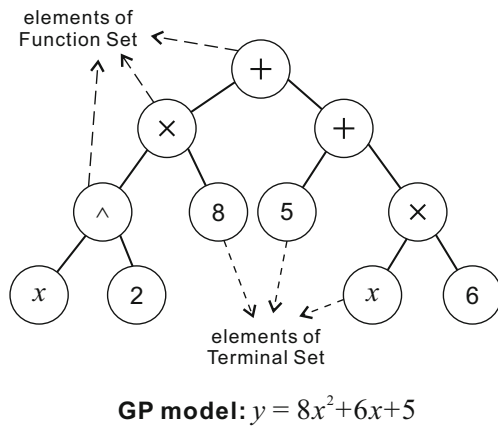


Fig. 6 Example of GP model

and lead to the over-fitting. λ and ε are defined by the user where ε is the tolerance level or band width of the model (Vapnik 1995). $(|y_i - f(x)|_\varepsilon)$ is the ε -insensitive loss function given in Eq. (13). If the predicted values $f(x)$ lies within the defined tolerance width ε , the loss function is zero while for points outside ε , the loss function equal to the absolute magnitude of difference between the values predicted by the SVR model and tolerance width ε .

The points on the margin lines defined by $(y = f(x) \pm \varepsilon)$ are called the support vectors, whereas those outside these lines are known as the error set (Fig. 5).

Increasing ε decreases support vectors and thus lead to the data reduction.

In this study, we will use the LS-SVM toolbox (Pelckmans et al. 2002) built in MATLAB for implementing the SVR

method. Least squares support vector machine, originally proposed by Suykens et al. (2001), is a variant of SVM. LS-SVM transforms the inequality constraints of the standard SVM to equality constraints. The recent extensive applications of this toolbox (Salgado and Alonso 2007; Salgado et al. 2009; Çaydas and Ekici 2012; Saptoro et al. 2012; Garg et al. 2014d) for solving symbolic regression problems in the varying nature show that the chosen method is reliable. Huang et al. (2012) and Chen et al. (2015) also used LS-SVM for modeling of variogram.

The implementation and performance of the SVR-based variogram model is discussed in “Case study”.

MGGP-based method

Except for SVR, the promising variogram modeling technique without assuming their basic structure is MGGP. In order to understand the concept of the proposed methodology MGGP-based variogram modeling, we first discuss about GP.

GP is considered to be the most famous for solving symbolic regression problems and is widely used in modeling process of varying nature (Koza 1994; Madár et al. 2005; Wang et al. 2011; Garg et al. 2013a, b). GP based on the Darwinian’s theory of “survival of fittest” finds the optimal model automatically by mimicking the process of the evolution in nature (Koza 1994). GP works on the principles of genetic algorithm (GA) but there are several differences between GP and GA (Garg et al. 2014a, b, c). The solutions in GP are usually represented by the tree structures of varying

Fig. 7 Crossover operation of GP model

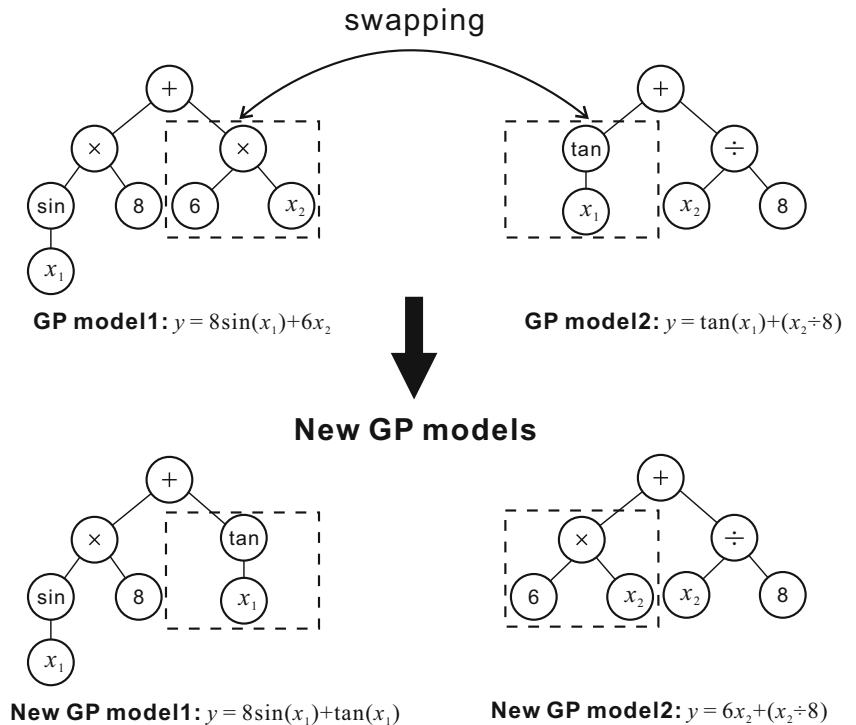
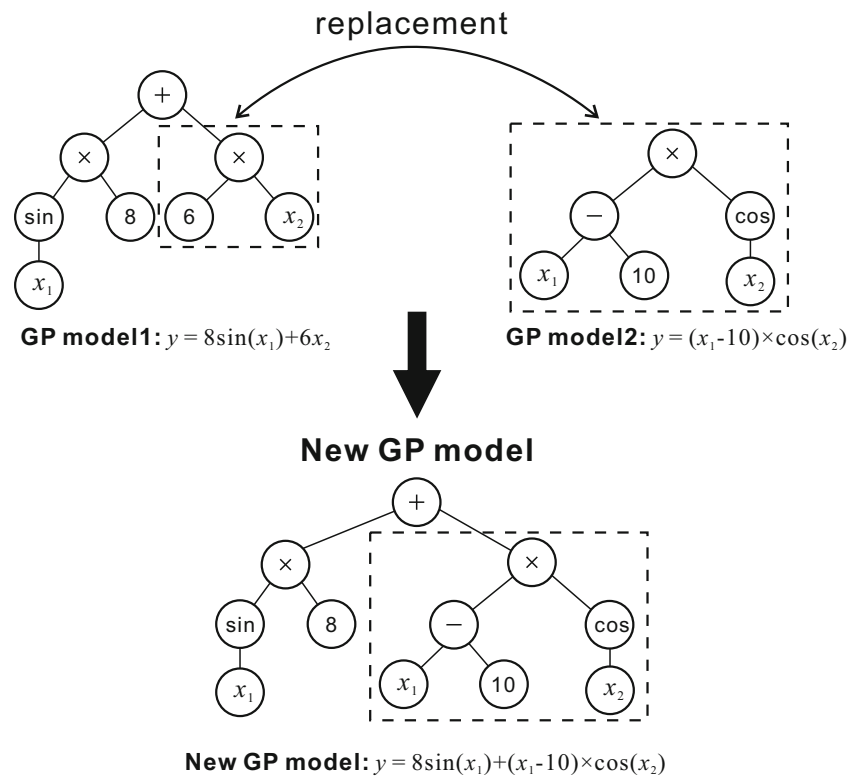


Fig. 8 Mutation operation of GP model



sizes while the solutions in GA are represented by strings (binary or real number) of fixed length. Thus researchers said that GA is the parameter optimization method whereas GP is the structure optimization method. GP generates both of model types and its coefficients automatically based on the given input data. The main advantage of GP over the other regression analysis and statistical modeling techniques is to have the ability of generating the mathematical expressions without assuming any prior form of the existing relationships.

GP algorithm is started by generating the models randomly. The numbers of the generated models are represented by the population size. The models are encoded in form of a tree, each tree node representing a function, variable or a constant number by combining the elements randomly from the functional and terminal set. The function set F usually comprise of elements such as arithmetic operators (+, -, ×, /, etc.), non-linear functions (sin, cos, tan, exp, tanh and log), Boolean operators (AND, OR, etc.), or other operators as defined by the user. The terminal set T consists of the elements such as the random numerical constants and input variables of the process. An example of the GP model is shown in Fig. 6. The performance of the initial population in the models is evaluated on the training data according to the fitness function, namely, root mean square error (*RMSE*), given by

$$RMSE = \sqrt{\frac{\sum_{i=1}^N |G_i - A_i|^2}{N}} \times 100 \tag{14}$$

where G_i is the predicted value of i th data sample by the GP model, A_i is the actual value of the i th data sample and N is the number of training samples.

Based on the performance on training data, the algorithm selects models for the genetic operations such as reproduction, mutation and crossover. The selection methods such as tournament selection, rank selection, and roulette wheel are used to select the individuals for the genetic operations. The most commonly used method is the tournament selection, which is well known for maintaining the genetic diversity in the population. The purpose of performing the genetic operations is to form the new population which represents the new generation. The individuals with minimum fitness value would be reproduced in the next generation whereas the crossover and mutation operations are applied on the remaining selected

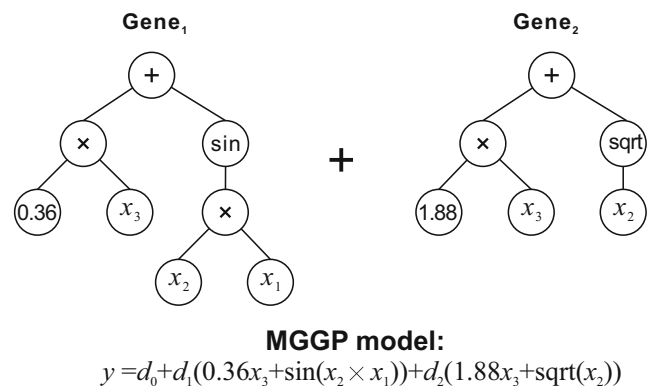


Fig. 9 Example of MGGP model

individuals. The subtree crossover operation is shown in Fig. 7, in which two branches are selected randomly from two trees and swapped. The subtree mutation operation is shown in Fig. 8, in which the node (terminal or functional) is selected at random from the given tree to be replaced by branch of the new generated random tree. This iterative algorithm of generating the new populations continues until the termination criterion is satisfied. The termination criterion can be the maximum number of generations or the threshold error of the model as specified by the user, whichever is achieved earlier.

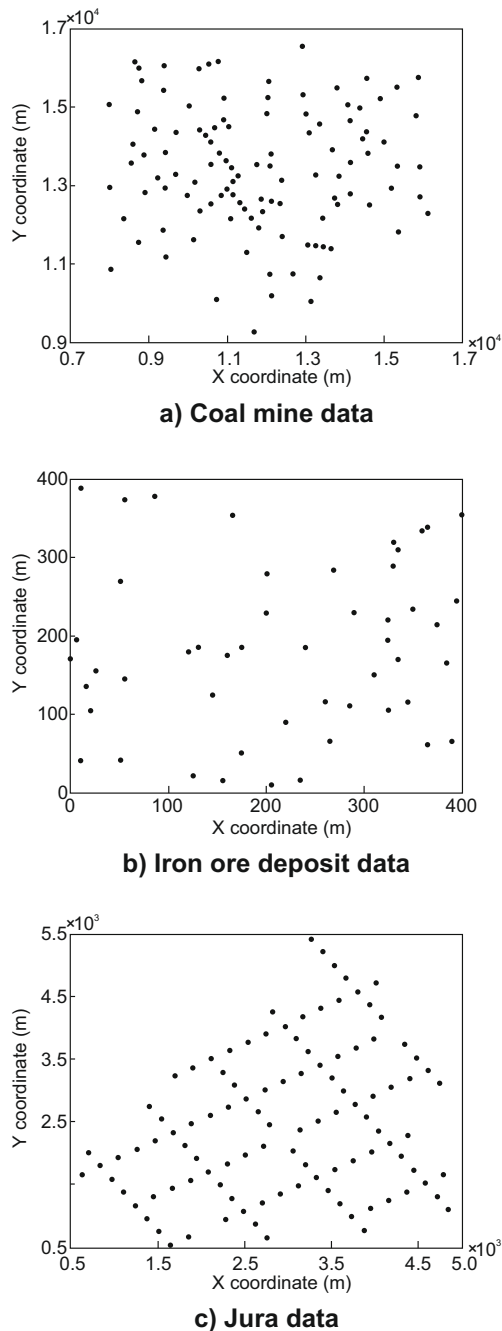


Fig. 10 Spatial distribution of sample points

MGGP is the robust variant of GP, which effectively combines the model structure selection ability of the standard GP with the parameter estimation power of the classical regression by using the new characteristic called ‘multi-gene’. In traditional GP method, the model is the single tree/gene expression whereas the model formed in MGGP is the linear combination of several low order non-linear trees/genes which each of them is the traditional GP tree (Searson et al. 2010). Recently, the MGGP have been used successfully for engineering modeling problems (Gandomi and Alavi 2012; Garg et al. 2014b). It has been shown that MGGP regression would be more accurate and efficient than the standard GP for modeling the nonlinear problems.

Specifically speaking, the key difference between GP and MGGP is that, in the latter, the model participating in the evolution is the combination of several sets of genes/trees. For the system with u input of dimension $R^{n \times m}$ to produce the model output y with dimension $R^{n \times 1}$, where n is the number of observations taken and m is the number of input variables, we could produce the tree structure which introduces the mathematical relationship:

$$\hat{y} = f(u_1, \dots, u_i) \quad (15)$$

In MGGP symbolic regression, each prediction of the output variable \hat{y} is formed by the weighted output of each of the trees/genes in the multi-gene individual plus the bias term. Each tree is the function of zero or more of the i input variables u_1, \dots, u_i . Mathematically, the MGGP model can be written as:

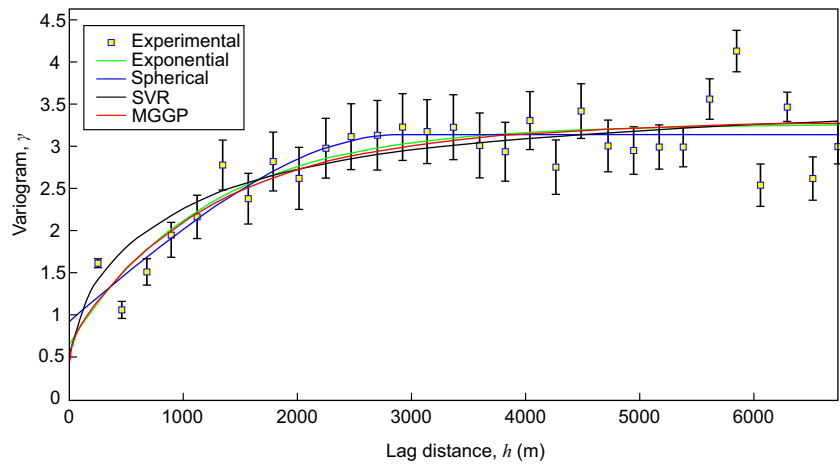
$$\hat{y} = d_0 + d_1 \times tree_1 + \dots + d_M \times tree_M \quad (16)$$

where d_0 represents the bias of offset term while d_1, \dots, d_M are the gene weights and M is the number of genes (i.e. trees) which constitute the available individual. The weights (i.e. regression

Table 1 Parameter setting for MGGP

Parameters	Setting values
Population size	300
Number of generations	200
Tournament size	20
Function set (F)	Multiply, plus, minus, protected divide, protected power, tanh, exp, atan, ifth
Terminal set (T)	h (i.e. lag distance), $[-2 \ 2]$
Crossover probability	0.85
Reproduction probability	0.10
Mutation probability	0.05

Fig. 11 Experimental variogram and variogram models using different methods in coal mine data



coefficients) are automatically determined by the least squares procedure for each multi-gene individual. In multi-gene symbolic regression, each symbolic model is represented by number of GP trees weighted by linear combination. Each tree is considered as the gene by itself. The typical example of MGGP model and its mathematical expression are shown in Fig. 9.

The MGGP algorithm is outlined as follows:

BEGIN

Step1: Formulate problem

Step2: MGGP algorithm

Begin

2.a Set initial parameters such as function and terminal set, number of generations, population size, maximum depth of gene, maximum number of genes to be combined, probability rate of genetic operators and termination criterion, etc.

2.b Randomly generate initial population of genes

2.c Form models by combining set of genes using least squares method

2.d Evaluate performance of models based on the fitness function

2.e Apply genetic operations and form the new population

2.f Cross-check the models performance against the termination criterion, if not satisfied, GO TO Step 2.e, and else if satisfied, select the best model

End;

END;

Returning to the aim of the present work, our goal is to get the variogram model from the experimental variograms by using MGGP. For this purpose, we can formulate problem as follows:

Input and output variables of the process are the lag distance h and variogram $\gamma(h)$, respectively. The training data can be composed of the lag distances h_1, \dots, h_N and experimental variograms $\hat{\gamma}_1, \dots, \hat{\gamma}_N$. The fitness

Fig. 12 Experimental variogram and variogram models using different methods in iron ore deposit data

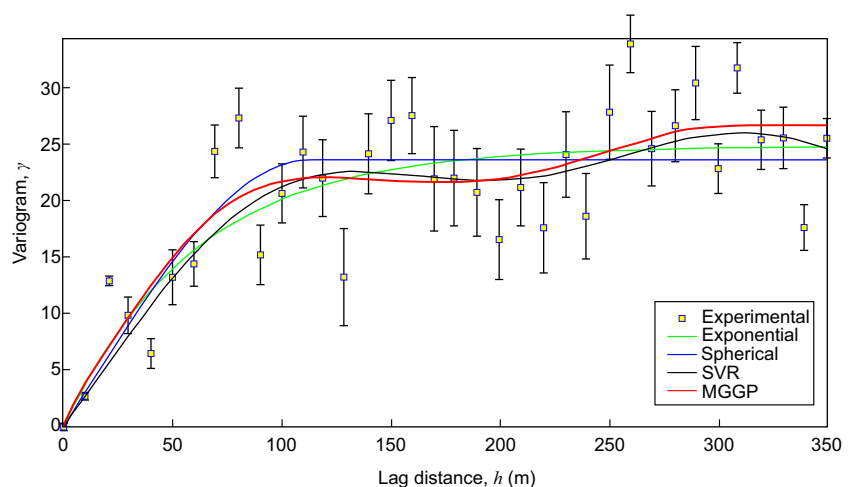
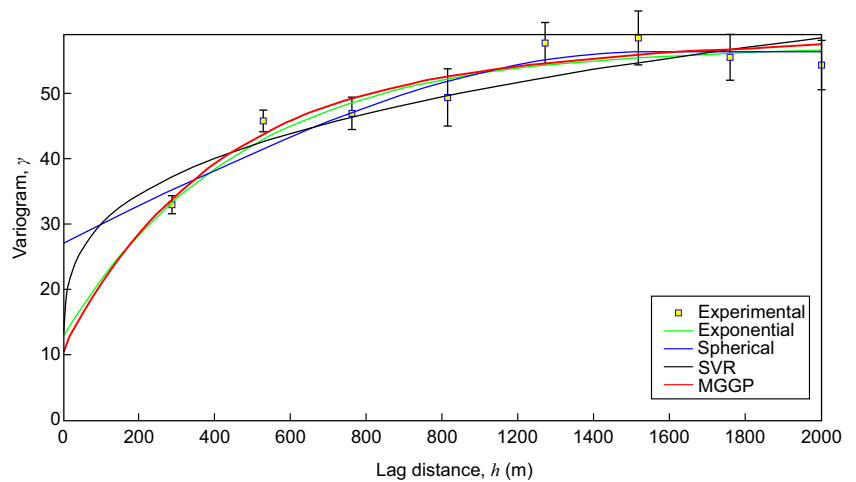


Fig. 13 Experimental variogram and variogram models using different methods in Jura data



function used for performance evaluation of population in the variogram modeling can be defined by:

$$fitness = \sqrt{\frac{\sum_{i=1}^N |\gamma_i - \hat{\gamma}_i|^2}{N}} \quad (17)$$

Where γ_i is the predicted value at the i th lag distance by the MGGP model, $\hat{\gamma}_i$ is the experimental variogram value at the i th lag distance and N is the number of training samples.

MGGP-based variogram model is selected based on minimum fitness on training data from all runs. This implementation and performance are discussed in “Case study”.

Case study

This section aimed to compare the proposed methods with the traditional method in the modeling of variogram to demonstrate their performance. Studies of Huang et al. (2012) and Chen et al. (2015) using SVR limited to variogram modeling

with no nugget effects, and also they did not consider condition about the size of input data for modeling variogram (i.e. number of discrete points corresponding to experimental variogram). It is well known that a good fit of the variogram near the origin is especially important (Cressie 1991; Stein 1988). In addition, AI method such as SVR depends greatly on the size of training data. For this illustration, we have selected three data taken from the geostatistical studies well-known in many of geoscientists. The first and second is the simulated coal mine and iron ore deposit data from the Clark’s geostatistical study (Clark 1979, 1983; Clark and Harper 2000), respectively, and the third data is nickel (Ni) concentrations in the topsoil of a region of the Swiss Jura, analyzed by Lark (2000).

The coal mine data based on a real coal seam in Southern Africa. Several measurements are made on each sample: width of coal seam (m); calorific value of the coal (KJ); and the vertical location of the top of the seam (elevation, m). Among of them, the calorific value of the coal is selected for the performance of the variogram modeled by various methods. It includes 116 borehole samples drilled into a coal

Fig. 14 Fourier transform result of MGGP variogram model in coal mine data

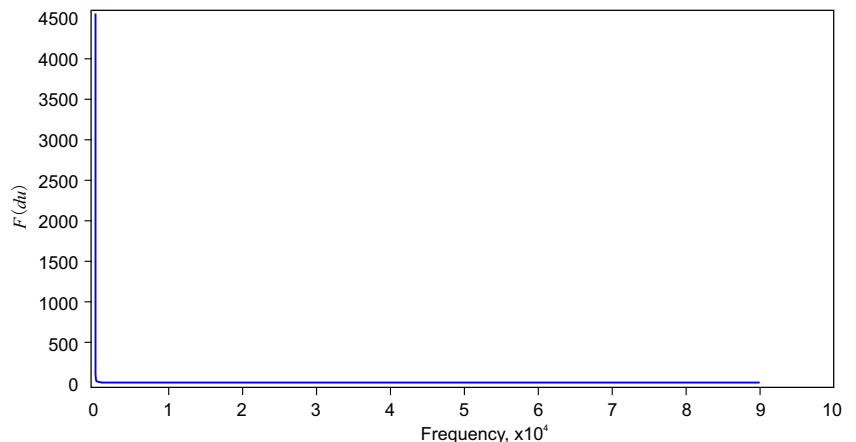
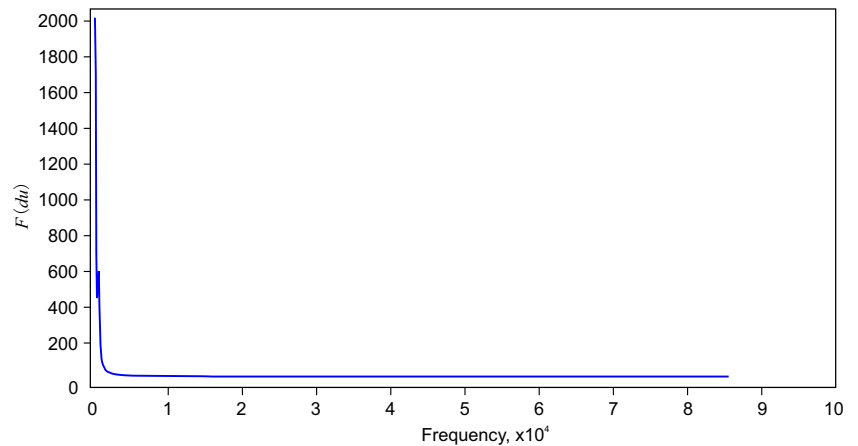


Fig. 15 Fourier transform result of MGGP variogram model in iron ore deposit data



seam. After outlier processing, 110 data are used for our analysis. The calorific value of the coal ranges from 19.89KJ to 26.90KJ. All coordinates are in meters.

The simulated data of low-grade iron ore deposit used in Clark’s geostatistical studies (Clark 1979, 1983), with an overall average of about 35 % Fe, has been sampled by means of 50 randomly positioned boreholes perpendicular to the dip of the ore body. This iron ore grades data are known to have a variogram model with no nugget effect.

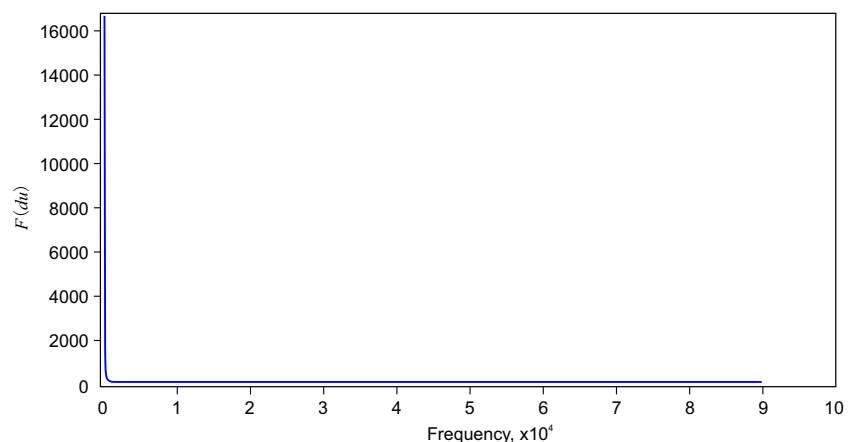
The coal mine data and iron ore deposit data can be accessed from <http://www.kriging.com/datasets/>.

Jura data were collected and described by Atteia et al. (1994) and analyzed fairly exhaustively by the authors (Atteia et al. 1994; Webster et al. 1994; Goovaerts et al. 1997; Lark 2000; Marchant and Lark 2007b). Following Lark (2000), the data set was divided into a group of 106 prediction data and one of 104 validation data. Among them, the former set consisted of 10 intersecting transects of different length, which would be used in our analysis. At each site the metal concentrations of the soil to 25 cm is measured, in which nickel value ranges from 5.24 to 43.68 mg/kg.

Jura data can be accessed from <http://home.comcast.net/~pgoovaerts/book.html/>.

The layouts of three data sets are shown in Fig. 10.

Fig. 16 Fourier transform result of MGGP variogram model in Jura data



The analyses of these data sets described in the literature show the little evidence of anisotropy. Therefore, the variograms in this study were modeled assuming isotropy.

Our work is carried out according to the flow process as shown in Fig 1.

Calculation of experimental variogram

The experimental variograms for various cases are calculated based on Eq. (3). It was already mentioned that the lag interval and bin width affect the reliability of the experimental variogram, as well as the sample size.

Experimental variogram for the coal mine data is calculated at lags $h_l = 225 \times l$ (meter), $l = 1, \dots, 30$. The lags for iron ore deposit data are $h_l = 225 \times l$ (meter), $l = 1, \dots, 30$. In Jura data, the calculation of experimental variogram was applied for lag classes centred on 250 m, 500 m, 750 m, ..., 2000 m.

Implementation of traditional WLS variogram estimator

Based on the principle as shown in “Standard variogram modeling”, the rational shape of the variogram model is selected and the weighted least-squares algorithm recommended by Cressie (1985) is applied in modeling of the variogram. In

Table 2 Mathematic expressions of different variogram models for coal mine data

Model	Mathematic expression
Spherical model	$\gamma(h) = \begin{cases} 0.9166 + 2.2190 \times \left(\frac{3h}{2 \times 2923.3} - \frac{h^3}{2 \times 2923.3^3} \right), & 0 \leq h \leq 2923.3 \\ 0.9166 + 2.2190, & h > 2923.3 \end{cases}$
Exponential model	$\gamma(h) = 0.6452 + 2.6102 \times \left(1 - e^{-\frac{h}{1204.7}} \right)$
MGGP model	$\gamma(h) = 2.6164 - 2.3376 \times (0.9996)^{2h} + 0.2595 \times h^{0.1082}$

all cases, the spherical or exponential model gave the best fit as judged by the Akaike information criterion (McBratney and Webster 1986; Oliver and Webster 2014).

SVR implementation

The selection of kernel function plays the key role in learning and minimizing the loss function efficiently since it affects the generalization ability of the SVR model. Huang et al. (2012) and Chen et al. (2015) used RBF kernel function in the modeling of variogram. But the select of the kernel function must be changed according to the practical demands. In this work, four kernel functions such as linear, polynomial, radial basis function and multi layer perceptron are chosen for the performance of SVR models. Among them, the Gaussian radial basis function (in the coal mine data and iron ore deposit data) and polynomial (in Jura data) kernel functions show the advantage for the faster and efficient training, and therefore, they are selected for our analysis.

The optimal kernel parameters of λ , σ^2 (radial basis function) and t (polynomial) are determined using a combination of coupled simulated annealing (CSA) and the grid search method. The CSA determines the good initial values of λ , σ^2 and t , and then, these are passed to the grid search method, which uses cross-validation to fine tune the parameters. Optimal parameters for coal mine data are $\lambda=42.8769$ and $\sigma^2=15938.4781$. $\lambda=31.0137$ and $\sigma^2=12133.0025$ for iron ore deposit data and $\lambda=1.6487$ and $t=1.6487$ for Jura data are set, respectively.

MGGP implementation

In modeling of the variogram, the implementation of MGGP method also requires adjustment of its parameters.

The parameter setting is important since it affects the generalization ability of the MGGP model. The parameters selected based on trail-and-error approach are shown in Table 1.

The elements in the function set are the broader set of functions so as to evolve the variety of the non-linear mathematical models. The elements in the terminal set consists of the one input process variable (i.e. lag distance h) and random constants chosen in the range $[-2 \ 2]$. The range of random constants is chosen so as to take into account the variance of measurement errors in the data collection.

The parameters like population size and number of generations fairly depends on the complexity of the regression problem. In generally, the population size and number of generations should be fairly small for training data of the large samples. Since the MGGP model is formulated from the set of genes, the model will have the higher complexity i.e. greater number of nodes along with the evolution, and may result in over-fitting. The restriction on the maximum number of genes and depth of the gene exerts influence over the complexity of the models and results in accurate and compact models. One of major goals in this study is to find the smooth curve that ignore the point-to-point erratic fluctuation based on the experimental variogram values so as to ensure the conditional negative definiteness. Therefore, in this study, the maximum number of genes and maximum depth of gene are kept at 3 and

Table 3 Mathematic expressions of different variogram models for iron ore deposit data

Model	Mathematic expression
Spherical model	$\gamma(h) = \begin{cases} 23.6191 \times \left(\frac{3h}{2 \times 113.1663} - \frac{h^3}{2 \times 113.1663^3} \right), & 0 \leq h \leq 113.1663 \\ 23.6191, & h > 113.1663 \end{cases}$
Exponential model	$\gamma(h) = 24.8005 \times \left(1 - e^{-\frac{h}{60.1255}} \right)$
MGGP model	$\gamma(h) = \begin{cases} -334.606 - 66.6977 \times (1.0206^h)^{0.2722} + 0.9650 \times h^{1.7705} + 319.68 \times \tan(1.0206^h) + 191.318 \times \tan(0.9758^h) - 0.1545 \times h^{1.6714}, & 0 \leq h \leq 305.4765 \\ 26.6630, & h > 305.4765 \end{cases}$

Table 4 Mathematic expressions of different variogram models for Jura data

Model	Mathematic expression
Spherical model	$\gamma(h) = \begin{cases} 27.1 + 29.2 \times \left(\frac{3h}{2 \times 1542} - \frac{h^3}{2 \times 1542^3} \right) & 0 \leq h \leq 1542 \\ 27.1 + 29.2 & h > 1542 \end{cases}$
Exponential model	$\gamma(h) = 12.6551 + 44.4545 \times \left(1 - e^{-\frac{h}{462.7579}} \right)$
MGGP model	$\gamma(h) = 52.8865 + 0.4094 \times (\text{atan}(h) + 0.0048h) - 42.3957 \times 0.9975^h$

2 for the coal mine data and Jura data and 3 and 5 for the iron ore deposit data, respectively.

Assessments of the various variogram models

The variogram modeling result obtained from the three modeling methods, traditional WLS, SVR and MGGP, are illustrated in Figs. 11, 12, and 13 and Tables 2, 3, and 4.

As shown in Figs. 11, 12, and 13, the coal mine data and Jura data have the variogram with nugget effect while the iron ore deposit data have one with no nugget effect. In addition, the number of discrete points for modeling of variogram in the iron ore deposit and coal mine data are relatively large (30 and 50, respectively) whereas one in the Jura data small (8 of discrete points).

In all cases, the SVR-based and MGGP-based variogram models are shown in Figs. 11, 12, and 13, which indicate these two models have impressively learned the non-linear relationship between spatial variation (output variables) and lag distance (input variables) without assuming their basic structure in comparison with the traditional WLS method.

Now, we have to test whether these specified functions satisfy the conditionally negative-definite. For this purpose, Bochner’s theorem can be applied directly to assert that $\gamma(h)$ is conditionally negative definite.

Theorem (Bochner): A continuous real function $-\gamma(h)$ defined in R^n is continuous and positive definite if and only if it is the Fourier transform of a positive bounded Borel measure $F(du)$:

$$-\gamma(h) = \int e^{2\pi i(u,h)} F(du) = \int \cos(2\pi(u, h)) F(du) \quad (18)$$

with

$$\int F(du) < \infty$$

where u represents the frequency, i the unit pure imaginary number. That is, we must take the Fourier transform of $-\gamma(h)$ and testify that it is positive bounded.

The obtained MGGP variogram models consist of the linear combination of the complicated functions with the positive or negative coefficients. This makes it difficult to take the Fourier transform of the obtained variogram model using the analysis method. Therefore, we used the numerical approximation method to see whether its Fourier transform have the positive bounded symmetric measure. The obtained results (Figs. 14, 15, and 16) show that it is positive bounded, summable $\int F(du) < \infty$ and then that the MGGP variogram models is a valid for satisfying the conditionally negative definiteness property.

When the size of input data is small (in the case of Jura data), the MGGP model is more robust than SVR model. In addition, MGGP approach completely surmounts the deficiency of SVR that does not provide explicit formulation of obtained variogram model (Table 2, 3, and 4).

In the cases of Coal mine and Jura data, MGGP model is almost close to the exponent model. However, in the case of Iron ore deposit, it reflects more exactly the non-linear relationship between the experimental points than in the traditional WLS method.

Once the variogram was estimated, it would be used in the kriging interpolation. Kriging predicts the unknown values from the weighted average of the sparse sampled values in the neighborhood of non-sampled location based on the stochastic model of the spatial variation i.e. the variogram model. The estimate of the error variance for each prediction is also generated by the kriging. Performance of the variogram model must be assessed with the results of kriging interpolation. For

Table 5 Hypothesis testing to compare the kriging prediction using MGGP and SVR models

95 % CI	p value					
	Coal mine		Iron ore deposit		Jura data	
	MGGP	SVR	MGGP	SVR	MGGP	SVR
Mean <i>t</i> test	0.4922	0.4827	0.4881	0.4905	0.4922	0.4979
Variance <i>F</i> test	0.9364	0.8632	0.7986	0.7914	0.8101	0.8320

Table 6 assessment of different variogram models by error statistics in the coal mine data

Model	RMSPE	MAPE	$\tilde{\theta}(x)$
Spherical model	1.5183	1.1872	0.7263
Exponential model	1.4656	1.1464	0.6283
SVR model	1.4399	1.1251	0.4327
MGGP model	1.4178	1.1056	0.4702

this purpose, here we used the ordinary kriging prediction to enable the fair comparison to be made between the variograms modeling by the traditional WLS, SVR and MGGP method. The kriging variances were also computed from the corresponding variogram models.

As the comparison of the kriging interpolation results, the cross-validation technique was used. That is, we attempt to validate our models by dropping out each observed values and cross estimating the value at that location from the neighboring residual samples. For each spatial location x_i , based on the set of observations without $Z(x_i)$, a predictor of $Z(x_i)$ is calculated as following:

$$\hat{Z}(x_i) = \sum_{j \neq i} \lambda_j Z(x_j) \quad (19)$$

And their kriging variance is obtained correspondingly.

Firstly, hypothesis tests are used to compare the goodness of kriging prediction using the MGGP and SVR models. These are t tests to test the mean and f tests for variance (Table 5). For the t and f tests, the p values of the two models are >0.05 , so there is not enough evidence to conclude that the observed values and predicted values from these two models differ. Therefore, all the two models have statistically satisfactory goodness of the kriging prediction from the sample points.

And then, the best method of variogram modeling to give the good kriging prediction is determined by comparing these modeling methods using the three statistics given by:

$$RMSPE = \sqrt{\frac{1}{N} \sum_{i=1}^N [Z(x_i) - \hat{Z}(x_i)]^2} \quad (20)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N |Z(x_i) - \hat{Z}(x_i)| \quad (21)$$

$$\theta(x) = \frac{\{Z(x) - \hat{Z}(x)\}^2}{\sigma_K^2(x)} \quad (22)$$

Where $Z(x)$ is the observed value at sampling location x , $\hat{Z}(x_i)$ is the predicted value at that location using the estimated variogram model, $\sigma_K^2(x)$ is the corresponding kriging variance and N is the number of points for cross-examination. *MAPE* can express the estimation accuracy generally, and *RMSPE* is the fundamental measurement for comparing the accuracy of different interpolation methods. *RMSPE* is smaller, the interpolation method is better. Error statistics combined with kriging variance is known as the useful statistic for validation of kriging (Fernández-Casal and Francisco-Fernández 2014; Lark 2000). We follow Lark (2000) and assess the variogram models with his statistic $\theta(x)$. If the correct variogram is modeled, $\bar{\theta}$ (the mean of $\theta(x)$) should be 1 since the kriging variances must be consistent with the observed variances (Lark 2000; Marchant and Lark 2007b; Fernández-Casal and Francisco-Fernández 2014). However, the outliers in the cross-validation data will influence $\theta(x)$ irrespective of their effect on the variogram estimate. Since outlier occurred at x will affect the $Z(x)$ term in Eq. (21) or the outlier occurred close to x will affect the $\hat{Z}(x)$ and $\sigma_K^2(x)$ term, Lark (2000) proposes $\tilde{\theta}$ the median of $\theta(x)$ for a more robust measure of the suitability of the estimated variogram model. If the correct variogram model is used in the kriging interpolation at the cross-validation locations, then the expectation of $\tilde{\theta}$ would be 0.455 (Lark 2000).

Table 7 assessment of different variogram models by error statistics in the iron ore deposit data

Model	RMSPE	MAPE	$\tilde{\theta}(x)$
Spherical model	3.6285	2.9096	0.5217
Exponential model	3.7194	3.0618	0.5063
SVR model	3.4879	2.7598	0.4684
MGGP model	3.3854	2.7108	0.4605

Table 8 assessment of different variogram models by error statistics in the Jura data

Model	RMSPE	MAPE	$\tilde{\theta}(x)$
Spherical model	6.4332	5.0686	0.5978
Exponential model	6.3246	5.0000	0.5436
SVR model	6.2257	4.8897	0.3996
MGGP model	6.2919	4.9641	0.4960

The results of error statistics analysis in all cases are shown in Table 6, 7, and 8, where the both of SVR and MGGP methods yielded the better results for three measures than the traditional WLS method. Especially, in the case of iron ore deposit, they yielded the best results.

In all cases, the value of $\tilde{\theta}(x)$ obtained from the WLS method was significantly larger than the expected value for the correct variogram. This indicates that the variance is underestimated. However, the values of $\tilde{\theta}(x)$ obtained from the SVR and MGGP variogram models were much closer to the expected value of 0.455 than those of the traditional WLS models. The accuracy of selected variogram model is known to be more sensitive to $\tilde{\theta}(x)$ than *RMSPE* or *MAPE* (Lark 2000; Marchant and Lark 2007b). This shows an improved performance of SVR and MGGP method against the traditional WLS method.

Compared with the MGGP and SVR model, in the case of Coal mine data and Iron ore deposit data the performance of SVR model is slightly lower than the MGGP model. In the case of Jura data, the value of *RMSPE* and *MAPE* obtained from SVR model is slightly better than MGGP model, whereas the value of $\tilde{\theta}(x)$ was smaller than the expected value for the correct variogram while those of MGGP is closer to one. It indicates that MGGP model is rarely affected than SVR model by the discrete degree and size of the input points for the variogram modeling.

Conclusions

The variogram modeling is the critical stage for the kriging interpolation because it becomes the mouthpiece of the spatial variation in the real field and its exact estimation can affect the interpolation accuracy. From the case studies, we can conclude as following:

Firstly, the performance of SVR-based variogram modeling method is tested for various cases such as the case with/without nugget effect, using relatively large/small number of the input data for modeling of the variogram.

Secondly, MGGP-based method as well as SVR-based method for variogram modeling has the ability to fit more exactly experimental variogram without assuming the basic model shape and reflect more objectively spatial variation of

the real field comparing with traditional WLS method, and improve the kriging interpolation precision significantly.

Thirdly, MGGP method overcomes the defect of the SVR that does not provide the explicit formulation for the variogram model. In addition, MGGP method is more flexible than SVR in the case the discrete degree and size of input data for modeling of variogram are changed.

In a word, the application of MGGP has the potential value in variogram modeling.

The present study is limited to the three set of case studies which have been done previously by other researchers. Future work perhaps would generate the more realizations to evaluate the benefits in implementing AI models such as MGGP and SVR for the modeling variogram and prediction of the ore grade.

Acknowledgments This study was jointly supported by the Research of High-grade Iron Ore Formation Mechanism and Prediction of China (2012CB416800) and the Fundamental Research Funds for the Graduate Innovation State Education Ministry (N110601002). We are very grateful to Dr Isobel Clark and Dr Pierre Goovaerts who made geostatistical data available on the internet. We are thankful to three of anonymous reviewers for their constructive comments. We are also very grateful to Earth Science Informatics editor, Dr. Hassan A. Babaie. Their careful review and editorial comments were very useful for the preparation of a revised version.

References

- Al-Ahmari A (2007) Predictive machinability models for a selected hard material in turning operations. *J Mater Process Technol* 190:305–311. doi:10.1016/j.jmatprotec.2007.02.031
- Atteia O, Dubois JP, Webster R (1994) Geostatistical analysis of contamination in the Swiss Jura. *Environ Pollut* 86:315–327
- Basak S, Dixit U, Davim J (2007) Application of radial basis function neural networks in optimization of hard turning of AISI D2 cold-worked tool steel with a ceramic tool. *Proc Inst Mech Eng B J Eng Manuf* 221:987–998
- Byvatov E, Schneider G (2003) Support vector machine applications in bioinformatics. *Appl Bioinforma* 2:67–77
- Çaydas U, Ekici S (2012) Support vector machines models for surface roughness prediction in CNC turning of AISI 304 austenitic stainless steel. *J Intell Manuf* 23:639–650. doi:10.1007/s10845-010-0415-2
- Çaydaş U, Hacıhalık A (2008) A study on surface roughness in abrasive waterjet machining process using artificial neural networks and regression analysis method. *J Mater Process Technol* 202(1):574–582. doi:10.1016/j.jmatprotec.2007.10.024

- Cevik A, Guzelbey IH (2007) A soft computing based approach for the prediction of ultimate strength of metal plates in compression. *Eng Struct* 29:383–394. doi:10.1016/j.engstruct.2006.05.005
- Cevik A, Sonebi M (2008) Modelling the performance of self-compacting SIFCON of cement slurries using genetic programming technique. *Comput Concr* 5:475–491
- Chen YK, Yang YG, Wu WW (2015) Coal seam thickness prediction based on least squares support vector machines and kriging method. *Electron J Geotech Eng* 20(1):167–176
- Chilès JP, Delfiner P (2012) *Geostatistics: Modeling spatial uncertainty*, 2nd edn. Wiley, New York
- Clark I (1979) *Practical geostatistics*. Applied Science Publishers, Essex
- Clark I (1983) Regression revisited. *Math Geol* 15(4):517–536
- Clark I, Harper WV (2000) *Practical geostatistics 2000*. Greyden Press, Ecosse North America
- Cressie N (1985) Fitting variogram models by weighted least squares. *Math Geol* 17(5):563–586
- Cressie N (1991) *Statistics for spatial data*. Wiley, New York
- Desassis N, Renard D (2013) Automatic variogram modeling by iterative least squares: univariate and multivariate cases. *Math Geosci* 45:453–470. doi:10.1007/s11004-012-9434-1
- Deutsch CV, Journel AG (1998) *GSLIB: geostatistical software library and user's guide*, 2nd edn. Oxford University Press, New York
- Emery X (2010) Iterative algorithms for fitting a linear model of coregionalization. *Comput Geosci* 36:1150–1160. doi:10.1016/j.cageo.2009.10.007
- Fernández-Casal R, Francisco-Fernández M (2014) Nonparametric bias-corrected variogram estimation under non-constant trend. *Stoch Environ Res Risk Assess* 28:1247–1259. doi:10.1007/s00477-013-0817-8
- Gandomi AH, Alavi AH (2011) Multi-stage genetic programming: a new strategy to nonlinear system modeling. *Inf Sci* 181:5227–5239. doi:10.1016/j.ins.2011.07.026
- Gandomi AH, Alavi AM (2012) A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems. *Neural Comput Applic* 21:189–201. doi:10.1007/s00521-011-0735-y
- Gandomi AH, Alavi AH, Mousavi M, Tabatabaei SM (2011) A hybrid computational approach to derive new ground-motion prediction equations. *Eng Appl Artif Intell* 24:717–732. doi:10.1016/j.engappai.2011.01.005
- Garg A, Tai K (2012) Review of genetic programming in modeling of machining processes. In: *Proceedings of the 2012 I.E. International Conference on Modelling, Identification and Control (ICMIC2012)*. Wuhan, China, 24–26 June 2012, pp. 653–658
- Garg A, Tai K (2013) Comparison of statistical and machine learning methods in modeling of data with multicollinearity. *Int J Model Identif Control* 18(4):295–312
- Garg A, Tai K (2014) Stepwise approach for the evolution of generalized genetic programming model in prediction of surface finish of the turning process. *Adv Eng Softw* 78:16–27. doi:10.1016/j.advengsoft.2014.08.005
- Garg A, Bhalerao Y, Tai K (2013a) Review of empirical modeling techniques for modeling of turning process. *Int J Model Identif Control* 20(2):121–129. doi:10.1504/IJMIC.2013.056184
- Garg A, Rachmawati L, Tai K (2013b) Classification-driven model selection approach of genetic programming in modeling of turning process. *Int J Adv Manuf Technol* 69:1137–1151. doi:10.1007/s00170-013-5103-x
- Garg A, Garg A, Tai K, Sreedeeep S (2014a) An integrated SRM-multi-gene genetic programming approach for prediction of factor of safety of 3-D soil nailed slopes. *Eng Appl Artif Intell* 30:30–40. doi:10.1016/j.engappai.2013.12.011
- Garg A, Garg A, Tai K (2014b) A multi-gene genetic programming model for estimating stress-dependent soil water retention curves. *Comput Geosci* 18:45–56. doi:10.1007/s10596-013-9381-z
- Garg A, Tai K, Gupta AK (2014c) A modified multi-gene genetic programming approach for modeling true stress of dynamic strain aging regime of austenitic stainless steel 304. *Meccanica* 49:1193–1209. doi:10.1007/s11012-013-9873-x
- Garg A, Tai K, Lee CH, Savalani MM (2014d) A hybrid M⁵-genetic programming approach for ensuring greater trustworthiness of prediction ability in modeling of FDM process. *J Intell Manuf* 25(6):1349–1366. doi:10.1007/s10845-013-0734-1
- Genton MG (1998) Variogram fitting by generalized least squares using an explicit formula for the covariance structure. *Math Geol* 30(4):323–345
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, New York
- Goovaerts P, Webster R, Dubois JP (1997) Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environ Ecol Stat* 4:31–48
- Gupta AK (2008) Predictive modeling of turning operations using response surface methodology, artificial neural networks and support vector regression. *Int J Prod Res* 48:763–778. doi:10.1080/00207540802452132
- Hadi Y, Ahmed SG (2006) Assessment of surface roughness model for turning process. *Knowledge enterprise: intelligent strategies in product design. Manuf Manag* 207:152–158
- Hearst MA, Dumais S, Osman E, Platt J, Scholkopf B (1998) Support vector machines. *Intell Syst Appl IEEE* 13:18–28
- Hilário M, Manuela SM (2011) Combining robustness with efficiency in the estimation of variogram. *Math Geosci* 43:363–377. doi:10.1007/s11004-010-9308-3
- Huang ZS, Wang HZ, Zhang R (2012) An improved kriging interpolation technique based on SVM and its recovery experiment in oceanic missing data. *Am J Comput Math* 2:56–60. doi:10.4236/ajcm.2012.21007
- Jian X, Olea RA, Yu YS (1996) Semivariogram modeling by weighted least squares. *Comput Geosci* 22:387–397
- Journel AG, Huijbregts CJ (1978) *Mining geostatistics*. Academic, London
- Kecman V (2001) *Learning and soft computing: support vector machines, neural networks, and fuzzy logic models*. MIT Press, Cambridge
- Kerry R, Oliver MA (2007) Sampling requirements for variograms of soil properties computed by the method of moments and residual maximum likelihood. *Geoderma* 140(4):383–396. doi:10.1016/j.geoderma.2007.04.019
- Kitanidis PK (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resour Res* 19(4):909–921
- Kitanidis PK (1997) *Introduction to geostatistics: applications to hydrology*. Cambridge University Press, New York
- Koza JR (1994) *Genetic Programming. II: Automatic Discovery of Reusable Programs*. MIT, Cambridge
- Lahiri S, Lee Y, Cressie N (2002) On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters. *J Stat Plan Inference* 103(1):65–85. doi:10.1016/S0378-3758(01)00198-7
- Lark RM (2000) Estimating variogram of soil properties by the method-of-moments and maximum likelihood. *Eur J Soil Sci* 51:717–728
- Lark RM (2012) Towards soil geostatistics. *Spat Stat* 1:92–99. doi:10.1016/j.spasta.2012.02.001
- Lark RM, Webster R (2006) Geostatistical mapping of geomorphic surfaces in the presence of trend. *Earth Surf Process Landf* 31(7):862–874. doi:10.1002/esp.1296
- Lark RM, Cullis BR, Welham SJ (2006) On optimal prediction of soil properties in the presence of trend: the empirical best linear unbiased predictor (E-BLUP) with REML. *Eur J Soil Sci* 57(6):787–799. doi:10.1111/j.1365-2389.2005.00768.x

- Li SH, Lu WT (2010) Automatic fit of the variogram. Third International Conference on Information and Computing (ICIC) 4:129–132. doi:10.1109/ICIC.2010.303
- Madár J, Abonyi J, Szeifert F (2005) Genetic programming for the identification of nonlinear input–output models. *Ind Eng Chem Res* 44:3178–3186
- Marchant BP, Lark RM (2007a) The Matérn variogram model: implications for uncertainty propagation and sampling in geostatistical surveys. *Geoderma* 140(4):337–345. doi:10.1016/j.geoderma.2007.04.016
- Marchant BP, Lark RM (2007b) Robust estimation of the variogram by residual maximum likelihood. *Geoderma* 140(1):62–72. doi:10.1016/j.geoderma.2007.03.005
- Mardia KV, Marshall RJ (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* 71(1):135–146
- Matheron G (1962) *Traite de geostatistique appliqué, vol I. Memoires du bureau de recherches geologiques et minières, no 14. Editions Technip, Paris*
- Matheron G (1963) Principles of geostatistics. *Econ Geol* 58:1246–1266
- McBratney AB, Webster R (1986) Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. *J Soil Sci* 37:617–639
- Minasny B, McBratney AB (2005) The Matern function as a general model for soil variograms. *Geoderma* 128:192–207
- Oliver MA (2010) Ch. B6: the variogram and kriging. In: Fischer MM, Getis A (eds) *Handbook of applied spatial analysis: software tools, methods and applications*. Springer, Berlin, pp 319–352. doi:10.1007/978-3-642-0364-7_17
- Oliver MA, Webster R (2014) A tutorial guide to geostatistics: computing and modeling variograms and kriging. *Catena* 113:56–59. doi:10.1016/j.catena.2013.09.006
- Pannatier Y (1996) *Variowin: software for spatial analysis in 2D*. Springer, New York
- Pardo-Igúzquiza E, Dowd PA (2001) Variance-covariance matrix of the experimental variogram: assessing variogram uncertainty. *Math Geol* 33:397–419
- Pelckmans K, Suykens JAK, Vangestel T, De Brabanter D, Lukas J, Hamers B et al (2002) LS-SVMlab: a matlab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT, Leuven
- Pyrzcz MJ, Deutsch CV (2006) Semivariogram models based on geometric offsets. *Math Geol* 38(4):475–488. doi:10.1007/s11004-005-9025
- Ricardo AO (2006) A six-step practical approach to semivariogram modeling. *Stoch Environ Res Risk Assess* 20:307–318. doi:10.1007/s00477-005-0026-1
- Salgado DR, Alonso FJ (2007) An approach based on current and sound signals for in-process tool wear monitoring. *Int J Mach Tools Manuf* 47(14):2140–2152. doi:10.1016/j.ijmactools.2007.04.013
- Salgado DR, Alonso FJ, Cambero I, Marcelo A (2009) In-process surface roughness prediction system using cutting vibrations in turning. *Int J Adv Manuf Technol* 43:40–51. doi:10.1007/s00170-008-1698-8
- Saporo A, Tade MO, Vuthaluru H (2012) A modified Kennard-Stone algorithm for optimal division of data for developing artificial neural network models. *Chem Prod Process Model* 7(1):13. doi:10.1515/1934-2659.1645
- Searson DP, Leahy DE, Willis MJ (2010) GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS 2010), Hong Kong, 1, pp. 77–80
- Shi D, Gindy NN (2007) Tool wear predictive model based on least squares support vector machines. *Mech Syst Signal Process* 21(4):1799–1814. doi:10.1016/j.ymsp.2006.07.016
- Stein ML (1988) Asymptotic distributions of minimum norm quadratic estimators of the covariance function. *Ann Stat* 16:55–63
- Suykens JAK, Vandervalle J, Moor BD (2001) Optimal control by Least squares support vector machine. *Neural Netw* 14(1):23–35. doi:10.1016/S0893-6080(00)00077-0
- Vapnik V (1995) *The nature of statistical learning*. Springer, New York
- Wackernagel H (2003) *Multivariate geostatistics, 3rd edn*. Springer, Berlin
- Wang B, Yao M, Zhu R (2011) Gene expression programming with multiple chromosomes. *Int J Model Identif Control* 14:235–241. doi:10.1504/IJMIC.2011.043144
- Webster R, Oliver MA (2007) *Geostatistics for environmental scientists, 2nd edn*. Wiley, Chichester
- Webster R, Atteia O, Dubois JP (1994) Coregionalization of trace metals in the soil in the Swiss Jura. *Eur J Soil Sci* 45:205–218
- Zhang H, Zimmerman DL (2007) Hybrid estimation of semivariogram parameters. *Math Geol* 39(2):247–260. doi:10.1007/s11004-006-9070-8
- Zimmerman DL (1989) Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Math Geol* 21(7):655–672