AN UNCERTAINTY ANALYSIS PROCEDURE FOR SPATIALLY JOINT SIMULATIONS OF MULTIPLE ATTRIBUTES

Shoufan Fang

Guangxing Wang

George Z. Gertner

Alan Anderson

See next page for additional authors

Recommended Citation

Author Information

Shoufan Fang, Guangxing Wang, George Z. Gertner, and Alan Anderson

AN UNCERTAINTY ANALYSIS PROCEDURE FOR SPATIALLY JOINT SIMULATIONS OF MULTIPLE ATTRIBUTES

Shoufan Fang[a], Guangxing Wang[a], George Z. Gertner[a]*, and Alan Anderson[b]

[a]Department of Natural Resources and Environmental Sciences
University of Illinois at Urbana-Champaign
W503 Turner Hall
1102 S. Goodwin Ave.
Urbana, IL 61801
USA

[b]Construction Engineering Research Laboratories
P.O.Box 9005
Champaign, IL 61826-9005
USA

*Corresponding author: gertner@uiuc.edu, Phone: (217) 333-9346; Fax: (217) 244-3219

Abstract

In this study, an uncertainty analysis procedure for joint sequential simulation of multiple attributes of spatially explicit models was developed based on regression analysis. This procedure utilizes information obtained from joint sequential simulation to establish the relationship between model uncertainty and variation of model inputs. Using this procedure, model variance can be partitioned by model input parameters on a pixel by pixel basis. In the partitioning, the correlation of neighboring pixels is accounted for. With traditional uncertainty analysis methods, this is not possible. In a case study, spatial variation of soil erodibility from a joint sequential simulation of soil properties was analyzed. The results showed that the regression approach is a very effective method in the analysis of the relationship between variation of the model and model input parameters. It was also shown for the case study that (1) uncertainty of soil erodibility of a pixel is mainly propagated from its own soil properties, (2) soil properties of neighboring pixels contribute negative uncertainty to soil erodibility,   (3) it is sufficient for uncertainty analysis to include the nearest three neighboring pixel groups, and (4) the largest uncertainty contributors vary by soil properties and location.

Keywords: joint simulation, regression analysis, Revised Universal Soil Loss Equation (RUSLE), soil erodibility, uncertainty analysis.

1. Introduction

Soil loss in the USA is currently predicted using either the Universal Soil Loss Equation (USLE) (Wischmeier and Smith, 1978) or the Revised Universal Soil Loss Equation (RUSLE) (Renard et al., 1997) for the purpose of agricultural, rangeland and environmental management. Both USLE and RUSLE are related to rainfall erosivity factor (R), soil erodibility factor (K), slope length factor (L), slope steepness factor (S), cover management factor (C), and support practice factor (P). These equations consist of empirical model sets derived from an extensive database and their model parameters contain uncertainty when using these equations for a specific area. Moreover, these factors vary over space and time, depend on other variables, and may be correlated with each other. Additionally, errors in sampling, measuring, and modeling will lead to uncertainty in the estimate of these factors. These uncertainty sources should not be neglected, since they will propagate into the prediction of soil loss. Uncertainty might be important, particularly when local estimates are required for management planning. Thus, there is a strong need to develop a general methodology for the spatial assessment of uncertainty for users of such systems.

In recent years we have been working on a large project related to the prediction and uncertainty analysis of soil erosion using the USLE and RUSLE. Our objectives are to develop a general procedure for spatially and temporally predicting soil loss, identifying various errors, modeling their propagation, and generating error budgets in order to provide guidelines for error reduction and management planning. As one part of the overall of uncertainty analysis for soil loss prediction, the study reported here focuses on a method uncertainty assessment of soil erodibility factor K used in the USLE when spatially explicit joint simulation methods are used for modelling the soil erodibility factor K. The methodology presented here will be extended to the full USLE and RUSLE.

Uncertainty analysis is essential to evaluate and control the quality of model simulations. It is the first step to efficiently improve the quality of simulation predictions. There are many uncertainty analysis methods for non-spatial simulation models, such as methods based on Taylor series, Fourier series, Monte Carlo simulation, and computational experiment design (Cox, 1977; Cukier et al., 1973; Dawning et al., 1985; Fang, 2000; Fang and Gertner, 2000; Gertner et al., 1996; Jansen et al., 1994; Parysow and Gertner, 2000; Matsumoto et al. 1994; Sobol, 1993). Functional models are often developed to predict certain resources and properties based on mapped attributes. These mapped attributes are input parameters (inputs) of the functional models. Often the mapped attributes are correlated spatially. Some of the above uncertainty analysis methods have been extended to these spatially explicit functional models (Bachmann and Allgöwer; 2000; Fang et al. 2000; Heuvelink, 1998; Wang et al., 2000a). However, in these extensions, the uncertainty analysis methods were applied under the assumption that mapped attributes were independent in space.

Joint sequential spatial simulations of multiple attributes have been widely used in modelling natural resources, ecological and environmental attributes (Goovaerts, 1997, Mowrer, 1997, Wang et al., 2000b). With this type of spatial simulator, it is possible to simultaneously account for the spatial correlation of individual attributes, as well as for the correlation between attributes. (Gómez-Hernández and Cassiraga, 1994; Goovaerts, 1997). With sequential simulations, when the attributes of interest are correlated, the variability of an attribute of interest

at a pixel comes from spatial variability of neighboring pixels as well as variability between the attributes. The variation (variance and covariance) of the attribute vector for a pixel is determined by auto and cross correlation, and variation of the attribute vectors for neighboring pixels.

When the attributes estimated using sequential joint simulation are used as inputs of a functional model, the uncertainty of model predictions on a pixel basis comes from the variation of inputs at the same pixel, from variation of neighboring pixels, and from the spatial correlation of attributes. When the uncertainty in predictions is assessed, both spatial and non-spatial uncertainty sources should be accounted for.

All existing uncertainty analysis methods are deficient when the spatial correlation and neighboring information is considered in uncertainty analysis. The structure of the functional model and computational cost can make it very difficult to apply the existing methods. When the structure of a functional model does not contain any terms dealing with neighboring information, the computationally-efficient uncertainty analysis methods, such as Taylor series and Fourier series based methods, can not be used to partition spatial uncertainty contributions. Since sequential joint simulation algorithms are usually computationally very intensive, uncertainty analysis methods that are based on Monte Carlo simulations can be computationally too expensive. Those uncertainty methods based on experimental design are impractical since it is very difficult to systematically control sequential joint simulation algorithms to assess uncertainty.

The objective of this study is to develop an uncertainty analysis procedure for spatially explicit functional models whose inputs (attributes) are estimated using joint sequential simulation. As a case study, the procedure was applied to a model used to spatially predict soil erodibility.

2. Brief review of joint sequential simulation

In joint sequential simulation of multiple attributes, an expected vector and estimation variance matrix of the attributes are obtained for each unknown location of a study area. The expected estimate and variance for each attribute are derived from a set of estimates drawn from conditional cumulative density functions (ccdfs). Commonly the density functions are assumed to be multiGaussian. Given this assumption, for each draw the ccdf is updated by an estimate and error variance yielded using a cokriging estimator with the modeled spatial variability of the attribute, sample data and previously simulated values given a neighborhood. Updating the ccdfs is thus critical for the joint sequential simulation.

Suppose that a study area is divided into N nodes of a grid and P attributes are estimated. In this area, a sample is drawn and the sample data set { $z_p(u_p)$, $u_p = 1, 2, ..., n_p$, $p = 1, 2, ..., P$} is obtained, and n_p is the number of sample data for the attribute p. The expectation and variance for the attribute p are m_p and σ_p^2. The simple cokriging estimate $Z_p^{sck}(u)$ and error variance $\sigma_p^{sck}(u)$ at the location u for the attribute p are (Gómez-Hernández, and Journel, 1992, Goovaerts, 1997):

$$Z_p^{sck}(u) = m_p + \sum_{u_1=1}^{n_1(u)} \lambda_{u_1}[z_1(u_1) - m_1] + ... + \sum_{u_p=1}^{n_p(u)} \lambda_{u_p}[z_p(u_p) - m_p]$$

$$+ ... + \sum_{u_P=1}^{n_p(u)} \lambda_{u_P}[z_p(u_p) - m_p] \tag{1}$$

$$\sigma_p^{sck}(u) = \sigma_p - \sum_{u_1=1}^{n_1(u)} \lambda_{u_1} C_{p1}(u_1 - u) - ... - \sum_{u_p=1}^{n_p(u)} \lambda_{u_p} C_{pp}(u_p - u)$$

$$- ... - \sum_{u_P=1}^{n_p(u)} \lambda_{u_P} C_{pp}(u_P - u) \tag{2}$$

where p = 1, 2, ..., P; $n_p(u)$ is the number of sample data and previously simulated values available at the location u for the attribute p given a neighborhood; and the number varies from location to location. λ_{u_p} is the weight of the data at location u_p and $C_{pp'}(u_p - u)$ is the cross covariance between the estimated attribute k and attribute p' (p' = 1, ..., P). Let h be the separation distance of data given a direction, for example, $h = u_p - u$, and $C_{pp'}(0)$ the covariance at the separation distance of zero. Then, $C_{pp'}(h) = C_{pp'}(0) - \gamma_{pp'}(h)$. $\gamma_{pp'}(h)$ is an auto or cross semivariogram measuring spatial variability of attributes p and p'.

The weights λ_{u_p} are the solutions of a linear equation system consisting of the co-variances and cross co-variances. For the solutions, the semivariograms should be permissible covariance functions such as spherical and Gaussian model, and P+1 auto and cross co-variances matrices should be positive semi-definite. A linear model of coregionalization is thus needed. If the data of attributes are not normally distributed, moreover, a normal score transformation is needed for each attribute. Then, the transformed data are used in simulation, and the simulated values are transformed back to the original data form.

In a joint sequential simulation, first define a random path to visit each node of the grid once. At the uth node to be visited, determine the mean and variance of the ccdf for attribute p given the n_p original data and all (u -1) simulated values at the locations previously visited, and from that distribution, draw a value which will become the conditional datum for all subsequent drawings. The simulation is done when all N nodes are visited and provided with simulated values. Repeating the joint sequential simulation process many times with different visiting paths leads to a set of estimates at each location for each attribute. Finally, an expected vector and co-variance matrix for P attributes at each location is calculated.

For more details of the joint sequential simulation algorithm, see relevant sections in Gómez-Hernández, and Journel (1992) and Goovaerts (1997).

3. Uncertainty Analysis Procedure

Based on attribute maps generated using joint sequential simulation, a functional model can be used to make predictions on a pixel basis. Let the functional model be:

$$y = f(Z_1, Z_2, \cdots, Z_p)$$

where y is the prediction and $Z_1, Z_2, \cdots, Z_p$ are the inputs (attributes). Figure 1 illustrates the relationship between attributes and prediction maps. The uncertainty of attributes $Z_1, Z_2, \cdots, Z_p$ determines the uncertainty of y.

When using joint sequential simulation to estimate attributes, the uncertainty of the attributes at one pixel are related to the field observations, spatial variability in terms of auto and cross semivariograms, and information from neighboring pixels. Therefore, in uncertainty partitioning on a pixel basis, besides the uncertainty of the attributes at the concerned pixel (host pixel), the spatial correlation and effect of neighboring pixels should also be considered.

In a joint sequential simulation, usually several hundred replications are used to estimate the attributes at each pixel in a given study area. The mean and variation (variance and covariance) of all attributes and the corresponding model prediction can be estimated. From these replications, a large amount of information about variation of both attributes and model prediction can be extracted from the simulations.

Gertner et al. (1996) and Parysow and Gertner (2000) proposed a regression analysis approach to partition uncertainty of a non-spatial functional model when dealing with large amounts of information on model uncertainty from computational experiments. The dependent variable for the regression model was the variance of the functional model prediction and the independent variables were the variances of the inputs of the functional model. After the coefficients of the regression model were estimated, variance contribution of each functional model input was determined by plugging its variance into the regression model and setting all other terms to be zero.

The regression approach can be extended to investigate spatial uncertainty propagation of a functional model when its inputs are estimated using the joint sequential simulation. Because it is very difficult to design a computational experiment when joint sequential simulation is used, the regression approach of Gertner et al. (1996) and Parysow and Gertner (2000) needs to be modified to meet the properties of spatial uncertainty from the joint sequential simulation.

Joint sequential simulation itself can provide needed data to develop a model for partitioning spatial uncertainty based on the regression approach. A random sample drawn from the variation of the predictions and inputs of the functional model obtained from the joint sequential simulation are similar to data sets obtained from computational experiments of Parysow and Gertner (2000). For spatial uncertainty analysis, the random sample should contain not only the variation at the sampled pixels (host pixels) but also that of their neighboring pixels. Such a random sample can be used in the regression analysis of spatial uncertainty. Correspondingly, the regression model should include the terms for the spatial variability and variation of neighboring pixels. The auto and cross covariance terms reflect the effect of both spatial variability and variation of neighboring pixels. Since auto and cross covariance are dependent on distance, neighboring pixels can be grouped according to the distance to their host pixel. The location and number of each neighbor group are shown in Figure 2.

The regression model used for spatial uncertainty partitioning will be a non-intercept model:

$$\text{var}(y) = \sum_{k=0}^{k_a} \sum_{i=1}^{p} \sum_{j \geq i}^{p} b_{ijk} \, \text{cov}[Z_i(u), Z_j(u+h_k)] + \varepsilon \qquad (4)$$

where var(y) is the variance of the functional model prediction at the host pixel u. Z_i and Z_j are the ith and j^{th} inputs of the spatially explicit functional model; k_a is the number of neighbor groups (based on Figure 2) included in the regression model; P is the number of inputs of the functional model; b's are the coefficients of the regression model; h_k is the distance between the host pixel and the neighbor pixels of the kth group; and when k=0 and h_0=0. The $cov[Z_i(u), Z_j(u+h_k)]$ is the sum of auto or cross covariance of all neighboring pixels of the kth group when h_k >0, otherwise it is the traditional covariance or variance (when i=j). The ε is the error term. In most realistic simulations, many terms in the regression model will not be significant (Eq. 4). For such situations, stepwise regression can be used to select significant terms.

The contribution of a variation term $cov[Z_i(u), Z_j(u+h_k)]$ to the uncertainty of the functional model is denoted as $var(y)_{cov_k(Z_i, Z_j)}$ and it is:

$$var(y)_{cov_k(Z_i, Z_j)} = b_{ijk} \, cov[Z_i(u), Z_j(u+h_k)] \quad, \quad i=1, ..., P; \; j=1, ..., P;$$

$$K=0, 1, ..., k_a$$

Based on an assumption of Fang (2000), $var(y)_{cov_k(Z_i, Z_j)}$ can be decomposed as the contributions of Z_i and Z_j, denoted respectively as $var(y)_{Z_i^k \leftarrow cov_k(Z_i, Z_j)}$ and $var(y)_{Z_j^k \leftarrow cov_k(Z_i, Z_j)}$ The $var(y)_{Z_i^k \leftarrow cov_k(Z_i, Z_j)}$, as well as $var(y)_{Z_j^k \leftarrow cov_k(Z_i, Z_j)}$, can be obtained by:

$$var(y)_{Z_i^k \leftarrow cov_k(Z_i, Z_j)} = w_{ik}(Z_i, Z_j) \cdot var(y)_{cov_k(Z_i, Z_j)}$$

where $w_{ik}(Z_i, Z_j) = \begin{cases} \dfrac{var(y)^{0.5}_{cov_k(Z_i, Z_i)}}{var(y)^{0.5}_{cov_k(Z_i, Z_i)} + var(y)^{0.5}_{cov_k(Z_j, Z_j)}}, & \text{if} \quad var(y)^{0.5}_{cov_k(Z_i, Z_i)} + var(y)^{0.5}_{cov_k(Z_j, Z_j)} \neq 0 \\ \dfrac{1}{2}, & \text{otherwise} \end{cases}$

Therefore, the contribution of $Z_i(u+h_k)$ is [denoted as $var(y)_{Z_i^k}$]:

$$var(y)_{Z_i^k} = \sum_{j=1}^{p} var(y)_{Z_i^k \leftarrow cov_k(Z_i, Z_j)} \tag{5}$$

Thus,

$$var(y) = \sum_{k=0}^{k_a} \sum_{i=1}^{p} var(y)_{Z_i^k} \tag{6}$$

When model inputs are considered as the uncertainty sources, Eq. 6 is used to partition spatial uncertainty of the predictions of the functional model.

4. Case Study

The uncertainty analysis procedure for joint sequential simulation was used to investigate uncertainty propagation in a case study related to spatial prediction of soil erosion. The spatial prediction of soil erodibility was made based on five soil properties (attributes) estimated using

Applied Statistics in Agriculture

joint sequential simulation. Spatial uncertainty partitioning of the soil erodibility is presented below.

Model

In Revised Universal Soil Loss Equation (RUSLE), soil erodibility factor (K factor) is used to reflect the impact of soil properties on soil erosion due to water. The soil properties include permeability (PE), structure (ST), and percent of organic matter (OM), sand (SA), and silt (SI). Permeability and structure were originally continuous and categorical variables, respectively, but they were treated as discrete variables when computing K factor (Wischmeier and Smith, 1978). The other three properties, organic matter, sand and silt, are continuous and their summation needs to be smaller than 100%. The values of soil erodibility factor K were computed from these soil properties according to the following function (Wischmeier and Smith, 1978):

$$K = \frac{1}{100}\{2.1 \cdot 10^{-4} \cdot (12 - OM) \cdot [SI \cdot (SA + SI)]^{1.14} + 2.5 \cdot (PE - 3) + 3.25 \cdot (ST - 2)\} \quad (7)$$

The (English) unit of K factor is (ton·acre·hour)/(hundreds of acre·foot·tonf·inch) and tonf is ton force.

In the case study, the five inputs were jointly estimated using the joint sequential simulation, and soil erodibility values were derived using Eq. 7.

Study Area and Data

The study area of 87,890 hectares is located in east Texas. Its geographical location is between the longitude 97°55'01.1" and 97°30'01.1" W and latitude 31°25'00.6" and 31°00'00.7" N. The soil is generally shallow to moderately deep and clayey, underlain by limestone bedrock. The elevation of this area ranged from 180 m to 375 m above sea level with 90 percent of the area below 260 meters (Tazik et al. 1993). The slopes ranged from 0° to 33° with most the slopes in the range of 1° to 3°. The dominant vegetation includes grass, shrub and tree.

Soil samples were collected from 192 plots over the entire area and analyzed in a laboratory in terms of soil properties. Measured soil properties include all inputs needed for Eq. 7. Descriptive statistics for the sampled soil attributes are listed in Table 1.

Spatial Simulation

The study area was divided into 440 x 440 pixels. The size of each pixel was 100 x 100 m^2. The joint sequential simulation described above was used for the spatial prediction of five soil properties. The estimation of permeability and structure was performed as continuous variables and was converted into integer values in computing the K factor based on their marginal distributions and the generated values. The joint simulations were replicated 500 times. Across the 500 replications, the expected values and covariance matrix of the inputs and predictions using Eq. 7 were obtained at each location.

The omni-directional auto and cross semivariogram models measuring spatial correlation of the five soil properties were derived based on the field observations using spherical variogram models (Table 2). Each of the semivariogram models obtained consisted of two spatial structures with the first range of 30 x 100 meters and the second range of 250 x 100 meters. The semivariogram models had different nugget parameters and sill parameters. The different sill parameters indicated that the spatial variability in structure varied depending on the soil properties (Table 2). Figure 3 shows examples of some of the auto and cross semivariograms.

Spatial distribution maps of the estimated mean and the variance of K factor are displayed in Figure 4. Estimated K factor ranged from 0.095 to 0.449 and its variance ranged from 0.0000 to 0.0099. The estimates of K factor and its variances were very large at the north and west edge of the study area, and both estimates and variances were low in patches in the south, east and northwest areas. In most of the study area, the estimated K factor was about 0.25 and the corresponding variance ranged from 0.005 to 0.006 (see Figure 4).

Sampling Spatial Uncertainty

A systematical random sample was drawn from the results of joint simulation to develop the regression partitioning model. Among the 440 x 440 pixels used in the simulation, 484 pixels were selected from a 22-row by 22-column grid. For each of the selected pixels, neighboring pixels were selected as illustrated in Figure 2. At each host pixel, both the variance of K factor and all variance and covariance terms of the five soil properties were recorded from the 500 joint simulations. For the neighboring pixels, the auto and cross covariance terms of the soil properties were also recorded. Table 3 lists some of the descriptive statistics from the systematic sample.

Uncertainty partitioning

Equation 4 was the general form of the regression model used for partitioning. Its dependent variable was the variance of K factor, and the independent variables were the covariances of soil properties: OM, SA, SI, PE, and ST. Six regression models were specified for comparison of the effects of neighboring groups. Model 1 only considered the host pixel and neighbor group 1 variables. Model 2 included the variables in Model 1 and, in addition, neighbor group 2 variables. This was continued until six neighbor groups were included in the regression model (see Table 4). The stepwise regression was carried out using SAS $^{®}$. The estimated coefficients of the regression models are listed in Table 4. The R-squares of all the regression models were larger than 0.999 and the p-values of all listed coefficients were smaller than 0.024.

Except for var(ST) (variance of soil structure, i.e., covariance of ST and ST), all variance and covariance terms of the soil properties from the host pixels were selected in all six regression models. In Model 3, all variance and covariance terms from the host pixels and most auto and cross covariance terms from the neighbor groups were selected. In Models 4, 5, and 6, some auto or cross covariance terms from the closer neighbor groups were not selected but their corresponding terms from the farther neighbor groups were selected (Table 4).

The influence of those immediate neighboring pixels to the host pixel should first be considered since spatial correlation decreases as distance increases. The most reasonable regression model should be the model that selected most of the terms from the immediate neighbor groups, since there was almost no difference in the quality in terms of goodness of fit of Models 1 to 6. Therefore, Model 3 was used to partition spatial uncertainty of K factor in this case study.

The error budget based on the average of the systematic sample showed that the uncertainty of K factor mainly came from the host pixels. The neighboring pixels had a negative uncertainty contribution (Table 5), implying that the neighboring information reduced uncertainty of the host pixel. At the host pixels, the largest uncertainty contributor was var(SI), which contributed over 48% uncertainty to the K factor. However, SI was the second largest contributor of the total uncertainty due to the large negative contribution from its covariance with other soil properties (Table 6). ST contributed the smallest proportion (about 13%) of total uncertainty. In neighbor groups, SI and ST were the largest negative uncertainty contributors.

The product of the estimated mean and variance of K factor was used to select "hot" and "cold" spots for uncertainty analysis. The pixels with the largest (hot spot) and smallest (cold spot) product of mean and variance were selected. Spatial uncertainty partitioning showed that the uncertainty at both spots came from their corresponding host pixels (Figure 5). Due to different cross covariance, the contributions from the neighboring pixels were different in sign (positive or negative) and amount of uncertainty even when they belonged to the same neighbor group. All neighboring pixels at the cold spot had a negative contribution, while about half of the neighboring pixels at the hot spot had positive contribution (Figure 5). The spatial uncertainty contributions of the host pixels at the interested spots are listed in Table 7.

A transect was selected in the study area to assess the spatial uncertainty contribution across pixels. All host pixels were positive uncertainty contributors and neighboring pixels were mainly negative uncertainty contributors along the selected transect (Figure 6). Negative uncertainties from the neighboring pixels ranged from 5% to 10% of the positive uncertainties from the host pixels. The closer neighbor groups produced a larger amount of absolute uncertainty (Figure 6).

Figures 7 and 8 illustrated the contributions of the five soil properties at the host pixels and neighbor groups along the transect line. The largest contributors varied by soil properties and location.

5. Discussion and Conclusion

In this study, a joint sequential simulation was integrated with an uncertainty analysis procedure to partition uncertainty from spatially estimated multiple attributes. The cross-spatial variability among the attributes was introduced into the joint simulation, which should be the basis of spatial uncertainty analysis. We did not show the details of the results from the joint sequential simulation, since the focus of this paper was on building spatial uncertainty budgets. However, the joint simulation reproduced well the joint spatial statistics of the attributes.

The results of the joint sequential simulation for several soil properties in this case study showed that both the variances of a spatially explicit model and the variation of the model inputs are close to being symmetric and approximately normally distributed. The assumption of normality can thus be acceptable for the distribution of variation of spatial simulation. The approximately normal distribution makes it possible to analyze the relationship between variation of models and that of model inputs using regression analysis.

Both the final regression models and spatial uncertainty partitioning showed that the first three neighbor groups are sufficient for an initial regression model. The final regression model obtained can explain the uncertainty propagation of the spatially explicit model from its inputs. The coefficients express the sensitivity of the corresponding independent variables of the regression model. Since there is no intercept, the uncertainty propagated from the variation of an input to the model is the product of the variance (or covariance or cross covariance) of the input and the corresponding coefficient.

The integration of the joint sequential simulation with the uncertainty analysis procedure in this study has made it possible to take into account the spatial correlation of multiple attributes and the effect of neighborhood when partitioning uncertainty. Most of the uncertainty of a model comes from the variation of the model inputs at the concerned (host) pixel. The spatial

correlation among model inputs may contribute positive or negative uncertainty. Discarding the spatial correlation among model inputs may result in large bias in predicting model variance. On the other hand, the neighbors of a host pixel usually contribute negative uncertainty through cross correlation, indicating a reduction in total uncertainty of the host pixel, although the uncertainty contribution from neighboring pixels occasionally is positive. This implies that neglecting the cross-spatial correlation in regression analysis may lead to overestimating the uncertainty contribution of model inputs for most pixels of a study area. The uncertainty contribution of neighboring pixels could be totally different even when they have the same distance to a host pixel. The largest and smallest uncertainty contributors vary by soil properties and location.

Generally, uncertainty of a spatially explicit model comes from the distribution of model inputs, sampling, measurements, data processing, and spatial simulation. In this study it was assumed that there were no errors in sampling, measurements, and data processing. The only uncertainty sources of a spatially explicit model considered here were the distributions of its inputs and the simulation process.

6. Acknowledgment

7. References

Bachmann, A. and B. Allgöwer, 2000. Error propagation in wildfire behavior modelling. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (eds. G.B.M. Heuvelink and M.J.P.M. Lemmens). Amsterdam, July 2000. Delft University Press.

Cox, N. D., 1977. Comparison of two uncertainty analysis methods. Nuclear Science and Engineering, 64:258-265.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly, 1973. Study of the Sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory. Journal of Chemical Physics 59: 3873-3878.

Downing, D. J., R. H. Gardner, and F. O. Hoffman, 1985. An Examination of response-surface methodologies for uncertainty analysis in assessment models. Technometrics, 27(2):151-163.

Fang, S., 2000. Uncertainty analysis of biological nonlinear models based on Bayesian Estimation. Ph.D. Dissertation, University of Illinois at Urbana-Champaign, U.S.A.

Fang, S. and G. Z. Gertner, 2000. Uncertainty analysis of a pipe model based on correlated distributions. Proceedings of Twelfth Annual Kansas State University Conference on Applied Statistics in Agriculture (Edtor G. A. Milliken). Manhattan, Kansas. April 30 to May 2, 2000.

Fang, S., S. Wente, G. Z. Gertner, G, Wang, and A. Anderson, 2000. Uncertainty analysis of predicted disturbance from off-road vehicular traffic in complex landscapes. Environmental Management (in review).

Gertner, G. Z., P. Parysow and B. Guan, 1996. Projection variance partitioning of a conceptual forest growth model with orthogonal polynomials. Forest Science 42: 474-486.

Gómez-Hernández, J.J. and A.G. Journel, 1992. Joint sequential simulation of multiGaussian fields. In Geostatistics Tróia 1992 (editor A. Soars). Vol(1):85-94. Kluwer Academic Publishers, Dordrecht.

Gómez-Hernández, J.J. and E. Cassiraga, 1994. Theory and practice of sequential simulation. In Geostatistical simulation (eds. M. Armstrong and P.A. Dowd). Kluwer Academic Publishers.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press.

Heuvelink, G.B.M., 1998. Error propagation in environmental modelling. Taylor & Francis.

Jansen, M. J. W., W. A. H. Rossing, and R. A. Daamen, 1994. Monte Carlo estimation of uncertainty contributions from several independent multivariate sources. In Proceedings Congress Predictibility and Nonlinear Modelling in Natural Sciences and Economics, 5-7 April 1993, The Netherlands

Matsumoto, K., T. Takeda, and T. Masuda , 1994. Approximate calculation method for 2nd order S. C.. Journal of Nuclear Science and Technology 31 (11): 151-1159.

Mowrer, H.T., 1997. Propagating uncertainty through spatial estimation processes for old-growth subalpine forests using sequential Gaussian simulation in GIS. Ecological Modelling. 98:73-86.

Myers, J.C., 1997. Geostatistical Error Management. Van Nostrand Reinhold, New York, NY.

Parysow, P., G.Z. Gertner, and J. Westervelt, 2000. Efficient approximation for building error budgets for large and computationally-intensive process models. Ecological Modeling. 135:111-125.

Pebesma, E.J. and G.B.M. Heuvelink, 2000. Error propagation in wildfire behavior modelling. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (eds. G.B.M. Heuvelink and M.J.P.M. Lemmens). Amsterdam, July 2000. Delft University Press.

Sobol, I. M., 1993. Sensitivity estimates for nonlinear mathematical models. Mathematical Modeling and Computational Experiments Vol.1(4):407-414.

Tazik, D.J, S.D. Warren, V.E. Diersing, R.B. Shaw, R.J. Brozka, C.F. Bagley, and W.R. Whitworth. U.S. Army Land Condition-Trend Analysis (LCTA) Plot Inventory Field Methods, Technical Report N-92/03/ADA247931 (USACERL, February 1992).

Wang, G., S. Fang, G.Z. Gertner, and A.B. Anderson, 2000a. Uncertainty propagation and partitioning in spatial prediction of topographical factors of RUSLE. Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences (eds. G.B.M. Heuvelink and M.J.P.M. Lemmens). Amsterdam, July 2000. Delft University Press.

Wang, G., G. Gertner, P. Parysow, and A.B. Anderson, 2000b. Spatial prediction and uncertainty analysis of topographical factors for the Revised Universal Soil Loss Equation (RUSLE). J. Soil and Water Conservation. Third Quarter 374-384.

Wischmeier, W.H. and D.D. Smith, 1978. Predicting rainfall erosion losses: A guide to conservation planning. U.S. Dep. Agric. Handbook Number. 537.

Table 1. Description of the measured soil properties. OM, SA, SI, PE, and ST represent organic matter, sand, silt, permeability, and structure, respectively.

	OM (%)	SA (%)	SI (%)	PE	ST
Mean	6.44	13.80	52.48	4.35	3.43
Standard deviation	4.48	11.30	7.12	1.12	0.66
Minimum	0.3	1.0	30.0	2	2
Maximum	45.1	54.0	70.0	6	4

Table 2. Estimated parameters of auto and cross semivariogram models based on observations of soil properties: percentage of organic matter (OM), sand (SA), silt (SI), class of permeability (PE), and structure (ST). Unit of ranges: 100 meter.

Soil property		Nugget	Sill(1)	Range(1)	Sill(2)	Range(2)
OM	OM	6.154129	10.11377	30	0.794887	250
OM	SA	-7.60947	0.438051	30	-5.56359	250
OM	SI	6.740964	-0.60269	30	-1.4687	250
OM	PE	-1.73727	-0.02959	30	-0.22453	250
OM	ST	-0.84803	-0.2821	30	-0.2474	250
SA	SA	84.64614	3.837834	30	39.02297	250
SA	SI	-37.2626	5.137836	30	10.2965	250
SA	PE	0.579084	-0.42573	30	1.557183	250
SA	ST	0.411975	-0.3042	30	1.795053	250
SI	SI	37.85565	7.018695	30	2.717107	250
SI	PE	-0.31864	-0.57219	30	0.411934	250
SI	ST	-0.11901	-0.37802	30	0.470051	250
PE	PE	0.856336	0.047263	30	0.065929	250
PE	ST	0.309037	0.033279	30	0.058792	250
ST	ST	0.211705	0.030194	30	0.126057	250

Table 3. Descriptive statistics of sampled uncertainty from spatial simulation.

Variation*	Mean	SD**	Skewness	Variation	Mean	SD	Skewness
Var(K)	0.0057	0.001	-0.816	Var(OM)2	-22.776	0.891	-0.718
Var(OM)	1.175	0.248	-0.624	Cov(OM,SA)2	21.825	1.625	0.489
Cov(OM,SA)	-2.115	0.602	0.063	Cov(OM,SI)2	-22.156	1.186	1.172
Cov(OM,SI)	1.136	0.395	0.351	Cov(OM,PE)2	5.427	0.256	0.136
Cov(OM,PE)	-0.383	0.076	0.144	Cov(OM,ST)2	2.618	0.141	0.744
Cov(OM,ST)	-0.214	0.041	0.527	Var(SA)2	39.437	20.095	-0.988
Var(SA)	94.872	8.358	-3.229	Cov(SA,SI)2	59.322	7.709	-0.362
Cov(SA,SI)	-21.832	3.844	0.293	Cov(SA,PE)2	2.845	0.906	0.025
Cov(SA,PE)	1.274	0.455	0.146	Cov(SA,ST)2	3.156	0.616	-0.088
Cov(SA,ST)	1.176	0.285	-0.203	Var(SI)2	27.970	6.518	-0.624
Var(SI)	45.302	3.648	-3.855	Cov(SI,PE)2	-0.549	0.593	0.184
Cov(SI,PE)	-0.492	0.297	0.145	Cov(SI,ST)2	0.203	0.373	-0.134

Contined next page

(Table 3 continued)

Cov(SI,ST)	-0.087	0.182	-0.041	Var(PE)2	0.425	0.119	-1.655
Var(PE)	0.965	0.071	-5.106	Cov(PE,ST)2	0.058	0.063	-0.509
Cov(PE,ST)	0.325	0.031	-2.281	Var(ST)2	0.548	0.060	-1.133
Var(ST)	0.351	0.028	-4.121	Var(OM)3	-23.941	0.883	-0.613
Var(OM)1	-21.906	0.888	-0.728	Cov(OM,SA)3	21.769	1.603	0.528
Cov(OM,SA)1	21.818	1.685	0.190	Cov(OM,SI)3	-22.068	1.138	1.027
Cov(OM,SI)1	-22.149	1.142	1.083	Cov(OM,PE)3	5.423	0.255	-0.099
Cov(OM,PE)1	5.413	0.256	0.198	Cov(OM,ST)3	2.647	0.138	0.593
Cov(OM,ST)1	2.590	0.140	0.662	Var(SA)3	38.542	20.434	-1.255
Var(SA)1	41.035	21.304	-1.010	Cov(SA,SI)3	59.012	7.684	-0.347
Cov(SA,SI)1	59.787	8.273	-0.509	Cov(SA,PE)3	2.891	0.902	-0.065
Cov(SA,PE)1	2.832	0.931	0.085	Cov(SA,ST)3	3.155	0.590	-0.485
Cov(SA,ST)1	3.151	0.649	-0.171	Var(SI)3	26.788	6.938	-1.306
Var(SI)1	28.670	6.350	-1.626	Cov(SI,PE)3	-0.481	0.605	-0.212
Cov(SI,PE)1	-0.553	0.631	-0.062	Cov(SI,ST)3	0.247	0.384	0.021
Cov(SI,ST)1	0.141	0.383	0.125	Var(PE)3	0.412	0.132	-1.931
Var(PE)1	0.428	0.120	-1.139	Cov(PE,ST)3	0.053	0.068	-1.038
Cov(PE,ST)1	0.064	0.065	-0.400	Var(ST)3	0.545	0.064	-1.837
Var(ST)1	0.551	0.063	-1.624				
Var(OM)4	-48.838	1.722	-0.741	Cov(SA,ST)5	3.193	0.567	-0.023
Cov(OM,SA)4	43.421	2.889	0.948	Var(SI)5	26.137	5.936	0.144
Cov(OM,SI)4	-44.141	1.947	1.082	Cov(SI,PE)5	-0.333	0.575	-0.059
Cov(OM,PE)4	10.850	0.474	0.354	Cov(SI,ST)5	0.298	0.345	0.073
Cov(OM,ST)4	5.323	0.263	0.894	Var(PE)5	0.415	0.110	0.115
Var(SA)4	78.234	36.440	-1.151	Cov(PE,ST)5	0.053	0.062	0.195
Cov(SA,SI)4	116.803	11.454	-0.316	Var(ST)5	0.540	0.056	-0.530
Cov(SA,PE)4	5.938	1.203	-0.045	Var(OM)6	-25.942	0.875	-0.552
Cov(SA,ST)4	6.435	0.901	-0.399	Cov(OM,SA)6	21.745	1.651	0.251
Var(SI)4	52.964	11.182	-2.300	Cov(OM,SI)6	-22.003	1.119	1.111
Cov(SI,PE)4	-0.948	0.833	0.126	Cov(OM,PE)6	5.430	0.253	-0.078
Cov(SI,ST)4	0.484	0.532	0.038	Cov(OM,ST)6	2.707	0.136	0.533
Var(PE)4	0.819	0.189	-3.018	Var(SA)6	38.207	21.186	-1.250
Cov(PE,ST)4	0.108	0.107	-1.214	Cov(SA,SI)6	57.468	7.964	-0.440
Var(ST)4	1.086	0.113	-2.036	Cov(SA,PE)6	2.986	0.937	-0.241
Var(OM)5	-25.608	0.858	-0.682	Cov(SA,ST)6	3.213	0.639	-0.359
Cov(OM,SA)5	21.722	1.574	0.493	Var(SI)6	25.686	7.196	-2.082
Cov(OM,SI)5	-21.979	1.065	1.007	Cov(SI,PE)6	-0.322	0.623	0.104
Cov(OM,PE)5	5.431	0.250	-0.039	Cov(SI,ST)6	0.320	0.362	0.217
Cov(OM,ST)5	2.695	0.136	0.617	Var(PE)6	0.397	0.133	-2.379
Var(SA)5	38.317	18.721	-0.723	Cov(PE,ST)6	0.043	0.068	-1.116
Cov(SA,SI)5	57.394	7.693	-0.279	Var(ST)6	0.534	0.063	-2.369
Cov(SA,PE)5	2.909	0.812	0.079				

*The numbers following variance or covariance terms indicate that they are auto or cross covariance of the order of neighbor groups. **SD represents standard deviation.

Table 4. Estimated coefficients of the regression models in which different neighbor groups were considered. When a term in a model has a coefficient with the symbol "~", it indicates that this term was not selected in regression analysis, and the symbol "NA" indicates that this term was not considered in the model.

Model	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Candidate terms in model	Host pixel in addition N.G.1	Model 1 in addition N.G.2	Model 2 in addition N.G.3	Model 3 in addition N.G.4	Model 4 in addition N.G.5	Model 5 in addition N.G.6
Selected Term*	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
Var(OM)	0.001020	0.000995	0.000977	0.000994	0.001010	0.000986
Cov(OM,SA)	-0.000204	-0.000200	-0.000209	-0.000205	-0.000204	-0.000211
Cov(OM,SI)	-0.000518	-0.000504	-0.000507	-0.000512	-0.000516	-0.000520
Cov(OM,PE)	-0.001480	-0.001500	-0.001480	-0.001410	-0.001400	-0.001440
Cov(OM,ST)	-0.001280	-0.001210	-0.001210	-0.001360	-0.001330	-0.001330
Var(SA)	0.000015	0.000015	0.000014	0.000014	0.000014	0.000014
Cov(SA,SI)	0.000051	0.000051	0.000053	0.000053	0.000052	0.000052
Cov(SA,PE)	0.000140	0.000156	0.000153	0.000149	0.000151	0.000151
Cov(SA,ST)	0.000215	0.000180	0.000194	0.000205	0.000210	0.000205
Var(SI)	0.000059	0.000058	0.000061	0.000063	0.000063	0.000063
Cov(SI,PE)	0.000406	0.000410	0.000401	0.000406	0.000406	0.000411
Cov(SI,ST)	0.000450	0.000438	0.000420	0.000427	0.000426	0.000421
Var(PE)	0.000568	0.000608	0.000547	0.000518	0.000524	0.000530
Cov(PE,ST)	0.002090	0.001540	0.001750	0.002390	0.002420	0.002280
Var(ST)	~	0.001110	0.001000	~	~	~
Cov(SA)1	0.000004	0.000004	0.000003	0.000002	0.000002	0.000002
Cov(SA,SI)1	-0.000003	~	-0.000003	-0.000003	-0.000002	-0.000003
Cov(SI)1	-0.000007	-0.000006	-0.000006	-0.000006	-0.000006	-0.000005
Cov(ST)1	-0.000672	-0.000542	-0.000415	~	~	~
Cov(SI,ST)2	NA	~	-0.000044	-0.000048	-0.000047	-0.000053
Cov(ST)2	NA	-0.000718	-0.000517	-0.000335	-0.000332	~
Cov(SA)3	NA	NA	0.000002	0.000002	0.000002	0.000002
Cov(PE,ST)3	NA	NA	-0.000362	-0.000326	-0.000345	-0.000324
Cov(ST)4	NA	NA	NA	-0.000270	-0.000207	-0.000236
Cov(SA)5	NA	NA	NA	NA	~	0.000002
Cov(SA,SI)5	NA	NA	NA	NA	-0.000003	~
Cov(SA,ST)5	NA	NA	NA	NA	~	-0.000032
Cov(SA,SI)6	NA	NA	NA	NA	NA	-0.000002
R-square	0.9994	0.9995	0.9995	0.9995	0.9995	0.9995
Max{P-value}	0.0008	0.0109	0.0164	0.0170	0.0232	0.0135

*The number k (k=1, ..., 6) behind a term indicates that this term was from neighbor group k.

Table 5. Uncertainty contribution of variation of inputs of K factor model based on the average of the systematically sampled spatial uncertainty (in Table 3).

Variation*	Host pixel Contribution	N.G. 1** Contribution	N.G.2 Contribution	N.G.3 Contribution
Var(OM)	0.001148	~	~	~
Cov(OM,SA)	0.000443	~	~	~
Cov(OM,SI)	-0.000576	~	~	~
Cov(OM,PE)	0.000567	~	~	~
Cov(OM,ST)	0.000259	~	~	~
Var(SA)	0.001312	0.000104	~	0.000084
Cov(SA,SI)	-0.001166	-0.000149	~	~
Cov(SA,PE)	0.000194	~	~	~
Cov(SA,ST)	0.000228	~	~	~
Var(SI)	0.002743	-0.000173	~	~
Cov(SI,PE)	-0.000197	~	~	~
Cov(SI,ST)	-0.000037	~	-0.000009	~
Var(PE)	0.000528	~	~	~
Cov(PE,ST)	0.000569	~	~	-0.000019
Var(ST)	0.000351	-0.000229	-0.000283	~
Subtotal	0.006366	-0.000447	-0.000292	0.000065
	111.84%	-7.86%	-5.13%	1.15%
Total var(K)= 0.005691				

*Variation terms are traditional variance/covariance and auto/cross covariance in host and neighboring pixels, respectively. **N.G. represents neighbor pixel group. The numbers follow N.G. are the orders of neighbor pixel groups.

Table 6. Spatial error budget of the average pixel based on the sampled spatial uncertainty.

Soil property		Host pixel Contribution	N.G. 1 Contribution	N.G. 2 Contribution	N.G. 3 Contribution
OM	Var(OM)	20.17%	~	~	~
	Cov(OM,X)	8.80%	~	~	~
	Subtotal	28.97%	~	~	~
SA	Var(SA)	23.05%	1.82%	~	1.48%
	Cov(SA,X)	0.37%	-1.15%	~	0.00%
	Subtotal	23.42%	0.68%	~	1.48%
SI	Var(SI)	48.20%	-3.04%	~	~
	Cov(SI,X)	-21.14%	-1.48%	~	~
	Subtotal	27.06%	-4.52%	~	~
P	Var(PE)	9.28%	~	~	0.00%
	Cov(PE,X)	9.80%	~	~	-0.34%
	Subtotal	19.08%	~	~	-0.34%
ST	Var(ST)	6.17%	-4.02%	-4.98%	0.00%
	Cov(ST,X)	7.15%	0.00%	-0.15%	-0.34%
	Subtotal	13.31%	-4.02%	-5.13%	-0.34%

Table 7. Uncertainty contribution of the selected spots of the study area from their corresponding host pixels.

| Soil property		Cold Spot		Hot Spot	
OM	Var(OM)	0.000688	25.15%	0.002391	30.35%
	Cov(OM,X)	0.00028	10.23%	0.000924	11.72%
	Subtotal	0.000967	35.38%	0.003314	42.07%
SA	Var(SA)	0.000372	13.61%	0.001492	18.94%
	Cov(SA,X)	-0.000096	-3.49%	0.000127	1.61%
	Subtotal	0.000276	10.11%	0.001619	20.55%
SI	Var(SI)	0.002354	86.11%	0.002459	31.22%
	Cov(SI,X)	-0.001144	-41.86%	-0.001511	-19.18%
	Subtotal	0.001210	44.25%	0.000948	12.03%
P	Var(PE)	0.000481	17.60%	0.000497	6.30%
	Cov(PE,X)	0.000465	17.02%	0.000635	8.06%
	Subtotal	0.000946	34.62%	0.001132	14.36%
ST	Var(ST)	0.000368	13.44%	0.000269	3.41%
	Cov(ST,X)	0.000362	13.23%	0.000403	5.11%
	Subtotal	0.000729	26.67%	0.000671	8.52%
Total		0.002734	151.03%	0.007878	97.54%
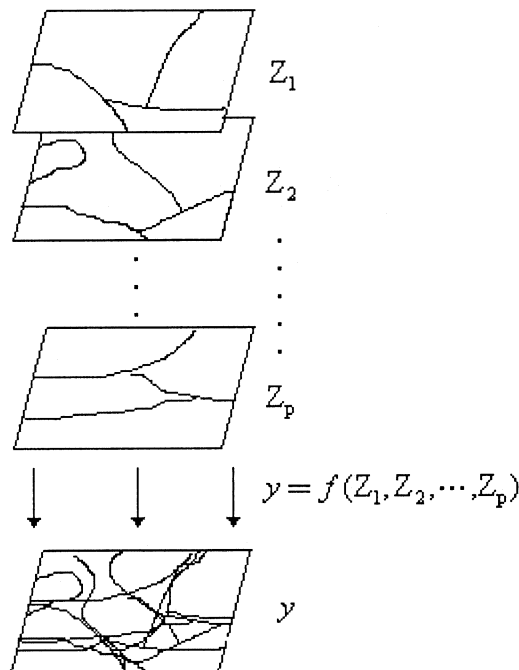
Figure 1. Relationship between attributes (information levels) $(Z_1, Z_2, \cdots, Z_p)$ and model prediction (y).

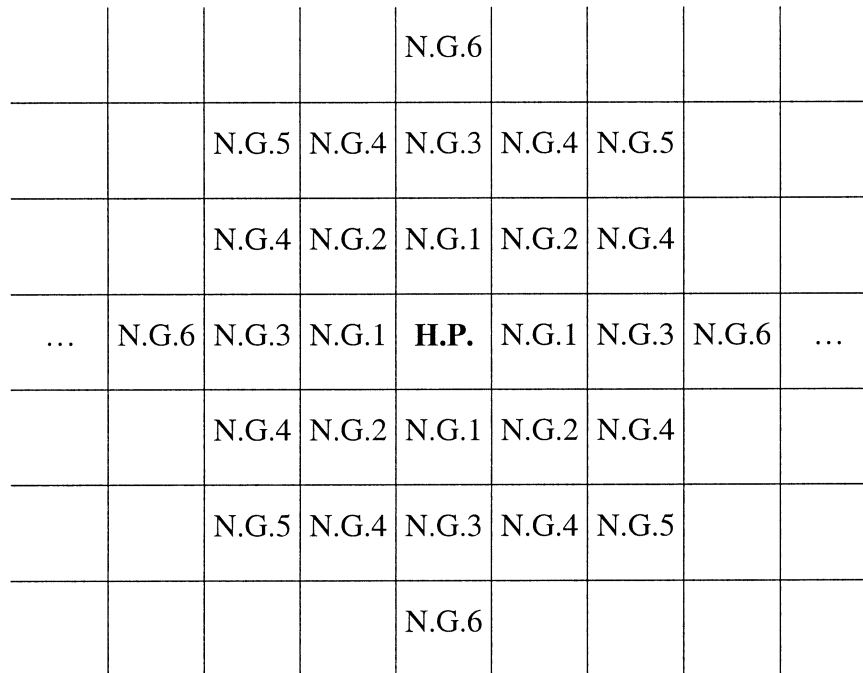|			N.G.6			
	N.G.5	N.G.4	N.G.3	N.G.4	N.G.5	
	N.G.4	N.G.2	N.G.1	N.G.2	N.G.4	
... N.G.6	N.G.3	N.G.1	**H.P.**	N.G.1	N.G.3	N.G.6 ...
	N.G.4	N.G.2	N.G.1	N.G.2	N.G.4	
	N.G.5	N.G.4	N.G.3	N.G.4	N.G.5	
			N.G.6			

Figure 2. Location of pixels of neighbor groups (N.G.'s) of a concerned pixel (host pixel, H.P.). The host pixel is in the center.
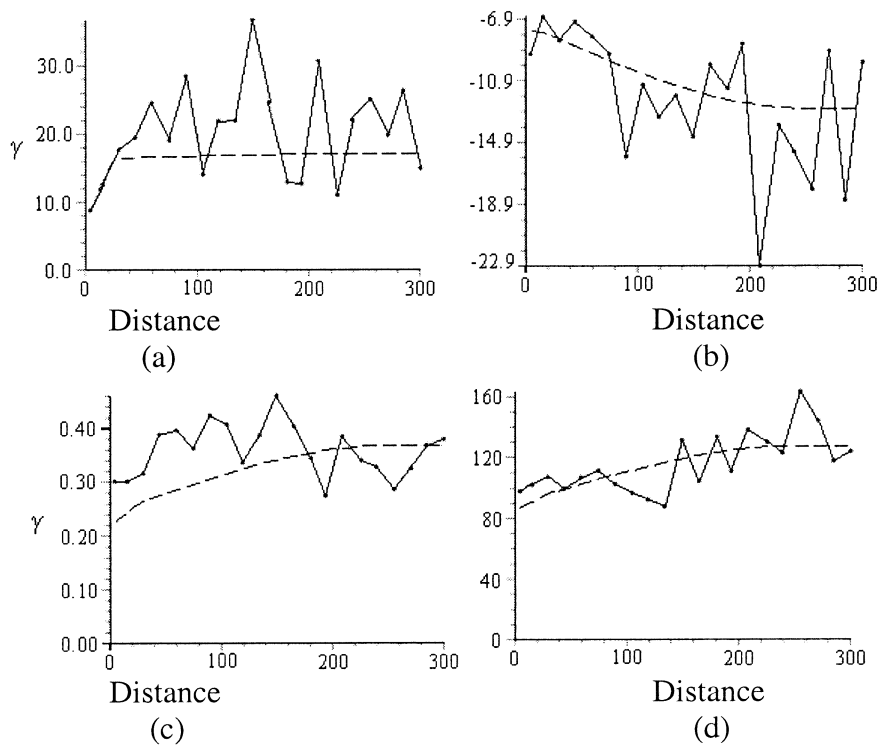

Figure 3. Auto and cross semivariogram models of (a) OM, (b) OM and SA, (c) ST, and (d) SA. Unit of distance: 100 meter.
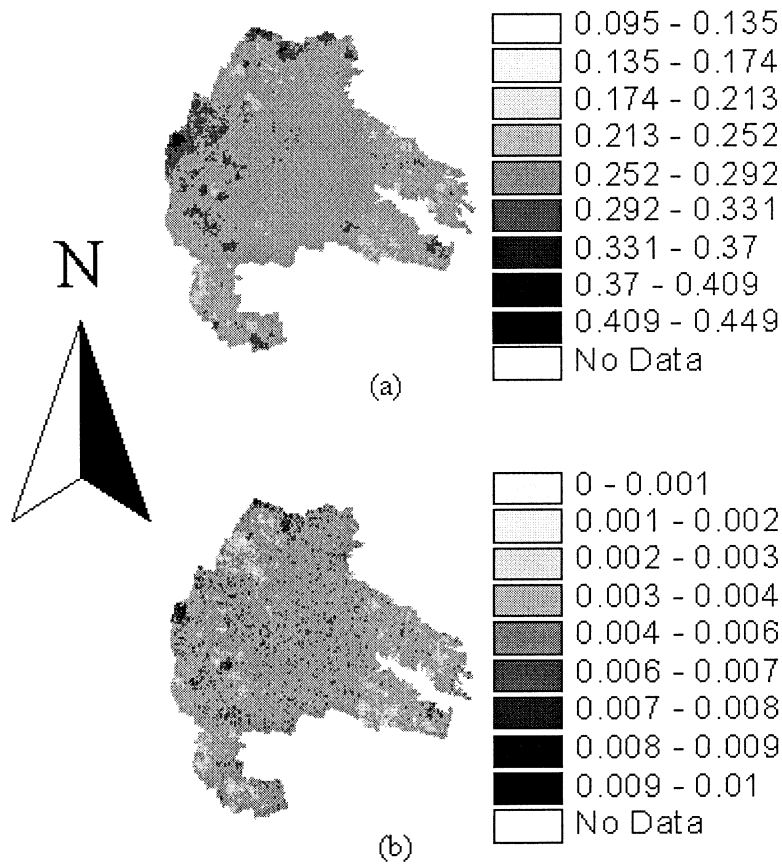
(a)

	0.095 - 0.135
	0.135 - 0.174
	0.174 - 0.213
	0.213 - 0.252
	0.252 - 0.292
	0.292 - 0.331
	0.331 - 0.37
	0.37 - 0.409
	0.409 - 0.449
	No Data

(b)

	0 - 0.001
	0.001 - 0.002
	0.002 - 0.003
	0.003 - 0.004
	0.004 - 0.006
	0.006 - 0.007
	0.007 - 0.008
	0.008 - 0.009
	0.009 - 0.01
	No Data

Figure 4. Spatial distribution of (a) estimated K factor [unit: (ton·acre·hour)/(hundreds of acre·foot·tonf·inch)] and (b) variance of K factor obtained using joint sequential simulation.

(a)   (b)
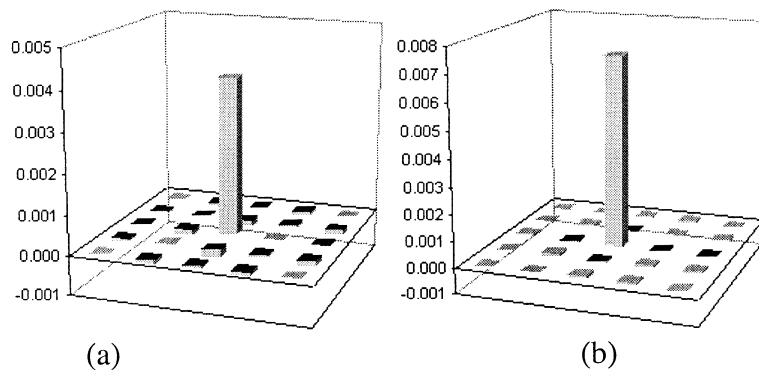
Figure 5. Spatial variance contribution to K factor at the cold spot (a) and the hot spot (b). The center bar in each figure is the host pixel, while the neighboring bars are the neighboring pixels contributions. A black cap on a bar indicates a negative uncertainty contribution to the K factor of the host pixel.
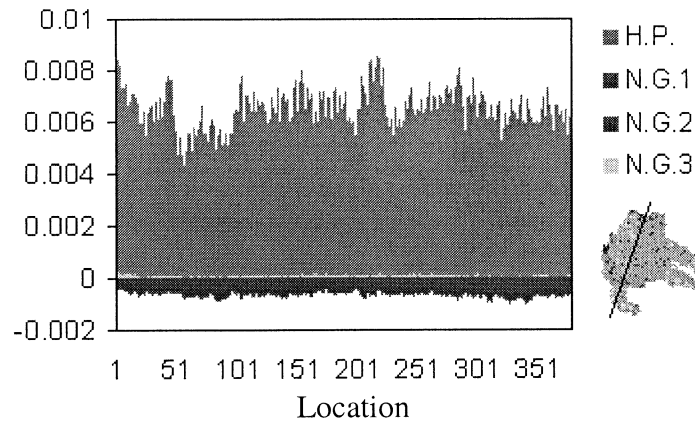
Figure 6. Spatial uncertainty propagation of the selected transect. H.P. represents host pixels. N.G. represents neighbor group. The numbers following N.G. are the orders of neighbor groups. The small image under the legend shows the location of the transect in the study area.
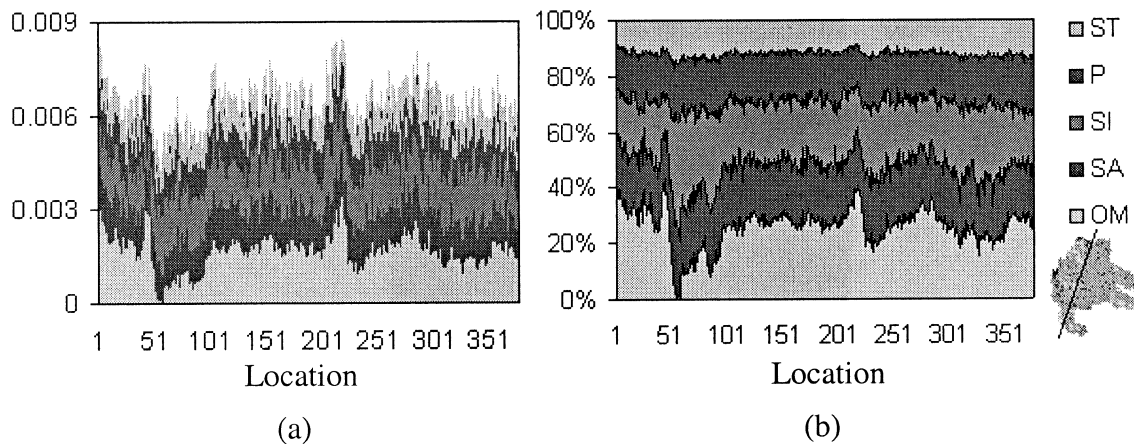

(a)   (b)

Figure 7. Uncertainty of K factor from soil properties at the host pixels along the selected transect. (a) and (b) are actual and percentage uncertainty contributions, respectively.
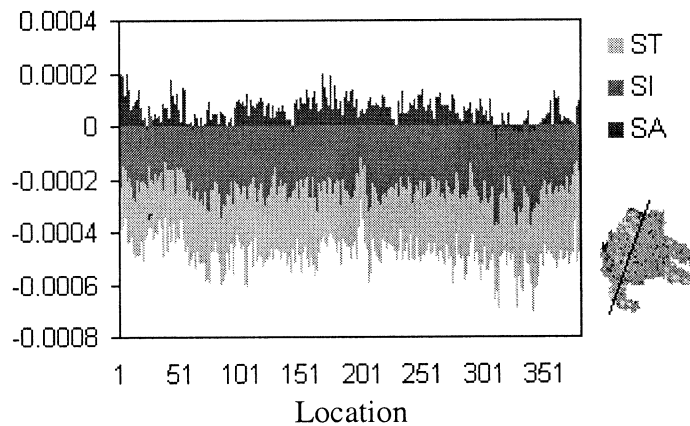

Figure 8. Uncertainty contributed from soil properties at the pixels of neighbor group 1.