Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 185 (2005) 255–269
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

Latin hypercube sampling and geostatistical modeling of spatial uncertainty in a spatially explicit forest landscape model simulation

Chonggang Xu[a,c,*], Hong S. He[b], Yuanman Hu[a], Yu Chang[a], Xiuzhen Li[a], Rencang Bu[a]

[a] *Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang 110016, China*
[b] *School of Natural Resources, University of Missouri at Columbia, Columbia, MO, USA*
[c] *Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign, IL, USA*

Received 28 May 2004; received in revised form 27 November 2004; accepted 13 December 2004

Abstract

Geostatistical stochastic simulation is always combined with Monte Carlo method to quantify the uncertainty in spatial model simulations. However, due to the relatively long running time of spatially explicit forest models as a result of their complexity, it is always infeasible to generate hundreds or thousands of Monte Carlo simulations. Thus, it is of great importance to generate a relatively small set of conditional realizations capturing most of the spatial variability. In this study, we introduced an effective sampling method (Latin hypercube sampling) into a stochastic simulation algorithm (LU decomposition simulation). Latin hypercube sampling is first compared with a common sampling procedure (simple random sampling) in LU decomposition simulation. Then it is applied to the investigation of uncertainty in the simulation results of a spatially explicit forest model, LANDIS. Results showed that Latin hypercube sampling can capture more variability in the sample space than simple random sampling especially when the number of simulations is small. Application results showed that LANDIS simulation results at the landscape level (species percent area and their spatial pattern measured by an aggregation index) is not sensitive to the uncertainty in species age cohort information at the cell level produced by geostatistical stochastic simulation algorithms. This suggests that LANDIS can be used to predict the forest landscape change at broad spatial and temporal scales even if exhaustive species age cohort information at each cell is not available.

Keywords: Latin hypercube sampling; Uncertainty analysis; Geostatistical stochastic simulation; LU decomposition; Forest landscape model

1. Introduction

In the last decade many spatially explicit forest models have been developed to simulate forest landscape

* Corresponding author. Tel.: +86 24 83970350; fax: +86 24 83970351.
E-mail address: xuchongang@yahoo.com (C. Xu).

0304-3800/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.12.009

changes (e.g. Mladenoff et al., 1996; Mladenoff and He, 1999; Pacala et al., 1993, 1996; Urban et al., 1999). Most of these models employ a raster data format and the forest landscape is conceptualized as a grid of equal-sized cells or sites. Each cell requires the input of dominant canopy tree species, secondary tree species and/or other species related information (e.g. age). However, forest inventory data is often sparsely distributed across the landscape and no forest inventory can provide all the information at the cell level. This is especially true when the simulated area has millions of cells. Thus, interpolation or extrapolation is often necessary to derive species and related data for cells where information is missing, based on the inventory data.

There are two main groups of interpolation techniques: deterministic and geostatistical (ESRI, 2001). Deterministic techniques include polynomial, inverse distance weighted, and radial based functions for interpolation. Geostatistical techniques include ordinary kriging, simple kriging, universal kriging, probability kriging, indicator kriging and disjunctive kriging. Because kriging methods quantify the spatial autocorrelation among measured points and account for the spatial configuration of the sample points around the prediction location, they have been widely used in soil science and forest science (e.g. Biondi et al., 1994; Goovaerts, 1999a, 1999b). However, kriging interpolation algorithms produce maps of best local estimates and tend to smooth out local details of the spatial variation of the attribute (Goovaerts, 1997). Thus, there emerges a set of geostatistical stochastic simulation algorithms. Instead of a map of local best estimates, geostatistical stochastic simulation algorithms provide multiple conditional realizations of the spatial distribution of attribute values reproducing statistics deemed consequential for the problem at hand (e.g. the data values at their location, sample histogram and spatial dependence of the attribute value) (Goovaerts, 1997). They do not show the smoothing effect characteristic of kriging interpolated map and looks more "realistic".

In practice, geostatistical stochastic simulation is always combined with Monte Carlo method to quantify the uncertainty in spatial model simulations (Isaaks, 1990; Pachepsky and Acock, 1998; Finke et al., 1999; Goovaerts Et al., 2001; Saito and Goovaerts, 2001; Van Meirvenne and Goovaerts, 2001; Viscarra Rossel et al., 2001). First, a set of conditional realizations is generated by geostatistical stochastic simulation from the available sample data set. Each conditional realization is then fed into the model of interest and a set of model results are produced. The analysis of the variability in the set of model results will give us insights into the uncertainty in model simulation results. In Monte Carlo simulation, it is important that geostatistical stochastic simulation algorithms generate a large set of conditional realizations that fully capture the spatial variability of attributes. However, due to the complexity and the relatively long running time of spatially explicit forest models, it is often infeasible to generate hundreds or thousands of Monte Carlo simulations (Xu et al., 2004). Thus, it is necessary to generate a relatively small set of conditional realizations capturing most of the variability.

The geostatistical stochastic simulation needs to build the local probability distributions for each no-data location. For each conditional realization, the simulated value is randomly drawn from the probability distribution. The general sampling method is the simple random sampling method. In order to reduce the number of conditional realizations needed to capture the spatial uncertainty, we introduced an effective sampling method (Latin hypercube sampling) into a geostatistical stochastic simulation algorithm (LU decomposition). Latin hypercube sampling is first compared with a common sampling procedure (simple random sampling) in LU decomposition simulation. Then it is applied to the investigation of uncertainty in the simulation results of a spatially explicit forest model (LANDIS). Simple as the application is, it will give us general insights about which model results are robust given the uncertainty introduced by interpolation.

2. Review of LU decomposition

A large number of geostatistical stochastic simulation algorithms are available: sequential Gaussian simulation, sequential indicator simulation, *p*-field simulation, simulated annealing, and LU decomposition simulation (Goovaerts, 1997). Each algorithm has its advantages and disadvantages and much work has been done to compare different stochastic simulation algorithms (Deutsch, 1994; Gotway and Rutherford, 1994; Srivastava, 1996; Goovaerts, 1999a, 1999b). The consensus is that there is no algorithm best suited for all cases, but a toolbox of alternative algorithms which

provide choices to accommodate specific problems at hand (Gómez-Hernández, 1997; Goovaerts, 2001).

Since the introduction of LU decomposition algorithm into geostatistics (Davis, 1987; Alabert, 1987), it became an attractive method due to its efficiency, simplicity and simultaneous conditioning to available data during the simulation. In LU decomposition simulation, the data value is first transformed into normal score data using standard normal cumulative distribution function (CDF). Then, the covariance matrix C between all conditioning data locations (n) and simulated locations (N) is built:

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \tag{1}$$

where C_{11} is the $n \times n$ data-to-data covariance matrix, C_{22} is the $N \times N$ simulated-to-simulated covariance matrix, $C_{12} = C_{21}^{\mathrm{T}}$ is the data-to-simulated covariance matrix. The covariance matrix C is subsequently decomposed into a lower triangular matrix (L) and an upper triangular matrix (U):

$$C = L \cdot U = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \cdot \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}. \tag{2}$$

Then a conditional realization (Y_N^{nst}) of simulated locations (N) is generated as the sum of a simple estimate plus a random component,

$$Y_N^{\mathrm{nst}} = L_{21} \cdot L_{11}^{-1} \cdot Y_n^{\mathrm{nst}} + L_{22} \cdot W, \tag{3}$$

where Y_N^{nst} is the normal score transformed data vector and W is a vector of N independent standard normal deviates. The variation among different vectors of standard normal deviates informs the variation among different conditional realizations. For notational convenience, Eq. (3) can be written as

$$Y_N^{\mathrm{nst}} = f(W), \tag{4}$$

where $W = [x_1, x_2, \ldots, x_N]$ can be seen as the input of model f. All individual element of W is independent with each other and conforms to the standard normal distribution. At last, Y_N^{nst} is back normal score transformed to the simulated data vector (Y_N),

$$Y_N = G^{-1}(Y_N^{\mathrm{nst}}) = G^{-1}(f(W)), \tag{5}$$

where G^{-1} represents the back normal score transformation function.

3. Simple random sampling versus Latin hypercube sampling

We define a general spatially explicit forest model,

$$Z = g(Y_N, Y_n), \tag{6}$$

where Y_n is the conditional data vector, Y_N is the simulated data vector and Z is a result vector of model g. Eq. (6) can be combined with Eqs. (4) and (5) and the model can be transformed to

$$Z = g(G^{-1}(f(W)), Y_n). \tag{7}$$

Because Y_n is a constant vector, Eq. (6) can be simplified as

$$Z = g(G^{-1}(f(W))). \tag{8}$$

For notational convenience, the probability space of W, Y_N^{nst}, Y_N and Z is defined as $(\varphi_w, \vartheta_w, p_w)$, $(\varphi_{y_N^{\mathrm{nst}}}, \vartheta_{y_N^{\mathrm{nst}}}, p_{y_N^{\mathrm{nst}}})$, $(\varphi_{y_N}, \vartheta_{y_N}, p_{y_N})$ and $(\varphi_z, \vartheta_z, p_z)$. In probability theory, φ is a set that contains everything that could occur in the particular situation under consideration. ϑ is a subset of φ for which probability p will be defined. p defines the probability for the elements of ϑ.

In Monte Carlo analysis, a probability based sample procedure is used to map model input (Y_N) to model output (Z). Because Y_N is derived from W, it is actually to map the vector of N independent standard normal deviates (W) to model output (Z). Specifically, a sample

$$W_i = [x_{1i}, x_{2i}, \ldots, x_{Ni}], \quad i = 1, 2, \ldots, N_s, \tag{9}$$

of size N_s is generated from φ_w. Then, each sample vector W_i is input to model f to generate a sample from $\varphi_{y_N^{\mathrm{nst}}}$

$$Y_{Ni}^{\mathrm{nst}} = [y_{1i}^{\mathrm{nst}}, y_{2i}^{\mathrm{nst}}, \ldots, y_{Ni}^{\mathrm{nst}}], \quad i = 1, 2, \ldots, N_s. \tag{10}$$

Each sample vector Y_{Ni}^{nst} is back normal score transformed and a sample is generated from φ_{y_N}

$$Y_{Ni} = [y_{1i}, y_{2i}, \ldots, y_{Ni}], \quad i = 1, 2, \ldots, N_s. \tag{11}$$

Each sample vector Y_{Ni} is subsequently fed to model g to generate a sample from φ_z,

$$Z_i = [z_{1i}, z_{2i}, \ldots, z_{Ni}], \quad i = 1, 2, \ldots, N_s \tag{12}$$

the variability in which is used to assess the uncertainty in model output. In order to avoid confusion, for the rest

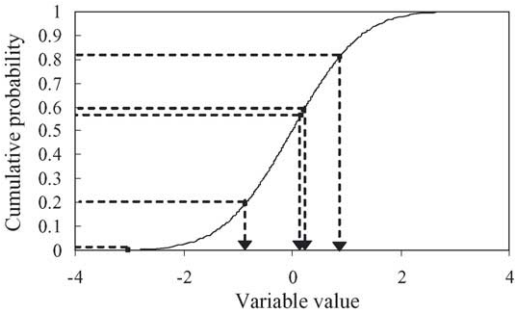

Fig. 1. Sample from standard normal cumulative distribution by simple random sampling.

of the article, the sample in Eqs. (10)–(12) is termed as the conditional realization ensemble and the sample size is referred as number of realizations.

Two main sampling procedures can be used to generate the sample in Eq. (9): simple random sampling and Latin hypercube sampling (Helton and Davis, 2003). Simple random sampling draws random values from the standard normal CDF for each individual variable in W by generating a random number from a uniform distribution on [0, 1] (Fig. 1). However, there is no assurance that a sample element will be generated from any particular subset of the sample space φ_w (Helton and Davis, 2003). In view of this, Latin hypercube sampling was introduced into Monte Carlo analysis (McKay et al., 1979). Latin hypercube sampling firstly stratified the range of each variable (i.e. x_i) into N_s disjoint intervals of equal probability and then a random value is drawn at each interval (Fig. 2). Finally, one of the N_s random values for each variable are randomly selected to form a sampling element in Eq. (9). A number of

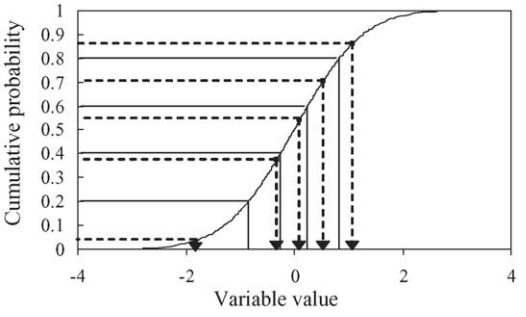

Fig. 2. Sample from standard normal cumulative distribution by Latin hypercube sampling.

studies have shown that Latin hypercube sampling can more exhaustively explore model parameter space (in this study φ_w) than simple random sampling (McKay et al., 1979; Iman, 1999; Helton and Davis, 2003). Thus, in this study, we will not focus our attention on how Latin hypercube sampling can affect the sample in Eq. (9) but on how Latin hypercube sampling can affect the sample in Eq. (10), which will be back normal score transformed and then be used as an input of our model of interest. Because the sample in Eq. (10) will not be affected by the specific back normal score transformation function G^{-1}, our results will be of general importance.

In order to assess the effects of Latin hypercube sampling on the sample in Eq. (10), we first generated a 500 m × 500 m random age map of a certain species (cell size 10 m × 10 m) using unconditional LU decomposition simulation algorithm (Fig. 3a). LU decomposition was implemented under the condition that the species age conforms to a normal distribution with an average of 100 years and a standard deviation of 20 and the normal score transformed values honor a spherical semi-variogram model with a sill of 1, a nugget of 0 and a range of 100. The generated map was used as reference map in our study. The simulated species age map has an average of 95.3 years and a standard deviation of 20.6 (Fig. 3c). The empirical semi-variogram of the normal score transformed value of the simulated age map fits to spherical model with a sill of 1.1, a nugget of 0 and a range of 97.1 (Fig. 3d). Twenty-five randomly selected locations are then used as the sample data in this study (Fig. 3b).

Both LU decomposition simulation by Latin hypercube sampling and LU decomposition simulation by simple random sampling are used to generate a series of conditional realization ensembles with size of N_r ($N_r = 1, 2, \ldots, 50$). For each ensemble of size N_r ($N_r = 1, 2, \ldots, 50$), the LU decomposition simulation is repeated 50 times. Since one desirable property of the model of uncertainty are accuracy (true but unknown outcome should be included in the probability distribution, i.e. belongs to the 95% probability interval) (Goovaerts, 1999a, 1999b), we ruled out the extreme values from the ensemble by sampling the normal distribution of individual variable in Eq. (9) from the 99.98% probability interval. In each conditional realization ensemble, we calculate the standard deviation of value for each cell to capture cell-specific variability among realizations. We

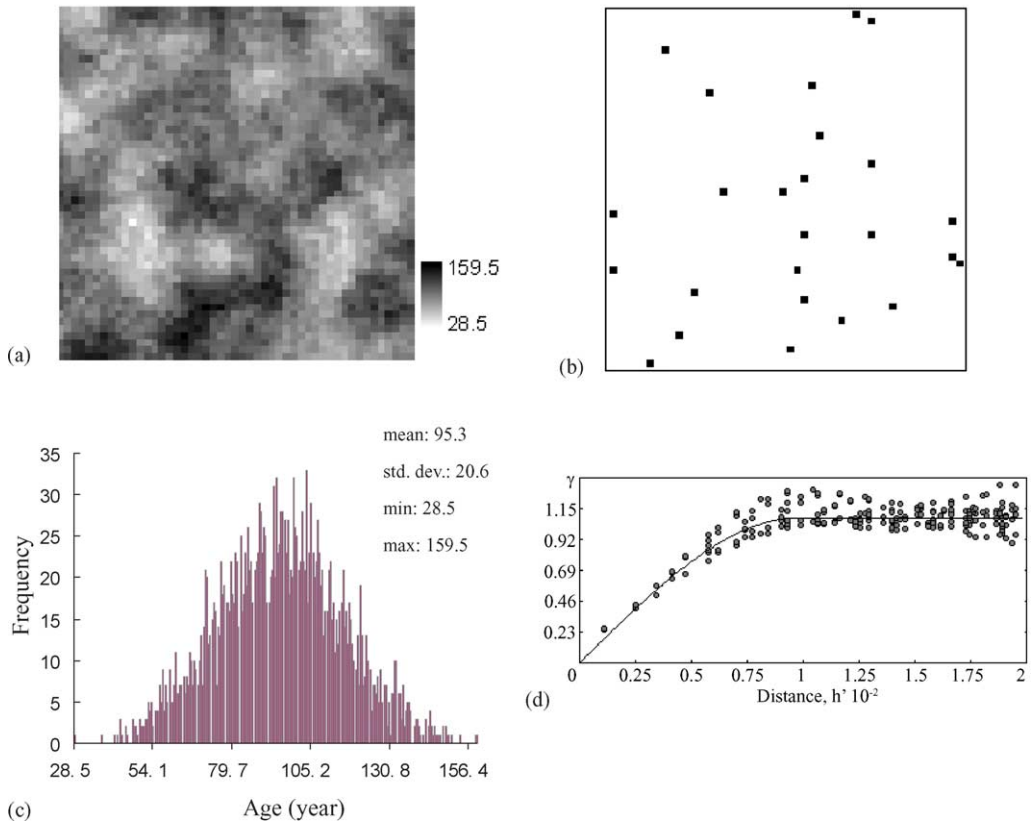

Fig. 3. Reference age map of a certain species with an average age of 95.3 years and a standard deviation of 20.6. (a) The reference map; (b) the randomly sampled map; (c) the histogram of values in the reference map; (d) the empirical semi-variogram of the normal scored transformed value from the reference map which is a spherical model with a sill of 1.1 and a range of 97.1.

also calculate the average of standard deviation to capture the overall cell-specific variability of attribute on the entire map. The average, 10 percentile and 90 percentile of the average standard deviation on the whole map are derived for each conditional realization ensemble over the 50 repeated runs.

The standard deviation generally captures the spread of simulated values in the ensemble for individual cells. However, it does not capture how exhaustive is the simulated values for the underlying sample space (i.e. exhaustiveness of realization). In this study, we use the Shannon entropy index (Shannon, 1948) to construct a realization exhaustiveness index (REI),

$$\mathrm{REI} = \frac{-\sum_{i=1}^{N_\mathrm{r}} p_i \ln p_i}{\ln(1/N_\mathrm{r})} \tag{13}$$

where N_r is conditional realization ensemble size, p_i is the proportion of the realization ensemble falling in the ith subset of the N_r non-overlapping and equal probability subset of the sample space for a certain cell and $\ln(1/N_\mathrm{r})$ is the maximum Shannon entropy index. The most exhaustive case of realization is that there should be at least one realization value from each of the N_r non-overlapping and equal probability subset of the corresponding sample space for individual cell. REI equals to 1 for the most exhaustive case. Because the original data value is always normal score transformed before conducting LU simulation, the sample space for each cell is assumed to be standard Gaussian. For each realization ensemble, we calculate the REI on each cell to capture the cell-specific realization exhaustiveness. We also calculate the average of REI (AREI) to capture the overall cell-specific realization exhaustiveness

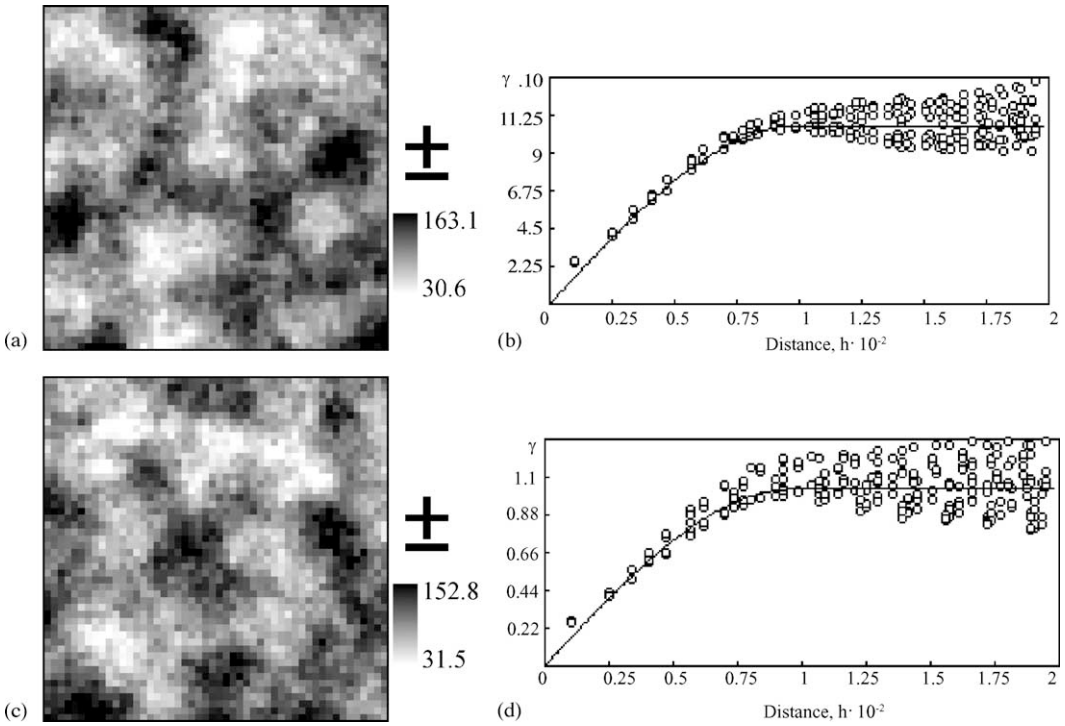

Fig. 4. One of the realizations generated by LU decomposition simulation. (a) A map generated by LU decomposition simulation using simple random sampling; (b) the experimental semi-variogram of the normal score transformed value from the generated map (a) which fits to a spherical model with a sill of 1.1 and a range of 99.7; (c) a map generated by LU decomposition simulation using Latin hypercube sampling; (d) the experimental semi-variogram of the normal score transformed value from the generated map (c) which fits to a spherical model with a sill of 1.0 and a range of 96.1.

on the entire map. The average, 10 percentile and 90 percentile of the AREI are derived for each realization ensemble over the 50 repeated runs.

Results show that both Latin hypercube sampling and simple random sampling can reproduce the spatial dependence (Fig. 4). However, Latin hypercube sampling produces larger average of standard deviation than simple random sampling with the same number of simulations (Fig. 5). When using Latin hypercube sampling, average standard deviation reaches a plateau when number of realization equals to 9. However, when using simple random sampling, average of standard deviation reaches a plateau when number of realization equals to 15. This suggests that Latin hypercube sampling can capture more variability in the sample space $\varphi_{y_N^{\text{nst}}}$ than simple random sampling especially when the sample size is small. At the same time, Latin hypercube sampling results in higher AREI than simple random sampling with the same number of realizations (Fig. 6).

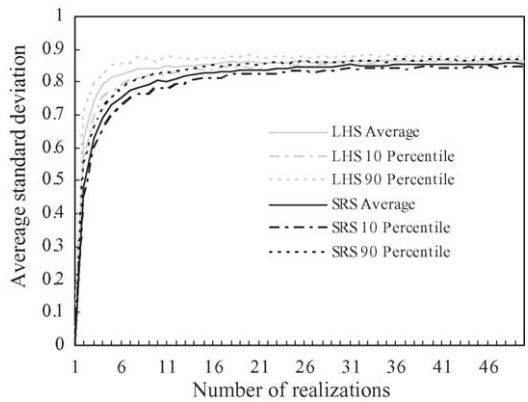

Fig. 5. Impact of number of realizations on the average of cell-specific standard deviation using Latin hypercube sampling (LHS) and simple random sampling (SRS).

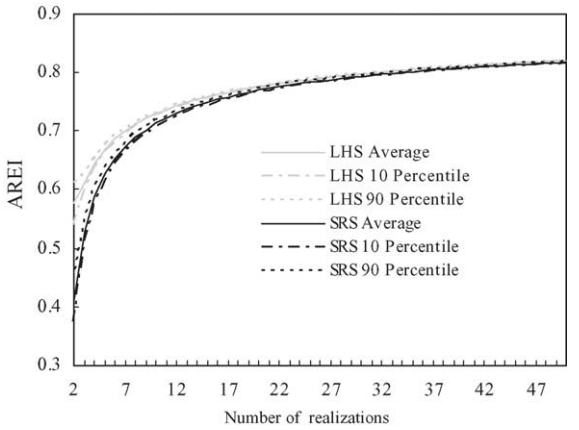

Fig. 6. Impact of number of realizations on the average of cell-specific realization exhaustiveness index (AREI) using Latin hypercube sampling (LHS) and simple random sampling (SRS).

The smaller is the number of realizations, the more evident is the difference between the two sample techniques. When the number of realization approaches 20, there is only little difference of AREI between these two sample techniques. The above results suggest that Latin hypercube sampling can produce more exhaustive realization ensemble than simple random sampling especially when the number of realizations is relatively small.

The results of the study suggest that, when using Monte Carlo method to assess the uncertainty produced by LU decomposition simulation algorithm, Latin hypercube sampling should be recommended especially when the allowed number of realizations is small.

4. Application

LU decomposition simulation using Latin hypercube sampling and simple random sampling is applied to simulate the age of a species in a forest landscape model simulation using LANDIS. The simulation is conducted on a homogeneous landscape with one early succession species (species 1) and one late succession species (species 2). The life history attributes of species 1 and species 2 are listed in Table 1. Species 1 and species 2 are assumed to be present at all cells in the map. In order to simplify the situation for the analysis of uncertainty propagation, the age of the species 2 is assumed to be a constant (50 years for all cells). The age of species 1 is assumed to have a spatial dependence. Both LU decomposition simulation by Latin hypercube sampling and random simple sampling is used to generate 10 simulated maps of age from the random sample (Fig. 3b). The purpose of this simple application is to assess how the difference of Latin hypercube sampling and simple random sampling will affect the LANDIS simulation results uncertainty when the sample size is small. At the same time, we would identify which model results are robust given the uncertainty of age introduced by the geostatistical stochastic simulation.

Table 1
Life history attributes for species 1 and species 2

Species	LONG	MTR	ST	FT	ED	MD	VP	MVP
1	200	20	3	4	30	80	0	0
2	100	15	1	2	100	−1	0.8	30

LONG, longevity (years); MTR, age of maturity (years); ST, shade tolerance; FT, fire tolerance; ED, effective seeding distance (m); MD, maximum seeding distance (m); VP, vegetative reproduction probability; MVP, minimum age of vegetative reproduction (years); −1 represents unlimited seeding range.

4.1. Description of LANDIS

LANDIS is a cell-based spatially explicit forest landscape model of disturbance, succession and management (Mladenoff et al., 1996; Mladenoff and He, 1999). It simulates species-level forest dynamics by tracking the presence/absence of species age cohorts (cohort of trees of certain species with their age in a 10-year interval) at 10-year time steps under natural and anthropogenic disturbances including fire, windthrow, insect and disease, harvesting, and fuel management. Detailed descriptions of various LANDIS components can be found in Mladenoff and He (1999), He and Mladenoff (1999a, 1999b), He et al. (1999) and Gustafson et al. (2000).

4.1.1. Seed dispersal and seedling establishment process

Seed dispersal and seedling establishment process in LANDIS are simulated as three steps: dispersal, light condition checking and site condition checking

(Mladenoff and He, 1999; He and Mladenoff, 1999a). First, the seed sources are identified by locating cells where there exists species whose age is older than the maturity age. Then seeds from the source stochastically disperse. The seed dispersal probability is modeled as a function of its effective and maximum seeding distance. Effective seed distance (ED) is the distance within which seed has the highest probability of reaching a site (e.g. $P > 0.95$). Maximum seed (MD) dispersal distance is the distance beyond which there is little possibility of reaching (e.g. $P < 0.01$). Seed dispersal probability (P) between ED and MD follows a negative exponential distribution:

$$P = e^{-b(x/\mathrm{MD})}, \quad \mathrm{ED} < x < \mathrm{MD} \tag{14}$$

where x is a given distance from the seed source and b is a coefficient, which is set to 1 in LANDIS 3.6. A random number (P_r) (from a uniform random number pool ranging from 0 to 1) is generated to compare with P to decide if the seed can successfully disperse to a specified site. The seed successfully disperses to a site if $P_r < P$. Once the seed successfully reaches the site, the light condition checking procedure is implemented. The shade tolerance rank of the arrived species is compared to that of the species already established on the site to check if the site favors the species arrived. In LANDIS, the shade tolerance of species is divided into five classes with class 5 corresponding to the most tolerant and class 1 the least tolerant. For arrived species whose shade tolerance rank is lower than class 5, the site favors the species if its shade tolerance rank is higher than or equal to the highest shade tolerance rank of the already established species. For arrived species whose shade tolerance rank equals to class 5, the site favors the species only if the oldest cohort in the arrived site has an age older than the minimum age of cohort growth required before enough shade is created so that a shade tolerance five species can seed into the site, which is an input of LANDIS. If the light condition on the site favors the species, the site condition checking procedure is implemented. A random number (P_r) is generated to compare with the arriving species' establish coefficient (C). The seedling successfully established if $P_r < C$. Establish coefficient is a floating number from 0 to 1 used to represent the relative scaling of how environmental conditions (e.g. moisture, climate and nutrient) favor various species (Mladenoff et al., 1996; Mladenoff and He, 1999).

4.1.2. Fire disturbance

LANDIS uses a stochastic simulation approaches to simulate the fire disturbance (He and Mladenoff, 1999b). The fire probability (P) of each cell is determined by the following formula:

$$P = B \times \mathrm{IF} \times \mathrm{MI}^{-(e+2)} \tag{15}$$

where MI is the mean fire-return interval of a given landtype, B is the fire probability coefficient designed for model calibration, and IF is the time since last fire. In order to simulate the fire disturbance, LANDIS firstly locates the ignition point on each landtype randomly. The number of the ignition points (N_{ip}) is determined from the ignition coefficient, which is user-defined (N_{ip} = ignition coefficient × total cell number of each landtype). Secondly, LANDIS calculates the fire probability (P) of the cell where the ignition point locates by Eq. (2). A random number (P_r) is then generated to check if this cell will be ignited. The fire is successfully ignited given $P_r < P$. Thirdly, if the cell were ignited, LANDIS simulates the fire spread. The coordinates of the four adjacent cells are entered in a priority queue in a random order. The fire probabilities of all the four adjacent cells are calculated using Eq. (2) and a random number is generated to check if the fire will be spread to the adjacent cell. Fire spreads until either the desired fire size is reached, or the surrounding cells cannot burn, or non-forest surrounds the cell.

Important LANDIS outputs include an age cohort map for each species and the species distribution map at each 10-year time interval. If there are multiple age cohorts for a species, only the oldest age cohort is output. The species distribution map records presence/absence of species at each cell.

4.2. Uncertainty quantification

In LANDIS, the species composition map records the species age cohort information for each cell. For each simulated age map for species 1, a corresponding species composition map was generated by fitting the species age into the 10-year age interval. For example,

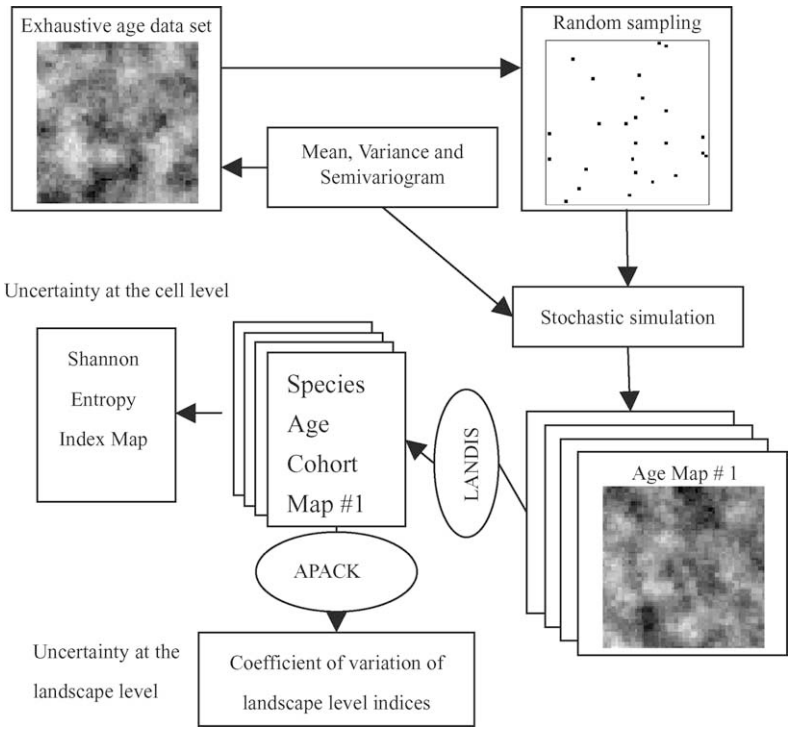

Fig. 7. Design of numerical experiment to assess uncertainty in LANDIS simulation resulted from stochastic simulation. See text for explanations.

if a cell is assigned certain species with an average age range from 10 to 20 years (including an age of 20 years), a 10–20-year species age cohort is generated for that cell.

We use the Monte Carlo method to quantify the uncertainty in model results produced by the LU decomposition. The Monte Carlo method assesses uncertainties through repeated model runs, with parameters drawn from specified probability distributions. The numerical experimental design is illustrated in (Fig. 7). Each species composition map, generated from the simulated age map for species 1 using LU decomposition, is fed into LANDIS to simulate forest landscape change for 1000 years. For each species, the running of LANDIS results in a series of species age cohort map. For notational convenience, we refer to the set of species age cohorts for a certainty cell from the species age cohort maps by the Monte Carlo simulations as Monte Carlo species age cohort ensemble. We quantify the uncertainty of LANDIS output at the cell level and at the landscape level.

4.2.1. Uncertainty at the cell level

For each species, Shannon entropy index (SEI) was used to quantify the uncertainty in species age cohort information for each individual cell,

$$\mathrm{SEI} = -\sum_{i=1}^{N_c} p_{ci} \ln p_{ci}, \tag{16}$$

where N_c is the total number of species age cohort present in the Monte Carlo species age cohort ensemble and p_{ci} is the ith species age cohort frequency (0–1). Higher SEI indicates higher uncertainty. The calculation of SEI for all individual cells results in a SEI map capturing the species age cohort information uncertainty at the cell level. In order to capture the overall uncertainty in species age cohort information at the cell level, we also calculated the average SEI (ASEI) of the species age cohorts on the whole map.

4.2.2. Uncertainty at the landscape level

In order to assess the uncertainty in model simulation results at the landscape level, we calculated percent

area (PA) and aggregation index (AI) for each species from the species distribution map in Monte Carlo runs. PA is the percent of the area occupied by a certain species in the study area and AI is a class specific landscape index used to qualify aggregation level of spatial pattern (He et al., 2000). When AI equals 1, the landscape has the highest level of aggregation; when AI equals 0, the landscape has the lowest level of aggregation. For each species, differences among PAs and AIs in the control run and Monte Carlo runs was used to quantify the variability in overall landscape pattern (or uncertainty in species information at landscape level). Both PA and AI were calculated using APACK, software for calculating landscape metrics (Mladenoff and DeZonia, 2000).

4.3. Results

4.3.1. Uncertainty at the cell level

Results show that, the ASEI behaves like a dampen triangle wave with a wavelength of about the longevity of species 1 (200 years) (Fig. 8a). The increase of uncertainty (shown by increase of ASEI) is produced by the stochastic seed dispersal and fire event (for details, please refer to Xu et al., 2004). The decrease of uncertainty (shown by decrease of ASEI) is produced by the removal of species age cohorts by mortality on cells where there is relatively high uncertainty. When the uncertainty increases, there is less difference between uncertainty of the newly simulated and the original species age cohort. Thus, the amplitude of increase of uncertainty by stochastic seed dispersal and fire event and increase of uncertainty by species mortality tends to decrease. At last, the ASEI reaches a plateau at a relative high value (1.89), which indicates that there are at least six alternatives output for each cell for the 10 simulations. This suggests that LANDIS simulation produces high uncertainty in species age cohort information at the cell level.

Results also show that the Latin hypercube sampling results in higher ASEI at the first 500 simulation years (except for the simulation year between 350 and 400) (Fig. 8a). After simulation year 500, there is no evidence that Latin hypercube sampling producing high uncertainty than simple random sampling. The reason is that both stochastic seed dispersal and stochastic fire event built in LANDIS have neutralized the initial uncertainty introduced by stochastic geostatistical simulation. However, it would still pay in uncertainty analysis to use Latin hypercube sampling in view of the higher ASEI for Latin hypercube sampling for the first 500 simulation years.

4.3.2. Uncertainty at the landscape level

The PA and AI derived from Monte Carlo runs do not differ much in the 10 Monte Carlo simulations for both Latin hypercube sampling and simple random sampling (Figs. 9a and b and 10a and b). Due to the

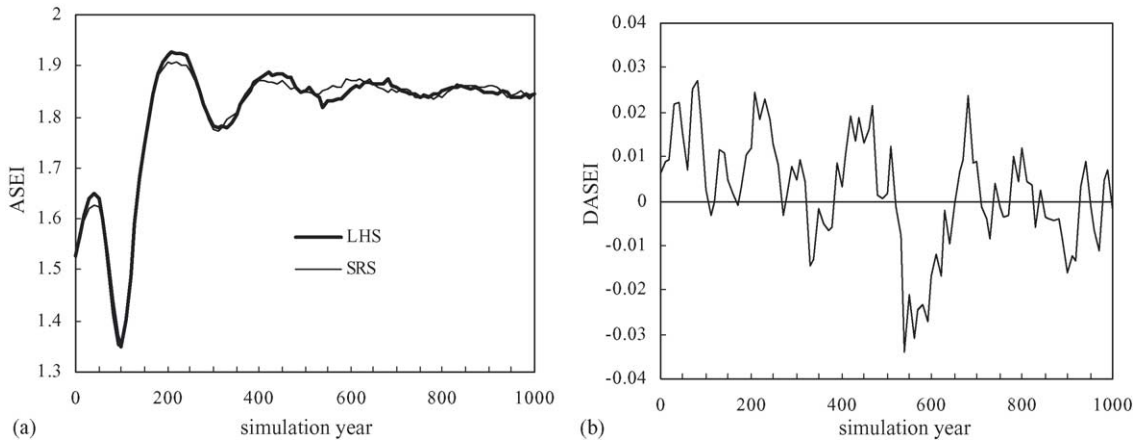

Fig. 8. Dynamics of average Shannon entropy index (ASEI) for Latin hypercube sampling (LHS) and simple random sampling (SRS) (a) and dynamics of their differences in ASEI (DASEI) (b) for species 1. Differences are calculated by subtracting the ASEI for LHS by the ASEI for SRS.

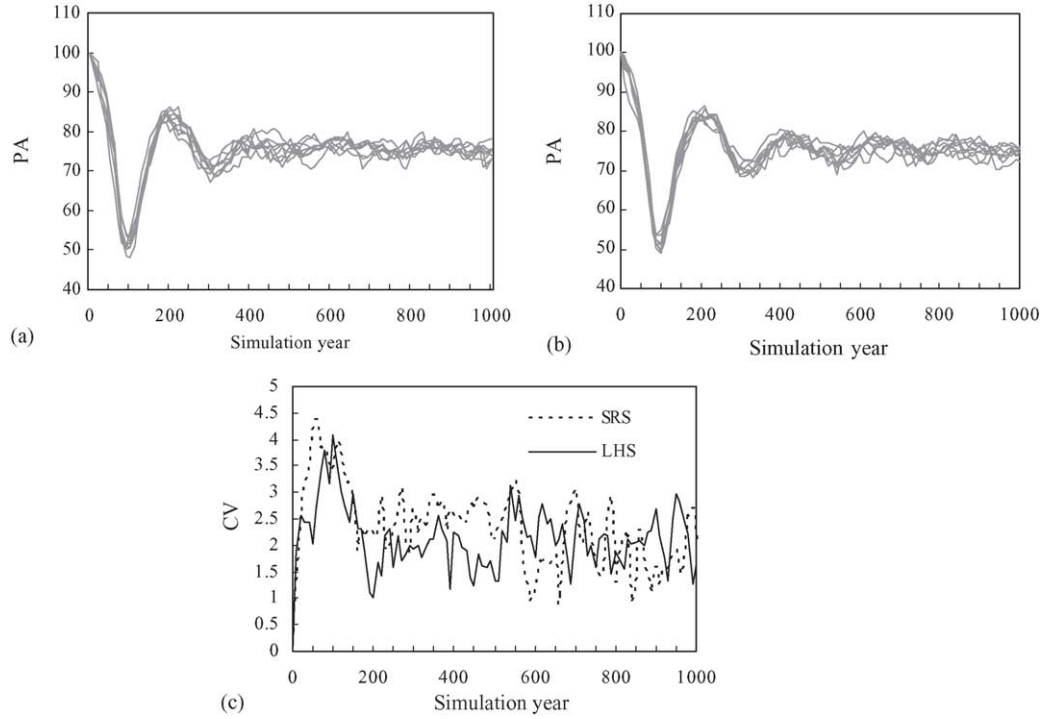

Fig. 9. Percent area of species 1 in LANDIS Monte Carlo runs. (a) Monte Carlo runs using simple random sample (SRS); (b) Monte Carlo runs using Latin hypercube sample (LHS); (c) coefficient of variation (CV) of PA in the Monte Carlo runs using SRS or LHS.

stochastic fire event in LANDIS, coefficient of variation of PA and AI from both Latin hypercube sampling and simple random sampling irregularly fluctuates with simulation year. However, they do not increase with simulation year and are less than 5% throughout the simulation (Figs. 9c and 10c). This suggests that, although the uncertainty of species age cohorts at the cell level for each species is high, species percent area and their spatial pattern (measured by the aggregation index) are not substantially affected. Consequently, there is no strong evidence that Latin hypercube sampling generate more uncertainty than simple random sampling (Figs. 9c and 10c).

5. Discussion

5.1. Method implications

Our results show that LU decomposition simulation using Latin hypercube sampling can effectively decrease the number of simulations required to capture the spatial variability. Therefore, Latin hypercube sampling should be preferred over simple random sampling in LU decomposition simulation in uncertainty analysis.

The Latin hypercube sampling can also be used in the sequential Gaussian simulation (Pebesma and Heuvelink, 1999). However, because the Latin hypercube sampling is conducted by shifting of simple random sampling, there will be meaningful deviations when the sample size is small (Pebesma and Heuvelink, 1999). Since the purpose of this study is to use a relatively small sample to capture the uncertainty, it would be infeasible to use the Latin hypercube sampling technique by shifting of simple random sampling which will results in meaningful deviations. From this point of view, LU decomposition simulation should be preferred over sequential simulations.

One well-known limitation of LU decomposition simulation is that it can only be applied to a few thousand grid nodes due to the computational requirements

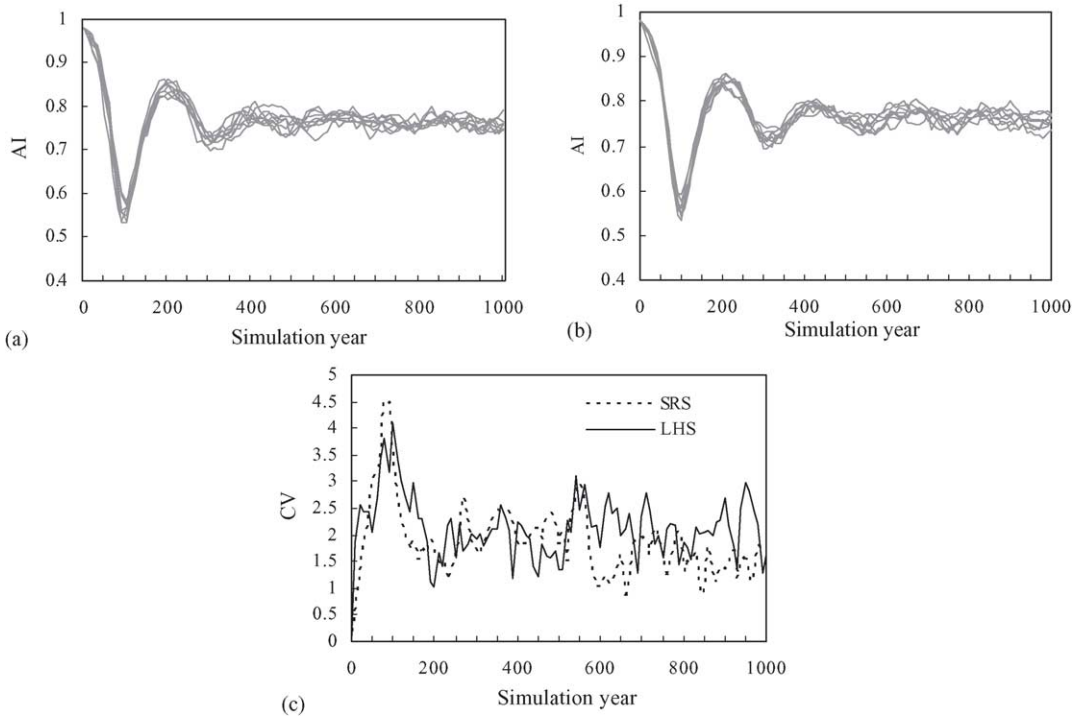

Fig. 10. Aggregation index (AI) of species 1 in LANDIS Monte Carlo runs. (a) Monte Carlo runs using simple random sample (SRS); (b) Monte Carlo runs using Latin hypercube sample (LHS); (c) coefficient of variation (CV) of AI in the Monte Carlo runs using SRS or LHS.

posed by the size of matrix *L*. However, with the development of computer technology and the use of super computer, the computer limitation can be solved. At the same time, Vargas-Guzman and Dimitrakopoulos (2002) extended LU decomposition simulation to grid nodes of any size based on a novel partitioning of the *L* matrix. For each subset of locations to be simulated, a similar LU decomposition simulation is implemented. This would relieve the limitation of LU decomposition.

In this study, the LU decomposition using Latin hypercube sampling is used to assess the uncertainty in a forest landscape model simulation. However, this method is applicable for uncertainty analysis of any kind of spatially explicit model where the exhaustive cell-level information needs to be interpolated and it is time consuming to run the model. This kind of models may include spatially explicit hydrological model (e.g. Voinov et al., 1999), spatially explicit spatial plant–herbivore model (e.g. Oom et al., 2004) and spatially explicit logistic regression model (e.g. Horssen et al., 2002).

5.2. Result implications

A challenge in ecology is to understand broad scale patterns emerging from the complexity of interaction at lower scales (O'Neil, 1989; Wiens, 1989; Wiens and Milne, 1989; Levin, 1992; Levin et al., 1997). Spatially explicit forest models present a potential to meet the challenge. However, the natural system is always overwhelmed with all kinds of uncertainty (Clark et al., 2001). In building or using these models, it is important to determine how much detail at the fine level is essential to more macroscopic regularities (Levin et al., 1997). Certain fine scale details and uncertainty associated with them are not essential and will not have much effect on regularities at large spatial and time scales. Pacala et al. (1996) showed that the community level predictions (succession dy-

namic, intraspecifically clumped and interspecifically segregated spatial distributions) by a spatially explicit forest model (SORTIE) were robust given the level of sampling uncertainty in the study. Deutschman et al. (1999) showed that SORTIE was surprisingly insensitive to the amount of detail used in the calculation of the local resource and light. Xu et al. (2004) showed that the LANDIS simulation results at landscape level were not sensitive to the uncertainty at the cell level produced by a stand based assignation approach.

Simple as the application is, it gives us general insights about which model results are robust given the uncertainty introduced by stochastic simulation. The results of this study show that LANDIS simulation results at the landscape level (species percent area and their spatial pattern measured by an aggregation index) is not sensitive to the uncertainty in species age cohort information at cell level produced by geostatistical stochastic simulation algorithms. This suggests that LANDIS can be used to predict the forest landscape change at broad spatial and temporal scales even if accurate species age cohort information at each cell is not available. However, results also show that, except for species 2 at the beginning of simulation, uncertainty in species age cohort information at cell level is high throughout simulation. Thus, it would be infeasible to predict the species age cohort distribution at the cell level using LANDIS. Just as Levin et al. (1997) have pointed out, "such models should not be expected to predict where every tree will be at each point in time; only aggregate statistical properties can be reliably predicted typically over broad spatial and temporal scales".

Generally, the LANDIS should be simulated on a larger grid size (more than 500×500). In order to simply our simulation, we only used a small grid size in our study (50×50). However, the result derived from our result should also be hold at larger grids. The reasons are (1) in this study we have only 25 sample points (1% of study area grid), which is a sample size proportional to that at larger grid and thus a similar uncertainty input at large grid; (2) LANDIS simulates with the same mechanism at the small grid as that at large grid, which would result in the same uncertainty propagation process; (3) our study reached similar result to the previous study that the uncertainty of LANDIS simulation (introduced by a stand based assignation approach) at landscape level were not sensitivity to the uncertainty at the cell level (Xu et al., 2004).

Acknowledgements

This study was funded by the National Science Foundation of China (Program code: 40331008), the GIS Mission Enhancement Program at University of Missouri and the Spaeth & Boggess graduate fellowship from the Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana-Champaign. We thank two anonymous reviewers for their helpful comments.

References

Alabert, F.G., 1987. The practice of fast conditional simulation through the LU decomposition of the covariance matrix. Math. Geol. 19, 369–386.

Biondi, F., Meyers, D.E., Avery, C.C., 1994. Geostatistically modeling stem size and increment in an old growth forest. Can. J. For. Res. 24, 1354–1638.

Clark, J.S., Carpenter, S.R., Barber, M., Collins, S., Dobson, A., Foley, J.A., Lodge, D.M., Pascual, M., Pielke Jr., R., Pizer, W., Pringle, C., Reid, W.V., Rose, K.A., Sala, O., Schlesinger, W.H., Wall, D.H., Wear, D., 2001. Ecological forecasts: an emerging imperative. Science 27, 657–660.

Davis, M.W., 1987. Production of conditional simulations via the LU decomposition of the covariance matrix. Math. Geol. 19, 91–98.

Deutschman, D.H., Levin, S.A., Pacala, S.W., 1999. Error propagation in a forest succession model: the role of fine-scale heterogeneity in light. Ecology 80, 1927–1943.

Deutsch, C.V., 1994. Algorithmically-defined random function models. In: Dimitrakopoulos, R. (Ed.), Geostatistics for the Next Century. Kluwer Academic Publishing, Dordrecht, pp. 422–435.

ESRI, 2001. Using ArcGIS^{TM} Geostatistical Analyst. ESRI Inc., Redlands, CA, USA.

Finke, P.A., Wladis, D., Dros, J., Pebesma, E.J., Reinds, G.J., 1999. Quantification and simulation of errors in categorical data for uncertainty analysis of soil acidification modelling. Geoderma 93, 177–194.

Gómez-Hernández, J.J., 1997. Issues on environmental risk assessment. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong'96. Kluwer Academic Publishing, Dordrecht, pp. 15–26.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York, 483 pp.

Goovaerts, P., 1999a. Impact of the simulation algorithm, magnitude of ergodic fluctuations and number of realizations on the spaces of uncertainty of flow properties. Stochast. Environ. Res. Risk Assess. 13, 161–182.

Goovaerts, P., 1999b. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1–45.

Goovaerts, P., 2001. Geostatistical modelling of uncertainty in soil science. Geoderma 103, 3–26.

Goovaerts, P., Semrau, J., Lontoh, S., 2001. Monte Carlo analysis of uncertainty attached to microbial pollutant degradation rates. Environ. Sci. Technol. 35 (19), 3924–3930.

Gotway, C.A., Rutherford, B.M., 1994. Stochastic simulation for imaging spatial uncertainty: comparison and evaluation of available algorithms. In: Amstrong, M., Dowd, P.A. (Eds.), Geostatistical Simulations. Kluwer Academic Publishing, Dordrecht, pp. 1–21.

Gustafson, E.J., Shifley, S.R., Mladenoff, D.J., Nimerfro, K.K., He, H.S., 2000. Spatial simulation of forest succession and timber harvesting using LANDIS. Can. J. For. Res. 30, 32–43.

He, H.S., Dezonia, B., Mladenoff, D.J., 2000. An aggregation index (AI) to quantify spatial patterns of landscapes. Landscape Ecol. 15, 591–601.

He, H.S., Mladenoff, D.J., Boeder, J., 1999. An object-oriented forest landscape model and its representation of tree species. Ecol. Model. 119, 1–19.

He, H.S., Mladenoff, D.J., 1999a. The effects of seed dispersal on the simulation of long-term forest landscape change. Ecosystems 2, 308–319.

He, H.S., Mladenoff, D.J., 1999b. Spatially explicit and stochastic simulation of forest landscape fire disturbance and succession. Ecology 80, 81–99.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab. Eng. Syst. Safety 81, 23–69.

Horssen, P.W.V., Pebesma, E.J., Schot, P.P., 2002. Uncertainties in spatially aggregated predictions from a logistic regression model. Ecol. Model. 154, 93–101.

Iman, R.L., 1999. Latin Hypercube Sampling. Encyclopedia of Statistical Science Update, vol. 3. Wiley, New York, pp. 408–411.

Isaaks, E.H., 1990. The application of Monte Carlo methods to the analysis of spatially correlated data. PhD Thesis. Stanford University.

Levin, S.A., 1992. The problem of pattern and scale in ecology. Ecology 73, 1943–1967.

Levin, S.A., Grenfell, B., Hastings, A., Perelson, A., 1997. Mathematical and computational challenges in population biology and ecosystems science. Science 275, 334–342.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21, 239–245.

Mladenoff, D.J., DeZonia, B., 2000. APACK 2.14 Users Guide. Department of Forest Ecology and Management, University of Wisconsin, Madison, WI, USA.

Mladenoff, D.J., He, H.S., 1999. Design and behavior of LANDIS, an object-oriented model of forest landscape disturbance and succession. In: Mlandenoff, D.J., Baker, W.L. (Eds.), Advances in Spatial Modeling of Forest Landscape Change: Approaches and Applications. Cambridge University Press, Cambridge, UK, pp. 1–13.

Mladenoff, D.J., Host, G.E., Boeder, J., Crow, T.R., 1996. LANDIS: a spatial model of forest landscape disturbance, succession and management. In: Goodchild, M.R., Steyaert, L.T., Parks, B.O. (Eds.), GIS and Environmental Modeling: Progress and Research Issues. GIS World Books, Fort Collins, CO, pp. 175–180.

O'Neil, R.V., 1989. Perspectives in hierarchy and scale. In: Roughgarden, J., May, R.M., Levin, S.A. (Eds.), Perspectives in Ecological Theory. Princeton University Press, Princeton, NJ, USA, pp. 140–156.

Oom, S.P., Beecham, J.A., Legg, C.J., Hester, A.J., 2004. Foraging in a complex environment: from foraging strategies to emergent spatial properties. Ecol. Complexity 1, 299–327.

Pacala, S.W., Canham, C.D., Saponara, J., Silander Jr., J.A., Kobe, R.K., Ribbens, E., 1996. Forest models defined by field measurements: estimation, error analysis and dynamics. Ecol. Monogr. 66, 1–43.

Pacala, S.W., Canham, C.C., Saponara, J., Sliander Jr., J.A., 1993. Forest models defined by field measurement. I. The design of a northeastern forest simulator. Can. J. For. Res. 23, 180–188.

Pachepsky, Y., Acock, B., 1998. Stochastic imaging of soil parameters to assess variability and uncertainty of crop yield estimates. Geoderma 85, 213–229.

Pebesma, E.J., Heuvelink, G.B.M., 1999. Latin hypercube sampling of Gaussian random fields. Technometrics 41 (4), 303–312.

Saito, H., Goovaerts, P., 2001. Accounting for source location and transport direction into geostatistical prediction of contaminants. Environ. Sci. Technol. 35, 4823–4829.

Shannon, C.E., 1948. A mathematical theory of communications. Bell Syst. Tech. J. 27, 379–623.

Srivastava, M.R., 1996. An overview of stochastic spatial simulation. In: Mowrer, H.T., Czapleweski, R.L., Hamre, R.H. (Eds.), Proceedings of the Second International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences, General Technical Report RM-GTR-277, US Department of Agriculture, Forest Service, Fort Collins, pp. 283–295.

Urban, D.L., Acevedo, M.F., Garman, S.L., 1999. Scaling fine-scale processes to large scale patterns using models derived from models: meta-models. In: Mladenoff, D.J., Baker, W.L. (Eds.), Spatial Modeling of Forest Landscape Change: Approaches and Applications. Cambridge University Press, Cambridge, UK, pp. 70–98.

Van Meirvenne, M., Goovaerts, P., 2001. Evaluating the probability of exceeding a site-specific soil cadmium contamination threshold. Geoderma 102, 75–100.

Vargas-Guzman, J.A., Dimitrakopoulos, R., 2002. Conditional simulation of random fields by successive residuals. Math. Geol. 34 (5), 597–610.

Viscarra Rossel, R.A., Goovaerts, P., McBratney, A.B., 2001. Assessment of the production and economic risks of site-specific liming using geostatistical uncertainty modelling. Environmetrics 12 (8), 699–711.

Voinov, A.A., Voinov, H., Costanza, R., 1999. Surface water flow in landscape models. 2. Patuxent watershed case study. Ecol. Model. 119, 211–230.

Wiens, J.A., 1989. Spatial scaling in ecology. Funct. Ecol. 3, 385–397.

Wiens, J.A., Milne, T.T., 1989. Scaling of "landscapes" in ecology, or landscape ecology from a beetle's perspective. Landscape Ecol. 3, 87–96.

Xu, C.G., He, H.S., Hu, Y.M., Chang, Y., Larsen, D.R., Li, X.Z., Bu, R.C., 2004. Assessing the effect of cell level uncertainty on a forest landscape model simulation in northeastern China. Ecol. Model. 181, 57–72.