Soil Use and Management, September 2008, **24,** 263–272 doi: 10.1111/j.1475-2743.2008.00165.x

Regional-scale modelling of the spatial distribution of surface and subsurface textural classes in alluvial soils using Markov chain geostatistics

C. Zhang & W. Li
Center for Environmental Sciences and Engineering and Department of Geography, University of Connecticut, Storrs, CT 06269, USA

Abstract

Soil texture is directly associated with other soil physical and chemical properties and can affect crop yield, erodibility and water and pollutant movement. Thus, maps of soil textural class are valuable for agricultural management. Conventional spatial statistical methods do not capture the complex large-scale spatial patterns of multi-class variables. Markov chain geostatistics (MCG) was recently proposed as a new approach for the conditional simulation of categorical variables. In this study, we apply an MCG algorithm to simulate the spatial distribution of textural classes of alluvial soils at five different depths in a 15-km^2 area on the North China Plain. Soil texture was divided into five classes – sand, sandy loam, light loam, medium loam and clay. Optimal prediction maps, simulated maps and occurrence probability maps for each depth were generated from sample data. Simulated results delineated the distribution of the five soil textural classes at the five depths and quantified related spatial uncertainties caused by limited sample size (total of 139 points). These results are not only useful for understanding the spatial distribution of soil texture in alluvial soils, but also provide valuable quantitative information for precision agriculture, soil management and studies on environmental processes affected by surface and subsurface soil textures.

Keywords: soil textural class, alluvial soil, predictive mapping, Markov chain geostatistics, transiogram, soil management

Introduction

The North China Plain is the largest plain in China and is one of the major agricultural regions. The soils in the plain are derived from recent alluvial deposits and are typically composed of layers originally deposited under flood conditions. Thus, soil texture may be very varied with depth. Moreover, rivers frequently changed their courses over time which adds complexity to the spatial structure of soils to explain why the spatial distribution of surface and subsurface soil texture in the plain is complicated (Shi *et al*., 1986). At the regional scale, soil profiles may markedly differ between different sites in terms of textural classes, sequences, numbers of textural layers and thickness, a consequence of variability in previous flood events (Li *et al*., 1997). It is therefore difficult to map soil texture accurately especially for subsurface depths from sparse field observations. Although remotely sensed data may be used to predict topsoil texture, such as clay content (Odeh & McBratney, 2000), their use for mapping subsoil textural variation is not possible (Vitharana *et al*., 2006). Soil maps in China are usually hand drawn and based on very sparse field observations; thus, they do not reveal detailed variation in soil texture. Maps showing variation in subsurface soil texture are unavailable for the North China Plain which poses a problem to the selection of appropriate precision agricultural methods.

Soil texture is a categorical variable with several classes. The non-parametric indicator kriging-based sequential indicator simulation technique has been used in simulating categorical variables from sample data (Journel & Isaaks, 1984; Bierkens & Weerts, 1994; Deutsch, 1998; Deutsch & Journel, 1998). However, this approach is essentially linear and has

Correspondence: C. Zhang. E-mail: zhangchuanrong@gmail.com
Received May 2007; accepted after revision May 2008
Editor: Donald Davidson

proved difficult for incorporating interclass relationship information for practical purposes (Goovaerts, 1997; Bogaert, 2002; McBratney *et al*., 2003). It usually generates dispersed patterns (i.e. short-range variations and noise) and geologically unrealistic transitions between categories in simulations (Deutsch, 1998). Interclass relationships, such as cross-correlations, neighbouring relationships and directional asymmetries, are important aspects of heterogeneity and ignoring or insufficiently incorporating them in a simulation means that the results are not based on all spatial information as derived from the samples.

Markov chain geostatistics (MCG) has been proposed as a non-parametric approach with nonlinear estimators and the ability to incorporate interclass correlations (Li, 2007a). MCG refers to Markov chain models, the transiogram and related algorithms based on the recently proposed Markov chain random field (MCRF) theory (Li & Zhang, 2007). In this paper, concern is with multiple nominal classes which are difficult to cope with by other methods. MCG has distinct advantages over indicator kriging for simulating categorical variables. For example, it can objectively reduce spatial uncertainty in predicted results and generate more realistic patterns given the same sample data sets (Li & Zhang, 2007; Zhang & Li, 2007). Another special characteristic of MCG is that it can generate polygonal patterns similar to conventional area-class mapping.

We have used MCG to simulate the spatial distribution of textural classes in surface and subsurface soils in a typical alluvial soil in the North China Plain. The objectives were to: (i) analyse the spatial distribution of textural classes in alluvial soils in the North China Plain; and (ii) develop a practical method for mapping surface and subsurface soil textural distributions from sparse samples.

Methods and materials

Markov chain random field

The theoretical foundation of MCG is the MCRF theory as described by Li (2007a) which extends Markov chains into any dimension for geospatial modelling. The spatial measure for MCG is the transiogram which is based on transition probabilities in space for characterizing inter- and intra-correlations of classes.

Let Z be a categorical random variable with n classes, defined in a state space $S = (1, 2,\ldots, n)$ with n different states, and z be a specific state (e.g. textural type) of Z at a specific location. A transiogram is defined as a function of a transition probability over a continuous lag $\mathbf{h}$:

$$p_{ij}(\mathbf{h}) = \Pr(z(\mathbf{x}+\mathbf{h}) = j | z(\mathbf{x}) = i), \tag{1}$$

where $\mathbf{x}$ represents any specific location. An auto-transiogram $p_{ii}(\mathbf{h})$ represents the self-dependence (i.e. auto-correlation) of a single class i and a cross-transiogram $p_{ij}(\mathbf{h})$ ($i \neq j$) represents the cross-dependence of class j on class i. Here class i is called a head class and class j is called a tail class.

For two-dimensional spatial simulation, if only nearest data locations (called *nodes*) in four cardinal directions are considered, then the conditional probability distribution of Z to have a state k at an unobserved location $\mathbf{x}$ in a MCRF can be simply written as

$$p[\mathbf{x}, k|(N)] = \Pr[z(\mathbf{x}) = k | z(\mathbf{x}_1) = l, z(\mathbf{x}_2) = m, z(\mathbf{x}_3) = q, z(\mathbf{x}_4) = r] = \frac{p^4_{kr}(\mathbf{h}_4) p^3_{kq}(\mathbf{h}_3) p^2_{km}(\mathbf{h}_2) p^1_{lk}(\mathbf{h}_1)}{\sum_{f=1}^{n} [p^4_{fr}(\mathbf{h}_4) p^3_{fq}(\mathbf{h}_3) p^2_{fm}(\mathbf{h}_2) p^1_{lf}(\mathbf{h}_1)]}, \tag{2}$$

where N represents the conditioning data; any $p_{ij}(\mathbf{h})$ represents a transition probability from state i to state j with a lag $\mathbf{h}$; 1, 2, 3 and 4 represent the four cardinal directions; $\mathbf{h}_1$, $\mathbf{h}_2$, $\mathbf{h}_3$ and $\mathbf{h}_4$ represent the distances from the current location $\mathbf{x}$ being estimated to its nearest known neighbours $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ in the four cardinal directions, respectively, and k, l, m, q and r represent the states of the Markov chain at the five locations $\mathbf{x}$, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$, respectively, all defined in a state space $S = (1, 2, \ldots, n)$. In directions 2, 3 and 4, transitions are from the current location $\mathbf{x}$ to its nearest known neighbours, but in direction 1 (i.e. the coming direction of the Markov chain) the transition is from the nearest known neighbour $\mathbf{x}_1$ to the current location $\mathbf{x}$ to be estimated. From equation (2), we can see that transiograms are necessary to provide transition probabilities at required lags to the MCRF model for estimating the conditional probability distribution of a random variable at a location. In this study, equation (2) (for four nearest known neighbours) and its further simplified forms (for less than four nearest known neighbours) were used in the simulation algorithm.

Transiogram estimation and modelling

In practice, an experimental transiogram is estimated directly from sample data by counting the number of transitions from a class to itself or another class over a sequence of lags (e.g. numbers of pixel length for raster data) using the following equation:

$$\hat{p}_{ij}(\mathbf{h}) = F_{ij}(\mathbf{h}) / N_i(\mathbf{h}), \tag{3}$$

where $N_i(\mathbf{h}) = \sum_{j=1}^{n} F_{ij}(\mathbf{h})$ is the total of elements in the ith row in a transition frequency matrix; $F_{ij}(\mathbf{h})$ represents the number of transitions from class i to class j at the lag $\mathbf{h}$ and n is the total number of classes. For sparse samples, to acquire reliable experimental transiograms a lag tolerance of Δh is set around the lag $\mathbf{h}$, which may be decided by users according to the density of samples. If the variable Z is anisotropic, experimental transiograms have to be estimated directionally with a tolerance angle similar to estimation of variograms.

Transiogram models for simulation use may be acquired through joint model fitting of experimental transiograms with expert knowledge (Li, 2007b). This approach is widely applicable but relatively time consuming, and it is also difficult to fit the complex shapes of experimental transiograms. An efficient method is to use a method to interpolate experimental transiograms into continuous models (Li & Zhang, 2007). This method is reliable when samples are sufficient to generate reliable experimental transiograms. The linear interpolation method is expressed in the following equation:

$$p_{ij}(\mathbf{h}) = \frac{\hat{p}_{ij}(\mathbf{h}_k)(\mathbf{h}_{k+1} - \mathbf{h}) + \hat{p}_{ij}(\mathbf{h}_{k+1})(\mathbf{h} - \mathbf{h}_k)}{\mathbf{h}_{k+1} - \mathbf{h}_k} \quad (4)$$

where $\hat{p}_{ij}(\mathbf{h}_k)$ and $\hat{p}_{ij}(\mathbf{h}_{k+1})$ are two neighbouring estimated values in an experimental transiogram; $\mathbf{h}_{k+1}$ and $\mathbf{h}_k$ are the corresponding lags of the two neighbouring estimated values with $\mathbf{h}_{k+1} > \mathbf{h}_k$ and $p_{ij}(\mathbf{h})$ is the value to be interpolated (or estimated) at the lag $\mathbf{h}$ between $\mathbf{h}_{k+1}$ and $\mathbf{h}_k$.

Transiogram modelling for Markov chain simulation of categorical variables must meet three constraint conditions – non-negative, no nugget and transition probabilities with a common head class summing to 1 at used lags (Li, 2007b). The transition probabilities obtained through equation (4) can meet all three constraint conditions at used lags and are therefore valid. However, the third condition may not be met at very high lags (e.g. lags close to the extent of the simulation domain), which are normally unused in random-path simulation algorithms, but may be used in some fixed-path simulation algorithms. In the case where the third condition cannot be met, it is proper to set the sills (or heights) of interpolated transiogram models at the high-lag section to the proportions of their corresponding tail classes, as suggested in Zhang & Li (2008). The other way to exactly ensure the third condition is to simply infer one transiogram model from others in each subset with a common head class by the following equation

$$p_{ik}(\mathbf{h}) = 1 - \sum_{\substack{j=1 \\ j \neq k}}^{n} p_{ij}(\mathbf{h}) \quad (5)$$

Thus, the experimental transiogram $\hat{p}_{ik}(\mathbf{h})$ does not need to be interpolated. This method was proposed for joint modelling of experimental transiograms using mathematical models (Li, 2007b). It may be used, but is normally unnecessary when equation (4) is used. In this study, we used equation (4) to obtain transiogram models from experimental transiograms estimated from sample data.

Simulation algorithm

For simulation, the classes are labelled as 1, 2,..., n in an arbitrary sequence, thus forming a state space S = (1, 2, ..., n). These numbers do not justify magnitudes. The random-path MCG simulation algorithm for point samples is called Markov chain sequential simulation (MCSS) (Li & Zhang, 2007). In simulation, the four cardinal directions are replaced by the four sectors of a search circle (i.e. quadrants), each covering a quarter of the search circle. Thus, in each sector a nearest known neighbour (labelled node, including sampled locations and previously simulated locations) is searched. The final nearest known neighbours found in the four sectors constitute the specific Markov chain model for estimating the unlabelled node. In the case where no labelled node can be found in some sectors (e.g. at the beginning of simulation or on boundaries), the number of nearest known neighbours found may be less than four (Li & Zhang, 2007). Choice of a search radius is a decision a user makes based on sample density, but the search radius should not be so small that the search circle covers no nearest known neighbour frequently at the initial stage of a simulation.

The simulation domain is divided into N nodes (or pixels) on a regular grid, and all nodes at sampled locations are labelled. The MCSS algorithm includes the following steps:

Step 1. Select a search radius for a simulation and split all the nodes of the simulation domain into two sets, the labelled and unlabelled nodes.

Step 2. Randomly pick a node from the unlabelled node set.

Step 3. Search for at most four nearest labelled nodes within the search circle, one from each sector of the search circle if there are labelled nodes in the sector.

Step 4. Estimate the conditional probability distribution of the state of the unlabelled node using the corresponding Markov chain model provided in equation (1) or its simplified forms, depending on the number of nearest known neighbours.

Step 5. Draw a specific state (i.e. label) for the unlabelled node from the cumulative conditional probability distribution by Monte Carlo sampling.

Step 6. Add the newly simulated node to the labelled node set for conditioning in subsequent simulations of other unlabelled nodes and delete it from the unlabelled node set.

Step 7. Repeat steps 2–6 until all unlabelled nodes are visited and every unlabelled node has been assigned a simulated value (i.e. a label).

Study site and sample data

The study site (36°51′N, 115°3′E) has an area of about 15 km^2 at Quzhou Experimental Station, China Agricultural University, Quzhou County, Hebei Province. The topography is virtually flat with an elevation from 34.5 to 37.3 m; most of the area is a slight depression on an alluvial fan of the Zhang River. A tributary of the Zhang River crosses the middle of the area from north to south. An old course of the Zhang River passes through the west part of the area

from north to south (Shi *et al.*, 1986; Li *et al.*, 1997). Soils are saline to varying degrees. Variation in soil texture is typical of alluvial soils with three to six textural classes within the first 2 m and the surface texture is dominantly sandy loam, clay and light loam.

In all, 142 sample points were selected on a triangular grid with a sampling interval of 350 m (Figure 1); three points had to be excluded because they occurred within villages. Soil textures were recorded by excavating pits to a depth of 2 m. Differences in soil texture in each soil profile were recorded by field texturing supported by laboratory analysis of a small number of soil samples from previous soil survey data (Li *et al.*, 1999). Although the determination of soil texture at many sampling locations and at different depths is costly, the results are of long-term value because changes in soil texture below the plough layer are likely to be minimal.

The initial soil textural classes were sand, sandy loam, light loam, medium loam, heavy loam and clay according to the soil textural system used in China (Li *et al.*, 1999). However, because the incidence of heavy loam was very small, this textural class was combined with medium loam. Thus, we used five textural classes – sand, sandy loam, light loam, medium loam and clay, indicated by numbers 1, 2, 3, 4 and 5, respectively. We chose five soil depths for modelling the spatial distribution of textural classes in the horizontal two dimensions – 0 (i.e. surface), 50, 100, 150 and 200 cm. For each depth, we had data for 139 sample points. The surface soil texture essentially represents the texture of the ploughed layer in which any differences in texture had been destroyed. The ploughed layer is usually 30 cm deep; thus, the texture at 50-cm depth is below this zone. The textures below the plough zone affect water and solute transport to groundwater and the growth of deep-rooted crops.

For simulation, the area was subdivided into a 239 × 97 grid with an individual pixel size of 25 × 25 m^2. Omni-directional experimental transiograms were estimated from sample data

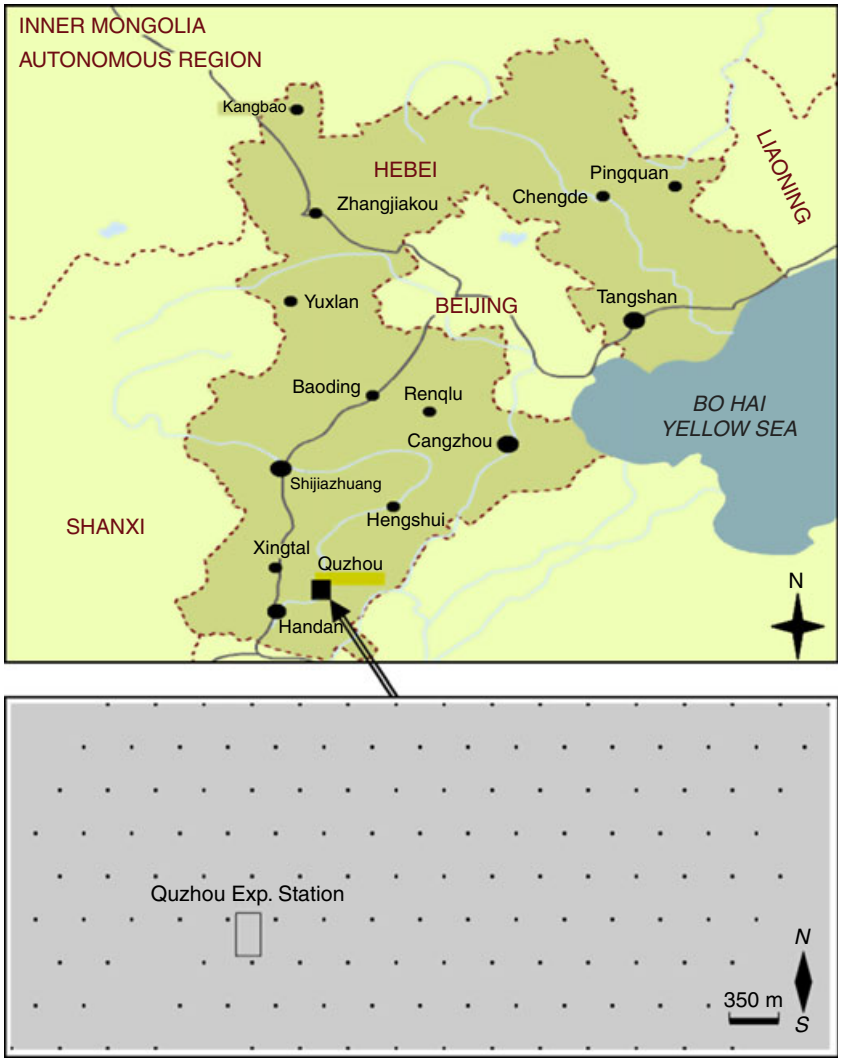

Figure 1 The study site and locations of observed soil profiles in the study area.

for each of the five sample depths and interpolated into continuous models using equation (4) for use in simulations. A lag tolerance of six pixel lengths was chosen for estimating experimental transiograms. All observed data from the 139 points were used as conditioning data in simulations.

Results

Auto- and cross-correlations

Figure 2 displays a subset of the omni-directional experimental transiograms and interpolated models with head class 3 (the light loam class) for 50-cm depth, and Figure 3 shows the same with head class 5 (clay) for 100-cm depth. Because the sample size was only 139, experimental transiograms are irregular and have only a small number of estimated values (nine values within the 2500 m lag). In addition, periodicities appeared in some experimental transiograms, particularly those with head classes 2, 3 and 4. However, we ignored them in estimating correlation ranges because of their usual irregularity with shallow amplitudes. We approximately inferred the auto- and cross-correlation ranges of classes from those experimental transiograms. Table 1 provides the approximate correlation ranges of classes at 100-cm depth.

An auto-transiogram represents the auto-correlation of a single class and its correlation range to some extent also reflects the boundary spacing (polygon length or width) of the class. At 100-cm depth, it can be seen from Table 1 that clay and sand have long auto-correlation ranges of 900 and 1000 m, respectively. This means that they occur in larger patches and this will result in larger boundary spacings in simulations. Other classes have a shorter auto-correlation range of about 500 m which means they occur in smaller patches. Cross-transiograms indicate the interdependencies and asymmetries between classes. Short cross-correlation ranges usually occur between a class with large patches and neighbouring class with small patches, and long cross-correlation ranges normally occur between classes that are often located distantly. At 100-cm depth, light loam and clay have the longest cross-correlation range of 1200 m; sandy loam and clay have a short cross-correlation range of 500 m.

Sills of transiograms or heights of experimental transiograms at the lags beyond correlation ranges are reflections of tail class proportions. The sills can be approximated from the experimental transiograms in Figures 2 and 3 and they are close to the corresponding tail class proportions provided in Table 2. Because we did not infer transiogram models by using mathematical models to fit experimental transiograms, no exact parameters (sill, range and model type) of transiogram models were used in this study.

Optimal prediction

For each of the five sample data sets, 100 simulated maps were generated which were used to estimate occurrence probabilities of each class at all locations. For each sample

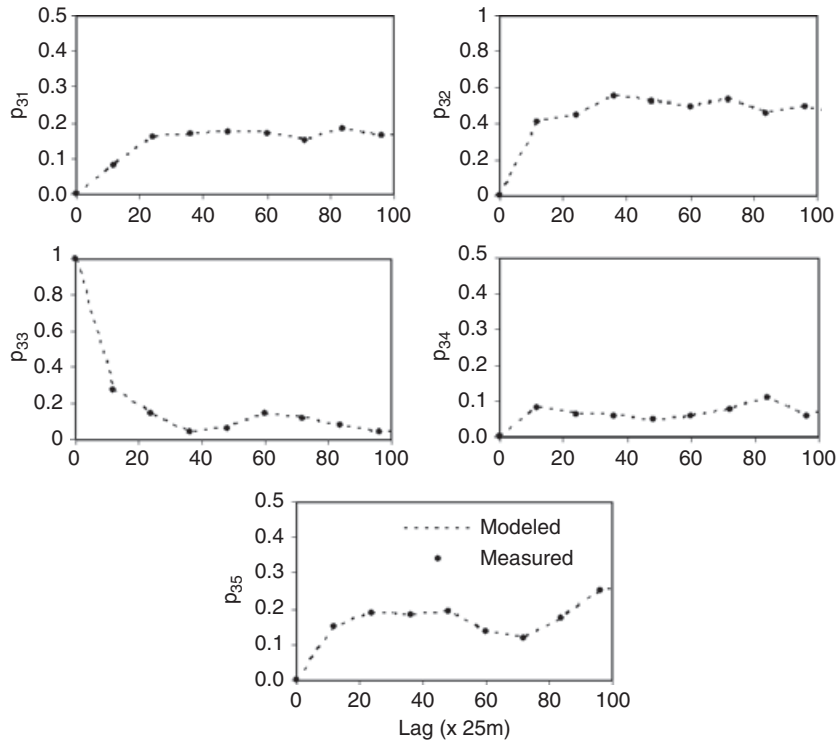

Figure 2 Experimental transiograms and interpolated models with head class 3 (light loam), estimated from the sample data of soil texture at 50-cm depth. The vertical ordinate represents transition probabilities (e.g. p_{31} refers to transition probabilities from class 3 to class 1).

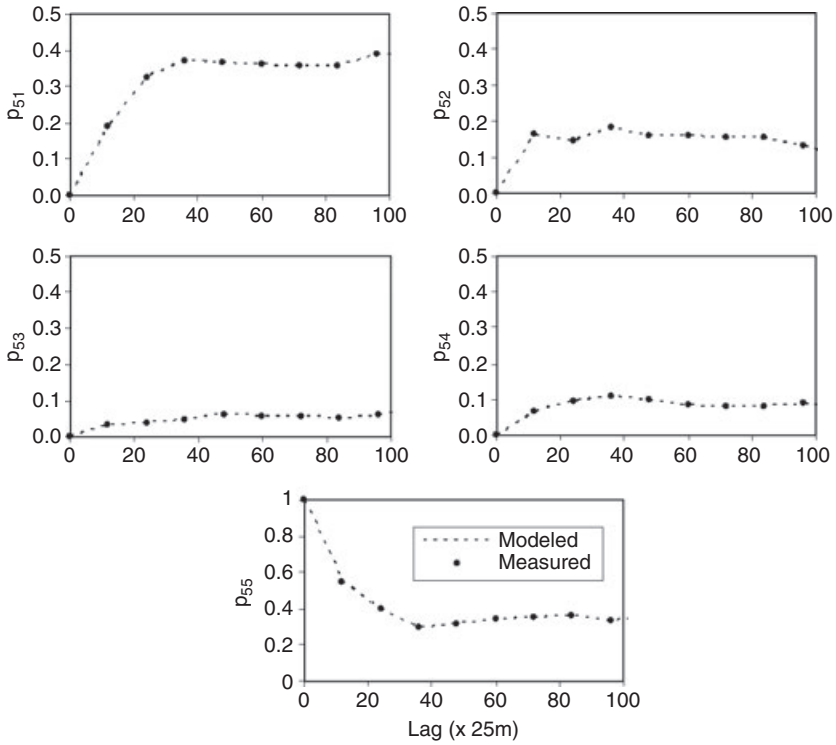

Figure 3 Experimental transiograms and interpolated models with head class 5 (clay), estimated from the sample data of soil texture at 100-cm depth. The vertical ordinate represents transition probabilities (e.g. p_{51} refers to transition probabilities from class 5 to class 1).

Table 1 Approximate auto- and cross-correlation ranges of different soil textural classes inferred from experimental transiograms estimated from the soil textural sample data set at 100-cm depth

	Approximate correlation range (m)				
Soil textural class	1 Sand	2 Sandy loam	3 Light loam	4 Medium loam	5 Clay
1. Sand	900	400	500	500	900
2. Sandy loam	400	500	500	400	500
3. Light loam	500	500	500	700	1200
4. Medium loam	500	400	700	500	900
5. Clay	900	500	1200	900	1000

data set, an optimal prediction map was further acquired by assigning class labels to unobserved locations where corresponding classes have maximum occurrence probabilities. This represents a way of optimal prediction which differs from interpolation and is practical for efficient simulation algorithms such as MCSS. Figure 4 displays the optimal prediction maps for soil textural classes at five depths.

From Figure 4, it can be seen that there is no textural class 'sand' in the surface layer and 'sandy loam' occupies most of the area in the central and eastern parts. One reason may be that no surface soil is sand and another may be that tillage and manuring have homogenized the texture of surface soils even if there was sand previously at some places. Clay and medium loam occur at the surface in the western part of the study area and several small areas of light loam occur within the larger one of sand loam.

At 50-cm depth, sandy loam is still the dominant textural class (Figure 4), but areas of sand and medium loam also occur. At this depth, the soil texture has effects on root extension and nutrient leakage. At 100-cm and deeper depths, sand becomes the dominant textural class and clay the second with other textural classes occurring in small patches.

Spatial pattern and uncertainty

Although optimal prediction maps show which class has the maximum probability to occur at a location, they have some common deficiencies: (i) they cannot characterize the real variation (or patterns) of soil textural classes such as the spatial correlation observed in the data; (ii) they tend to underestimate weakly auto-correlated (small-patch) classes and overestimate strongly auto-correlated (large-patch) classes because of the accompanying smoothing effect, and (iii) they do not provide a visual and quantitative measure of the uncertainty of the estimated variables. These characteristics are commonly shared by various geostatistical approaches. However, problems (i) and (ii) are not evident in our study (Table 2) which may be partially related to the incorporation of interclass relationships in MCG.

Table 2 Estimated areal proportions of different soil textural classes at different depths in the study area

	Soil textural class														
	Sand			Sandy loam			Light loam			Medium loam			Clay		
Depth (cm)	SA	RE	OP	SA	RE	OP	SA	RE	OP	SA	RE	OP	SA	RE	OP
0	0.00	0.00	0.00	0.77	0.73	0.75	0.11	0.11	0.08	0.04	0.08	0.08	0.08	0.08	0.08
50	0.15	0.16	0.13	0.50	0.46	0.52	0.09	0.11	0.10	0.06	0.07	0.06	0.19	0.20	0.19
100	0.36	0.36	0.41	0.14	0.16	0.12	0.06	0.07	0.05	0.09	0.10	0.07	0.35	0.33	0.34
150	0.45	0.40	0.43	0.06	0.08	0.07	0.05	0.06	0.05	0.14	0.13	0.10	0.31	0.33	0.36
200	0.50	0.47	0.53	0.02	0.03	0.03	0.07	0.09	0.08	0.12	0.13	0.10	0.29	0.28	0.26

SA, estimated from sample data for each depth; RE, estimated from 100 simulations for each depth; OP, estimated from the corresponding optimal prediction map for each depth.

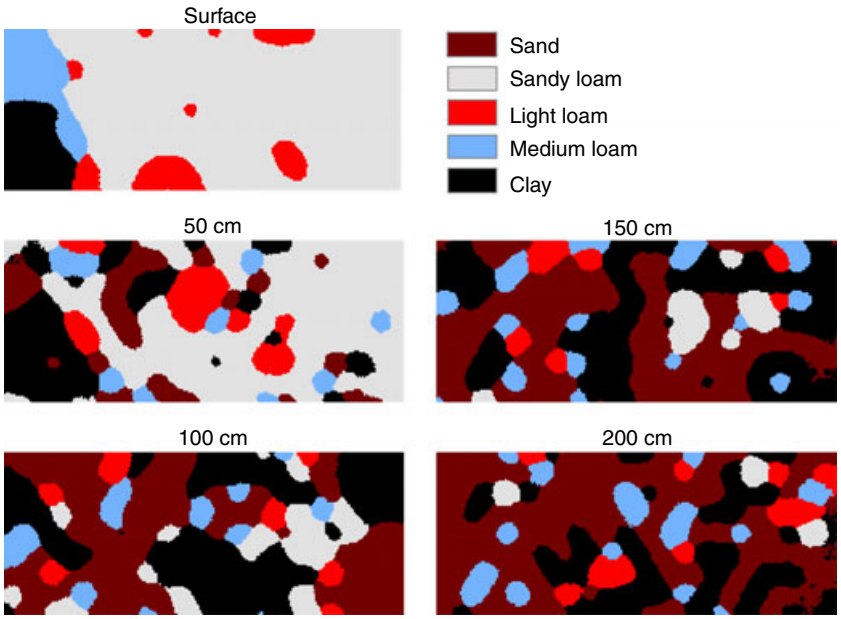

Figure 4 Optimal prediction maps of soil textural classes at five depths, based on maximum occurrence probabilities estimated from 100 simulations.

Figure 5 displays three simulations of textural class distribution for each of the five soil depths. Each map shows one possible spatial distribution of soil textural classes at the corresponding depth under the spatial uncertainty caused by incomplete observation. It can be seen that these spatial patterns are more complex than the corresponding optimal maps shown in Figure 4. It can also be seen that the general sizes of boundary spacings of single classes are in accordance with their corresponding auto-correlation ranges and the spatial distribution structures of class pairs are obviously related to their cross-correlation ranges. For example, at 100-cm depth, clay and sand have large boundary spacings which correspond to their long auto-correlation ranges (Table 1) and clay and sandy loam tend to occur as close neighbours which accords with their short cross-correlation range.

While multiple simulations may demonstrate spatial uncertainty in patterns, occurrence probability maps of classes can represent spatial uncertainty more efficiently, more accurately and in a clearer manner. Figures 6 and 7 provide single-class occurrence probability maps and maximum occurrence probability maps of soil textural classes at depths of 50 and 100 cm. The occurrence probability map of a class clearly indicates the estimated occurrence probability of the class at every location. It can be seen that at 50-cm depth, sandy loam dominates and occurs mainly in the central and the eastern parts, and clay appears mainly in the west (Figure 6). However, at 100-cm depth, sand and clay dominate, whereas the former tends to occur in the west and the latter tends to occur in the east (Figure 7). These single-class occurrence probability maps may be directly used as data input in risk assessment and uncertainty analysis in the modelling of affected variables. The maximum occurrence probability maps can be used to assess the quality of corresponding optimal prediction maps.

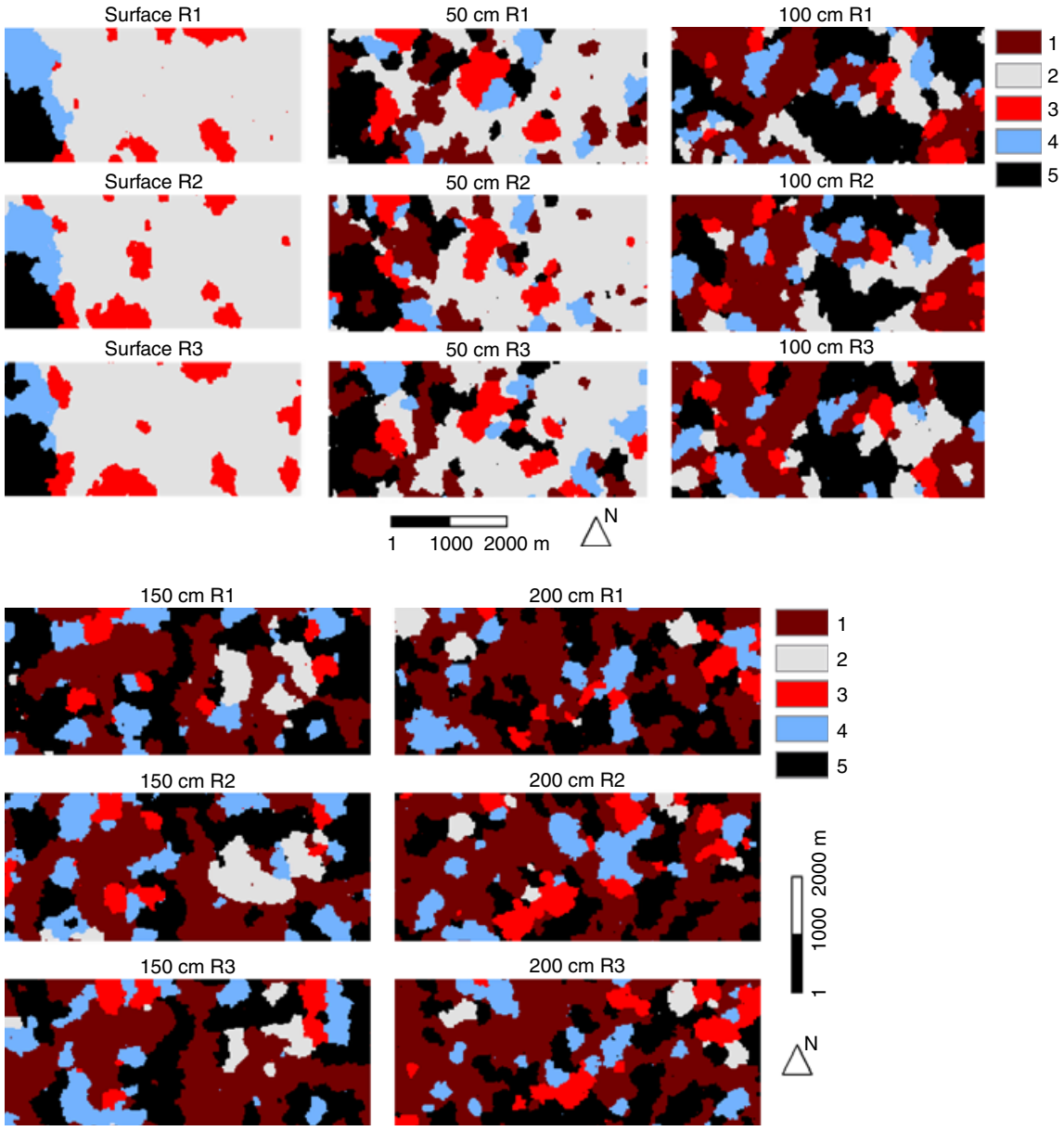

Figure 5 Simulated maps of soil textural classes at five depths. Numbers 1, 2, 3, 4 and 5 in the legend refer to sand, sandy loam, light loam, medium loam and clay, respectively. R1, R2 and R3 in labels refer to three different simulated maps.

A larger average maximum probability generally means that the optimal prediction map has higher quality (i.e. less uncertainty) (Li & Zhang, 2007).

Conclusions

We applied the random-path MCG simulation algorithm to simulate the spatial distribution of textural classes of alluvial soils at five different depths in an area of 15 km^2 in the North China Plain. Transiograms were used to estimate the auto-correlation of single textural classes and the interclass relationships between different textural classes. The simulations proved to accurately reflect actual variations in soil textural classes at different depths. Uncertainty in the spatial distribution of different soil textural classes was quantified and demonstrated by occurrence probability maps. The simulations showed that sandy loam dominates the upper soils (surface and 50 cm depth), whereas sand and clay dominate the lower soils (100-, 150- and 200-cm depths). The results aid the understanding of the spatial structure of soil texture in alluvial soils and also provide valuable quantitative information for precision agriculture, site-specific soil management and studies on environmental processes which are affected by surface and subsurface soil textures.

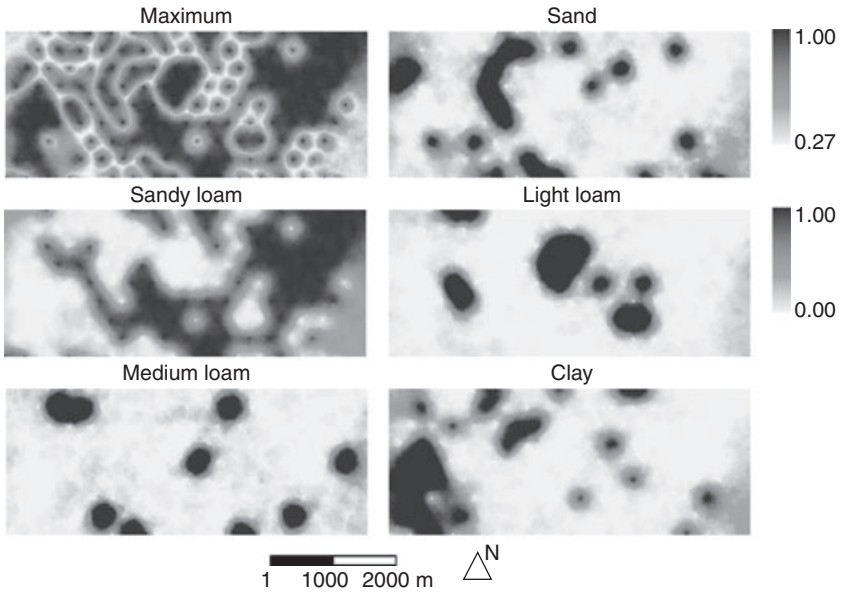

Figure 6 Occurrence probability maps of soil textural classes at 50-cm depth.

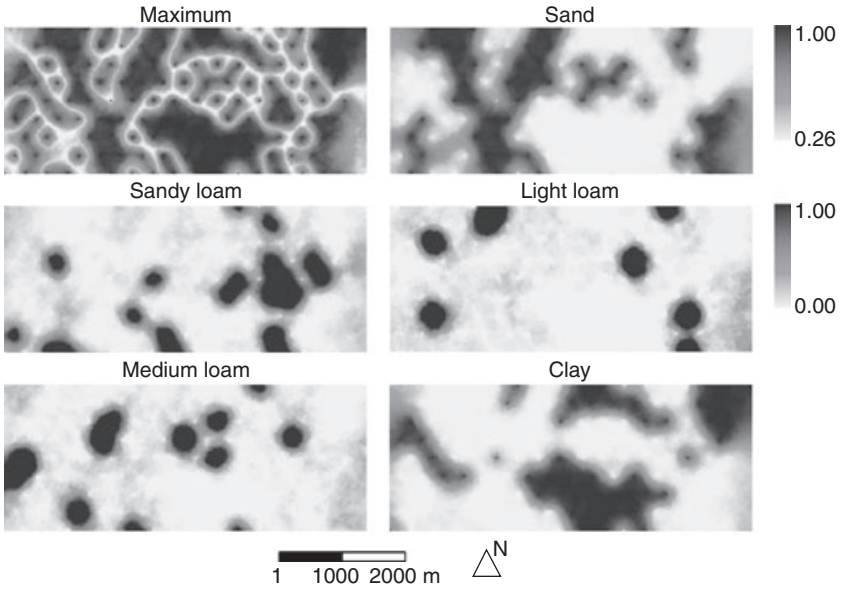

Figure 7 Occurrence probability maps of soil textural classes at 100-cm depth.

Acknowledgements

We thank the Editor, Professor Donald Davidson, for his constructive comments and kind revision, and the anonymous reviewers for their constructive comments.

References

Bierkens, M.F.P. & Weerts, H.J.T. 1994. Applications of indicator simulation to modeling the lithological properties of a complex confining layer. *Geoderma*, **62,** 265–284.

Bogaert, P. 2002. Spatial prediction of categorical variables: the Bayesian maximum entropy approach. *Stochastic Environmental Research and Risk Assessment*, **16,** 425–448.

Deutsch, C.V. 1998. Cleaning categorical variable (lithofacies) realizations with maximum a-posteriori selection. *Computers and Geosciences*, **24,** 551–562.

Deutsch, C.V. & Journel, A.G. 1998. *GSLIB: geostatistics software library and user's guide*. Oxford University Press, New York.

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

Journel, A.G. & Isaaks, E.H. 1984. Conditional indicator simulation: application to a Saskatchewan uranium deposit. *Journal of the International Association of Mathematical Geology*, **16,** 685–718.

Li, W. 2007a. Transiograms for characterizing spatial variability of soil classes. *Soil Science Society of America Journal*, **71,** 881–893.

Li, W. 2007b. Markov chain random fields for estimation of categorical variables. *Mathematical Geology*, **39,** 321–335.

Li, W. & Zhang, C. 2007. A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. *Soil Science Society of America Journal*, **71,** 656–668.

Li, W., Li, B., Shi, Y. & Tang, D. 1997. Application of the Markov chain theory to describe spatial distribution of textural layers. *Soil Science*, **162,** 672–683.

Li, W., Li, B. & Shi, Y. 1999. Markov-chain simulation of soil textural profiles. *Geoderma*, **92,** 37–53.

McBratney, A.B., Mendonca Santos, M.L. & Minasny, B. 2003. On digital soil mapping. *Geoderma*, **117,** 3–52.

Odeh, I.O.A. & McBratney, A.B. 2000. Using AVHRR images for spatial prediction of clay content in the lower Namoi Valley of Eastern Australia. *Geoderma*, **97,** 237–254.

Shi, Y.C., Li, Y.Z. & Lu, J.W. 1986. *Water and salt movement in salinized soils*. Beijing Agricultural University Press, Beijing (in Chinese).

Vitharana, U.W.A., Van Meirvenne, M., Cockx, L. & Bourgeois, J. 2006. Identifying potential management zones in a layered soil using several sources of ancillary information. *Soil Use and Management*, **22,** 405–413.

Zhang, C. & Li, W. 2007. Comparing a fixed-path Markov chain geostatistical algorithm with sequential indicator simulation in categorical variable simulation from regular samples. *GIScience & Remote Sensing*, **44,** 251–266.

Zhang, C. & Li, W. 2008. A comparative study of nonlinear Markov chain models for conditional simulation of multinomial classes from regular samples. *Stochastic Environmental Research and Risk Assessment*, **22,** 217–230.