Math Geosci (2015) 47:123–148
DOI 10.1007/s11004-014-9553-y

Bayesian Markov Chain Random Field Cosimulation for Improving Land Cover Classification Accuracy

Weidong Li · Chuanrong Zhang · Michael R. Willig · Dipak K. Dey · Guiling Wang · Liangzhi You

Received: 23 February 2014 / Accepted: 28 July 2014 / Published online: 26 August 2014

Abstract This study introduces a Bayesian Markov chain random field (MCRF) cosimulation approach for improving land-use/land-cover (LULC) classification accuracy through integrating expert-interpreted data and pre-classified image data. The expert-interpreted data are used as conditioning sample data in cosimulation, and may be interpreted from various sources. The pre-classification can be performed using any convenient conventional method. The approach uses the recently suggested MCRF cosimulation algorithm (Co-MCSS) to take a pre-classified image as auxiliary data while performing cosimulations conditioned on expert-interpreted data. It was tested using a series of expert-interpreted data sets and an image data set pre-classified by the supervised maximum likelihood (SML) algorithm. Results show that with the density

W. Li (✉) · C. Zhang
Department of Geography, University of Connecticut, Storrs, CT 06269, USA
e-mail: weidong.li@uconn.edu; weidongwoody@gmail.com

C. Zhang · M. R. Willig · G. Wang
Center for Environmental Sciences and Engineering, University of Connecticut, Storrs, CT 06269, USA

M. R. Willig
Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA

D. K. Dey
Department of Statistics, University of Connecticut, Storrs, CT 06269, USA

G. Wang
Department of Civil and Environmental Engineering, University of Connecticut, Storrs, CT 06269, USA

L. You
International Food Policy Research Institute, Washington, DC 20006, USA

of the interpreted data (pixel labels) increasing from 0 to 1.81 % of total pixels, the accuracy of optimal classification maps from Co-MCSS improves by 8.49 to 20.96 %, being much higher than that generated by SML and those purely conditioned on expert-interpreted data. This means that expert-interpreted data may largely contribute to the accuracy of LULC classification from remotely sensed imagery, and inversely the pre-classified image data also can largely contribute to the accuracy of LULC classes simulated by the MCRF approach based on expert-interpreted data. Therefore, the proposed approach can improve classification accuracy of pre-classified maps as long as some expert-interpreted data are available. The main advantage of this approach is that it may comprehensively utilize a variety of available information seamlessly to LULC classification through expert interpretation and spatial cosimulation.

Keywords Land cover classification · Remotely sensed image · Expert interpretation · Bayesian inference · Cosimulation · Markov chain · Geostatistics

1 Introduction

Land use/land cover (LULC) maps are the basic data that are widely used for social or environmental planning and studies. LULC changes, due to their significant impact on regional and global climate changes, are important forcing factors to include in both past climate simulation and future climate change predictions. Therefore, the accuracy of LULC maps and associated uncertainty data are crucial to the success of those activities using LULC maps as input data. However, due to a variety of reasons such as spectral confusion caused by high intra-class variability and low inter-class separability, clouds contamination, landscape complexity and/or algorithm limitations, LULC maps derived from remotely sensed imagery contain a large amount of errors (or uncertainties) and may not have a quality sufficient for many applications (Foody 2002; Gong and Howarth 1990). Consequently, it is necessary to explore effective approaches for improving LULC classification accuracy. During last two decades, significant effort has been devoted to developing various classification techniques for improving classification performance (Lu and Weng 2007), such as neural networks (Benediktsson et al. 1990), genetic algorithms (Tseng et al. 2008), Markov random fields (Solberg et al. 1996), and support vector machines (Foody and Mathur 2004).

The reasons for low quality of many LULC maps classified from remotely sensed imagery are usually complex. A pervasive reason may be that many scientists accomplish classification by only using one variable, that is, spectral characteristic (color) or black and white tone (Jensen 2005). In addition, some methods are not capable of sufficiently dealing with the complexity and low quality of many remotely sensed images caused by reasons mentioned above, or are too sophisticated for use by non-experts. An intuitive approach is to use expert knowledge, because experts may integrate a variety of available information (particularly visible patterns on images and ancillary data) to judge to which class a pixel or small area should belong. Therefore, expert knowledge should be incorporated into remotely sensed image classification for LULC mapping. The key challenge is how to incorporate expert knowledge in a practical and feasible manner.

Recently, Li and Zhang (2011) suggested using a Markov chain random field (MCRF) method for LULC classification with uncertainty assessment. The method used a Markov chain sequential simulation (MCSS) algorithm (Li and Zhang 2007) to conduct conditional stochastic simulations on expert-interpreted pixel label data. After experts interpreted a data set of pixels that are distinguishable from high resolution images and other available information, the remaining uninformed pixels were treated as uncertain areas waiting for further processing through conditional simulation. The aim was to make use of both expert knowledge and context information in remotely sensed LULC classification. Different from specific expert systems that usually reclassify pre-classified images with expert-made decision rules into LULC maps (e.g., Stefanov et al. 2001; Huang and Jia 2012), the MCRF method only uses expert-interpreted pixel class label data as sample data in conditional simulation for LULC classification; expert interpretation itself does not produce LULC maps. Experts may use various ways and implicit rules to interpret sample pixel data from remotely sensed images and other available information based on personal judgment. How to interpret sample data confidently is not the concern of the method. A case study showed that a small percentage of interpreted pixels might be able to generate a high quality land cover map. This explorative study suggested that it is crucial to integrate expert knowledge with multi-source information into one comprehensive framework for LULC classification.

Although still not popular in real application practice, studies on applications of geostatistics to remotely sensed imagery have been conducted since 1980s, and there is a large literature in this subject, as reviewed recently by Van der Meer (2012). There are a few studies on using geostatistical methods to improve land cover classification accuracy pre-classified by conventional methods such as the classical supervised maximum likelihood (SML) algorithm. For example, De Bruin (2000) used a co-located indicator cokriging based sequential indicator simulation algorithm to predict the areal extent of olive vegetation through integrating soft indicator data from classified imagery and hard data interpreted from digital color orthophotography. Carvalho et al. (2006) used a direct sequential cosimulation algorithm to improve accuracy of land cover classification by integrating ground data and remotely sensed land cover proportional data classified by the SML algorithm. Other applications of conventional geostatistics in remote sensing studies include, for example, deriving image texture measures to aid classification (Balaguer et al. 2010), evaluating large area forest inventory (Meng et al. 2009), and restoring missing or contaminated pixels (Zhang et al. 2014). Recently, Tang et al. (2013) used the multiple-point geostatistical algorithm (Strebelle 2002) to post-process a remotely sensed land cover classification generated by SML for accuracy improvement by taking the pre-classified image as a training image, the SML probabilities as soft conditioning data and the training samples as hard conditioning data. Other applications of multiple-point geostatistics to remote sensing may include sub-pixel mapping of coarse satellite remote sensing images (Boucher 2009) and extracting linear objects from remotely sensed imagery (Ge and Bai 2011).

The MCRF approach is a Bayesian spatial statistical method mainly for the purpose of simulating categorical fields (Li 2007a; Li et al. 2013). It provides a general framework for a new, non-linear non-kriging geostatistical approach. It is different from existing spatial statistical methods because it is based on a single spatial Markov

chain running in a space with local Bayesian updating at the neighborhood level (Li et al. 2013). The MCRF approach has the following merits: (1) it can deal with complex interclass relationships, (2) it treats all classes as labels rather than as binary values, (3) it is intuitive as a dynamic spatial Markov chain and uses transition probabilities as parameter inputs, and (4) it is nonlinear in estimators.

The basic MCSS algorithm used in Li and Zhang (2011) did not incorporate auxiliary variables. Thus, it may be excessively labor-intensive to interpret a densely distributed pixel data set in order to obtain a highly accurate classification. In addition, with only sparse sample data, it is also difficult to reproduce features with complex shapes. However, classified images using conventional methods based on spectral reflectance data may, to some extent, capture these kinds of features. Although classified LULC images using conventional methods may not be sufficiently accurate, there is no doubt that the classified data are normally strongly correlated with the expert-interpreted sample data in the same area. Thus, if we regard a classified LULC image using a conventional method to be a pre-classified image and use it as an auxiliary data set in a cosimulation conditioned on accurate expert-interpreted sample data, the classification accuracy may be greatly improved. Therefore, the recently developed Markov chain sequential cosimulation (Co-MCSS) algorithm (Li and Zhang 2012; Li et al. 2013) is employed in this paper so that a pre-classified image data set can be incorporated in a cosimulation for more accurate classification. That is, the cosimulation serves as an information updating step with new data layers (i.e., expert-interpreted sample data and spatial correlation parameters) to improve or finalize the LULC classification.

The objectives of this study are: (1) to introduce the MCRF cosimulation method for improving LULC classification accuracy by integrating pre-classified data and expert-interpreted sample data, and (2) to demonstrate the workability of the method through a simple case study with multiple sample data sets and a pre-classified image data set generated by the SML algorithm. The ultimate goal of this study is to provide a more practical and useful method for accurate LULC classification from remotely sensed imagery with ancillary information.

2 Methods

2.1 Strategy and Classification Procedure

Reclassifying pre-classified images with additional information (or rules) is a traditional strategy for improving LULC classification accuracy, especially in expert systems. The proposed MCRF cosimulation approach follows this general strategy, but its specific methodology is different, as presented below.

An expert can interpret a sparse data set of class labels about the LULC classes within a study area from a variety of related data (e.g., remotely sensed images, google earth images, current local land use maps). In the meantime, a pre-classified image data set also can be obtained from an available remotely sensed image with relatively good quality using any conventional fast classification method. Thus, a co-located cosimulation conditioned on the two data sets can be further conducted to achieve a high-accuracy LULC classification in the study area. This means that a pre-classified

image data set is used as an exhaustive auxiliary data set in our cosimulation so that the spectral information of pixels without expert interpretation is utilized for improving the remotely sensed LULC classification accuracy. In other words, an expert-interpreted sample data set is used as an additional and accurate data set in our cosimulation to update or finalize the pre-classified image data set to produce a more accurate classification. The number and physical meanings of classes in the pre-classified image data set may not be identical to those of the expert-interpreted data set. The former usually should be classified according to the spectral separability of the image; however, the latter may be decided based on the need of the study, availability of ancillary data and interpretability of the classes by experts. Here, the expert-interpreted data for cosimulation to improve the classification are different from the training data that are used in supervised pre-classification of a remotely sensed image. Although expert-interpreted data can be clusters or small areas of pixels, this kind of data is not suitable to use in parameter (i.e., transiogram models in this study) estimation without declustering; in addition, data of mixed pixels (i.e., class proportions in single pixels) are also difficult to use for parameter estimation but may be used as conditioning data in simulation.

We hypothesize that pre-classified image data using a conventional classification method (usually fast but may not be sufficiently accurate) are highly correlated to the real but unknown LULC classes, and hence may be used as an auxiliary data set to improve the classification of LULC through spatial cosimulation. Auxiliary data from other sources, such as legacy map data, if highly correlated to the LULC classes to be classified, may also be incorporated into the classification. The general framework of the MCRF cosimulation approach for improving LULC classification accuracy is illustrated in Fig. 1. Such a general approach is an improvement over Li and Zhang (2011), which did not include auxiliary data. The proposed approach in this study includes the following procedures:

1. interpret a sparse data set of pixel class labels from one or more related remotely sensed images (e.g., a time series, multiple sources) and other auxiliary data based on expert knowledge;
2. pre-classify a remotely sensed image (usually choose one with better quality among available images) of the study area using a conventional fast method such as SML to serve as the auxiliary data for cosimulation;
3. combine the interpreted pixel label data with possibly other high quality data (e.g., ground observation data) as the final sample data set, and estimate a set of transiogram models using the sample data (these transiogram models describe the spatial auto and cross correlations among land cover classes in the sample data set);
4. estimate the cross-field transition probability matrix from the sample data to the auxiliary data set (or each set of auxiliary data if multiple sets are used) (these transition rules describe the co-located cross correlations between the land cover classes in the expert-interpreted sample data and the land cover classes in the pre-classified image data);
5. conduct cosimulation conditioned on both the sample data and auxiliary data using the Co-MCSS algorithm; and

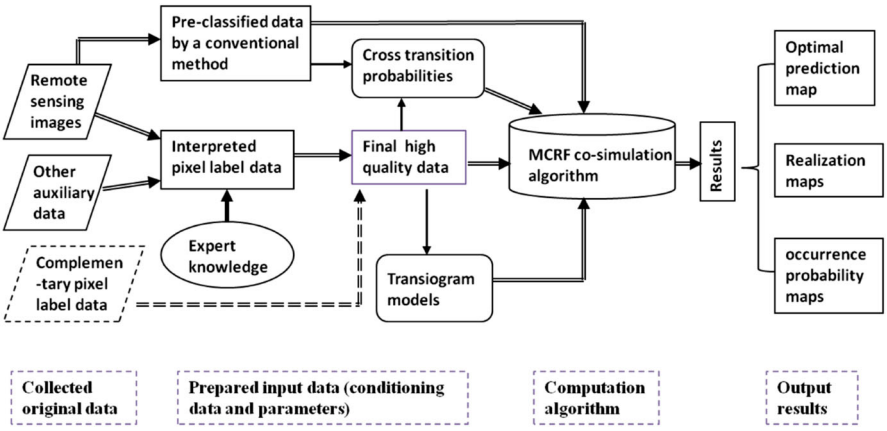

Fig. 1 Flowchart of the MCRF cosimulation approach for improving LULC classification from remotely sensed imagery with expert-interpreted data. Expert knowledge may also be used in pre-classification (e.g., in choosing training data) and parameter estimation. Complementary pixel label data refer to those obtained from other ways (e.g., ground observations)

6. analyze and visualize the simulated results into various maps, including optimal classification maps, simulated realization maps, and occurrence probability maps.

Based on a number of simulated realizations, the probabilities of each uninformed pixel belonging to different classes can be obtained. Because a variety of correlated data and image information may be used through expert interpretation or as auxiliary data, such a LULC classification approach is more comprehensive than are previous methods and may generate more accurate classification results with quantified uncertainty.

2.2 Transiograms

Theoretically, a transiogram refers to a two-point transition (or conditional) probability function over a lag distance

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

where $z(u)$ represents a realization of the random variable $z(u)$, $\mathbf{u}$ represents a specific location and $\mathbf{h}$ refers to the separate distance between the two spatial points. Under the second-order stationarity assumption, $p_{ij}(\mathbf{h})$ is dependent only on $\mathbf{h}$, not on the specific location $\mathbf{u}$. An auto-transiogram $p_{ii}(\mathbf{h})$ represents the self-dependence (i.e., autocorrelation) 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. However, experimentally a transiogram represents a series of transition probability values of the same pair of classes (i.e., states) versus corresponding spatial separations, estimated from sample data. In addition, transiograms also can be calculated from a single-step transition probability matrix if available, and such transiograms are called idealized transiograms because they are smooth curves based on the first-order stationary Markov assumption (Li 2007b).

The word "transiogram" was abbreviated from the phrase "*transi*tion probability dia*gram*". Under the inspiration of the variogram in classical geostatistics (Matheron 1963, 1973), the transiogram was initially proposed as a concept for the MCRF approach (Li 2007a) and also as a method for estimating multi-step transition probabilities or transition probability functions from sparse point sample data for two-dimensional Markov chain simulation. It also served the purpose of directly conditioning the simplified MCRF models to sparse sample data. One can see that when the simplified general solution of MCRFs (Li 2007a) was derived, the transiogram emerged naturally as an accompanying two-point spatial measure. Nevertheless, such a spatial measure is related to pioneer studies that involved transition probability curves (or functions) and model fitting of observed transition probability values (e.g., Schwarzacher 1969; Lou 1996; Carle and Fogg 1997; Ritzi 2000), particularly the works of Carle and Fogg (1996, 1997) in transition probability-based indicator geostatistics. Thus, the transiogram concept may cover all of these theoretical and experimental lagged transition probability curves. Li (2007b) addressed the concepts, types, basic properties, physical meanings, estimation, and model fitting issues of transiograms based on the Markov chain theory and pioneer studies.

2.3 MCRFs

A MCRF is a single spatial Markov chain that moves or jumps in a space (one to multiple dimensions) and decides its state at any uninformed location by interactions with its nearest neighbors in various directions. The theory of a generalised MCRF was initially proposed by Li (2007a), and further described from the Bayesian updating perspective by Li et al. (2013). It is consistent with the Bayesian inference principle (Bayes 1763; Gelman et al. 1995). Although the MCRF approach can be regarded as a Bayesian spatial statistics (or geo-statistics), it is different from other existing methods because Bayesian updating in MCRF models aims to approach the local conditional probability distribution using the evidence of nearest data within a neighborhood rather than to estimate parameters, and the updating process is performed in a single sweep. The single-chain idea of MCRFs initially resulted from the effort to revise the flaws of the coupled Markov chain model suggested by Elfeki and Dekking (2001).

For a random variable $Z(\mathbf{u})$ characterizing a MCRF, assuming i_1 to i_m are the states of the nearest known neighbors in different directions around an uninformed location $\mathbf{u}_0$, the local conditional probability distribution of $Z(\mathbf{u})$ at location $\mathbf{u}_0$ can be denoted as $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$. Here, a state in a Markov chain is equivalent to a class of land cover, and m is the number of nearest neighbors in different directions within a neighborhood. Emphasizing the single chain nature of a MCRF and its last visited location, the local conditional probability distribution of $Z(\mathbf{u})$ at current uninformed location $\mathbf{u}_0$ can be factorized as

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)] = \frac{1}{A} p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0), \ldots, i_{m-1}(\mathbf{u}_{m-1})] \cdots p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0), i_1(\mathbf{u}_1)] p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1)] \quad (2)$$

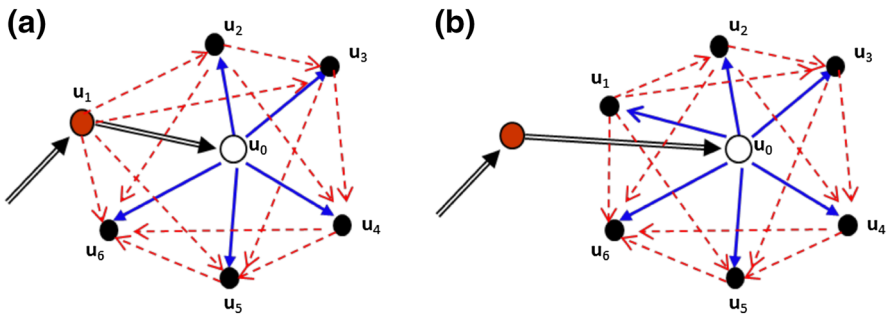

Fig. 2 Illustrations of neighborhood structures and sequential Bayesian updating processes in basic MCRFs: **a** assuming $\mathbf{u}_1$ to be the last visited location; **b** assuming the last visited location is far away (outside the neighborhood). Data interactions across the uninformed location $\mathbf{u}_0$ being estimated are ignored based on the Markov property

based on Bayes' theorem (Bayes 1763), where $A = p[i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]/p[i_1(\mathbf{u}_1)]$ is a constant, and $\mathbf{u}_1$ is the last visited location or the location that the spatial Markov chain goes through to the current location $\mathbf{u}_0$ (Fig. 2a) (Li 2007a). Equation (2) is the full general solution of the local conditional probability distribution of MCRFs. Here, $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m)]$ is the posterior probability distribution, $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1)]$ is the prior probability distribution, and the remaining part of the right hand side of Eq. (2) excluding the constant is the likelihood component. The likelihood component comprises multiple likelihood functions, each for one nearest known neighbor except for the last visited location. Therefore, Eq. (2) is essentially in the form of

$$\text{posteior} \propto \text{likelihood}\ [i_m(\mathbf{u}_m)] \times \cdots \times \text{likelihood}\ [i_2(\mathbf{u}_2)] \times \text{prior}, \qquad (3)$$

where the likelihood functions represent a process of sequential Bayesian updating on different nearest data within a Markov-type neighborhood. Each nearest datum represents new evidence that may be used to update previous belief in the local probability distribution of $Z(\mathbf{u})$ at uninformed location $\mathbf{u}_0$ to make it closer to the truth. How close the final local probability distribution (i.e., the final posterior) is to the truth depends on how much related evidence is available. The sequential updating process starts from nearest neighbor $i_2(\mathbf{u}_2)$ and ends at nearest neighbor $i_m(\mathbf{u}_m)$ within a Markov-type neighborhood around the uninformed location $\mathbf{u}_0$ (Fig. 2a). The updating process may not need to follow a fixed ordering of nearest neighbors, because earlier considered nearest neighbors become the conditioning data of later updatings, and all updatings are conditioned on the datum $i_0(\mathbf{u}_0)$ being estimated. Such a sequential Bayesian updating on nearest spatial data within a neighborhood is one of the major characteristics of MCRFs. It should be clarified that "sequential Bayesian updating" here is just an explanation of Eq. (2) using the Bayesian view, rather than an iterative process in computation.

If the last visited location is far away and the spatial Markov chain is first-order stationary (i.e., the transition probabilities from the last visited location to the current location decay to corresponding marginal probabilities), the last visited location in Eq. (2) can be ignored. Supposing that there are still m nearest neighbors in different directions around the uninformed location $\mathbf{u}_0$, we have

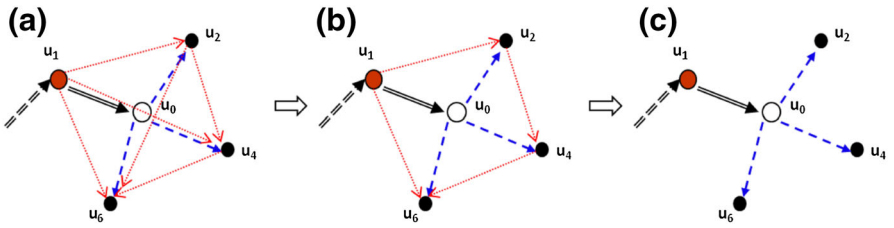

Fig. 3 Spatial structures of a size-4 neighborhood in a MCRF and its simplified forms based on Markov property and conditional independence assumption: **a** four nearest neighbors approximately located in four cardinal directions; **b** data interactions across the uninformed location $\mathbf{u}_0$ are ignored; and **c** data interactions between nearest neighbors are ignored

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)] = \frac{1}{A}p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0),\ldots,i_{m-1}(\mathbf{u}_{m-1})] \cdots p[i_1(\mathbf{u}_1)|i_0(\mathbf{u}_0)]p[i_0(\mathbf{u}_0)], \quad (4)$$

where $A = p[i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)]$ is a constant, $\mathbf{u}_1$ is a nearest neighbor, and $p[i_0(\mathbf{u}_0)]$ represents the marginal probability distribution of state or class i_0. Compared to Eq. (2), the prior probability distribution of Eq. (4) becomes $p[i_0(\mathbf{u}_0)]$, and it has one more likelihood function -$p[i_1(\mathbf{u}_1)|i_0(\mathbf{u}_0)]$. Therefore, Eq. (4) is in the form of

$$\text{posteior} \propto \text{likelihood}\,[i_m(\mathbf{u}_m)] \times \cdots \times \text{likelihood}\,[i_2(\mathbf{u}_2)] \times \text{likelihood}\,[i_1(\mathbf{u}_1)] \times \text{prior}. \quad (5)$$

The Bayesian updating process starts from nearest neighbor $i_1(\mathbf{u}_1)$ and ends at nearest neighbor $i_m(\mathbf{u}_m)$ in a Markov-type neighborhood around the uninformed location $\mathbf{u}_0$ (Fig. 2b).

The full solution of MCRFs, Eq. (2), and its special case, Eq. (4), both comprise a series of two to m+1 point spatial structures that involve the uninformed location $\mathbf{u}_0$, such as $p[i_m(\mathbf{u}_m)|i_0(\mathbf{u}_0),\ldots,i_{m-1}(\mathbf{u}_{m-1})]$ and $p[i_2(\mathbf{u}_2)|i_0(\mathbf{u}_0),i_1(\mathbf{u}_1)]$. Therefore, they are essentially spatial models of multiple-point statistics at the neighborhood level and these multiple-point statistics are in the form of multiple-point conditional probabilities. Even though only four nearest known neighbors are considered, they still involves two to five point structures (Fig. 3a), that are functions of directional lag distances because these points may not be adjacent in a sparse or inexhaustive data space. Therefore, without simplification the local conditional probability distribution functions of $Z(\mathbf{u})$ for a specific unsampled location are difficult to compute from sparse sample data. However, it is worth mentioning that multiple-point statistics had been used in the multiple-point geostatistical methods emerged in recent years, mainly for mapping subsurface curvilinear features, where multiple-point statistics were directly estimated from training images using defined data templates (Strebelle 2002); more recently, Dimitrakopoulos et al. (2010) suggested a kind of high-order (i.e., multiple-point) statistics - spatial cumulants, and provided their computation methods. These studies may provide the prospects for implementing the MCRF approach as a new multiple-point geostatistical approach.

Applying the conditional independence assumption of nearest known neighbors to Eq. (2), a simplified general solution for MCRFs, which is directly computable from sparse sample data, can be derived as

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)] = \frac{p_{i_1 i_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{i_0 i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n} [p_{i_1 f_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{f_0 i_g}(\mathbf{h}_{0g})]}, \quad (6)$$

where $p_{i_0 i_g}(\mathbf{h}_{0g})$ represents a transition probability function (i.e., transiogram) from class i_0 at location $\mathbf{u}_0$ to class i_g at location $\mathbf{u}_g$ with a distance lag $\mathbf{h}_{0g}$; $i_1(\mathbf{u}_1)$ represents the nearest known neighbor from or across which the Markov chain moves to the current uninformed location $\mathbf{u}_0$; m represents the number of nearest known neighbors including the last visited location; i and f represent states (or classes) in the state space $S = (1,\ldots,n)$. Equation (6) is still in the Bayesian inference formula, with $p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)]$ being the posterior distribution, $p_{i_1 i_0}(\mathbf{h}_{10})$ being the prior distribution, the denominator being the constant and the remaining part of the right hand side being the likelihood component. Figure 3 provides a case of MCRF models with $m = 4$ and its simplified forms using the Markov chain property and the conditional independence assumption.

Similarly, with the conditional independence assumption of nearest known neighbors, Eq. (4) becomes

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m)] = \frac{p_{i_0} \prod_{g=1}^{m} p_{i_0 i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n} [p_{f_0} \prod_{g=1}^{m} p_{f_0 i_g}(\mathbf{h}_{0g})]}, \quad (7)$$

where p_{i_0} is the marginal probability of class i_0. Equation (7) also can be simply derived from Eq. (6) by assuming $p_{i_0 i_1}(\mathbf{h}_{01}) p_{i_0} / p_{i_1} = p_{i_1 i_0}(\mathbf{h}_{10})$. Therefore, Eq. (7) is a special case of Eq. (6). Compared to Eq. (6), the prior distribution in Eq. (7) becomes p_{i_0}. With second-order spatial stationarity assumption and non-unidirectional transiograms, Eqs. (6) and (7) are identical.

Equations (6) and (7) do not consider the clustering effect of data. Apparently clustered nearest neighbors largely deviate from the conditional independence assumption and may impact the estimates of local conditional probability distributions. Therefore, dealing with this issue is necessary in algorithm design with either of these two equations.

2.4 MCRF Cosimulation Model

Similar to cokriging models in classical geostatistics (Matheron 1971; Goovaerts 1998), MCRF cosimulation models (Co-MCRFs) also can be established through extending the MCRF solutions. Bayesian thinking has been used in classical geostatistics for incorporating soft data since 1980s (Omre 1987; Christakos 1990). Based on the Bayesian inference principle (Bayes 1763), a Co-MCRF model may be regarded as the Bayesian update of a MCRF model based on new evidence of auxiliary data. In this study, our main focus is the MCRF cosimulation model.

Assume X is a target categorical variable to be estimated and E is an auxiliary data set, the Bayesian inference formula for this knowledge updating can be written as

$$P(X|E) = \frac{P(E|X)P(X)}{P(E)} = \frac{P(E|X)P(X)}{\sum_X P(E|X)P(X)} = C^{-1} \cdot P(E|X) \cdot P(X) \quad (8)$$

where $C = \sum_X P(E|X)P(X)$ is a constant; $P(X)$ is the local probability distribution of X, which serves as the prior here and can be estimated by Eq. (2) or its simplified forms; $P(E|X)$ is the probability of observing E given X, that is, the likelihood term to update the prior with new evidence of E. Because a MCRF model is a spatial Bayesian model at the neighborhood level with sequential Bayesian updating on different nearest data, the incorporation of an auxiliary data set in a Co-MCRF model is essentially a further Bayesian update to a MCRF model. Ecan be a set of auxiliary variables. For n mutually independent auxiliary variables $\mathbf{E} = \{e_1, \ldots, e_n\}$, we have

$$P(X|\mathbf{E}) = \frac{P(\mathbf{E}|X)P(X)}{\sum_X P(\mathbf{E}|X)P(X)} = \frac{P(X)\prod_k P(e_k|X)}{\sum_X P(X)\prod_k P(e_k|X)}. \quad (9)$$

Based on above Bayesian principle, the simplified general solution, Eq. (6), can be expanded into a Co-MCRF model to incorporate auxiliary data through cosimulation. The contributions of auxiliary variables may be incorporated through different methods. In this study, the data of auxiliary variables are regarded as nearest neighbors of the uninformed location $\mathbf{u}_0$ in other variable spaces. Only the co-located cosimulation case is considered, as illustrated by Fig. 4a, because it is what we need for improving LULC classification while the involved auxiliary variables provide exhaustive data. The co-located Co-MCRF model with k auxiliary variables can be written as

$$\begin{aligned} &p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1), \ldots, i_m(\mathbf{u}_m); r_0^{(1)}(\mathbf{u}_0^1), \ldots, r_0^{(k)}(\mathbf{u}_0^k)] \\ &= \frac{p_{i_1 i_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{i_0 i_g}(\mathbf{h}_{0g}) \prod_{l=1}^{k} q_{i_0 r_0^{(l)}}}{\sum_{f_0=1}^{n} [p_{i_1 f_0}(\mathbf{h}_{10}) \prod_{g=2}^{m} p_{f_0 i_g}(\mathbf{h}_{0g}) \prod_{l=1}^{k} q_{f_0 r_0^{(l)}}]}, \end{aligned} \quad (10)$$

where $r_0^{(k)}$ represents the state of the kth auxiliary variable at the co-location $\mathbf{u}_0^k$, and $q_{i_0 r_0^{(k)}}$ represents the cross transition probability from class i_0 at location $\mathbf{u}_0$ in the primary variable space to class r_0 at the co-location $\mathbf{u}_0^k$ in the kth auxiliary variable space. The cross transiograms from the primary variable to auxiliary variables reduce to cross transition probabilities due to the co-location property. This kind of cross transition probability (or transiogram) between classes of two different spatial variables is called a cross-field transition probability (or transiogram) (Li et al. 2013). Because the co-located datum of an auxiliary variable is in another variable space, it may be assumed to be conditionally independent of the nearest neighbors of the uninformed location $\mathbf{u}_0$ in the primary variable space. The cross-field transition probabilities from the primary variable to each auxiliary variable, however, must be estimated separately. In this equation, the cross correlations between auxiliary variables are not dealt with and they are practically assumed to be independent of each other.

In this study, only one auxiliary variable, a pre-classified remotely sensed image data set generated by a conventional classification method, is considered. Hence, Eq. (10) further reduces to

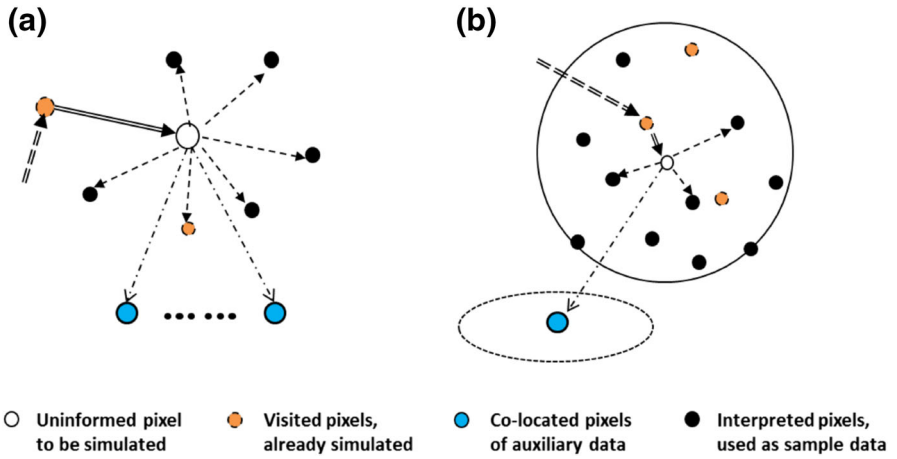

Fig. 4 Illustrations of the simplified MCRF cosimulation model with conditional independence assumption (**a**) and the co-located Co-MCSS algorithm with one auxiliary variable (**b**). *Double arrows* represent the moving directions of the spatial Markov chain. *Dash arrows* represent the interactions of the spatial Markov chain with nearest known neighbors and co-located auxiliary data

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_m(\mathbf{u}_m);r_0(\mathbf{u}_0)] = \frac{q_{i_0r_0}p_{i_1i_0}(\mathbf{h}_{10})\prod_{g=2}^{m}p_{i_0i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n}[q_{f_0r_0}p_{i_1f_0}(\mathbf{h}_{10})\prod_{g=2}^{m}p_{f_0i_g}(\mathbf{h}_{0g})]}. \quad (11)$$

This model still can be regarded as a Bayesian inference model. Compared to Eq. (6), this Co-MCRF model uses one more data layer—an auxiliary data set, where each co-located datum may be regarded as new evidence for further updating the local conditional probability distribution.

2.5 MCRF Co-Located Cosimulation Algorithm

In real applications, it is unnecessary and difficult to consider many nearest known neighbors in different directions. For those nearest known neighbors outside correlation ranges and those approximately located in the same directions with small separate angles but not the closest to the uninformed location $\mathbf{u}_0$ being estimated, they can be canceled in Eq. (11). Thus, m can be much smaller than the real number of nearest known neighbors in different directions. In addition, Eq. (11) was based on a general conditional independence assumption that does not hold for clustered data because it ignores data configuration. Thus, dealing with data clustering effect may be necessary in simulation algorithm design.

For dealing with pixel data from remotely sensed imagery, considering four cardinal directions can be sufficient. If only the four nearest known neighbors in the four cardinal directions within a search circle are considered, Eq. (11) can be further simplified to

$$p[i_0(\mathbf{u}_0)|i_1(\mathbf{u}_1),\ldots,i_4(\mathbf{u}_4);r_0(\mathbf{u}_0)] = \frac{q_{i_0r_0}\,p_{i_1i_0}(\mathbf{h}_{10})\prod_{g=2}^{4}p_{i_0i_g}(\mathbf{h}_{0g})}{\sum_{f_0=1}^{n}[q_{f_0r_0}\,p_{i_1f_0}(\mathbf{h}_{10})\prod_{g=2}^{4}p_{f_0i_g}(\mathbf{h}_{0g})]}. \tag{12}$$

In the above equation, it is assumed that the last visited location of the spatial Markov chain is always within the four nearest neighbors; if it is not, we assume the spatial Markov chain comes through one of them. Equation (12) assumes conditional independence of the four nearest neighbors in the four cardinal directions (Fig. 3c). Such a conditional independence assumption holds for lattice data in some Markov random fields, such as Pickard random fields (Pickard 1980; Rosholm 1997), and is applicable to sparse data situation (Li 2007a).

Because the interpreted sample pixel label data are usually sparse, the nearest known neighbors in a neighborhood may not be exactly located along cardinal directions. Therefore, relaxation in cardinal directions is needed. To cover the whole space of a search circle, sectors can be used to replace cardinal directions, that is, seeking one nearest known neighbor from each sector to compose the neighborhood. For the four cardinal directions, the corresponding sectors are four quadrants. There may be no data in some quadrants within a search range at the boundary areas or at the beginning of a simulation when sample data are too sparse; thus, the size of the neighborhood may sometimes be less than four. Equation (12) can adapt to the situation. In case no data can be found in the entire search circle, we assume the spatial Markov chain comes from a location outside the search range.

Sequential simulation (Johnson 1987) means that at an unsampled location, the estimation of the local conditional probability distribution incorporates both original sample data and previously simulated data within a neighborhood so that simulated data are directly related to each other, but the specific estimation model may be various. Such a simulation strategy has been widely used in conventional geostatistics for stochastic simulations, such as the sequential indicator simulation algorithm (Deutsch and Journel 1992; Goovaerts 1996). The MCSS algorithm was developed based on the above quadrant search method for nearest data and the sequential simulation strategy using the simplified MCRF model (i.e., Eq. (6)) as its estimator, and it was effective in stochastic simulations of categorical fields (Li and Zhang 2007). For the co-located Co-MCSS algorithm, the sequential simulation procedure is similar to that for the MCSS algorithm, except that an auxiliary data set is conditioned through the co-located Co-MCRF model (Fig. 4b). Fortran PowerStation computer code was developed for the computation algorithm, but image visualization has to use ArcGIS.

2.6 Transiogram Models and Cross-Field Transition Probability Matrix

To perform simulation using the Co-MCRF model (i.e., Eq. (12)) requires a set of transiogram models. To get transiogram models, we need to first estimate the experimental transiograms from sample data. An experimental transiogram can be estimated using the following equation

$$\hat{p}_{ik}(\mathbf{h}) = \frac{F_{ik}(\mathbf{h})}{\sum_{j=1}^{n}F_{ij}(\mathbf{h})}, \tag{13}$$

where $F_{ik}(\mathbf{h})$ represents the frequency of transitions from class i to class k at the lag distance $\mathbf{h}$, and n is the total number of classes in a categorical field.

Different methods were suggested to obtain transiogram models from experimental transiograms (Li and Zhang 2010). One is non-parametric, using an interpolation or smoothing method to get continuous models from experimental transiograms. The second is parametric, using mathematical models to fit most experimental transiograms, with one in each row of a transiogram matrix to take the left potion of unity. The non-parametric linear interpolation method is suitable when samples are relatively abundant and experimental transiograms are reliable. The parametric method is widely applicable, but it is relatively tedious when many classes are involved. In this study, the linear interpolation method (Li and Zhang 2010) was used because it is efficient and the data set of the expert-interpreted pixel labels is usually sufficient for estimating reliable experimental transiograms in real applications. Carle and Fogg (1996, 1997) suggested two methods for joint modeling of experimental transiograms: One is the transition rate method, which was used in application studies (Weissmann and Fogg 1999), and the other is the random transition probability model (Carle and Fogg 1996, p. 463). However, these two methods are oversimplified for complex landscapes like LULC patterns, and are not as efficient as the linear interpolation method.

For co-located cosimulation conditioned on one auxiliary variable, a cross-field transition probability matrix (CTPM) from the sample data set to the co-located auxiliary data set is needed (Li et al. 2013). A transition probability in the CTPM can be estimated by counting the point-to-point frequency of a pair of classes from the sample data of the primary variable to the co-located data of the auxiliary variable using the following equation

$$\hat{q}_{ik} = \frac{f_{ik}}{\sum_{j=1}^{n} f_{ij}}, \tag{14}$$

where f_{ik} represents the frequency of transitions from class i of the primary variable to class k of the auxiliary variable, and n is the number of classes of the auxiliary variable.

3 Test Data and Parameters

3.1 Data

The case study was performed in a small area of 1.035 km^2, which is located in the suburb of Wuhan city, Hubei, China (Li and Zhang 2011). Here, only one high-quality high-resolution remotely sensed image was used to test the proposed method. For the convenience of display, the remotely sensed raster image used for testing is a small piece, having a 234 $\times$ 192 grid lattice with a pixel size of 4.8 m $\times$ 4.8 m (Fig. 5a). In the image, three LULC classes can be easily differentiated by human eyes. These classes are (1) farmlands (mainly planted with vegetables), (2) built areas (i.e., villages), and (3) waterbodies (i.e., ponds). Therefore, these three LULC classes will be considered in the study area. However, farmlands comprise numerous patches with different vegetables and different management practices including greenhouses, which result in different spectral responses at different places. In villages, buildings

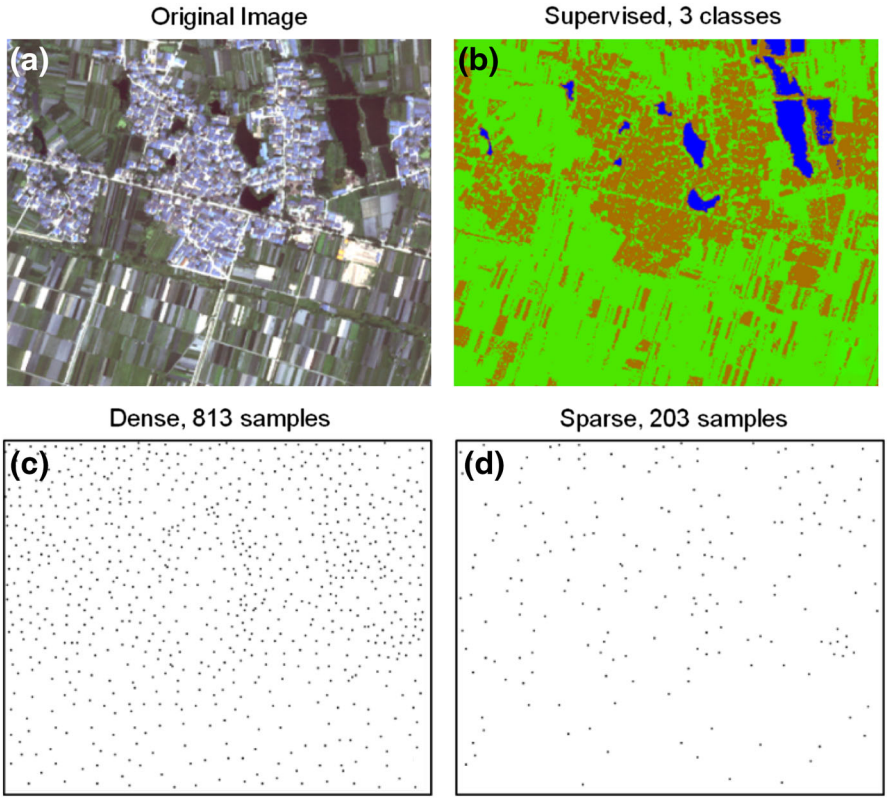

Fig. 5 The original remotely sensed image (**a**), pre-classified image using the supervised maximum likelihood algorithm (**b**), and expert-interpreted pixel label data sets (**c** a dense data set, **d** a sparse data set)

speckled by trees also show varied spectral responses. The spectral responses of water bodies are relatively identical, except that some small shallow ponds are not very clear. Although the general pattern is simple, it is still difficult to accurately classify image pixels into the three LULC classes purely based on confusing and overlapping spectral responses.

To test the proposed Co-MCSS algorithm, 1591 data points were interpreted from the high resolution image using ArcGIS and used as sample data. This extra-dense data set was then split into two subsets—one dense data set (813 data points, Fig. 5c) for conditional cosimulation and the other data set (778 data points) for validation. From the dense data set, a medium data set (406 data points), a sparse data set (203 data point, Fig. 5d), an extra-sparse data set (100 data points) and an extreme-sparse data set (50 data points) were extracted for conditional cosimulation. Unconditional cosimulation was regarded as a conditional cosimulation on a zero data set. These six data sets accounted for 0 to 1.81 % of the total image pixels in the study area and were assumed to be expert-distinguishable pixels that could be interpreted confidently from the remotely sensed imagery. They were regarded as hard data (similar to ground observations), and were used for transiogram estimation. Cosimulations were subsequently conducted on each of these data sets.

The same image was pre-classified into three LULC classes with the same names using the SML classifier (Fig. 5b), and the pre-classified data served as an auxiliary exhaustive data set. One may obtain such an auxiliary data set using any conventional classification method (including unsupervised methods) to pre-classify the image, where the class number of auxiliary data may be different from that of expert-interpreted sample data and the classes need not have the same physical meanings as those of sample data have. However, the classes of the pre-classified image should be well-correlated with the classes of the expert-interpreted land cover sample data for improving classification accuracy.

These data sets were only used for testing the proposed method, and were not a real application study. One may assume that in a real application, such sample data could be interpreted by experts from one or multiple images with other auxiliary information (e.g., local land use maps) for the same area, and the image used for pre-classification is a relative good and current one, selected from those available remotely sensed images for the study area.

3.2 Parameters

Omni-directional experimental transiograms were estimated from each of the dense and medium data sets, and were interpolated into continuous transiogram models, some of which were shown in Li and Zhang (2011). Because there is no identical anisotropy and asymmetry direction for all classes in the study area, it is unnecessary to consider estimating multiple sets of directional transiograms. The sparser data sets were too small to generate reliable experimental transiograms, so for the method testing purpose the transiogram models estimated from the dense data set were used for the simulations conditioned on sparser data sets. It is also preferable to use identical parameters for comparison purpose. In real applications, if samples are too sparse to generate reliable transiogram models, expert knowledge should be used to obtain transiogram models (Li and Zhang 2010).

The CTPM from the classes of expert-interpreted sample data to the classes of the pre-classified data set is provided in Table 1. In the pre-classified data set, the expert recognized farmlands and built areas classes were not effectively differentiated from each other, but none of them were classified into waterbodies, and the expert recognized waterbodies class was completely discriminated from farmlands, although some

Table 1 Cross-field transition probability matrix from classes of expert-interpreted sample data (in first column) to classes of pre-classified data (pre-classification was conducted using the supervised maximum likelihood classifier) (in first row)

Class	A1 (farmlands)	A2 (built areas)	A3 (waterbodies)
S1 (farmlands)	0.8325	0.1675	0.0000
S2 (built areas)	0.2684	0.7316	0.0000
S3 (waterbodies)	0.0000	0.2736	0.7264

S1 Class 1 of expert-interpreted sample data, *A1* Class 1 of pre-classified image data

areas of this class were still classified into built areas. Due to the relatively accurate discrimination of the waterbodies class in the pre-classification, it is expected that this auxiliary data set will provide stronger supportive information for the classification of waterbodies in the co-located cosimulations.

4 Results and Analysis

4.1 Supervised Maximum Likelihood Classification

The SML classification generated relatively good results due to the simplicity of the image (Fig. 5b). However, some patches of farmlands were still classified as the class of built areas due to spectral overlap, and some non-building areas within or around villages (mainly woods) were classified as farmlands. In addition, some shallow ponds were classified as other two classes. The accuracy of this LULC map classified by SML is 77.63 % based on the validation data set of 778 interpreted pixel label data. This supervised classification accuracy is normal because the spectral responses over the areas of the three LULC classes have overlaps.

4.2 Simulated Results by Co-MCSS

A simulation search radius of 30 pixel-lengths was chosen for the dense and medium sample data sets, and that of 60 pixel-lengths for the sparser sample data sets, after some testing simulations. Because only four nearest neighbors in four quadrants are considered in the final simulation algorithm, a large radius is not necessary when the density of conditioning data is not very low. If conditioning data are very sparse, the search radius should approximate the largest correlation range among all auto/cross transiogram models of land cover classes. For the pixel label data in this study, the radius length of 60 pixels used for sparse sample data sets reached or surpassed the correlation ranges of most transiogram models (Li and Zhang 2011). For each of the six interpreted sample data sets (from dense to zero), 100 realisation maps were simulated using Co-MCSS and a set of occurrence probability maps for all LULC classes were further estimated. An optimal classification map was obtained for each sample data set on the basis of maximum occurrence probabilities estimated from the 100 simulated realisations. The computation time for a simulation (i.e., generating 100 realisations and other related maps) is usually a few minutes.

The six optimal classification maps (Fig. 6) show that when the expert-interpreted sample data is dense, all three LULC classes were well-captured compared to the original image (Fig. 5a). With decreasing density of sample data, the built areas gradually deviated from their real shapes at boundaries. However, the general pattern was not much altered and water bodies were always well-classified, except for two shallow ponds that were not captured by the SML classification. Even when no sample data were used for conditioning, the general pattern of the LULC classes and the approximate shapes of water bodies were captured in classification. This should be attributed to the contribution of the auxiliary data set, that is, the pre-classified image by SML, and the spatial correlation information provided by transiogram models. Comparing

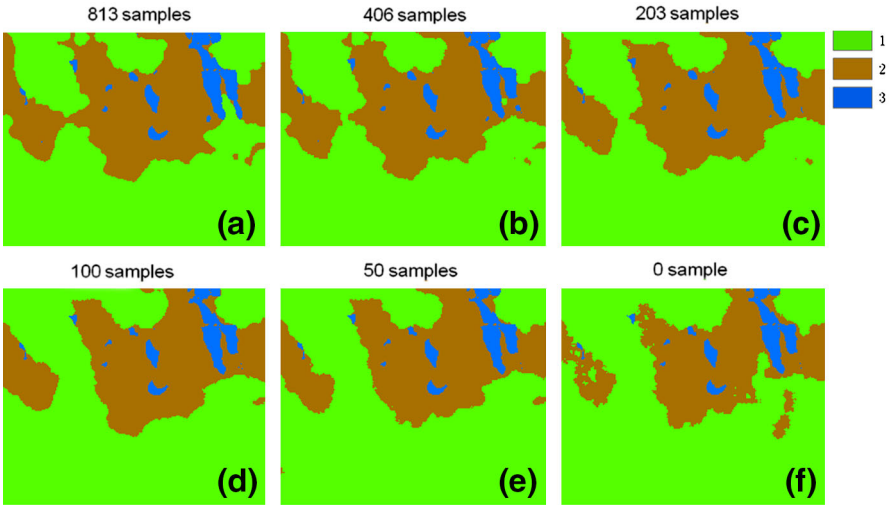

Fig. 6 Optimal classification maps based on maximum occurrence probabilities generated by Co-MCSS. Simulations were conditioned on pre-classified image data and respectively six different expert-interpreted sample data sets (*green* farmlands, *brown* built areas, *blue* waterbodies)

their classifications against the validation data indicates that these optimal classification maps have high accuracies (i.e., percentages of correctly classified pixels), ranging from 86.12 to 98.59 %, demonstrating much improvement over the conventional SML classification (Table 2). Even when very sparse (e.g., 0.2 % of the total pixels) or zero sample data are available, Co-MCSS still can improve the classification accuracy (10 % or so for optimal classification maps) compared to the SML method. Interpreting a sparse sample data set from remotely sensed imagery with other auxiliary data (e.g., existing land use maps, google earth images) for a study area may not be very difficult and the gain in classification accuracy seems huge. This indicates the possible practicality of the MCRF approach in LULC classification from remotely sensed imagery and auxiliary information. Moreover, the incorporation of pre-classified image data through cosimulation makes the approach less labor-intensive to acquire higher classification accuracy over conventional classification methods.

Generating a number of simulated realizations is the advantage of stochastic simulation methods. Simulated realization maps do not have the smoothing effect of optimal maps and are usually regarded as "possible realities" of a spatial variable under study. Differences in simulated patterns of a number of realizations indicate the uncertainty in spatial distribution of LULC classes in the study area, which is mainly caused by the ambiguity of pixel signatures. Therefore, the aim of a stochastic simulation is to capture the greatest spatial heterogeneity rather than to pursue high classification accuracy. A major value of simulated realizations is that they can be used as input data to study the propagation of uncertainty through ecological or hydrological models (Goovaerts 1996). Figure 7 shows some simulated realizations based on the dense, sparse and extreme-sparse data sets of the interpreted pixel labels. These realizations indicate more details than their corresponding optimal classification maps because the latter are essentially reflections of classes with maximum occurrence

Table 2 Classification accuracies (i.e., percentages of correctly classified locations) of simulated maps (i.e., realization and optimal maps) using the Co-MCSS and MCSS algorithms

Data set	Informed pixels		Classification accuracy based on the validation data set (%)								
	Number	Percentage (%)	MCSS simulation		Co-MCSS simulation		Improvement by Co-MCSS				SML
							Over MCSS		Over ML		
			R maps	O map	R maps	O map	R maps	O map	R maps	O map	
Dense	813	1.81	97.18	99.23	97.76	98.59	0.58	−0.64	20.13	20.96	
Medium	406	0.90	91.82	96.79	95.23	96.91	3.41	0.12	17.60	19.28	
Sparse	203	0.45	81.84	88.13	89.84	92.29	8.00	4.16	12.21	14.66	
Extra-sparse	100	0.22	73.07	83.16	87.99	91.90	14.92	8.74	10.36	14.27	
Extreme-sparse	50	0.11	60.64	73.52	84.35	90.75	23.71	17.23	6.72	13.45	
Zero	0	0.00	38.38	42.29	75.63	86.12	37.25	43.83	−2.00	8.49	
Validation	778	1.74									
Image	44,736										77.63

R maps realization maps, *O map* optimization map based on 100 realizations, *SML* supervised maximum likelihood classifier, *MCSS* Markov chain random field sequential simulation algorithm, *Co-MCSS* Markov chain random field sequential cosimulation algorithm

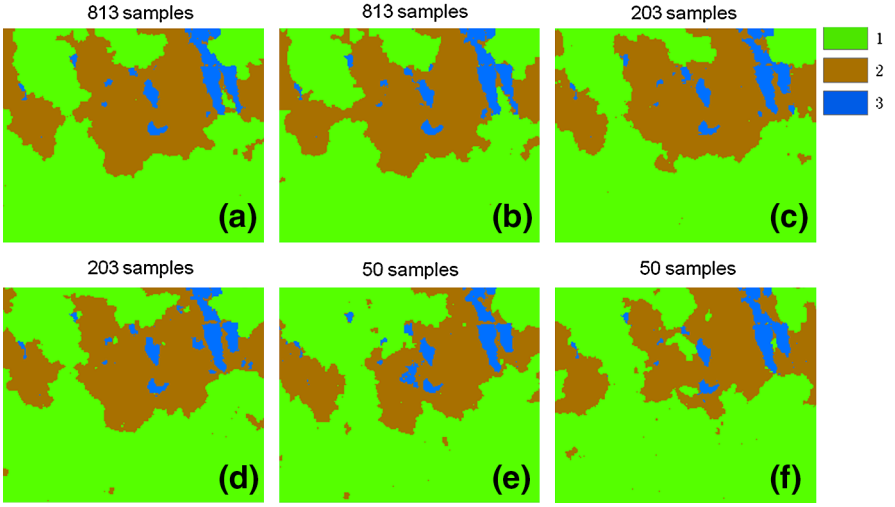

Fig. 7 Simulated realization maps by Co-MCSS. Simulations were conditioned on pre-classified image data and respectively three different expert-interpreted sample data sets (note that **a** and **b**, **c** and **d**, and **e** and **f** are three pairs, with each pair of realizations being conditioned on the same sample data set). *Color codes* as in Fig. 6

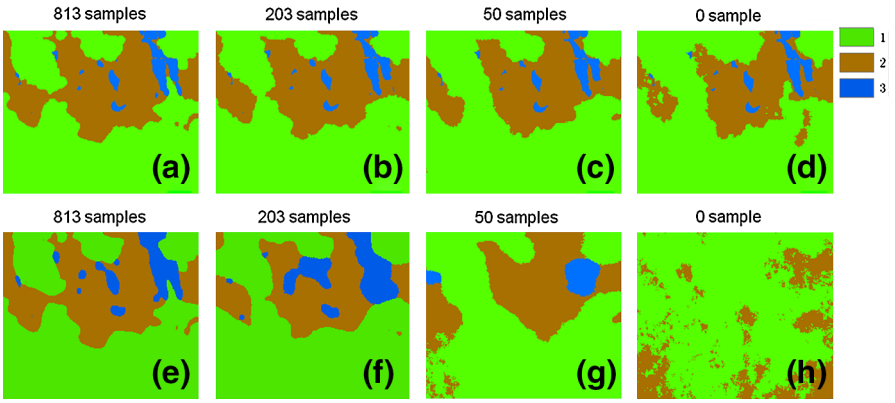

Fig. 8 Comparison of optimal classification maps generated by Co-MCSS (**a**–**d**) and MCSS (**e**–**h**). Simulations by Co-MCSS were conditioned on pre-classified image data and different expert-interpreted sample data sets. Simulations by MCSS were only conditioned on different expert-interpreted sample data sets. *Color codes* as in Fig. 6

probabilities. Differences in pattern details between realizations can be seen and also increase with decreasing number of interpreted pixels. Consequently, the average classification accuracy of realizations decreases from 97.76 to 75.63 % with decreasing density of interpreted sample data from 1.81 to 0 % (Table 2). Although the classification accuracies of realization maps are lower than those of corresponding optimal maps, they are still higher than the accuracy of the SML classification map as long as some interpreted sample data are available. Realization maps generated by Co-MCSS represent improvements in accuracy over the SML classification map.

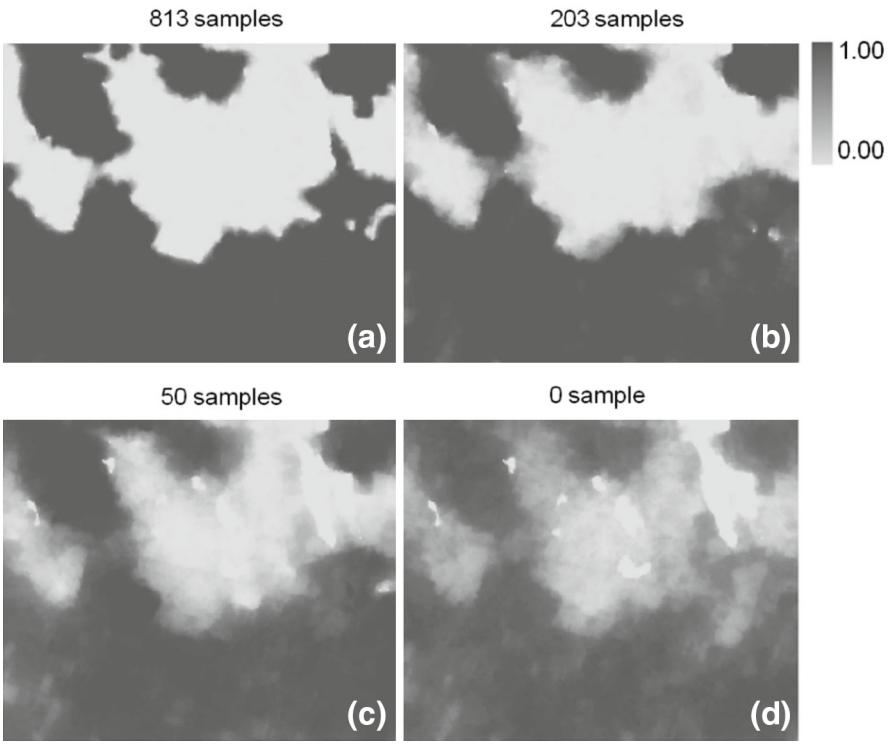

Fig. 9 Occurrence probability maps of land use class 1 (i.e., farmland), estimated from one hundred of simulated realizations by Co-MCSS

4.3 Comparison with Simulated Results by MCSS

The incorporation of pre-classified data represents an important boost to the practicality of the MCRF approach for LULC classification. To demonstrate the improvement, we made a comparison with the simulated results by the MCSS algorithm, which does not incorporate pre-classified image data. Figure 8 shows some optimal classification maps generated by both Co-MCSS (Fig. 8a–d) and MCSS (Fig. 8e–h). When sample data are dense, both methods generate similar optimal classification maps. When sample data are very sparse, Co-MCSS accurately capture the pattern of LULC classes, while MCSS do not. When no sample data are available, Co-MCSS captures the general pattern of the three classes, but MCSS cannot generate meaningful data. In addition, because the waterbodies class in the pre-classified data set is relatively well-captured, the feature of this class is always effectively reproduced by Co-MCSS. Consequently, classification accuracies by Co-MCSS show significant improvements over those by MCSS, and the improvement is particularly large when sample data are sparse (Table 2). Apparently, this is attributed to the contribution of the good pre-classified data. The MCSS algorithm for LULC classification is relatively labor intensive for real world applications. It mainly demonstrated the promising prospect of the development of the MCRF approach in remotely sensed image classification.

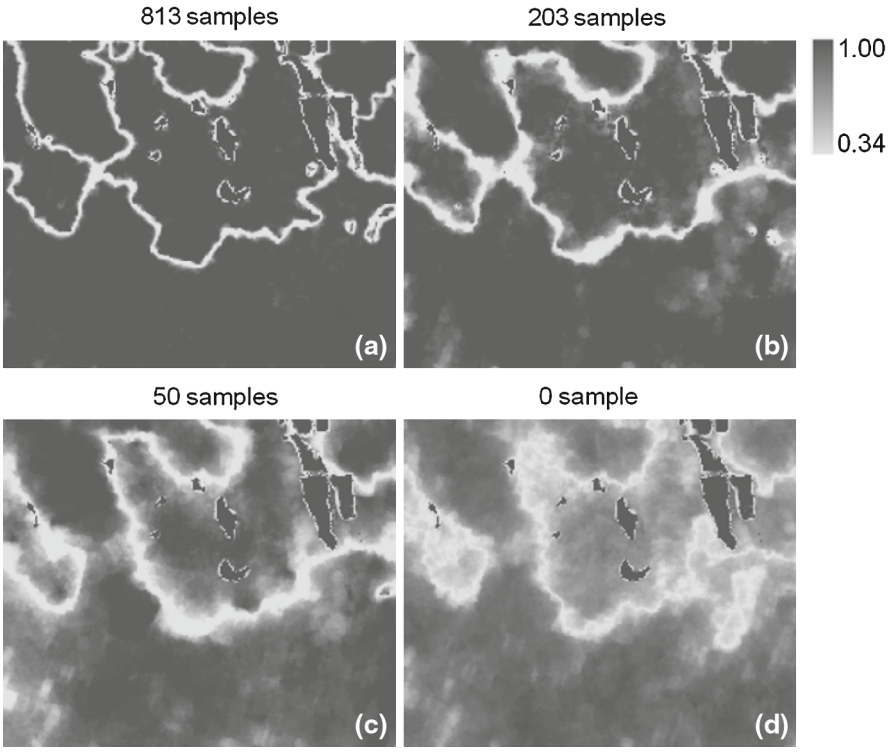

Fig. 10 Maximum occurrence probability maps of land use/land cover classes, estimated from a hundred simulated realizations by Co-MCSS, conditioned on pre-classified image data and different sample data sets

However, the addition of the cosimulation function, that is, the Co-MCSS algorithm, to the MCRF approach improves performance.

4.4 Uncertainty Assessment

Quantifying spatial uncertainty is the major advantage of stochastic simulation methods. Co-MCSS can generate a large number of simulated realizations, from which occurrence probabilities of each class at every location can be estimated. Realization maps only qualitatively indicate the spatial uncertainties associated with a classification. The occurrence probability maps quantitatively illustrate where and with how much probability (or how large proportion) a LULC class may occur, and also indicate the areas of clustered occurrence of a class (i.e., the dark areas). Figure 9 shows the occurrence probability maps of farmlands (i.e., class 1) conditioned on four different sample data sets. They display where this class has more chances to occur. The uncertainty increases with decreasing sample data. Even when no sample data are available, the correlation information with the auxiliary data still can support classification and make occurrence probabilities of the class vary over the study area. Therefore, occur-

rence probability maps are more efficient and vivid in representing uncertainty than are realization maps.

Figure 10 shows maximum occurrence probability maps conditioned on four different sample data sets. A maximum occurrence probability map comprises maximum single-class occurrence probabilities among all classes at all locations. It quantifies the credibility of the corresponding optimal classification map at every pixel. The higher the probability at a pixel, the less uncertainty that is associated with the pixel class label. It also clearly indicates the transition zones (i.e., shallow gray stripes) between clustered areas of pixels of different classes. Apparently, overall uncertainty increases with decreasing density of sample data.

5 Conclusions and Discussion

A MCRF cosimulation approach for improving classification accuracy of LULC classes from remotely sensed imagery was introduced. The approach employs the Co-MCSS algorithm for integrating a pre-classified image data set generated by a conventional classification method and an expert-interpreted sample data set to obtain an improved LULC classification. The approach was tested using a case study, including six expert-interpreted LULC sample data sets and a pre-classified image data set generated by the SML algorithm. Classification accuracies were calculated using a validation data set that was also expert-interpreted pixel class label data. Compared to the pre-classification, which had an accuracy of 77.63 %, the classification accuracy of optimal prediction maps generated by the Co-MCSS algorithm improved by 8.49 to 20.96 % with increasing density of expert-interpreted sample data (from 0 to 1.81 % of total pixels), generally reaching an accuracy of above 90 % when one thousandth of total pixels were interpreted. A comparison with the MCSS algorithm, which does not incorporate a pre-classified image data set, shows that Co-MCSS may capture the geometric shapes of LULC class patterns to some extent and can lead to substantial improvement in classification accuracy, especially when expert-interpreted sample data are sparse. In fact, through the integration of expert-interpreted sample data, estimated transiogram models, and pre-classified data, Co-MCSS essentially may utilize expert knowledge, contextual information, spectral information, textural information and multi-source data. Consequently, it can perform highly accurate LULC classifications. Co-MCSS also can quantify the spatial uncertainty associated with a classification as a set of occurrence probability maps, that provide additional valuable information.

The approach may be understood from different perspectives. If we focus on the stochastic cosimulation and the expert-interpreted sample data, the pre-classified image data set serves as an auxiliary data set to improve simulation accuracy. However, if we focus on the pre-classified image data set, the expert-interpreted sample data may be regarded as an accurate data layer for updating the pre-classification. In general, the approach may combine the Co-MCSS algorithm with any conventional LULC classification method to use expert-interpreted sample data to improve a classification. Even when no sample data or very few sample data are available, it still may improve classification accuracy due to the contribution of the spatial correlation information

provided by transiogram models if they are available (e.g., from expert knowledge). Considering that LULC maps generated by conventionally used LULC classification methods from remotely sensed imagery often cannot reach sufficient quality and that it is not difficult for experts to interpret a highly accurate sample data set of LULC classes for a study area from various available data sources, the proposed approach may have great promise to be a valuable and practical method for improving LULC classification accuracy. Nevertheless, data collection is always time-consuming. Parameter estimation, simulation, and data processing also take time. This labor and time cost is the trade-off for an improved classification accuracy and uncertainty assessment, which have been difficult to achieve by faster classification methods purely based on image spectral information.

Besides the popular SML classifier which was used to create the pre-classified data set from a remotely sensed image in the case study, any other existing classifiers can be used to perform the pre-classification. Thus, the classification accuracy generated by the proposed approach depends on the expert-interpreted sample data as well as the quality of the pre-classification. If the pre-classification is already of very high quality, the accuracy improvement added by the cosimulation with expert-interpreted sample data may not be appreciable. Under this situation, users must decide whether they will use the proposed cosimulation approach to make further classification improvement. Considering that some LULC classes (e.g., farmlands with different cropping situations), particularly the second level LULC classes, are difficult to differentiate automatically from remotely sensed imagery, the proposed approach may be a good tool for experts to use to revise a classification. Thus, the approach may be more useful in areas where remotely sensed imagery is difficult to classify accurately without expert involvement or even field observations. In addition, the approach used expert-interpreted sample data as hard data in cosimulation, while they may also contain uncertainty. Thus, the uncertainty brought by expert interpretation was not taken into account in uncertainty assessment.

Although only the interpreted sparse pixel label data were used as sample data, other data types (e.g., small area data, mixels' class proportion data, and ground observations) can also be incorporated into the cosimulations as conditioning data. The land cover classification accuracy data provided in Table 2 are the results of the simple case study for testing and demonstrating the proposed method. For a large area characterized as a complex landscape with many classes, the classification accuracy may not be so high given the same density of expert-interpreted sample data, due to the corresponding low accuracy of pre-classified image data.

Acknowledgments This study is partially supported by the National Science Foundation of United States (Award numbers 1049017 and 1414108).

References

Balaguer A, Ruiz LA, Hermosillas T, Recio JA (2010) Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. Comput Geosci 36:231–240

Bayes T (1763) An essay towards solving a problem in the doctrine of chances. Phil Trans R Soc Lond 53:330–418

Benediktsson JA, Swain PH, Ersoy OK (1990) Neural network approaches versus statistical methods in classification of multisource remote sensing data. IEEE Trans Geosci Remote Sens 28:540–552
Boucher A (2009) Sub-pixel mapping of coarse satellite remote sensing images with stochastic simulations from training images. Math Geosci 41:265–290
Carle SF, Fogg GE (1996) Transition probability-based indicator geostatistics. Math Geol 28:453–477
Carle SF, Fogg GE (1997) Modeling spatial variability with one- and multi-dimensional continuous Markov chains. Math Geol 29:891–918
Carvalho J, Soares A, Bio A (2006) Improving satellite images classification using remote and ground data integration by means of stochastic simulation. Int J Remote Sens 27:3375–3386
Christakos G (1990) A Bayesian/Maximum-Entropy view to the spatial estimation problem. Math Geol 22(7):763–777
De Bruin S (2000) Predicting the areal extent of land-cover types using classified imagery and geostatistics. Remote Sens Environ 74:387–396
Deutsch C, Journel A (1992) GSLIB: Geostatistical software library. Oxford University Press, New York
Dimitrakopoulos R, Mustapha H, Gloaguen E (2010) High-order statistics of spatial random fields: Exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena. Math Geosci 42:65–99
Elfeki AM, Dekking FM (2001) A Markov chain model for subsurface characterization: theory and applications. Math Geol 33:569–589
Foody GM (2002) Status of land cover classification accuracy assessment. Remote Sens Environ 80: 185–201
Foody GM, Mathur A (2004) A relative evaluation of multiclass image classification by support vector machines. IEEE Trans Geosci Remote Sens 42:1335–1343
Ge Y, Bai H (2011) Multiple-point simulation-based method for extraction of objects with spatial structure from remotely sensed imagery. Int J Remote Sens 32:2311–2335
Gelman A, Carlin J, Stern H, Rubin DB (1995) Bayesian data analysis. Chapman and Hall, London
Goovaerts P (1996) Stochastic simulation of categorical variables using a classification algorithm and simulated annealing. Math Geol 28:909–921
Goovaerts P (1998) Ordinary cokriging revisited. Math Geol 30(1):21–42
Gong P, Howarth PJ (1990) The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. Photogramm Eng Remote Sens 56:67–73
Huang Z, Jia X (2012) Integrating remotely sensed data, GIS and expert knowledge to update object-based land use/land cover information. Int J Remote Sens 33:905–921
Jensen JR (2005) Introductory digital image processing: a remote sensing perspective. Prentice Hall, Upper Saddle River
Johnson M (1987) Multivariate statistical simulation. Wiley, New York
Li W (2007a) Markov chain random fields for estimation of categorical variables. Math Geol 39:321–335
Li W (2007b) Transiograms for characterizing spatial variability of soil classes. Soil Sci Soc Am J 71: 881–893
Li W, Zhang C (2007) A random-path Markov chain algorithm for simulating categorical soil variables from random point samples. Soil Sci Soc Am J 71:656–668
Li W, Zhang C (2010) Linear interpolation and joint model fitting of experimental transiograms for Markov chain simulation of categorical spatial variables. Int J Geogr Info Sci 24:821–839
Li W, Zhang C (2011) A Markov chain geostatistical framework for land-cover classification with uncertainty assessment based on expert-interpreted pixels from remotely sensed imagery. IEEE Trans Geosci Remote Sens 49:2983–2992
Li W, Zhang C (2012) A Bayesian Markov chain approach for land use classification based on expert interpretation and auxiliary data. In: Xiao N, Kwan M-P, Lin H (eds.) GIScience 2012. Sept. 18–21, 2012, Columbus, OH. http://www.giscience.org/past/2012/proceedings/index.html. Accessed 17 Jul 2014
Li W, Zhang C, Dey DK, Willig MR (2013) Updating categorical soil maps using limited survey data by Bayesian Markov chain cosimulation. Sci World J. Article ID 587284. doi:10.1155/2013/587284
Lou J (1996) Transition probability approach to statistical analysis of spatial qualitative variables in geology. In: Forster A, Merriam DF (eds) Geologic modeling and mapping. Plenum Press, New York, pp 281–299
Lu D, Weng Q (2007) A survey of image classification methods and techniques for improving classification performance. Int J Remote Sens 28:823–870
Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266

Matheron G (1971) The theory of regionalized variables and its applications. Les Cahiers du Centre de Morphologie Mathematique de Fontainebleau, No. 5
Matheron G (1973) The intrinsic random functions and its applications. Adv Appl Probab 5:439–468
Meng Q, Cieszewski C, Madden M (2009) Large area forest inventory using Landsat ETM+: A geostatistical approach. ISPRS J Photogramm Remote Sens 64:27–36
Omre H (1987) Bayesian kriging—Merging observations and qualified guesses in kriging. Math Geol 19(1):25–39
Pickard DK (1980) Unilateral Markov fields. Adv Appl Probab 12:655–671
Ritzi RW (2000) Behavior of indicator variograms and transition probabilities in relation to the variance in lengths of hydrofacies. Water Resour Res 36:3375–3381
Rosholm A (1997) Statistical methods for segmentation and classification of images. Ph.D. dissertation, Technical University of Denmark, Lyngby
Schwarzacher W (1969) The use of Markov chains in the study of sedimentary cycles. Math Geol 1:17–39
Solberg AHS, Taxt T, Jain AK (1996) A Markov random field model for classification of multisource satellite imagery. IEEE Trans Geosci Remote Sens 34:100–113
Stefanov WL, Ramsey MS, Christensen PR (2001) Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. Remote Sens Environ 77:173–185
Strebelle S (2002) Conditional simulation of complex geological structures using multiple-point statistics. Math Geol 34:1–21
Tang Y, Atkinson PM, Wardrop NA, Zhang J (2013) Multiple-point geostatistical simulation for post-processing a remotely sensed land cover classification. Spat Stat 5:69–84
Tseng M-H, Chen S-J, Hwang G-H, Shen M-Y (2008) A genetic algorithm rule-based approach for land-cover classification. ISPRS J Photogramm Remote Sens 63:202–212
Van der Meer F (2012) Remote-sensing image analysis and geostatistics. Int J Remote Sens 33:5644–5676
Weissmann GS, Fogg GE (1999) Multi-scale alluvial fan heterogeneity modeled with transition probability geostatistics in a sequence stratigraphic framework. J Hydrol 226:48–65
Zhang D, Zhang C, Li W, Cromley R, Hanink D, Civco D, Travis D (2014) Restoration of the missing pixel information caused by contrails in multispectral remotely sensed imagery. J Appl Remote Sens 8(1):083698. doi:10.1117/1.JRS.8.083698