Combining Transition Probabilities in the Prediction and Simulation of Categorical Fields

Guofeng Cao and Phaedon C. Kyriakidis
Department of Geography, University of California Santa Barbara, U.S.A.

Abstract. Categorical spatial data, such as land cover classes or soil types, are important data sources in many scientific fields, including geography, geology and environment sciences. In geostatistics, indicator kriging (IK) and indicator coKriging (ICK) are typically used for estimating posterior probabilities of class occurrence at any location in space given known class labels at data locations within a neighborhood around that prediction location. In addition, IK and ICK constitute the core of the sequential indicator simulation (SIS) algorithm used for generating realizations of categorical fields. Both IK and ICK require a set of consistently specified indicator (cross)covariance or (cross)variogram models, whose parameter inference can become cumbersome. In addition, IK and ICK may yield estimated probabilities that do not satisfy fundamental probability constraints. To overcome these limitations, transition probability diagrams have been used as an alternative measure of spatial structure for categorical data. More recently, a Spatial Markov Chain (SMC) model was developed for combining transition probabilities into posterior probabilities of class occurrence, under the conditional independence assumption between neighboring data.

This paper surveys alternative approaches for combining pre-posterior (two-point) auto- and cross-transition probabilities of class occurrence between any datum location and a prediction or simulation location into conditional or posterior (multi-point) such probabilities. Advantages and disadvantages of existing approaches are highlighted. Last, a proposal is made to synthesize elements of geostatistical and Markov Chain approaches for combining transition probabilities for prediction and simulation of categorical fields.

Keywords: indicator cokriging, transition probability, Markov Chain, geostatistics, spatial statistics

1. Introduction

In geostatistics, indicator kriging (IK) [11] and indicator coKriging (ICK) [6] are the most frequently used methods for predicting the probability of class occurrence in space from nearby data on different classes. Both IK and ICK are based on a characterization of spatial class structure through indicator (cross)covariance or (cross)variogram models. Although covariances and variograms are suitable for continuous sample spaces, such as elevation fields or pollutant concentrations, the complex spatial patterns found in categorical data render the interpretation of such covariances and variograms less intuitive with associated implications for kriging derived probabilities of class occurrence.

Different from (but linked to) indicator covariances and variograms, transition probabilities are naturally designed to quantify spatial variability in categorical data. In one dimensional stochastic process, rich elegant theoretical results have been derived based on transition probabilities and the Markov assumption. But the difficulties to distinguish “past” and “future” in multidimensional cases have limited their usage in spatial data. Multidimensional stochastic processes have been discussed in the statistics literature [1, 13], but most of these discussions rely on strong and impractical assumptions, which often impede their application. Recently, a Spatial Markov Chain (SMC) model was developed [9, 10] for combining transition probabilities into posterior probabilities of class occurrence, under the conditional independence assumption between neighboring data. The SMC model applies only to data on a regular grid, and only the 4 nearest neighbors are used (rook’s interaction) for estimating conditional probabilities of class occurrence. Due to the conditional independence assumption, there is no requirement that the auto- and cross-transition probabilities be jointly fitted by a linear model of coregionalization (LMC) [7], as is the case with indicator coKriging.

* Corresponding author. E-mail address: cao@geog.ucsb.edu.
This paper surveys existing approaches for combining transition probabilities in the prediction and simulation of categorical fields. More precisely, IK, ICK and SMC (under conditional independence) are considered, and the intrinsic relationships among them are revealed. A case study using unconditional simulation is used to showcase the advantages and disadvantages on the different methods. Based on these analyses, we propose to synthesize elements of the above methods to account for cross-dependencies between multiple categories and avoid the conditional independence assumption.

2. Methods

2.1. Indicator Kriging (IK)

As stated in [11], the application of IK in categorical data modeling could be dated back to the mid sixties when Switzer [12] used it to estimate the spatial distributions of pollutant concentration. As with kriging for continuous data, IK can also have many variants depending on different assumptions on the local mean value or prior probability of class occurrence [7]. In this paper, only simple IK and simple ICK are discussed, and in particular their dual forms which better reveal their links to Markov Chain approaches. Last, for simplicity in notation, we do not differentiate between a global and a local prediction.

Consider a categorical random variable (RV) \( C(x) \) which can take one out of \( K \) mutually exclusive and collectively exhaustive states \( c(x) \in \{1, \ldots, K\} \) at any arbitrary location with coordinate vector \( x_n \); in the absence of any other information, the probability mass function (PMF) of RV \( C(x) \) can be assumed stationary and populated by the \( K \) global class proportions \( \{\pi_1, \ldots, \pi_K\} \). A central task in prediction and simulation of categorical fields is the estimation of the conditional PMF of \( C(x) \) in the presence of observed or previously simulated class labels available at \( N \) locations \( \{x_1, \ldots, x_N\} \). For \( K \) classes, one can define the \((NK \times 1)\) indicator vector \( \mathbf{d} = \text{vec}(\mathbf{i}_k, k = 1, \ldots, K) \) where \( \mathbf{i}_k = [i_k(x_1), n = 1, \ldots, N] \) is the \((N \times 1)\) indicator vector for the \( k \)-th class with \( i_k(x_1) = 1 \), if \( c(x_1) = k \), 0 if not, and vec\( (\cdot) \) denotes the operator that stacks the columns of a matrix one below the other. Our task can now be re-stated as that of estimating the conditional PMF \( P(C(x) = k | \mathbf{d}, \pi_k) \), \( k = 1, \ldots, K \) of \( C(x) \).

Following [7], in simple IK, the \( k \)-th class conditional probability \( P(C(x) = k | \mathbf{d}, \pi_k) \) depends only on the \((N \times 1)\) indicator vector \( \mathbf{i}_k \), and on the global proportion \( \pi_k \) for the same \( k \)-th class; in other words, in simple IK \( P(C(x) = k | \mathbf{d}, \pi_k) \) is approximated by \( P(C(x) = k | \mathbf{i}_k, \pi_k) \). More precisely, the dual simple IK estimate \( \hat{P}_k(x) = \hat{P}_k(x) | \mathbf{i}_k, \pi_k) \) is defined as:

\[
\hat{P}_k(x) = \hat{P}_k(x) | \mathbf{i}_k, \pi_k) = \sum_{n=1}^{N} w_{n,k}^{IK} \left( x_n - x_0 \right)
\]

(2.1)

where \( w_{n,k}^{IK} \) is the dual IK weight pertaining to the same \( k \)-th class indicator auto-covariance \( \sigma_{kk} \left( x_n - x_0 \right) \) quantifying the interaction (correlation) between the unknown \( k \)-th class indicator \( i_k(x_0) \) at location \( x_0 \) and the known \( k \)-th class indicator \( i_k(x_n) \) at location \( x_n \).

The dual IK weights \( w_{n,k}^{IK} \) are obtained by requiring exact data (here indicator) reproduction by the kriging expression in (2.1) instead of minimizing the prediction error variance as done in primal kriging:

\[
i_k(x_n) = \pi_k + \sum_{n=1}^{N} w_{n,k}^{IK} \left( x_n - x_n \right), \quad n = 1, \ldots, N
\]

(2.2)

where \( \sigma_{kk} \left( x_n - x_n \right) \) is the \( k \)-th class indicator auto-covariance between two locations \( x_n \) and \( x_n \) with known class labels, hence known \( k \)-th class indicators \( i_k(x_n) \) and \( i_k(x_n) \). Both \( \sigma_{kk} \left( x_n - x_n \right) \) and \( \sigma_{kk} \left( x_n - x_n \right) \) are typically computed using a theoretical indicator covariance model \( \sigma_{kk} \left( h; \theta_k \right) \) with parameter vector \( \theta_k \) (range, sill, nugget) specific to the \( k \)-th class.

A major drawback of IK is the lack of modeling of inter-dependencies, i.e., for \( k \neq k' \). A related problem is that IK-derived probabilities might violate probability constraints; that is, IK-derived probabilities might lie outside the \([0,1]\) interval and/or not sum to 1. This is awkward, although one can always truncate and normalize the IK-derived probabilities to abide by these constraints. Last, one should consistently parameterize the \( K \) indicator auto-covariance models \( \left\{ \sigma_{kk} \left( h; \theta_k \right), k = 1, \ldots, K \right\} \), particularly when the different classes are defined by truncation of an underlying (latent) continuous field.
2.2. Indicator coKriging (ICK)

Unlike IK, ICK takes inter-class dependencies into account via cross-variogram or cross-covariance models. In ICK, \( P\{C(x_0) = k | d, \pi\} \) depends on all \( K \) indicator vectors \( d = vec\{i_k, k = 1, \ldots, K\} \), and all \( K \) global class proportions \( \pi = [\pi_1, \ldots, \pi_K] \). The dual simple ICK estimate \( \left[ \hat{\pi}_k(x_0) \right]_{ICK} \) for the \( k \)-th class is defined as:

\[
\left[ \hat{\pi}_k(x_0) \right]_{ICK} = \hat{P}\{C(x_0) = k | d, \pi\} = \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} w_{n,k}^{ICK} \sigma_{k'k}(x_n - x_0) \tag{2.3}
\]

where \( w_{n,k}^{ICK} \) is the dual ICK weight pertaining to the indicator covariance \( \sigma_{k'k}(x_n - x_0) \), an auto-covariance if \( k = k' \) and a cross-covariance if \( k \neq k' \), quantifying the correlation between the unknown \( k \)-th class indicator \( i_k(x_n) \) at location \( x_n \) and the known \( k \)-th class indicator \( i_k(x_0) \) at location \( x_0 \).

As with IK, the dual ICK weight \( w_{n,k}^{ICK} \) is calculated by requiring exact sample data (here indicator) reproduction by the ICK expression (2.3):

\[
i_k(x_n) = \pi_k + \sum_{k=1}^{K} \sum_{n=1}^{N} w_{n,k}^{ICK} \sigma_{k'k}(x_n - x_n), n = 1, \ldots, N, k = 1, \ldots, K \tag{2.4}
\]

where \( \sigma_{k'k}(x_n - x_n) \) is the \( k \) to \( k' \) class indicator auto- or cross-covariance between two informed (with class labels) locations \( x_n \) and \( x_n \). Both \( \sigma_{k'k}(x_n - x_n) \) and \( \sigma_{k'k}(x_n - x_n) \) are typically computed using the linear model of coregionalization (LMC), which specifies a set of \( (K \times K) \) indicator covariance models \( \{\sigma_{k'k}(h; \theta_{k'k})\}, k = 1, \ldots, K, k' = 1, \ldots, K \) with parameter vectors \( \theta_{k'k} \) (range, sill, nugget) specific to a particular combination of \( k \) and \( k' \) [7].

Because of its similarity with IK, ICK can not eliminate most of the problems existing in the former. The problem of estimated probabilities violating probability constraints still exists. In addition, the price of considering inter-class dependencies in ICK is the exponential increase in computational complexity. Equations (2.3) and (2.4) imply that the size of the left-hand side matrix increases dramatically when the number of classes increases. Theoretically, ICK should yield more accurate predictions; but in practice, the results of ICK improve little over those of IK: considering the increased complexity of ICK compared to IK, the gain achieved by the former does not compensate against its expense over the latter. This due to the fact that class indicators combined via indicator auto-covariances carry considerable information regarding class inter-dependencies; simply put, such inter-dependencies are not explicitly accounted for in IK.

2.3. Spatial Markov Chain (SMC)

Transition probability is not a new concept; but it is until recently that transition probability has been regarded as a spatial continuity measure and its relationship with the cross-variogram/covariance was discussed [3, 4]. More specifically, the \( k \) to \( k' \) class transition probability \( \pi_{k'k}(h) \) for lag \( h \) is defined as:

\[
\pi_{k'k}(h) = P\{C(x + h) = c_{k'} | C(x) = c_k\} = P\{I_{k'}(x + h) = 1 | I_k(x) = 1\} \tag{2.5}
\]

and is linked to the indicator cross-covariance as \( \pi_{k'k}(h) = \pi_k[\pi_{k'k}(h) - \pi_k] \); transition probabilities appear more suitable than indicator cross-covariances for quantifying spatial continuity in categorical data [3].

Based on the concept of transition probability, [10] revived interest in the transiogram as a measure of spatial continuity and modeled categorical data using 2D discrete Markov Chains [9]. The transiogram can be regarded as a model of transition probabilities as a function of \( h \), i.e., \( \pi_{k'k}(h; \theta_{k'k}) \). Similarly with variograms, transiograms can also be modeled from sample data. First, one needs to compute transition probabilities for different lags, i.e. estimate the experimental transiograms, and then fit such experimental transiograms with mathematical models.

Capitalizing on the notion of a transiogram, [10] developed a Spatial Markov Chain (SMC) model, whereby the conditional probability of class occurrence, is given as:

\[
\left[ \hat{\pi}_k(x_0) \right]_{SMC} = \hat{P}\{i_k(x_0) = 1 | i_v(west) = 1, i_v(north) = 1, i_v(east) = 1, i_v(south) = 1\}
\]

\[
\left[ \sum_{i} [\pi_{k'k}^I(h_i^1) \pi_{k'k}^I(h_i^2) \pi_{k'k}^I(h_i^3) \pi_{k'k}^I(h_i^4)] \right] \tag{2.6}
\]
where \( x, y \) represent the axes directions, \( k^{(1)}, k^{(2)}, k^{(3)} \) and \( k^{(4)} \) all represent the states of the Markov Chain (class indicators) at neighboring locations, \( x_1(west), x_2(north), x_3(east), x_4(south) \), and \( h_1, h_2, h_3, h_4 \) represent the distances between these four nearest neighbors and the current location \( x_0 \) where prediction or simulation is performed. Equation (2.6) provides an extremely simple way for merging pre-posterior (two-point) transition probabilities of class occurrence into posterior probabilities. Its simplicity is due to the assumption of conditional independence, which is difficult to corroborate in real-world cases.

### 3. Simulation Examples

To investigate the spatial patterns implied by the above three approaches, sequential simulation of categorical fields is performed without any conditioning data (unconditional simulation), so that the patterns generated by each algorithm are more easily revealed. The truncated Gaussian simulation (TGS) method [14] is a popular way for simulating categorical fields. This is partly due to the existence of multiple algorithms to generate realizations of Gaussian random fields, but mostly to the fact that TGS guarantees that the indicator auto- and cross-covariances of simulated realizations are all consistent with each other [14].

We consider \( K = 3 \) categories with labels \( k = 1, 2, 3 \) and global proportions \( \pi_1 = 0.35, \pi_2 = 0.4, \pi_3 = 0.25 \); simulation is performed at the grid nodes of a \( 300 \times 300 \) regular raster with unit spacing. The TGS method is applied to generate one simulated realization of class labels, from which we compute the indicator auto- and cross-variograms and their transition probability counterparts; these are displayed in Figure 1.

![Figure 1 Sample auto-/cross-semivariograms (solid lines) and sample transition probabilities (dashed lines) from TGS simulation](image)

These sample measures of spatial structure are fitted with theoretical models, and unconditional sequential simulation with IK and ICK, along with SMC simulation, is then conducted. One realization from each simulation algorithm is given in Figure 2. The sample auto-/cross-variograms of these simulation results are computed and compared in Figure 3 with the target models obtained via the TGS method.

---

28
3.1. Sequential Indicator Simulation with Indicator Kriging

For sequential indicator simulation using IK, three spherical auto-variogram models were fit to the diagonal plots of Figure 1. These indicator variogram models were then used along with IK in sequential simulation to estimate three conditional probabilities at each grid node. The entire procedure can be summarized as:

1. A random simulation path is defined.
2. For each node on the path:
   i. Search neighboring nodes to extract simulated class labels to be used as data, and compute the occurrence probability for each class label using IK. If there are no nearby informed nodes, or if this is the first node in the simulation path, use the global proportions as the estimated conditional probabilities. A class label is simulated from the computed probabilities and assigned to this node.
   ii. Proceed to the next node along the random path, and repeat step i until all nodes of the random path are visited once.

The entire procedure is repeated to generate another realization, possibly with a different random path associated with each such realization. Figure 2 (top right graph) displays one such realization generated with sequential simulation and IK. Figure 3 displays the ensemble average of indicator auto- and cross-variograms computed from 50 simulated realizations. One can easily appreciate that the nesting of different classes in the SISIM realization is not the same as the one found in the TGS realization. This is corroborated by the mismatch between the ensemble average indicator variograms (particularly for the transition between classes 1 and 3) and the target ones corresponding to the TGS realization. Note, however, that even if one does not inject into the simulation any information regarding inter-dependencies between classes, the ensemble averaged indicator cross-variograms of the SISIM realizations (apart from that between categories 1 and 3) are not too different from the indicator cross-variograms of the TGS realization.

3.2. Sequential Indicator Simulation with Indicator coKriging

In this case, a set of 9 auto- and cross-semivariogram functions must be jointly modeled; here, the linear model of coregionalization (LMC) is employed for this purpose [7]. More specifically, a model with 3 nested structures, a nugget model, a spherical model and an exponential model, was fit to the sample auto- and cross-semivariograms of Figure 1 using Goulard’s iterative method [15]. Although this iterative method does not necessarily converge, experience shows that it always converges and leads to similar results for different initial values.
Once the parameters of the LMC are estimated, the simulation procedure follows the same steps with the one used in the indicator kriging case. The only difference is that the dependencies between different class labels, which are ignored in indicator kriging, are now taken into account via indicator cross-semivariogram models. Figure 2 (bottom left graph) displays one simulated realization obtained from SISIM with indicator coKriging. One can easily appreciate that the differences in spatial texture between the target TGS realization and the SISIM generated one still exist. This is corroborated in Figure 3, where the ensemble average indicator variograms are even more different than the TGS-target ones. It is indeed, not uncommon in the literature to find references to poorer performance of indicator coKriging over indicator Kriging; this comes to re-instate the argument that the performance improvement associated with indicator coKriging does not suffice to justify the dramatic increase in both human and computational effort.

3.3. Spatial Markov Chain Sequential Simulation
Since the properties of transiogram and its mathematical forms are not that clear, the sample transition probabilities computed from the TGS realization are used directly, without fitting any theoretical model to them. Since these are sample transition probabilities, they satisfy probability constraints by construction. In addition, since no Kriging is involved in this algorithm, one need not consider the LMC.

The SMC simulation proceeds in the following steps:
1. For a node on the boundary of the simulation grid, a class label is simulated from the global class proportions and assigned to this node.
2. An alternating advancing path is defined; this essentially amounts to visiting the nodes of each grid row from different directions as one progresses from the top to the bottom of the grid.
3. For each node on this path:
   i. If the current node is not on boundary, the 4 nearest neighbors of this node are found (rook’s interaction), and the conditional probabilities of class occurrence are computed based on equation (2.6). A new class label is simulated from these conditional probabilities and assigned to this node.
   ii. Proceed to the next node along the alternating advancing path, and repeat steps i to ii.

Figure 2 (bottom right graph) displays a realization obtained using the SMC model. It is evident that in the absence of conditioning data, the SMC model has no ability to generate any realistic spatial pattern. As Figure 3 indicates, the ensemble averaged indicator variograms for the SMC case are the most different from target TGS-derived indicator variograms.

Figure 3 Sample auto-/cross-semivariogram from initial TGS simulation (solid lines). Ensemble average indicator auto- cross semivariograms generated using SISIM with indicator kriging (dashed line), indicator cokriging (dashed dotted lines) and SMC (dotted lines)

30
4. Discussion

In the SMC method, transition probabilities are combined using the naïve Bayes’ rule under the assumption of conditional independence among neighbors. This assumption is difficult to verify in real-world cases and is the reason for the bad performance of unconditional sequential simulation with naïve Bayes’ fusion. Although indicator Kriging methods account for the redundancy between neighbors, they do not yield satisfactory results, most probably due to the linear way in which individual transition probabilities are fused. In what follows, we propose a simple avenue for incorporating the non-linearity of Bayesian fusion with the redundancy information of neighbors accounted for by indicator coKriging.

According to Bayes’ rule and the definition of conditional probability, the posterior (multi-point or multi-source) conditional probability of \( k \)-th class occurrence at location \( x_0 \) can be decomposed as:

\[
P\{k(0) | k(n), n = 1, \ldots, N\} = \frac{P\{k(0) | k(n)\} P\{k(1) | k(0)\} \cdots P\{k(N) | k(0), \ldots, k(N-1)\}}{P\{k(n), n = 1, \ldots, N\}}
\]

(4.1)

where \( k(n) \) is the outcome \( k(n) = c(x_n) \in \{1, \ldots, K\} \) of the categorical RV \( C(x_n) \) at location \( x_n \).

In a spatial context, assuming conditional independence of \{\( C(x_n), n = 1, \ldots, N \)\} given \( C(x_0) = k(0) \), and stationary class proportions, the above conditional probability can be rewritten as:

\[
\hat{P}\{C(x_0) = k | d, \pi\} \propto \pi_k \prod_{n=1}^{N} \frac{\pi_{n,k} \cdot (h_{n,0})}{\pi_k}
\]

(4.2)

where the proportionality constant is comprised of the sum of the numerator evaluated across all \( K \) classes.

In real-world cases, it is difficult to verify the assumption of conditional independence among source data (here neighboring class labels). In order to account for the dependence or redundancy among data sources, as [8] suggested, different weighting coefficients could be assigned to each source (neighbor) according to a certain criterion. Equation (4.2) can then be changed to:

\[
\hat{P}\{C(x_0) = k | d, \pi\} \propto \pi_k \prod_{n=1}^{N} \frac{\pi_{n,k} \cdot (h_{n,0})^{\beta_n}}{\pi_k}
\]

(4.3)

The main problem here is the determination of the exponent \( \beta_n \). A solution might be to use the relationship between the covariance and the transition probability discussed in the Section 2. More precisely, using that link the dual ICK estimate of the conditional probability in Equation (2.3) becomes:

\[
\left[ \hat{p}_k \left( x_0 \right) \right]_{n,\text{ICK}} = \hat{P}\{C(x_0) = k | d, \pi\} = \pi_k + \sum_{k'=1}^{K} \sum_{n=1}^{N} w_{n,k,k'} \cdot \pi_{n,k} \cdot (h_{n,0}) - \pi_k
\]

(4.4)

In words, there appears to be a link between the exponent \( \beta_n \) associated with the ratio \( \pi_{n,k} \cdot (h_{n,0}) / \pi_k \), quantifying the information content of the \( n \)-th datum, and its corresponding dual ICK weight \( w_{n,k,k'} \). This link suggests an avenue for modeling the information redundancy in observed class labels. To avoid solving the indicator coKriging system of equations, hence the need to fit a LMC to all indicator auto- and cross-variograms, one could use simplifications of ICK, such as indicator principal component kriging [16] and min-max autocorrelation factors [17].

5. Conclusions and Future Work

Three approaches for combining two-point transition probabilities into posterior (multi-point) probabilities of class occurrence were surveyed in this paper, namely indicator kriging, indicator coKriging, and naïve Bayes’ rule. Unconditional simulation was performed using these three methods to reveal the spatial patterns implied by these methods, without the obscuring effect of conditioning data. Our results show that SMC performs worst among the three approaches, mostly due to its failure to model spatial interactions among neighbors. The SMC model relies on transition probabilities by assuming conditional independence, which is rare in spatial data. On the contrary, transition probabilities have clear advantages over indicator covariances and variograms when it comes to modeling categorical data. Transition probability based methods are therefore promising for prediction and simulation of categorical fields. Indicator Kriging and indicator coKriging are two methods that are based on transition probabilities, if one expresses these methods in their dual form. Kriging-based methods, however, do not constitute the best way of combining transition probabilities.
probabilities, as was indicated by the results of our case study. In particular, the linear (additive) combination of transition probabilities is in disagreement with probability fusion (e.g., Bayesian) rules that are typically multiplicative.

We proposed a simple way to proceed in searching for ways to combine transition probabilities, by synthesizing the advantages of the SMC and geostatistical methods. We would like to end this paper with a warning against the uninformed use of SMC under conditional independence: its application and possible generalization to higher dimensional, dynamic and scattered data cases should not proceed without a sound knowledge about spatial processes.

6. References


