Uncertainty Characterization in Remotely Sensed Land Cover Information

ZHANG Jingxiong  ZHANG Jinping  YAO Na

Abstract  Uncertainty characterization has become increasingly recognized as an integral component in thematic mapping based on remotely sensed imagery, and descriptors such as percent correctly classified pixels (PCC) and Kappa coefficients of agreement have been devised as thematic accuracy metrics. However, such spatially averaged measures about accuracy neither offer hints about spatial variation in misclassification, nor are useful for quantifying error margins in derivatives, such as the areal extents of different land cover types and the land cover change statistics. Such limitations originate from the deficiency that spatial dependency is not accommodated in the conventional methods for error analysis.

Geostatistics provides a good framework for uncertainty characterization in land cover information. Methods for predicting and propagating misclassification will be described on the basis of indicator samples and covariates, such as spectrally derived posteriori probabilities. An experiment using simulated datasets was carried out to quantify the error in land cover change derived from postclassification comparison. It was found that significant biases result from applying joint probability rules assuming temporal independence between misclassifications across time, thus emphasizing the need for the stochastic simulation in error modeling. Further investigations, incorporating indicators and probabilistic data for mapping and propagating misclassification, are anticipated.

Keywords  geostatistics; land cover change; misclassification; stochastic simulation
CLC number  P208; P237.1

Introduction

Information and analysis about occurrences and dynamics of land cover over space and time are important for many spatial applications, such as environmental and ecosystem modeling and global change research\(^{[1-6]}\). Sophistication of remote sensing and geographic information systems (GIS) has enhanced land cover mapping and the provision of collateral information on biophysical variables.

However, information derived from remote sensing often suffers from various errors because many biophysical processes underlying land cover and land cover change cannot be remotely monitored with adequate accuracy due to the difficulties of separating one class from another when both are showing similar spectral signatures and, in a changing environment, discriminating real changes from natural variations such as plant phenology\(^{[7, 8]}\). Besides, remote sensor data are subject to measurement error, and information extraction will suffer from various man-machine
limitations. Inaccuracy in remote sensing information will have complex implications on the resultant models concerning landscape dynamics and ecosystem processes, as uncertainty enters and creates weak links in the chain of information extraction and knowledge construction. As land cover mapping and change detection have been extended from local, regional, national, to global scale, it is required that error in remotely sensed land cover information and its propagation in derivative products be quantified and handled correctly\cite{9}.

There has been increasing research on uncertainty in land cover information and spatial analysis \cite{10-13}. Uncertainty for categorical maps can be approached from the inaccuracy in position, i.e., area-class boundary placement, and class labeling, respectively. Positional errors are processed by identifying and removing sliver polygons, i.e., mismatches between versions of polylines or polygon boundaries\cite{14}. Slivers removal is often done by carefully setting the tolerance of positional errors. Errors in classification are analyzed by using the error matrices where classification results and their references are compared, and the hits and misses between them are cross-tabulated. Error matrices provide the basis for computing percent correctly classified pixels and kappa coefficients of agreement \cite{15}. Methods for estimating the variance of classification accuracy are established well, and recent research has explored the use of spatially varying probabilities of misclassification evaluated at individual locations \cite{16}.

It is common to apply the law of variance and covariance propagation to quantify error in a derivative quantity given knowledge of the input variables’ variance and covariance, as well as the function form leading to the specific quantity. For error modeling with spatial data, the simple applications of the law of variance and covariance propagation assuming spatial and cross-variable independence will lead to biased quantification of standard errors in derivatives. This is because geo-processing, such as the summation of grid cells labeled with certain land cover types and the differencing of percent tree cover maps for change detection, often operates over certain neighborhoods and invokes cross-map variables.

Geostatistics provides a good framework for spatial uncertainty analysis. This paper attempts to extend geostatistics to uncertainty characterization in remotely sensed land cover information through spatially explicit quantification of misclassification and erroneous change detection. These methods can be used for a range of scales and facilitate error propagation in spatial predictive modeling, such as land use and land cover change.

1 Probabilistic mapping of misclassification

Suppose a domain is discretized into locations denoted by \( x \), which takes values in a chosen coordinate system that may be 2-dimensional, 3-dimensional, or 4-dimensional if a spatiotemporal frame is assumed. Let \( Z(x) = (Z_1(x), \ldots, Z_b(x)) \) be a covariate vector defined on a feature space of dimension \( b \) (\( b \) being a positive integer) and located at \( x \) (e.g., a Landsat TM pixel). \( Z(x) \) consists of sensor measurements, such as reflectance, and physical properties. At location \( x \), the prevailing class is denoted as \( C(x) \), which takes values in set of class codes \( \{1, \ldots, K\} \), where \( K \) stands for the total number of classes under consideration or for a vector of length \( K \), with individual components pertaining to some categorical likelihood.

Consider a population \( P \) comprised of elements \( T(x) = (Z(x); C(x)) \). A training sample of size \( n \) collected from \( P \) is denoted by \( T = \{t(x_1), \ldots, t(x_n)\} \). The posterior probability that \( x \) belongs to class \( k \), given the covariate vector \( Z \), is denoted by \( P(C=k \mid Z) \). Let \( \eta \) denote a classification rule trained on \( T \), and \( \eta(Z(x)) \) denote a prediction of \( C \) obtained by applying the classification rule to a pixel \( x \) with covariate \( Z(x) \).

A classifier can be seen as the mapping from measurement to class labels and expressed as

\[
\hat{C}(x) = \eta(Z(x)) = \arg \max_{k=1, \ldots, K} F_k(Z(x)) \tag{1}
\]

where \( F_k \) calculates measures of proximity to indicate categorical similarity to class \( k \), with the prevailing class at location \( x \) taking the maximum utility. For instance, the \( k \)-nearest neighbor (\( k \)-NN) classifier assigns \( Z(x) \) to the class with the largest plurality among the \( k \) nearest neighbors of \( x \) among the training data where the normal Euclidean distance be-
The $k$-NN classifier can also be formulated as an estimator of the $K$ posterior probabilities, which are the sample proportions of the $K$-nearest neighbors in the training sample that belong to $K$ classes. The classifier assigns $x$ to the class with the maximum posterior probability.

For every possible value of $Z$, there is a probability distribution expressing the relative likelihood of membership within each class. This hints on the notion of a **covariate space** in which the probability of group membership varies across the space. Classification errors occur because groups are not completely separated in the covariate space. In other words, the spatially varying probabilities of class attainment originate from the spatial variability of the underlying covariates, signaling inherent uncertainty in classification.

Clearly, there is varied likelihood of pixels being labeled as candidate classes, depending on their similarity to named classes in the covariate space. The probability of correctly classifying $x$ using the rule $\eta$ is denoted by $P(\eta(Z) = C \mid Z)$. The probability that $t = (Z, C)$ is correctly classified, given $Z$, is the maximum posterior probability\[16, 17\]:

$$p\left[\eta(Z(x))=C(x)\mid Z\right]=\arg\max_{k=1,\ldots,K} p(C(x)=k\mid Z(x))$$

The classifier $\eta$ allows the construction of estimators $P(C = k \mid Z)$, for each class $k$, thus classification accuracy.

The focus of this paper is misclassification, which can be thought of as one minus the maximum posterior probability of class allocation:

$$p(I_{c}(x) = 1\mid Z(x)) = 1 - p\left[\eta(Z(x)) = C(x)\mid Z(x)\right]$$

$$= 1 - \arg\max_{k=1,\ldots,K} p(C(x)=k\mid Z(x))$$

where $I_{c}(x)$ denotes an indicator variable for misclassification at $x$.

If land cover maps at times $t_1$ and $t_2$ are accompanied by surfaces of misclassification probability, it is still not straightforward to arrive at the estimation of probability of misclassification in the resultant maps of land cover change. To see this in detail, we first consider the issue about how to quantify uncertainty in land cover change at location $x$, given misclassification probability maps corresponding to the land cover maps at times $t_1$ and $t_2$, denoted as $P^1(I_c(x))$ and $P^2(I_c(x))$. It may be tempting to write the probability of misclassification in land cover change, $P^1\cup^2(I_c(x))$, upon overlaying the two maps of misclassification probability evaluated for land cover maps at times $t_1$ and $t_2$, as

$$P^1\cup^2(I_c(x)) = P^1(I_c(x)) + P^2(I_c(x)) - P^1(I_c(x))P^2(I_c(x))$$

where the calculation of joint probability is based on the intersection of independent events of misclassification on the pair of bi-temporal land cover maps.

However, there is often the case that the misclassified pixels or parcels on bi-temporal land cover maps tend to be positively correlated, especially at the unchanged pixels that often dominate the landscape (while changed pixels are relatively rare events). Thus, the evaluation of misclassification on the land cover change map should be conducted through a more elaborated procedure that takes account for co-occurrence of misclassification on the bi-temporal land cover maps\[12, 18\]. What is needed is an error model or mechanism that can simulate realized maps conforming to observed distributions of categories and their structures quantified by variogram models\[19\], as discussed in the next section.

## 2 Geostatistics for propagating misclassification

A fundamental concept in geostatistics is regionalized random variables, of which land cover, land cover change, and biophysical variables are examples. Spatial dependence plays an important role in inferential statistics for regionalized variables, which is effectively quantified by variogram models describing spatial (cross) covariance. Frequently, the stationarity in the moments of random variables is assumed to facilitate spatial prediction and stochastic simulation.

Spatial covariance, $\text{cov}_A(h)$, is defined as the expectation of the product of a variable $Z$'s (to be distinguished from the case of vector $Z$ discussed in the preceding section) deviates from local means $m_A(x)$
for locations separated by a lag of \( h \):

\[
\text{cov}_{z}(h) = E(Z(x_{i}) - m_{z}(x_{i}))(Z(x_{j}) - m_{z}(x_{j}))
\]

\[
\text{for } x_{i} - x_{j} = h
\]  

(5)

where \( Z(x_{i}) \) and \( Z(x_{j}) \) are the values of variable \( Z \) at locations \( x_{i} \) and \( x_{j} \), respectively, which are separated by a lag of \( h \).

Given a set of samples, the kriged estimate for a variable \( Z \) at location \( x \) may be pursued as linear combination of data values at sampled locations. With knowledge of the mean \( m_{z} \), simple kriging estimate for an unsampled location \( x \) is obtained as

\[
z(x)^{\text{s}} = m_{z} + \sum_{j=1}^{n} \lambda_{j}(z(x_{j}) - m_{z})
\]

(6)

where \( \lambda_{j} \) stands for the weight attached to the sample located at \( x(j = 1, 2, \ldots, n) \) within the search neighborhood, which are, in turn, determined by

\[
\sum_{j=1}^{n} \lambda_{j} \text{cov}_{z}(x_{j} - x_{i}) = \text{cov}_{z}(x - x_{i})
\]

(7)

where \( \text{cov}_{z}(x_{j} - x_{i}) \) elements are \( Z \) covariance between sampled locations \( x_{j} \) and \( x_{i} \), while \( \text{cov}_{z}(x - x_{i}) \) stands for the covariance between location \( x_{i} \) and the unsampled location \( x \).

The indicator variables can be derived from the analysis of sampled locations, where the misclassification can transformed into ones and zeros:

\[
i(x) = \begin{cases} 1, & \text{if } \eta(Z(x)) \neq C(x) \\ 0, & \text{otherwise} \end{cases}
\]

(8)

where \( \eta(Z(x)) \) and \( C(x) \) are the predicted and true class labels at location \( x \).

The indicator random variables, being binary, possess some special characters, which lend themselves to straightforward the evaluation of some statistics and probabilities. One is the expectation of indicators:

\[
E(I(x)) = 1 \cdot \text{prob}(I(x) = 1) + 0 \cdot \text{prob}(I(x) = 0) = p_{i}(x)
\]

(9)

and the other is the covariance of indicators at locations \( h \) apart, i.e., for \( x_{i} - x_{j} = h \):

\[
\text{cov}_{I}(h) = E((I(x_{i}) - p_{i})(I(x_{j}) - p_{j}))
\]

\[
= \text{prob}(I(x_{i}) = 1 \cap I(x_{j}) = 1)) - p_{i}^{2}
\]

(10)

where \( p_{i} \) is misclassification probability.

Analyzing Eqs.6 and 7, it is possible to determine the local probability of misclassification, conditional to existing data, using simple indicator kriging:

\[
p_{E}(x)^{s} = i(x)^{s} = p_{E} + \sum_{s=1}^{n} \lambda_{s}(i(x_{s}) - p_{E})
\]

(11)

where \( i(x_{s}) \) represents the indicator transform of a sample point \( x_{s} (s = 1, 2, \ldots, n) \); \( \lambda_{s} \) is the weight associated with the sample point \( x_{s} \), which are derived from solving a kriging system of the form shown in Eq.8 (replacing \( \text{cov}_{z} \) by \( \text{cov}_{I} \)); and \( p_{E} \) is the probability of misclassification inferred from sample data.

Suppose that there exist samples from validation samples (primary) and maximum posteriori probability in classification (secondary) for a land cover mapping project. To predict the probability of misclassification, drawing upon1

\[
\text{probits}(x) = p_{E} + \sum_{s=1}^{n} \lambda_{s}(i(x_{s}) - p_{E}) + \sum_{j=1}^{n} \lambda_{j}(j(x_{j}) - j_{E})
\]

(12)

where \( \lambda_{s} \) and \( \lambda_{j} \) are the weights assigned to \( s \)-th \( i \) datum and \( s \)-th \( j \) datum for prediction of misclassification probability at location \( x \), and \( p_{E} \) and \( j_{E} \) are the means of primary and secondary variables [19].

The provision of only location-specific class probabilities \( p_{i}(x) \) is, however, not sufficient for error propagation in categorizing land cover change, because there is no way to evaluate spatial or temporal joint probabilities, \( \text{prob}(I(x_{s})I(x_{j}) = 1) \), or two-point statistics, \( E(I(x_{s})I(x_{j})) \), unless independence between pixels but perfect positive spatial dependence within pixels is assumed. For the characterization of spatio-temporal uncertainty, therefore, it is crucial to devise models for generating map realizations honoring spatial structures observed in empirical data to compute joint distributions of (in)correctly classified locations.

Working with indicators, stochastic simulation proceeds by building a conditional cumulative distribution function (ccdf) as

\[
\text{ccdf}_{E}(x|\text{data}) = \sum_{k=1}^{K} p_{E}(x|\text{data}, k = 1, \ldots, K)
\]

(13)

where the conditional probabilities \( p_{E}(x|\text{data}) \) is estimated from the indicator kriging set forth in
Eqs.11 or 12 if both primary and secondary data are available, and the conditioning data consist of neighboring original indicator data and previously simulated indicators, with the former being withheld for unconditional simulation. Draw a random number $p$ uniformly distributed between 0 and 1. The category simulated at location $x$, $c(x)^{(0)}$, is the category that corresponds to the probability interval including $p$, i.e.,

$$c(x)^{(0)} = k \text{ if } p \in ( \text{ccdf}_{k-1}(x \mid (\text{data})), \text{ccdf}_{k}(x \mid (\text{data})) )$$

(14)

Add that simulated value to the conditioning data-set and continue the previous four steps for the next node along the random path until the problem domain is exhausted.

The stochastic simulation can be performed to generate the realizations for misclassifications for the bi-temporal land cover maps, with spatio-temporal variogram models specified and conditioning data incorporated into the process of ccdf quantification. Set the indicator to 1 if misclassification occurs or is observed at a location. One will be able to characterize how misclassifications in land cover change maps behave by applying a simple logical “OR” rule to combine the pairs of simulated bi-temporal maps of ones and zeros and proceeding to some statistical summary, such as mean of indicators $i(x)$, i.e., the probability of misclassification at location $x$ on a map of land cover change:

$$p_{e}^{12}(x) = \frac{1}{N} \sum_{z=1}^{k} i^{(0)}(x; e^{1} \cup e^{2})$$

(15)

where $N$ is the number of realizations generated for the purpose of quantifying misclassification in land cover change. The location set $\{x\}$ can be made to exhaust any patches of pixels or parcels of irregular polygons to facilitate the quantification of uncertainty in any complex queries about land cover information over space and time.

### 3 Experiment with simulated data

Because of the ease of controlled experiment, stochastic simulation was performed to generate equal-probable realizations of spatially and temporally correlated error surfaces to be superimposed on the mean maps to produce areal-class maps corrupted with prescribed errors. First, mean categorical maps depicting reference classes for two dates were generated from thresholding continuous fields of $Z_1$ and $Z_2$, which were simulated using the GSLIB software system, given variogram models$^{[20]}$. The simulation was run over a grid system of $100 \times 100$, each cell being unit square.

The two random fields have to be cross-correlated to emulate the typical land cover change scenarios commonly observed in reality, that is, change occurs less probably than nonchange, keeping the world evolving. This was made possible by linear combination of differing proportions of two simulated random fields, with a parameter to control their intercorrelation. In this study, the global percent change was about 32%. The simulated surfaces were smoothened by taking moving averages over local windows to ensure that they looked like typical areal-class maps.

Alternatively, it is possible to generate a realized $Z_1$ field with only its autocovariance model using unconditional simulation, followed by simulating a realized $Z_2$ field with both its auto- and cross-covariance models specified, conditional to the $Z_1$ field simulated beforehand. The efficiency in simulation can be achieved by having all subsequent realizations conditioned to the first versions of $Z_1$ and $Z_2$, respectively. All $Z_2$ realizations were made to conform to the cross-covariance model in addition to its autocovariance model. The results from applying the former method are shown in Fig.1 (a) and (b), where the mean class maps at two times are shown, each of three classes.

The error surfaces were then simulated independently on mean class surfaces. To reproduce the interdependence between errors observed with bi-temporal areal-class maps, 100 equal-probably realized error surfaces for time 1 were first generated from the Gaussian sequential simulation with only its autocovariance model using unconditional simulation. The error surfaces for time 2 were simulated with both its auto- and cross-covariance models specified, conditional to error surfaces simulated for time 1 beforehand. Another way for simulating error surfaces for time 2 was to add white noise to simulated error sur-
faces at time 1 through certain signal-to-noise ratio (SNR). As examples, error-distorted areal-class maps are shown in Fig.1 (c) and (d) for time 1 and time 2, respectively, whereas the SNR was set to 10.

The errors are interdependent through time because some classes are inherently more difficult to classify than others, and class boundaries are often mixtures. At nonchange locations, the errors tend to be more strongly correlated, as same spatial classes exhibit similar natural variations and uncertainty in measurement and analysis. This is clearly illustrated in Fig.1 (c) and (d). The 100 pairs of equal-probably realized error surfaces for time 1 and time 2 were superimposed upon the bi-temporal mean categorical maps to generate 100 pairs of error-contained categorical maps. These maps were then compared with their corresponding mean categorical maps (i.e., reference) to produce 100 pairs of binary maps, in which 1s indicate instances of misclassification. Each pair of binary maps were processed by the logical OR formulated in Eq.16 to get another binary map to emulate the distribution of error in a land cover change map. Three sets of simulated error maps were summarized to get three maps depicting misclassification probabilities in two single-date land cover maps and a land cover change map. These are shown in Fig.2 (a), (b), and (c), respectively, where the darker gray shades indicate the higher probability of errors (gray scale in the interval between 0 and 1).

For comparison, the probability surface of erroneous land cover change types is computed using Eq.4, and is shown in Fig.2 (d). To see this comparison clearly, histograms are shown in Fig.3 (a) and (b), where the former illustrates the overestimation of error propagation by assuming independence between errors in bi-temporal categorical maps at locations of non-change, while the latter shows nonsignificant biases at changed locations.

The stochastic simulation will be valuable addition to any endeavors that have a vision on spatial uncertainty. The spectrally derived posteriori probability should also find good use in accuracy assessment in land cover change analysis, as shown in Eq.12.

While experimentation with simulation data was reported in this paper, further studies with real data sets are required. A land use and land cover remote sensing project, utilizing Landsat TM, SPOT5, and map data, for the bi-temporal period of 2002 to 2005 in a Wuhan suburb is ongoing, in which a hierarchy of land cover data will be established so that training and testing data are sampled adequately to permit rigorous validation.

4 Conclusion

This paper has shown that geostatistics is capable
of quantifying spatial uncertainty in land cover information. In particular, the stochastic simulation has demonstrated its utility for propagating misclassification in land cover to error in change information through reproducing auto- and cross-covariance in misclassification and by honoring conditional data. Geostatistics will also offer sound techniques for addressing the issues of scale in heterogenous land cover data so that scalable land cover information products may be developed with uncertainty properly quantified[21]. This will lead to fruitful research investigating the interaction of spatio-temporal independence and scale and their implications for uncertainty.

References

[1] Skole D, Tucker C (1993) Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988[J]. Science, 260:1 905-1 909
[2] DeFries R S, Townshend J R G (1994) NDVI-derived land cover classification at a global scale[J]. International Journal of Remote Sensing, 15: 3 567-3 586
[3] Goodchild M F (1994) Integrating GIS and remote sensing for vegetation analysis and modeling: methodological issues[J]. Journal of Vegetation Science, 5:615-626
[4] Bruzzone L, Serpico B (1997) An interactive technique for the detection of land-cover transitions in multitemporal remote-sensing images[J]. IEEE Transactions on Geoscience and Remote Sensing, 35 (4): 858-867
[5] Tso B, Mather P M (2001) Classification methods for remotely sensed data[M]. London: Taylor & Francis
[6] Muchoney D, Strahler A (2002) Regional vegetation mapping and direct land surface parameterization from remotely sensed and site data[J]. International Journal of Remote Sensing, 23 (6):1 125-1 142
[7] Coppin P, Jonckheere I, Nackaerts K, et al. (2004) Digital change detection methods in ecosystem monitoring: a review[J]. International Journal of Remote Sensing, 25 (9):1 565-1 596
[8] Lambin E F, Ehrlich D (1996) The surface temperature-vegetation index space for land cover and land cover change analysis[J]. International Journal of Remote Sensing, 17: 463-487
[9] Stehman S V, Sohl T L, Loveland T R (2003) Statistical sampling to characterize recent United States land-cover change[J]. Remote Sensing of Environment, 86 (4): 517-529
[10] Hunsaker C T, Goodchild M F, Friedl M A, et al. (2001) Spatial uncertainty in ecology: implications for remote sensing and GIS applications[M].New York: Springer-Verlag
[11] Foody G M (2002) Status of land cover classification accuracy assessment[J]. Remote Sensing of Environment, 80 (1): 185-201
[12] Zhang J, Goodchild M F (2002) Uncertainty in geographical information[M]. London and New York: Taylor & Francis
[13] Kyriakidis P C, Zhang J (2003) A geostatistical framework for accuracy assessment of remotely sensed land-cover information[C]. Presented at Geocomputation, Southampton
[14] Chrisman N R (1989) Modeling error in overlaid categorical maps[M]/Goodchild M F, Gopal S (eds). Accuracy of Spatial Databases.London: Taylor & Francis
[15] Congalton R G (1991) A review of assessing the accuracy of classifications of remotely sensed data[J]. Remote Sensing of Environment, 37: 35-46
[16] Steele B M, Patterson D A, Redmond R L (2003) Towards estimation of map accuracy without a probability test sample[J]. Ecological and Environmental Statistics, 10: 333-356
[17] Ripley B D (1996) Pattern recognition and neural networks[M]. Cambridge: Cambridge University Press
[18] Arbia G, Griffith D, Haining R P (2003) Spatial error propagation when computing linear combination of spectral bands: the case of vegetation indices[J]. Environmental and Ecological Statistics, 10 (3): 375-396
[19] Journel A G, Huijbregts C J (1978) Mining geostatistics[M]. London: Academic Press
[20] Deutsch C V, Journel A G (1998) GSLIB: geostatistical software library and user’s guide[M]. New York: Oxford University Press
[21] Atkinson P M (2001) Geostatistical regularization in remote sensing[M]/Tate N J, Atkinson P M (eds). Modeling Scale in Geographical Information Science. Chichester: John Wiley & Sons