A Novel Method for Mapping Land Cover Changes: Incorporating Time and Space with Geostatistics *

Alexandre Boucher, Karen C. Seto, and André G. Journel
Stanford Center for Reservoir Forecasting
Department of Geological and Environmental Sciences

April 20, 2005

Abstract

Satellite images are a principal medium to detect and map changes in the landscape, both in space and time. The method proposed here aims to better exploit remotely sensed data by simultaneously taking into account the spatial and the temporal relations between land covers for multitemporal mapping and change detection.

At each location on a landscape, a time series of land cover classes is modeled with transition probabilities. That time series at any specific location is estimated with pixel-specific satellite measurements, the neighboring ground truth land cover data, and any neighboring previously estimated time series deemed well-informed by the satellite measurements. The spatial context is incorporated with indicator kriging and image segmentation which constrain the relevance of the spatial data.

When space and time are both integrated with a maximum likelihood classifier, the prediction accuracy of the time series improves significantly, increasing from 31% to 61%, compared with the likelihood classifier alone. The consideration of spatial continuity also reduced unwanted speckles in the classified images, removing the need for any post-processing.

Keywords
Kriging, Variogram, Edges, Context, Transitions probability, Combining priors.

1 Introduction

Mapping of land cover changes is intrinsically a space time operation, yet current image processing methods often do not consider simultaneously the spatial and temporal contexts when estimating land cover change.

A common method to incorporate spatial information is to model the spatial distribution of land cover classes as a Gibbs-Markov random field (GMRF) (Li, 2001; Tso and Mather, 2001). The GMRF calls for minimizing an energy function U computed from cliques, which are templates relating mutual neighbors. The energy is a global measure of the spatial distribution of the classes in the image. Classification is aimed at producing an image

*Submitted to IEEE Geosciences and Remote Sensing

where the classes are such that the energy is low. The common procedure is to perturb the classification to minimize the energy and stop when convergence is reached. That iterative procedure can be computationally demanding for large domains.

Geostatistics, especially indicator kriging, has also been used to incorporate spatial autocorrelation in estimating or simulating classes (Stein et al., 1998; Atkinson and Lewis, 2000; Brown et al., 2002; Goovaerts, 2002; Wang et al., 2004). Kriging has the advantage of not being iterative, but requires many conditioning data in order to provide useful contextual information. Contextual relations between classes require the modeling of a linear model of coregionalization (Goovaerts, 1997), a task that is both data and time-demanding.

An accurate cross-sectional classification is not sufficient to map land cover changes. At any specific locations, the time series of land covers must also be accurate. The temporal component becomes as important as the spatial component when one wants to know both when and where changes have occurred.

There are two main issues with change detection; (a) the combination of the images and (b) the classification of those images. The simplest change detection method classifies each image independently, changes are then mapped by identifying which pixels have changed. The problem lies in the errors associated with the mapping of changed classes. The final accuracy is approximately the product of accuracy associated with the classification performed at each time, often with poor results.

A second method consists of analyzing the images concurrently and classifying the class trajectories. For example, instead of classifying an image into label 1 or 2, all the possible transitions between those classes (1 to 1, 1 to 2, 2 to 1 and 2 to 2) are considered, expanding to potentially K^{N_t} transitions, where K is the number of classes, and N_T the number of images.

A third method models every pixel as a time series, where the time of change is estimated. For example Kaufmann and Seto (2001) utilize time series econometrics to detect dates of change with better results than when the changes are obtained from post-processing independently classified images.

Another common method to process multitemporal images is the cascade approach (Swain, 1978) which consists of analyzing the sequence of image in chronological order. Past classifications being used to condition future classifications.

Few applications have looked at change detection using spatial (contextual) information (Jeon and Landgrebe, 1992; Wang et al., 2004; Kasetkasem and Varshney, 1992; Melgani and Serpico, 2003, 2001). Wang et al. Wang et al. (2004) evaluate change in vegetation between two dates using cokriging and cosimulation, whereas Kasetkasem and Varshney (1992) and Melgani and Serpico (2003) use a Markov random field to detect change. In most cases, the examples are limited to change detection between only two images.

The methodology proposed here integrates the spatial correlation of the land cover classes with temporal information, thus improving the mapping of land cover changes. The aim is to improve existing methods of land cover change detection by considering prior knowledge about the class spatial and temporal patterns and efficiently integrating them into the classification procedure. Furthermore, that framework has to be flexible enough to use most of the sophisticated algorithm being developed to classify remotely sensed data at a single time. Finally, the computational complexity of the algorithm should not increases too drastically when one increases the number of classes and/or the length of the time series.

The framework is applied to mapping land cover changes in the Pearl River Delta, China. This region is undergoing urban growth and other landscape changes.

2 Notations

Consider a domain $D \subset \mathcal{R}^2$ sampled at different times $t_i, i = 1, ..., N_t \subset T$. The N_t measurements of D constitute a set of images denoted $\mathcal{I} = \{\mathcal{I}^{(t_1)}, ..., \mathcal{I}^{(t_{N_t})}\}$. Let $(\mathbf{u}, t)$ be a point in $D \times T$ informed by a vector of length n_B of continuous attributes, $\mathbf{Z}(\mathbf{u}, t) = \{Z_1(\mathbf{u}, t), ..., Z_{n_B}(\mathbf{u}, t)\}$. These attributes are the satellite measurements known as digital numbers (DN).

Each pixel $(\mathbf{u}, t)$ must be classified into one of K classes $\mathcal{L}_1, ..., \mathcal{L}_K$, for example K land cover types. Define $I_k(\mathbf{u}, t)$ an indicator variable indicating whether or not the pixel at location $(\mathbf{u}, t)$ has class $\mathcal{L}_k$

$$I_k(\mathbf{u}, t) = \begin{cases} 1 & \text{if } (\mathbf{u}, t) \in \mathcal{L}_k \\ 0 & \text{otherwise} \end{cases}$$

And let

$$\mathcal{L}(\mathbf{u}, t) = k \quad \text{if } I_k(\mathbf{u}, t) = 1$$

Furthermore, let Ω be the set of location $\mathbf{u}_\alpha, \alpha = 1, ..., n$ whose classes are known at all times (ground truth). $V(\mathbf{u}, t)$ is the set of known pixel data in an isochronous neighborhood of $\mathbf{u}$ at time t.

3 Coding and combining information

Following Serpico and Melgani (2000) and Brown et al. (2002), three sources of information are considered relevant for the mapping of land cover changes. The first and foremost source is the satellite measurements. The second is the spatial pattern that relates land cover classes to each other. The third source of information is the temporal pattern of classes. Based on three data sources, the available information at each unsampled location $\mathbf{u}$ is separated between isochronous (cross-sectional) and time series information. The isochronous information includes the satellite response and the neighboring land cover indicators at any specific time. The time series information consists of transition probabilities linking the land cover indicators through time. The classification at location $\mathbf{u}$ is then done by combining these two types of information in such a way to minimize misclassification over a given training set.

3.1 Time series transition probabilities

Denote by $p_k^{\mathrm{T}}(\mathbf{u}, t)$ the probability of having class $\mathcal{L}_k$ at location $(\mathbf{u}, t)$ given the collocated land cover indicators in the past ($\mathcal{L}(\mathbf{u}, t - \Delta_1 t)$) and/or future ($\mathcal{L}(\mathbf{u}, t + \Delta_2 t)$):

$$p_k^{\mathrm{T}}(\mathbf{u}, t) = \Pr\{I_k(\mathbf{u}, t) = 1 \mid \mathcal{L}(\mathbf{u}, t - \Delta_1 t), \mathcal{L}(\mathbf{u}, t + \Delta_2 t)\ \} \tag{1}$$

The probability $p_k^{\mathrm{T}}(\mathbf{u},t)$ is calibrated directly from ground truth data or determined as function of the transition probabilities $p_{kk'}(t_i,t_j)$ relating the probability of having class $\mathcal{L}_{k'}$ at time t_j given that $\mathcal{L}_k$ is observed at time t_i.

$$p_{kk'}(t_i,t_j) = \Pr\{I_{k'}(\mathbf{u},t_j)=1 \mid I_k(\mathbf{u},t_i)=1\}, \forall\ \mathbf{u},k,k' \tag{2}$$

The transition probabilities $p_{kk'}(t_i,t_j)$ are calibrated from ground truth or historical data.

3.2 Isochronous probabilities

The isochronous information at any specific time is obtained by combining the satellite response and the spatial information available at that time. All information is expressed in terms of probabilities. Denote by $p_k^{\mathrm{iso}}(\mathbf{u},t)$ the isochronous probability obtained by combining the probabilities $p^{\mathrm{DN}}(\mathbf{u},t)$ and $p^{\mathrm{S}}(\mathbf{u},t)$ obtained from satellite and spatial information respectively.

$$\begin{aligned} p_k^{\mathrm{iso}}(\mathbf{u},t) = &\ \Pr\{I_k(\mathbf{u},t)=1 \mid \mathbf{Z}(\mathbf{u},t), \mathcal{L}(\mathbf{u}',t), \mathbf{u}' \in V(\mathbf{u},t)\} \\ = &\ \phi(p_k^{\mathrm{DN}}(\mathbf{u},t), p_k^{\mathrm{S}}(\mathbf{u},t)) \end{aligned} \tag{3}$$

with

$$p_k^{\mathrm{DN}}(\mathbf{u},t) = \Pr\{I_k(\mathbf{u},t)=1 \mid \mathbf{Z}(\mathbf{u},t)\},\ \forall\ k \tag{4}$$

and

$$p_k^{\mathrm{S}}(\mathbf{u},t) = \Pr\{I_k(\mathbf{u},t)=1 \mid \mathcal{L}(\mathbf{u}',t), \mathbf{u}' \in V(\mathbf{u},t)\},\ \forall\ k \tag{5}$$

where $V(\mathbf{u},t)$ is a set of location in the vicinity of $(\mathbf{u},t)$. The combination algorithm ϕ is presented later. The isochronous probability is calculated independently for each time.

Satellite-derived probabilities

The conditional probability $p_k^{\mathrm{DN}}(\mathbf{u},t)$ (4) for the pixel at location $(\mathbf{u},t)$ to be assigned to class $\mathcal{L}_k$ given the satellite response is computed with a classifier $F(\cdot)$ calibrated from the known data $\{\mathbf{Z}(\mathbf{u}_\alpha,t), \mathcal{L}(\mathbf{u}_\alpha,t)\}$ (Richards and Jia, 1999). The function $F(\cdot)$ approximates the conditional expectation of $I_k(\mathbf{u},t)$ given the sole collocated satellite response. Furthermore, the classifier $F(\cdot)$ for time t_i is calibrated only from the ground truth data available at time t_i.

In this study, the conversion of Landsat TM measurements into land cover types probabilities is done with the conventional Gaussian maximum likelihood (ML) classifier (Richards and Jia, 1999), a generative algorithm. The principle is simple, each class is associated with a multiGaussian RF modeling the satellite bands. The probabilities $\Pr\{I_k(\mathbf{u},t) = 1 \mid \mathbf{Z}(\mathbf{u},t)\}, k=1,..,K$ are calculated from the training set using a Bayes' inversion

$$p_k^{\mathrm{DN}}(\mathbf{u},t) = \Pr\{I_k(\mathbf{u},t)\ =1 \mid \mathbf{Z}(\mathbf{u},t)=\mathbf{z}\} = \frac{\Pr\{\mathbf{Z}(\mathbf{u},t)=\mathbf{z} | I_k(\mathbf{u},t)=1\}\Pr\{I_k(\mathbf{u},t)=1\}}{\sum_{k'=1}^{K}\Pr\{\mathbf{Z}(\mathbf{u},t)=\mathbf{z} | I_{k'}(\mathbf{u},t)=1\}\cdot\Pr\{I_{k'}(\mathbf{u},t)=1\}}$$

Assuming the random vector $\mathbf{Z}(\mathbf{u},t)$ to be multiGaussian, its conditional probability is written as

$$\Pr\{\mathbf{Z}(\mathbf{u})=\mathbf{z}|I_k(\mathbf{u},t)=1\}= \frac{1}{(2\pi)^{N/2}|\Sigma_i|^{1/2}}e^{-\frac{1}{2}(\mathbf{z}-\mathbf{m}_i)^T\Sigma_i^{-1}(\mathbf{z}-\mathbf{m}_i)} \tag{6}$$

where $\mathbf{m}_i$ and Σ_i are the mean vector and covariance matrix of the DN values belonging to the training data with class $\mathcal{L}_i$.

Spatially-derived probabilities

Denote by $p_k^{\mathrm{S}}(\mathbf{u},t)$ the conditional probability of observing $\mathcal{L}_k$ at location $(\mathbf{u},t)$ given the isochronous class data found in the neighborhood $V(\mathbf{u},t)$ 5).

This spatial probability $p_k^{\mathrm{S}}(\mathbf{u},t)$ may be estimated from simple indicator kriging (Goovaerts, 1997). Simple indicator kriging is a linear interpolator that applied kriging weights to indicator data yielding the probability of belonging to a class given the neighborhood data, the marginal and the covariance model of that class.

In addition to the ground truth data, the neighboring data in $V(\mathbf{u},t)$ also include locations that are considered well informed by the sole satellite measurements. A measure of information content is used to determine which locations are well informed and which ones are not. Those well-informed pixels are locations where the DN measurements $\mathbf{Z}$ alone are deemed sufficient to label them. For example, a pixel where the classifier $F(\cdot)$, see expression (4), would indicate a probability of 0.98 or more to belong to a certain class would qualify as a well-informed node.

Those pixels, assumed to be informed adequately by the satellite information such that no spatial information is needed, are used as anchor for the less informed ones. This spreads information from high-confidence pixels to their neighbors. For example, if all the well-informed locations in an area are urban, the neighboring pixels are more likely to belong to the urban class. The indicator kriging from the well-informed classes performs just that.

Taking the neighborhing pixels has indicative of the presence or absence of a class assumes that those classes are somewhat directly related to the unknown pixel. There is, however, a risk to overextend the spatial relevance of the well-informed locations. The problem lies in the discontinuity of the landscape. For example a certain region may be predominantly urban, without forest or agriculture, yet the vegetated area could start abruptly a few pixels away. A well-informed water class located in a lake close to the shore does not say whether that shore is urbanized or vegetated, instead it tends to artificially increase the probability that the shore would belong to a water class.

To offset this problem of borders and discontinuities, the images are first segmented to find edges delineating those discontinuities. Then a data neighborhood that does not cross the edges is retained for the indicator kriging process. Interpolation (kriging) is thus limited to homogeneous neighborhoods, a schematic representation of that adaptive neighborhood is shown in Figure 1.

3.3 Posterior probability

The posterior probability $p_k(\mathbf{u},t)$ for class $\mathcal{L}_k$ to occur at location $(\mathbf{u},t)$ is computed by combining the isochronous probability $p_k^{\text{iso}}(\mathbf{u},t)$ and the time series probability $p_k^{\text{T}}(\mathbf{u},t)$

$$\begin{aligned} p_k(\mathbf{u},t) &= \Pr\{I_k(\mathbf{u},t)=1 \mid \text{all data}\} \\ &= \psi(\, p_k^{\text{iso}}(\mathbf{u},t), p_k^{\text{T}}(loc,t)\,),\ \forall\ k \end{aligned}$$

The proposed combination algorithm ψ is developed in the next section.

Finally, the class $\mathcal{L}(\mathbf{u},t)$ is estimated by taking the most probable class of the posterior distribution:

$$\mathcal{L}^*(\mathbf{u},t) = \arg\max_k \{p_k(\mathbf{u},t),\ k=1,..,K\} \tag{7}$$

The time series, $\{\mathcal{L}(\mathbf{u},t_1),...,\mathcal{L}(\mathbf{u},t_{N_t})\}$ at location $\mathbf{u}$ is generated with a modified cascade approach. The estimation sequence is not chronological, the time series is produced by estimating the classes starting from the best informed time, as measure from the satellite information, and then sequentially estimating the time before and after that starting time. The idea is that the starting time is very consequential for the estimation of the whole time series, that starting time is thus chosen to reduce the prediction error. If the first pixel in the time series is misclassified, it is quite likely that this misclassification will be propagated to the rest of the time series. The less informed times at any given location benefit from being conditioned on the better informed collocated times.

3.4 Combining probabilities

Consider the isochronous probability vector $p^{\text{iso}}(\mathbf{u},t)$ defined in 3 and the time series conditional probability $p_k^{\text{T}}(\mathbf{u},t)$ defined in 1 as two sources of information. Each of the those two probabilities can be transformed into a distance related to the likelihood of event $\mathcal{L}(\mathbf{u},t)=k$ occurring (Journel, 2002). Let those distances be

$$\begin{aligned} x_{\mathcal{L}_k}^{\text{iso}}(\mathbf{u},t) &= \frac{1-p_k^{\text{iso}}(\mathbf{u},t)}{p_k^{\text{iso}}(\mathbf{u},t)} \ \in [0,\infty] \\ x_{\mathcal{L}_k}^{\text{T}}(\mathbf{u},t) &= \frac{1-p_{\mathcal{L}_k}^{\text{T}}(\mathbf{u},t)}{p_{\mathcal{L}_k}^{\text{T}}(\mathbf{u},t)\}} \ \in [0,\infty] \end{aligned}$$

Consider also the distance related to the marginal probabilities

$$x_{\mathcal{L}_k}^{(0)} = \frac{1-\Pr\{(\mathbf{u},t)\in\mathcal{L}_k\}}{\Pr\{(\mathbf{u},t)\in\mathcal{L}_k\}},\quad \forall\mathbf{u}$$

The updated distance to the event $\mathcal{L}(\mathbf{u},t)=k$ occurring accounting for both information (1) and (3) is given by the "tau model":

$$x_{\mathcal{L}_k}(\mathbf{u},t) = x_{\mathcal{L}_k}^{(0)} \cdot \left(\frac{x_{\mathcal{L}_k}^{\text{iso}}(\mathbf{u},t)}{x_{\mathcal{L}_k}^{(0)}}\right)^{\tau_{\text{iso}}} \cdot \left(\frac{x_{\mathcal{L}_k}^{\text{T}}(\mathbf{u},t)}{x_{\mathcal{L}_k}^{(0)}}\right)^{\tau_{\text{T}}} \tag{8}$$

where τ_{iso} and τ_{T} are parameters measuring redundancy between the two information sources (Journel, 2002; Krishnan et al., 2004). The posterior probability is then retrieved by inverting from the updates distance (8):

$$\Pr\{(\mathbf{u},t) \in \mathcal{L}_k | \mathbf{Z}(\mathbf{u},t)\} = \frac{\frac{1}{1+x_{\mathcal{L}_k}(\mathbf{u},t)}}{\frac{1}{1+x_{\mathcal{L}_1}(\mathbf{u},t)} + \frac{1}{1+x_{\mathcal{L}_2}(\mathbf{u},t)} + ... + \frac{1}{1+x_{\mathcal{L}_K}(\mathbf{u},t)}} \in [0,1] \tag{9}$$

The integration of $p^{\text{DN}}(\mathbf{u},t)$ and $p^{\text{S}}(\mathbf{u},t)$ into $p^{\text{iso}}(\mathbf{u},t)$ is also done with expression 9 but using different tau parameters τ_{S} and τ_{DN}.

The tau-model integration has convenient properties. If one of the prior probabilities is zero or one (no uncertainty), the combined probability is also zero or one. The combined probability is also always admissible, i.e. included between zero and one.

Although mathematically similar to the logarithmic opinion pool from consensual theory (Benediktsson and Swain, 1992) and to the aggregating formula of Bordley (1982), the tau-model is conceptually different. The τ exponents are not a measure of the reliability of the information source, but serves to model the redundancy between these sources. The reliability of the information is assumed to be already coded into each of the prior probability distribution related to each source. More details about the tau-model can be found in Journel (2002), Krishnan et al. (2004) and Benediktsson and Swain (1992).

The tau model is derived from the exact representation of the conditional distribution. Consider the conditional distribution of an event A conditional to a series of event $D_1, ..., D_n$. Also denote $\mathbf{D_i}$ as the set $\{D_1, ..., D_i\}$. The conditional probability can be rewritten as

$$\Pr\{A|D_1, ..., D_n\} = \frac{\Pr\{A|D_1\}\Pr\{D_1\}\prod_{i=2}^{n}\Pr\{\mathbf{D}_i|A,\mathbf{D}_{i-1}\}}{\Pr\{D_1, ..., D_n\}} \tag{10}$$

Expression (10) can also be written with $\tilde{A}$, the complement of event A.

$$\Pr\{\tilde{A}|D_1, ..., D_n\} = \frac{\Pr\{\tilde{A}|D_1\}\Pr\{D_1\}\prod_{i=2}^{n}\Pr\{\mathbf{D}_i|\tilde{A},\mathbf{D}_{i-1}\}}{\Pr\{D_1, ..., D_n\}} \tag{11}$$

Dividing (11) by (10) yield:

$$x = x^{(1)}\prod_{i=2}^{n}\frac{\Pr\{\mathbf{D}_i|\tilde{A},\mathbf{D}_{i-1}\}}{\Pr\{\mathbf{D}_i|A,\mathbf{D}_{i-1}\}}$$

The ratio above can always be written as

$$\frac{\Pr\{\mathbf{D}_i|\tilde{A},\mathbf{D}_{i-1}\}}{\Pr\{\mathbf{D}_i|A,\mathbf{D}_{i-1}\}} = \left(\frac{\Pr\{\mathbf{D}_i|\tilde{A}\}}{\Pr\{\mathbf{D}_i|A\}}\right)^{\tau_i}$$

Furthermore

$$\frac{\Pr\{\mathbf{D}_i|\tilde{A}\}}{\Pr\{\mathbf{D}_i|A\}} = \frac{\Pr\{\tilde{A}|\mathbf{D}_i\}}{\Pr\{A|\mathbf{D}_i\}}\cdot\frac{\tilde{A}}{A} = \frac{x^{(i)}}{x^{(0)}}$$

The full conditional probability can then be written as

$$\frac{x}{x^{x(0)}} = \frac{x^{(1)}}{x^{x(0)}} \prod_{i=2}^{n} \left(\frac{x^{(i)}}{x^{(0)}} \right)^{\tau_i}$$

where rigorously

$$\tau_i = \frac{\log \frac{\Pr\{\mathbf{D}_i|\tilde{A},\mathbf{D}_{i-1}\}}{\Pr\{\mathbf{D}_i|A,\mathbf{D}_{i-1}\}}}{\log \frac{\Pr\{\mathbf{D}_i|\tilde{A}\}}{\Pr\{\mathbf{D}_i|A\}}} \tag{12}$$

The tau exponents can then be interpreted as the ratio at which the previous data $\mathbf{D}_{i-1}$ change the information content of $\mathbf{D}_i$. It is then easy to verify that setting all tau parameters to 1 is equivalent to data conditional independence.

The challenge is in finding a suitable function for τ_i $i = 1, ..., n$ for a given decomposition sequence of the data D_i. In the general case, the τs cannot be determined analytically and a heuristic function has to be used, potentially trained on a training set. This preminilary study assume all τs equal to 1, corresponding to conditional independence between the sources.

4 A case study, urbanization in the Pearl River Delta, China

This study focuses on detecting and mapping changes between between 1988 and 1996 using a time series of Landsat TM images. We acquired 6 images of the Pearl River Delta, China, dating from 1988,1989, 1992, 1994, 1995 and 1996 all taken around December. The scene consists of 1 917 870 pixels approximatively covering an area of size 45km by 45 km, with each pixel of dimension 30x30 meters.

The landscape is divided into K=7 classes: water, forests, agriculture, urban, fish pond, transition (land getting cleared for urban settlement) and shrub. The ground truth measurements consists of 1917 locations identified by expert interpretation or by field reconnaissance. At ground truth locations the classes are deemed known at all times. The prediction errors, the expected errors between the estimated class and the true class at any location, are estimated by a 5-fold cross-validation procedure (Hastie and Friedman, 2001). The known classes are divided five times, each time into a training set and a testing set such that all samples are used once for testing purposes. Each split is done such that 80% of the ground truth data belong to the training set and the remainder 20% to the test set.

4.1 Computing the transition probabilities

The time series transition probabilities $p_{kk'}(t_i, t_j)$ defined in expression 2 are assumed stationary in time, such that

$$p_{kk'}(t_i, t_j) = p_{kk'}(\Delta t)$$

where $\Delta t = t_j - t_i$. The $p_{kk'}(\Delta t)$ are computed from the training set by evaluating the proportions of transitions from class k to class k'. Notable characteristics of this transition probability matrix is that the urban land cover type is an absorbing state, and the transition land cover type only communicates with either itself or the urban land cover type. This means that once a pixel is urban, it will remain urban; if a pixel has a transition class, it can only remain in transition or become urban.

4.2 Computing isochronous probabilities

Satellite-derived probabilities

The probabilities $p_k^{\mathrm{S}}(\mathbf{u}, t)$, $k = 1, ..., K$ is computed with a maximum likelihood estimator, see expression 6.

Spatially-derived probabilities

The spatial continuity for each land cover type is measured with indicator variograms Goovaerts (1997) defined as

$$\gamma_k(\mathbf{h}) \;=\; \mathrm{E}\{[I_k(\mathbf{u}, t) - I_k(\mathbf{u} + \mathbf{h}, t)]^2\}$$

The variogram $\gamma(\mathbf{h})$ is linked to the covariance by

$$C(\mathbf{h}) \;=\; C(0) \;-\; \gamma_k(\mathbf{h})$$

where $C(0$ is the variance. It can also be linked to the transition probability of being in state $\mathcal{L}_k$ at location $(\mathbf{u} + \mathbf{h})$ given location $\mathbf{u}$ has class $\mathcal{L}_k$ with

$$\Pr\{I_k(\mathbf{u} + h) = 1 | I_k(\mathbf{u} + h) = 1\} = \frac{p_k(1 - p_k) - \gamma_k(h)}{\Pr\{I_k(\mathbf{u}) = 1\}}$$

where $p_k = \mathrm{E}\{I_k(\mathbf{u}) = 1\}$ is the marginal probability for class $\mathcal{L}_k$.

The indicator kriging system uses that measure of spatial continuity to optimally assigns weight to the indicator data.

The simple indicator kriging system is

$$\mathbf{K}\lambda = \mathbf{k}$$

Where $\mathbf{K}$ is the data-to-data covariance matrix, λ an unknown weight vector and $\mathbf{k}$ the unknown-to-data covariance vector. The weight vector λ is then found as

$$\lambda = \mathbf{K}^{-1}\mathbf{k}$$

Finally the conditional probability of having class $\mathcal{L}_k$ at location $(\mathbf{u}, t)$ given the the neighboring data is

$$\begin{aligned} &\Pr\{(\mathbf{u}, t) \in \mathcal{L}_k | I_k(\mathbf{u}'), \mathbf{u}' \in V(\mathbf{u})\} \\ &\qquad = \; \lambda^T \mathbf{d} + (1 - \lambda^T \mathbf{1}) \mathrm{E}\{I_k(\mathbf{u}, t)\} \end{aligned} \tag{13}$$

where $\mathbf{d}$ is the known indicator data vector and $\mathbf{1}$ is a column vector of one. The full posterior probability density function is obtained by computing and normalizing (13) for $k = 1, ..., K$ such that they sum to one.

The previous kriging in combination with the tau-model (9) assures that no ground truth (hard data) locations are misclassified. At any informed location $\mathcal{L}(\mathbf{u}, t) = k$, the kriging estimate returns a probability of one to belong to $\mathcal{L}_k$ and zero for all others classes. The tau model than ensures that the final probability remains so.

The spatial context is thus accounted for through the probabilities $p^{\mathrm{S}}(\mathbf{u}, t)$ estimated with simple indicator kriging using for conditioning data the time series at those locations deemed well informed by the satellite measurements. The information content of a time series at location ($\mathbf{u}$) is measured as the sum of the maximum satellite-derived probability at each times.

$$\mathrm{Inf}(\mathbf{u}) = \frac{1}{N_t} \sum_{i=1}^{N_T} \max(p_k^{\mathrm{DN}}(\mathbf{u}, t_i), k = 1, ..., K) \tag{14}$$

To ensures homogeneous spatial neighborhood, those well informed time series are included in the kriging neighborhood only if a straight line going from the center of the neighborhood to any well informed datum does not cross an edge. The edges are found by performing a Canny segmentation method (Canny, 1986). Figure 2 shows two examples of edge detection. The edges in Figure 2(a) represent the shore of a bay with a dam at the SW extremity. In Figure 2(b), the edges separate a port from the ocean and also segment homogeneous regions inside the port complex.

4.3 Algorithms

The algorithm proposed proceeds as follows

- Perform a segmentation of the scenes
- Compute $p^{\mathrm{DN}}(\mathbf{u}, t), \forall\ \mathbf{u}, t$, see (4)
- Calculate the information content $\mathrm{Inf}(\mathbf{u})$ $\forall \mathbf{u}$ for all time series
- For each well-informed time series $\mathbf{u}_\beta, \beta = 1, ..., N_{\mathrm{Inf}}$
  - Estimate time series with p^{DN}, and p^{T} (see next algorithm)
- For each remaining uninformed time series
  - Estimate time series using all information (p^{S}, p^{DN}, and p^{T})

The algorithm to estimate a time series at location $\mathbf{u}$ can be described as

- Find the times t_i most informed by the satellite measurements
- Assign $\mathcal{L}\mathbf{u}, t_i = k$ such as $k = \arg\max_j p_j^{\mathrm{iso}}(\mathbf{u}, t_i)$
- Estimate sequentially future times $t_{i+1}, ..., t_{N_t}$
  - Assign classes $\mathcal{L}(\mathbf{u}, t_{i+s}), s = 1, ..., N_t - i$ based on $p^{\mathrm{iso}}(\mathbf{u}, t_{i+s})$ (3) and $p^{\mathrm{T}}(\mathbf{u}, t_{i+s})$
- For all past times $t_{i-1}, ..., t_1$
  - Assign classes $\mathcal{L}(\mathbf{u}, t_{i-s}), s = 1, ..., i + 1$ based on $p^{\mathrm{iso}}(\mathbf{u}, t_{i-s})$ and $p^{\mathrm{T}}(\mathbf{u}, t_{i-s})$

4.4 Results

The results of the proposed method applied to the Shenzhen scene are compared to the accuracy resulting from the maximum likelihood (ML) classifier, see expression 6. The ML classification is done by assigning to a time-space location $(\mathbf{u}, t)$, the class that has the maximum probability $p_k^{\mathrm{DN}}(\mathbf{u}, t)$. This ML classification considers only the satellite responses thus ignores the temporal and spatial correlation between classes. The changes are mapped by comparing the ML classification performed independently at each time.

The results are validated using (a) an overall accuracy criterion, the percent of correctly classified pixels, and (b) a time series accuracy, the percent of locations which have their vector of classes **all** correct. A time series at location $\mathbf{u}$ is well classified only if its six classes have been correctly predicted. For change detection purposes, the time series accuracy is critical as it measures how well the changes are mapped in both time and space. This allows for a correct identification of the time when the landscape has changed, and from which class to which other class. Such information is necessary to study the dynamics of the landscape.

With the ML classifier, the accuracy from a five-fold cross validation yields an overall accuracy of 78%, but the time series accuracy is only 33%. The proposed method marginally improves the overall accuracy from 78% to 82%. However, the time series accuracy goes up to 61%, a considerable improvement.

Tables 1 and 2 show the confusion matrices of the classification using ML and then using the proposed method which utilizes both spatial and temporal information. The producer and consumer accuracy for both cases are shown in Table 3. For all classes the consumer accuracy is greater or at least equal when space and time are incorporated in the classification. The improvement is especially noticeable for the forest class, the consumer accuracy jump from 44% to 91% as less forest were misclassifed into shrub. The producer accuracy, however, remains low with both methods for the forest class. The shrub ground truth data are almost all correctly identified (a producer accuracy of 0.93), a significant improvement from 0.83 when only the ML classifier is used. The fish ponds class also greatly benefits from the proposed method as its consumer accuracy goes from 0.66 to 0.81 and its producer accuracy improves 20 points from 0.72 to 0.92. The proposed method reduced the misclassification of fish pond into the water class from 56 misclassifications to only 6. Only the producer accuracy of water and forest decreases with the integration of spatial and temporal information.

The indicator kriging step decreases the level of speckling in the images, producing smoother maps. For example, the ML tend to classify many shadow zones in mountainous areas as water; the integration of spatial information corrects many of those misclassified pixels. There is no need to post-process the classified images to remove the speckles.

Kriging based on well-informed locations with an adaptive neighborhood has some advantages over commonly used post-processing filters such as the majority filter. With that filter any pixel can be changed according to the neighboring pixels; hence there is a possibility that a pixel well informed by the satellite measurement, thus with a low probability of misclassification, gets changed based on potentially more uncertain pixels in its neighborhood. Such scenario cannot happen with our proposed kriging as the spatial context is created from those well-informed pixels and propagated to the poorly informed ones, never

the other way.

The maps in Figure 3 show, for each location, the year at which change first occurred. A comparison between Figure 4(a) and (b) shows that the proposed method preserves some spatial relationships for the land cover changes, exhibiting a structured evolution of the landscape. On the contrary, the ML method produces a salt and pepper texture where the physical evolution of the landscape is difficult to discern.

The proposed method considerably reduced the number of false positive. With the ML prediction, 35% of locations had changed more than once, a number that expertise visual inspection of the images does not validate. With proposed method only 9% of locations are predicted to change more than once. The ML also predicts that 22% of locations did not change while that percent goes up to 64% when both spatial and temporal information are combined for prediction.

5 Conclusion

This paper proposes a framework that integrates the spatial and temporal autocorrelation of classes in remote sensing applications. That integration results in a more accurate change detection map that better identifies when, where and what to the landscape had changed. This study uses only 6 images, the extension to longer time series would be straightforward as the complexity of the algorithm only increases linearly with additional images.

The algorithm is flexible as it can handle any spatial, temporal or satellite classifier as long as it provides the probability for a pixel to belong to any class. For example, the ML classifier could be replaced without any change to the algorithm by a probabilistic neural networks or any other suitable algorithm. In the same way, a Gibbs-Markov RF or multiple points geostatistical algorithm (Strebelle, 2002) could replace the indicator kriging to provide the spatial context. Furthemore, if integrated with a GIS, other sources of information (such as distance to road, topography) could be integrated if that information can be coded into probabilities. Optimizing the tau exponents through some kind of training algorithm could further increase the classification accuracy.

Importantly, the study shows a considerable increase of the time series accuracy with the proposed method. The evolution of the landscape display greater spatial continuity and appears more realistic.

References

P. M. Atkinson and P. Lewis. Geostatistical classification for remote sensing: an introduction. *Computers & Geosciences*, 26(4):361 – 371, MAY 2000.

J. A. Benediktsson and P. H. Swain. Consensus theoretic classification methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(4):688 – 704, JUL-AUG 1992.

R.F. Bordley. A multiplicative formula for aggregating probability assessments. *Management Science*, 28(10):1137–1148, 1982.

D. G. Brown, P. Goovaerts, A. Burnicki, and M. Y. Li. Stochastic simulation of land-cover change using geostatistics and generalized additive models. *Photogramm. Eng. Remote Sens.*, 68(10):1051 – 1061, Oct 2002.

J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI/8(6):679 – 98, 1986.

P. Goovaerts. Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data. *Journal of Geographical Systems*, 4(1):99–111, 2002.

Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Oxford University press, New York, 1997.

Trevor J. Hastie and Jerome Friedman. *The Element of Statistical Learning*. Springer, 2001.

B. Jeon and D.A. Landgrebe. Classification with spatio-temporal interpixel dependency conetxts. *IEEE Transactions on Geoscience and Remote Sensing*, 1992.

A. G. Journel. Combining knowledge from diverse sources: An alternative to traditional data independence hypotheses. *Mathematical Geology*, 34(5):573 – 596, July 2002.

T. Kasetkasem and P.K. Varshney. An image change detection algorithm based on markov random field models. *IEEE Transactions on Geoscience and Remote Sensing*, 1992.

R. K. Kaufmann and K. C. Seto. Change detection, accuracy and bias in a sequential analysis of Landsat imagery in the Pearl River Delta, China: econometric techniques. *Agriculture, Ecosystems and Environment*, 85:95–105, 2001.

S. Krishnan, A. Boucher, and A.G. Journel. Evaluating information redundancy through the tau model. In *In Proceedings of Geostatistical Congress 2004, Banff, Canada*, 2004.

Stan Z. Li. *Markov random field modeling in image analysis*. Springer, 2001.

F. Melgani and S. B. Serpico. A mutual approach based on markov random fields for multitemporal contextual classification of remote sensing images. In *IGARSS 9-13 July, Sydney, NSW, Australia*, pages 2949 – 51 vol.7. Piscataway, NJ, USA : IEEE, 2001.

F. Melgani and S. B. Serpico. A markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11): 2478 – 87, November 2003.

J. A. Richards and X. Jia. *Remote sensing digital image analysis*. Springer-Verlag, Berlin, 3 edition, 1999.

S. B. Serpico and F. Melgani. A fuzzy spatio-temporal contextual classifier for remote sensing images. In *IGARSS. 24-28 July, Honolulu, HI, USA*, pages 2438 – 40 vol.6. Piscataway, NJ, USA : IEEE, 2000.

A. Stein, W. G. M. Bastiaanssen, S. DeBruin, A. P. Cracknell, P. J. Curran, A. G. Fabbri, B. G. H. Gorte, J. W. VanGroenigen, F. D. VanderMeer, and A. Saldana. Integrating spatial statistics and remote sensing. *Int. J. Remote Sensing*, 19(9):1793 – 1814, JUN 1998.

S. Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. *Mathematical Geology*, 34(1):1–21, 2002.

P. H. Swain. Bayesian classification in a time-varying environment. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC/8(12):879 – 83, Dec 1978.

B. Tso and P.M. Mather. *Classification methods for remotely sensed data*. Taylor and Francis, London, 2001.

G. Wang, G. Gertner, S. Fang, and A.B. Anderson. Mapping vegetation cover change using geostatistical methods and bitemporal landsat tm images. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3):632–643, 2004.

classes	wat.	for.	agr.	urb.	pond	tra.	shr.
wat.	1584	9	18	12	60	40	9
for.	27	185	12	0	0	0	287
agr.	35	23	988	39	0	24	503
urb.	4	2	30	877	5	292	55
pond	56	0	1	2	165	3	2
tra.	36	1	23	257	9	2535	77
shr.	13	204	215	44	12	57	2670

Table 1: Confusion table for maximum likelihood

classes	wat.	for.	agr.	urb.	pond	tra.	shr.
wat.	1503	5	24	81	35	40	44
for.	23	168	26	7	0	0	287
agr.	39	6	1059	15	0	49	444
urb.	0	3	23	938	0	265	36
pond	6	1	5	1	210	4	2
tra.	38	0	23	239	10	2569	59
shr.	6	2	116	34	3	60	2991

Table 2: Confusion table with space/time consideration

Maximum likelihood							
	wat.	for.	agr.	urb.	pond	tra.	shr.
C. Acc	0.9	0.44	0.77	0.71	0.66	0.86	0.74
P. Acc	0.91	0.36	0.61	0.69	0.72	0.86	0.83
Proposed method							
C. Acc	0.93	0.91	0.83	0.71	0.81	0.86	0.77
P. Acc	0.87	0.33	0.65	0.74	0.92	0.87	0.93

Table 3: Consumer and producer accuracy for both methods

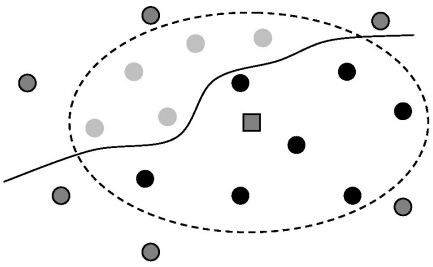
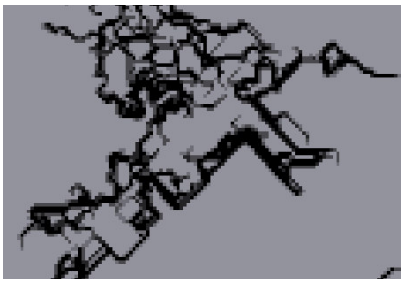


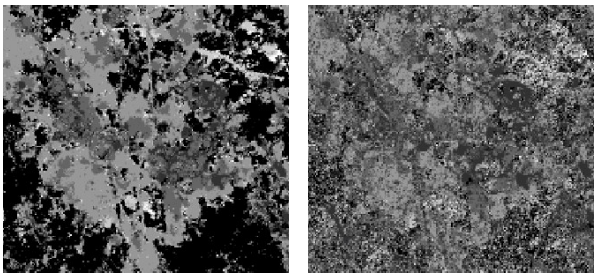
Figure 1: Example of an adaptive neighborhood (dashed ellipse) in presence of an edge (solid line). The square represents the unknown location to be mapped or estimated, the empty circle are data outside the search ellipsoid, hence not considered. The grey points are data inside the search ellipsoid but on the wrong side of the edge and are not taken into account. Only the black points are considered for kriging the square location.

(a) Bay and dam, 1988



(b) Edges between dock and water and internal division inside the dock complex

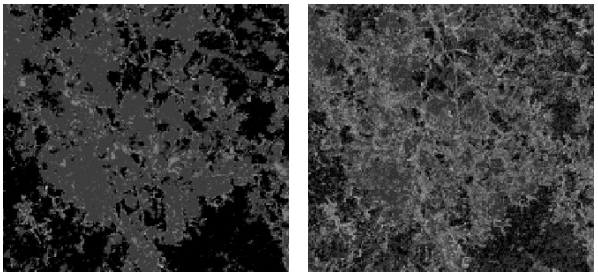
Figure 2: Examples of edge detection. In Figure (a), the edges capture the border of the bay and the dam at its extremity. In Figure (b), the edge define the contact between a port and the bay plus some internal divisions inside the port complex.



(a) Year of first changes for proposed method

(b) Year of first changes for ML

Figure 3: Map of predicted land cover changes representing the year at which the first change occurred. Figure (a) maps the year of change as predicted by the proposed method. Figure (b) does it for the ML method. Black indicates no changes, lighter tones indicates later times. Note the greater spatial resolution for the proposed method.



(a) Number of change for the proposed method

(b) Number of change for ML

Figure 4: Map of predicted number of land cover. Figure (a) maps the number of changes estimated for any pixels by the proposed method. Figure (b) does it for the ML method. Black indicates no changes, lighter tones indicates fewer times.