Change Detection from Temporal Sequences of Class Labels: Application to Land Cover Change Mapping

Varun Mithal* Ankush Khandelwal* Shyam Boriah* Karsten Steinhaeuser*
Vipin Kumar*

Abstract

Mapping land cover change is an important problem for the scientific community as well as policy makers. Traditionally, bi-temporal classification of satellite data is used to identify areas of land cover change. However, these classification products often have errors due to classifier inaccuracy or poor data, which poses significant issues when using them for land cover change detection. In this paper, we propose a generative model for land cover label sequences and use it to reassign a more accurate sequence of land cover labels to every pixel. Empirical evaluation on real and synthetic data suggests that the proposed approach is effective in capturing the characteristics of land cover classification and change processes, and produces significantly improved classification and change detection products.

1 Introduction

The physical surface of the earth is undergoing constant change: new cities are being built, existing cities are expanding to accommodate population increases, forests are being cleared for agricultural use, and lakes and other water bodies are changing in their extent [20]. The growth of urban areas, in particular, has received recent attention because the resulting large-scale changes have immediate impacts on a host of environmental factors [16]. Urban areas are now expanding twice as fast as their populations [16], and the United Nations estimated that more than half of the world's population lives in cities in its *2011 Revision of World Urbanization Prospects* report [19]. Thus, the ability to monitor the extent and rate of urbanization is critical due to its far-reaching consequences for the regional environment and beyond [1]. In particular, there is a strong need for accurate, timely, and regularly updated maps of global urban extent and dynamics.

Remote sensing has enabled the acquisition of multi-spectral imagery that can be used to study the earth surface. Data sets obtained via remote sensing are at a range of spatial and temporal resolutions (there is an inherent tradeoff between the two). Mapping urban extent and growth has traditionally been performed using moderate- to high-resolution multi-spectral data (250m to 30m). In most approaches in the literature [12], a classifier is built using manually selected training samples and is used to assign a land cover class from a finite set of classes to every pixel based on observed spectral values (the feature space). When a new image arrives for the latest time step, this classification process is repeated and every pixel is reassigned to a land cover class based on its current spectral values. Thus, every pixel is assigned a sequence of land cover class labels corresponding to the time period of image collection.

In this paper, we focus on the following two questions (in the region and time interval of interest):
Q1: Which pixels belong to an urban land cover class (or any other land cover class of interest)? (i.e. what is the urban *extent*?)
Q2: Which pixels changed from vegetation to urban land cover class (or any other land cover transition)? (i.e. where is land being *converted* to urban cover?)

The ability to answer the questions above depends on the accuracy of the underlying classification map, such as GL00 [3], IMPSA [5], GRUMP [4] or MOD500 [15]. The land cover research community has investigated many sophisticated approaches, including Random Forests [6] and Support Vector Machines [7]. However, satellite data has unique characteristics that make classification difficult: classes are often mixed, the feature space is unstable (due to variability in the sun angle, atmosphere, image registration, etc.), there is multimodality within classes (e.g. different kinds of trees), and there is a lack of high-quality training data. A recent study [12] showed that existing maps of global urban areas have disparities up to an order of magnitude, and discusses the lack of frequent map updates. Some classification maps (especially regional maps) are built with significant manual input from domain experts and are of high accuracy. However, even these detailed regional land cover class maps only have a classification accuracy between 85% to 95% [18].

To illustrate how classification accuracy will impact our ability to answer Q1 and Q2, let us consider a classifier built for a data set with two target classes and balanced training samples. Assume the classifier

*Dept. of Computer Science & Eng., University of Minnesota (mithal,ankush,sboriah,ksteinha,kumar)@cs.umn.edu

assigns a given test object to the incorrect class with probability ϵ and to the correct class with probability $1-\epsilon$. For the query Q1 (identify pixels of class *1* at any given time step t), the accuracy of the result set will be $1-\epsilon$. If ϵ is 0.1 (as is often the case for regional maps), the accuracy (also precision and recall) is 0.9. Therefore the result sets for Q1 are useful for the end-user looking for regions that belong to a particular land cover class.

Next, let us perform a similar analysis for query Q2 (identify pixels that changed from class *1* to class *2* between time t_1 and t_2), with the same classifier and error characteristics as the case above. Let us also assume the fraction of land surface that actually changed in this period is p. Due to classification error, ϵ fraction of pixels belonging to class *1* will be assigned class *2* in time t_2. Similarly, ϵ fraction of pixels belonging to class *2* will be assigned class *1* in time t_1. Thus, even when there is *no* land cover change between t_1 and t_2, ϵ fraction of class *1* pixels and ϵ fraction of class *2* pixels will be designated as changes from class *1* to class *2* and vice-versa. These incorrect labels will contribute to the false positives for a change detection query. Similarly, there will be $2\epsilon p$ false negatives. Ignoring the higher order terms in ϵ, the expected recall is $\frac{p-2\epsilon p}{p}$, and the expected precision is $\frac{p-2\epsilon p}{p+2\epsilon}$. However, changes in land cover typically occur in a very small portion of a large region of study; the area changed is often less than 1% of the total area ($p \approx 0.01$). Therefore, recall ≈ 0.8, and precision ≈ 0.05. Thus, even for high accuracy, state-of-the-art land cover classification products, the precision of change detection maps can be as poor as 5%. The analysis above shows that when land cover change mapping is done using post-classification comparison of images, even small amounts of classification inaccuracy can significantly lower precision.

In this paper, we present an innovative data mining approach to improve the class labels (associated with pixels) in land cover maps by using multi-temporal data. Our main intuition is to exploit the rich contextual information present in the temporal sequence of class labels which is not used by the classifier while generating class labels for images from different time steps. Our approach takes the original sequence of class labels as input; our task is to compute a new label sequence that is closer to the true land cover class of a pixel at the corresponding time step than the original classication sequence. Previous studies have used smoothing methods to improve classification using temporal context [17, 8] as well as spatial context [11]. In this study, we use a Hidden Markov Model (HMM) as the generative process for land cover label sequences and use the HMM for post-classification temporal smoothing to correct classification errors. The key contributions of this paper are as follows:

- We investigate the use of a latent variable model (HMM) for improving existing land cover classification products. Specifically, we formulate the problem as one of inferring the latent *true* land cover class sequence of a pixel from an input data set of inaccurate class label sequences.
- The ability of the proposed temporal smoothing to correct misclassifications depends on the fit of the HMM model to the observed and latent class label sequences. We introduce *confidence score* that measures the model fit for each pixel and identifies a set of pixels that are likely to have high classification accuracy after smoothing.
- We propose *change score* as a measure to assess the probability of a land cover change at a pixel between two given time steps. In contrast with bi-temporal comparison methods that give a boolean decision for change by looking at only two classified images, change score uses the label information from the entire sequence of the pixel to assign a probability of change.
- Our analysis with change detection results showed that many spurious changes tend to occur in pixels which have two or more land cover classes (eg. boundary pixels). We extend the temporal smoothing framework by introducing the concept of a *latent mixed class* and identify pixels that belong to it from sequences of only pure land cover class labels, i.e., with *no* training data for the mixed class. Our results show that modeling mixed class significantly improves change detection accuracy by correctly assigning pixels with multiple land cover classes to the mixed class.

Note to the reader: Most figures contained in this paper are best viewed in color.

2 Related Work

In Section 1 we discussed use of Hidden Markov Models (HMMs) for temporal smoothing of outliers in class label sequences. HMMs have also been used as models to infer the land cover state of pixels from continuous spectral time series data as input [14, 22]. This body of literature uses HMMs to leverage temporal context for land cover classification at each time step from multivariate time series data as input. The model requires specification of a class conditional data distribution and [14] uses a multivariate Gaussian distribution for the multivariate spectral data. In our problem setting we decouple the classification and temporal smoothing tasks. We

assume we are given a classifier that maps the multi-spectral data to a land cover class, and our objective is to use the temporal sequences of land cover class labels to correct any misclassification by incorporating temporal context. One advantage of decoupling these two tasks is that it allows usage of more powerful discriminative classifiers such as Decision Trees, SVM or Random forests that can learn complex decision boundaries in spectral data that generative models such as multivariate Gaussian distribution might fail to learn.

We model a latent, mixed land cover class and identify it from the temporal sequence of pure land cover class labels. Methods in remote sensing literature use spectral unmixing approaches [9] to find the fractional composition of mixed pixels in terms of pure classes. These approaches assume that the spectral data for mixed class pixels is a linear combination of spectral distributions of the pure classes. Our latent state modeling of a mixed pixel is different from these approaches as we do not assume that spectral distributions are additive in nature, but instead rely on the confusion between pure classes for a mixed pixel over multiple time steps. However, it is important to note that our approach only aims to identify mixed pixels and does not provide the exact proportion of pure land cover classes in the pixel.

3 Proposed Approach

In this section, we describe our approach for transforming an inaccurate class label sequence into a new, more accurate class label sequence. Furthermore, we describe two useful scores that can be derived from the proposed generative model for label sequences that will aid change detection queries by associating a notion of confidence with a given pixel's sequence of labels.

3.1 Definitions We begin by defining some terms and concepts related to the land cover mapping domain. A *pixel* is a fixed, regular-shaped spatial portion of an image. The entire image is divided into a mutually exclusive and exhaustive set of pixels. A land cover *class* or *label* comes from a finite set of classes used to categorize pixels on the earth surface (vegetation, water, etc.). A temporally ordered set of observed labels of a pixel is called a *label sequence* and represented as c^i. The sequence of "true" land cover state of a pixel is called its *state sequence* and represented as z^i.

The true land cover state may be different from the observed label due to noise in data or classifier inaccuracy. The difference between the observed label and the actual state of an object is referred to as *confusion*. The process of transformation of the material on the earth surface due to natural or human-induced actions such as wildfires and urbanization is known as land cover change. In this study, *land cover change* refers to a transition in the land cover state of a pixel occurring over a time period. Every object is assigned a land cover label based on its spectral values by the classifier (described in Section 4.2). We refer to the label sequence data of all the pixels in an image as the *original classification* (C^o). The proposed approach assigns every object a new land cover label. We refer to this new label sequence data of all the pixels in an image as the *new classification* (Z^1).

3.2 Observations The main intuition behind our methodology to modify class labels is based on the following observations in the land cover classification and change processes:
Observation 1: A pixel tends to remain in the same land cover states over time; therefore, change in land cover state is an infrequent phenomenon. This means that probability of transition to a different state is significantly smaller compared to self-transition.
Observation 2: The classification is more likely to be correct. This means that probability of misclassification is significantly smaller compared to correct classification.
Observation 3: The transition probability between some pairs of classes is higher than others. For example, vegetation to urban land cover change is more likely to happen than urban to vegetation.
Observation 4: The confusion probability between some pairs of classes is higher than others. For example, confusion between vegetation and urban is higher than confusion between vegetation and water.

Based on the above observations, the goal of our proposed method is to generate a new classification Z^1 using the original classification C^o as input, by modeling the observations about land cover classification and change processes. Our objective is that this method will assign correct labels to pixels that were misclassified in the original classification. In other words, Z^1 should have a higher classification accuracy than C^o.

3.3 Method Here, we describe the proposed generative model for the observed label sequence c^i of a pixel i. We assume that there exists a latent true land cover state z_t^i for each pixel i at time t. However, z_t^i is a latent variable and we can only observe c_t^i, i.e., the label assigned to it by the classifier based on the observed spectral signal. If the classifier was perfect and there were no data inaccuracies/incompleteness, then c_t^i will be sufficient to infer z_t^i. More precisely, we would set $z_t^i \equiv c_t^i$. However, classifiers make errors and satellite data has imperfections, and therefore c_t^i and z_t^i are sometimes different. Note that if we directly use the observed class label sequence for change detection, each confusion will be counted as a class transition, creating

several spurious land cover changes.

We model the stochastic process of class confusion in the classifier with a latent confusion matrix M, such that $m_{kl} = P(c_t^i = l | z_t^i = k)$, i.e., the observed land cover label is l when the actual land cover state is k. Based on Observation 2, we assume that m_{kl} is highest when $k = l$. Moreover, since we will always observe some label at each object, the following constraint is always satisfied: $\sum_l m_{kl} = 1$. Observation 3 discussed different transition probabilities for every pair of classes. We model this stochastic process as a transition matrix T, where each entry $t_{xy} = P(z_t^i = y | z_{t-1}^i = x)$, i.e., pixel in land cover state x transitions to state y. Based on Observation 1, t_{xy} is highest for $x = y$, since change in land cover class of a pixel is a rare event. Since each pixel in state x will transition to some state y, the following constraint is always satisfied: $\sum_y t_{xy} = 1$.

Thus, we see the above process is a doubly embedded stochastic process: an underlying stochastic process of transitions in true land cover state which is only observable through another stochastic process of class labels assigned by the classifier. Additionally, we make the following assumptions: (1) Land cover class change is a first order Markov process, i.e., z_t^i does not depend on $z_{t-2}^i ... z_1^i$ given z_{t-1}^i, and (2) the observed classification output c_t^i depends only on z_t^i. Motivated by the observations and assumptions above, we represent the generative model for land cover class sequences using a first-order Hidden Markov Model (HMM) with k states corresponding to the land cover states, s symbols corresponding to the class labels, a transition matrix T, and an emission matrix that corresponds to the latent confusion matrix M. Our next task is to infer the latent variables M, T, and Z^1 given the sequences in C^o, which we perform in two phases: parameter learning and inference.

3.3.1 Parameter Learning If the number of classes k is small, a domain expert may be able to provide good estimates for T and M. One can also provide an uninformative prior estimate for these matrices, with the final estimates being computed using an expectation-maximization algorithm that maximizes the posterior probability of state sequences [13]. Due to the issue of missing labels (e.g. because of missing spectral data), some of the c_t^i are assigned a missing class label. Therefore, we allow an additional observed label corresponding to missing data. In principle, a missing label has no information about the true state and observing a missing label is equally likely for every latent land cover state.

3.3.2 Inference Step We now discuss how to find the new class labels that are "most" likely given the observed label sequence c_i and the model parameters M and T. Note that there is no "true" sequence to be found and for a practical solution we look for a sequence that maximizes some objective function. Here, we consider the most likely state sequence as the one that maximizes the posterior probability $P(Z^1 | C^o, M, T)$. The solution of this formulation is discussed in detail in [13]. Other optimality criteria such as "most" probable individual states are less suited in our problem as they may assign states that form infeasible sequences due to the presence of unlikely transitions. Next, we use our probabilistic model for class label sequences to derive two useful scores that will aid change detection queries.

3.3.3 Confidence Score Our framework makes certain assumptions on the generative process of the land cover label sequences. If the observed sequences for pixels violate these assumptions, the possibility of classification errors in z^i is higher. Motivated by these issues, we define confidence score, a measure that indicates how well the HMM model fits the pair of an observed and hidden sequences. The score can also be used to partition the pixels into two subsets of high and low confidence scores. We are more confident that z_t^i represents the true state for the pixels with higher confidence compared to pixels with low confidence subset.

We propose to use the logarithm of the joint probability of observed and hidden sequence, i.e., $P(z^i, c^i | M, T)$ as a measure of our confidence on the reconstructed state sequence. *A high joint probability implies higher likelihood of z^i and c^i under our assumptions on M and T.* Therefore, pixels which have either many missing or undefined labels, or high confusion between two or more classes or a combination of these characteristics will be assigned low joint probability values. Under the Markovian assumptions in our model, it can be computed as

$$\text{Conf}(i) = \sum_{t=1}^{n} \log(m_{z_t^i, c_t^i}) + \sum_{t=0}^{n-1} \log(t_{z_t^i, z_{t+1}^i})$$

3.3.4 Change Score For any given classification product (C^o or Z^1), query Q2 finds the pixels for which the label is $c1$ at time step t_1 and the label is $c2$ at time step t_2. Often the cardinality of the query result set is large, while the end-user is only interested in seeing a few samples of the result. In such cases, it is more useful to provide a subset of results with a higher precision than a random subset from the result set.

We propose to use $P(z_{t1}^i = c1, z_{t2}^i = c2 | c^i, M, T)$ as the *change score* associated with every pixel of the query result. If $t_2 = t_1 + 1$, then this probability is equivalent to $\xi_{t1}(c1, c2)$, which is used in the Baum-Welch algorithm [13], and a dynamic programming-based solution

is known for computing it. However, for our purpose, we need to generalize $\xi_{t1}(c1, c2)$ to $\xi_{t1,t2}(c1, c2)$ which corresponds to $P(z^i_{t1} = c1, z^i_{t2} = c2|c^i, M, T)$. $\xi_{t1,t2}(c1, c2)$ can be computed as $\frac{P(c^i, z^i_{t1}=c1, z^i_{t2}=c2)}{P(c^i)}$. We derived a dynamic programming solution to compute the generalized $\xi_{t1,t2}(c1, c2)$ and have provided the algorithm for its computation in Algorithm 1 and evaluated its efficacy in S-4 in the supplement [10].

4 Data and Materials

Our proposed method takes as input a sequence of classified land cover maps. In this section, we describe the data and methods that we use in this study to generate the sequence of classified maps (i.e. the input to our method). We begin by describing the multi-spectral satellite imagery that serves as input for the classifier, and then describe the classifier and how it is trained.

4.1 Landsat Data We used satellite imagery from the Enhanced Thematic Mapper (ETM+) sensor on board the Landsat-7 satellite. Landsat-7 is the latest in a series of Landsat satellites, which have maintained a continuous record of the Earth's surface since 1972. ETM+ is a multi-spectral radiometric sensor that records eight spectral bands of data with varying spatial resolutions (30m spatial resolution for red, green, blue, near infrared, and two bands of medium infrared; 60m for thermal infrared; and a 15m panchromatic band). Landsat-7 data is available for public download from the U.S. Geological Survey [21]. For this study, we selected the city of Belo Horizonte, located in the state of Minas Gerais in southern Brazil. Belo Horizonte is the third largest city in Brazil and one of its fastest growing cities [2], thus providing a rich dataset for evaluation of the proposed algorithm. Images were collected bi-annually, corresponding to the dry season in that region which occurs in March-April and July-August, from the years 2003 through 2012. The extent of the region is $19.5°S$ to $20°S$ and $44.16°W$ to $43.75°W$, and it contains 2,868,495 pixels at 30m spatial resolution.

4.2 Base Classifier We implemented a classification module based on the methodology proposed by Taubenböck et al. [18]. This approach, recently developed at DLR (the German Aerospace Center), is a state-of-the-art method for urban mapping. The approach (henceforth called DLRclass) overcomes many of the challenges in classifying remote sensing imagery (e.g. multi-modality, lack of training data) using sophisticated innovations. The classifier in DLRclass is an ensemble of binary decision trees; each tree is trained for a different class using the labeled training samples. The binary trees are then ordered by land cover class using domain expertise. Finally, each Landsat image is independently classified into three categories: water ($\mathcal{W}$), vegetation ($\mathcal{V}$) and urban ($\mathcal{U}$). If the spectral data for a pixel is missing in an image or the pixel is not assigned a class by the ordered binary decision trees, then a missing label is assigned to that pixel for that image. Using DLRclass, Taubenböck et al. [18] were able to develop urban maps for 27 mega-cities globally with little training effort. (Though the authors also considered change maps in their paper, their focus was on general trends across decadal time scales, not pixel-level accuracy.)

4.3 Validation Imagery Land cover mapping research is often impeded by the lack of gold standard ground truth data, in our case pixel-level classification labels. To overcome this issue, in this study we take advantage of the high-resolution imagery available in Google Earth to interpret classified land cover labels, as other urban mapping studies have done [12]. The high-resolution imagery, which is orders of magnitude more detailed than Landsat, is tagged with the time of observation. For most regions of the globe, there is usually only a few (if any) high-quality cloud-free images available. To validate a given pixel's class label, we carefully examine the high-resolution imagery for agreement with the label; it is important to note that an additional criterion for agreement is that the time of observation of the validation imagery and classified image are similar. Henceforth, when we use the term *validation imagery*, we are referring to high-resolution imagery from Google Earth. Note that while imagery can be visualized in Google Earth, the underlying multi-spectral data is from commercial satellites and generally not freely available for analysis.

5 Evaluation and Discussion

5.1 Classification Accuracy In this experiment our goal is to show that (1) a sequence of land cover maps generated using DLRclass, while individually accurate ($\geq 90\%$), has errors which may lead to identification of spurious land cover class changes, (2) the temporal context is leveraged by our method to improve the class labels and thus avoid these spurious changes and (3) we are also able to assign a confidence score to every pixel and the classification accuracy of both C^o and Z^1 is higher for the subset of pixels with a high confidence score than the subset with low confidence score.

The data consists of two multi-spectral Landsat images one month apart (August and September 2008) for the city of Belo Horizonte. Each pixel of the two images is assigned a label from the target classes ($\mathcal{W}$, $\mathcal{V}$ or $\mathcal{U}$) by DLRclass based on their spectral attributes.

First, we consider the scheme of bi-temporal post-classification comparison to identify pixels that have

Classification data	Set of all pixels	Set of pixels with high confidence score
C^o	9.5%	7.12%
Z^1	1.89%	0.73%
Z^2	1.65%	0.95%

Table 1: Fraction of pixels with different land cover class in August and September, 2008.

different class labels in C^o. If these pixels (that were assigned different labels in August and September 2008) are treated as land cover changes, then Table 1 shows that 9.5% of the pixels changed their class in a period of one month. However, land cover change is a relatively rare event (across large spatial areas) and one expects the annual rate of land cover change to be under 1%. Thus, it is reasonable to consider these 9.5% pixels with land cover change in the short duration of a single month to be spurious due to classification errors.

In the absence of ground truth, for this experiment we rely on the percentage of changes in one-month period as a reasonable metric to compare the accuracy of any two land cover maps. Ideally, one expects this to be 0 (or close to 0) for a perfect classifier. For an imperfect classifier, we expect this percentage to *decrease* as the classification accuracy *increases*. (We previously discussed the connection between classification accuracy and spurious changes in Section 1.)

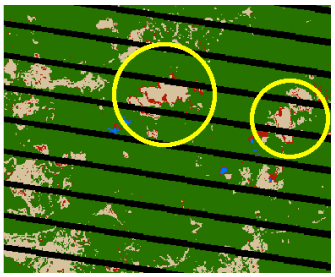
We used the proposed approach to obtain a new label sequence given a label sequence from DLRclass. Thus, the class label for a pixel for August, 2008 may be different from its label in the original classification. Next, we repeat this step for the same temporal sequence data with the class labels for August 2008 swapped with the class label from September 2008. September 2008 gets a new label which may also be different from its original label. If temporal context is playing a positive role in improving the class labels, then we expect that in case one of the labels (for August or September) was incorrect and is getting correctly reclassified in Z^1 the number of mismatches will reduce. Table 1 shows that the number of transitions in one month period is reduced from 9.5% to 1.89%. We observed similar results when we repeated this experiment using data from different years. Next, we partitioned the set of all pixels into high and low confidence score subsets and for 80% of the pixels which are in the high confidence subset our model improved classification accuracy and reduced the spurious transitions from 7.12% to 0.73%.

5.2 Correcting & Imputing Labels Due to Poor Data Remote sensing data is often plagued with noise (due to atmospheric interference such as clouds and aerosols) and missing data due to instrument malfunction. Therefore, a given classified image C^o often has labels that are both inaccurate (due to noise) and incomplete (due to missing data). To illustrate both these issues, Figure 1a shows a Landsat image in which we can see the presence of clouds (highlighted with the yellow circle) and missing data (shown as black stripes) on the earth surface. Figure 1b shows the classified map (C^o) corresponding to Figure 1a, where clouds have been misclassified as urban and Figure 1c shows the classified map (Z^1), where our method has reassigned the cloudy region to the $\mathcal{V}$ class and imputed missing labels using its temporal context.

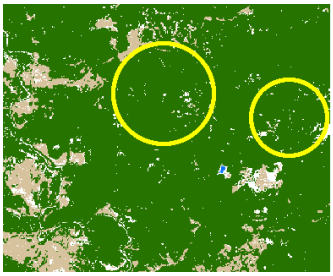
5.3 Change Detection Given two classification products, C^o and Z^1, one can create a change map for any pair of time steps (t_1, t_2). In this section, we will show that the change map prepared using Z^1 is better than one prepared from C^o. In particular, there is an improvement in both the number of false negatives and false positives.

Figures 2a and 2b show the validation imagery of a region in the city of Belo Horizonte that has instances of urbanization between the years 2004 and 2011. Figure 2a corresponds to the image from April 20, 2003 and Figure 2b corresponds to the image from August 22, 2011 for the same region. Figure 2d shows the change map produced using the bi-temporal post-classification comparison method for the images from years 2004 and 2011. This map has nine unique categories of transitions corresponding to the 3 classes ($\mathcal{V}$: vegetation, $\mathcal{U}$: urban and $\mathcal{W}$: water), but has only four

(a) Landsat image in Belo Horizonte.

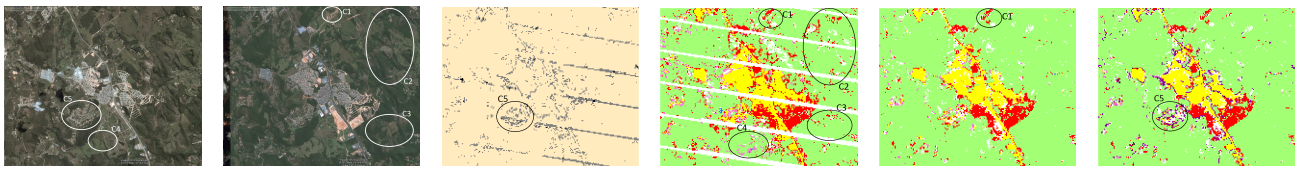


(b) Classified map generated by DLRclass.



(c) Classified map Z^1.

Figure 1: These figures show the issues of noise and missing data, and how our proposed method is able to correct the labels caused by these issues.



(a) Validation image for August 2003. **(b)** Validation image for August 2011. **(c)** Confidence score map. **(d)** Change map using original classification C^o. **(e)** Change map using multi-temporal context-based classification Z^1. **(f)** Change map using multi-temporal context-based classification Z^2.

Figure 2: These figures show the change detection for a region of study between 2003 and 2011. (Note that a larger version appears as Figure 4 in Section S-2 in supplement [10] due to space constraints.)

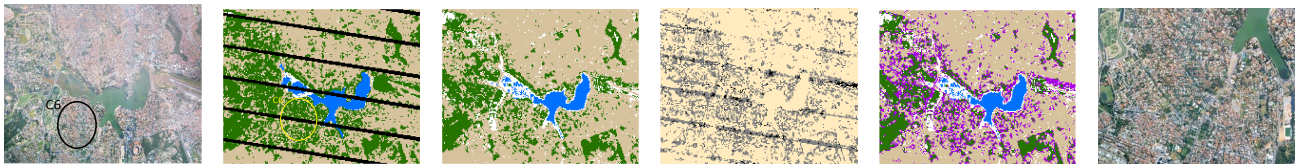
dominant categories: $\mathcal{V} \rightarrow \mathcal{V}$ (in green), $\mathcal{U} \rightarrow \mathcal{U}$ (in yellow), $\mathcal{V} \rightarrow \mathcal{U}$ (in red) and $\mathcal{U} \rightarrow \mathcal{V}$ (in pink). Here, we make some observations upon studying Figure 2d in the context of Figures 2a and 2b. DLRclass is usually accurate and we can see that the vegetated areas and urban areas (as they appear from the validation imagery) are typically assigned the $\mathcal{V} \rightarrow \mathcal{V}$ or $\mathcal{U} \rightarrow \mathcal{U}$ labels. We also see that this region is dominated by unchanged ($\mathcal{V} \rightarrow \mathcal{V}$ and $\mathcal{U} \rightarrow \mathcal{U}$) pixels. The missing data issue is seen as white stripes on the change map because change cannot be determined for a pixel if either of the two images have a missing label. Among the different change categories, $\mathcal{V} \rightarrow \mathcal{U}$ and $\mathcal{U} \rightarrow \mathcal{V}$ are the dominant change types, and between them, $\mathcal{V} \rightarrow \mathcal{U}$ change is significantly higher in number than $\mathcal{U} \rightarrow \mathcal{V}$. This is expected as typically vegetated areas are converted to urban land and not vice-versa.

This region has two large clusters (of several hundred pixels) that are marked as conversion from $\mathcal{V} \rightarrow \mathcal{U}$. The validation imagery for the corresponding region also shows that the land surface was vegetated in 2003 and was barren by 2011. In addition to these large changes, we see some moderate sized clusters consisting of 20-100 pixels (such as the example in C1) and of salt-and-pepper distribution of changes (such as the examples in C2 and C3). The other dominant category of change is $\mathcal{U} \rightarrow \mathcal{V}$ (marked as pink pixels). There are a few small clusters of 5-10 pixels that are marked as $\mathcal{U} \rightarrow \mathcal{V}$ (such as the examples in C4). Finally, we see that several pixels on the boundary of two land cover types have been marked as changed, either from $\mathcal{U} \rightarrow \mathcal{V}$ or vice-versa.

Next, we focus on Figure 2e which shows the change map produced by comparing the class labels of Z^1. Our first observation is that the number of changes (both $\mathcal{V} \rightarrow \mathcal{U}$ and $\mathcal{U} \rightarrow \mathcal{V}$) is smaller for the new change map compared to Figure 2d. The $\mathcal{V} \rightarrow \mathcal{U}$ changes reduce from 2880 to 1935 and the $\mathcal{U} \rightarrow \mathcal{V}$ reduce from 846 to 360. The relative reduction in $\mathcal{U} \rightarrow \mathcal{V}$ is higher because most of the changes in that category were spurious and these spurious changes are expected to decrease with a more accurate classification. To verify that the reduction in number of changes actually corresponds to a reduction of false positives and not an increase in false negatives, we examined validation imagery for the areas in circles C2 and C3 that are marked as changed from $\mathcal{V} \rightarrow \mathcal{U}$ in C^0 and are not changed in Z^1. In our analysis, we found that the validation imagery corroborates better with the new map from Z^1 and the pixels that were previously labeled as change (in C^0) were spurious transitions in most cases. Further evidence supporting this claim is the fact that most of the changes in large or moderate sized clusters such as C1 persist in the new change map, while most of the changes inside C2 and C3 are removed. C1 has a change from $\mathcal{V} \rightarrow \mathcal{U}$ which is also visible in the validation imagery, while C2 and C3 have a salt-and-pepper distribution of $\mathcal{V} \rightarrow \mathcal{U}$ changes with no clear evidence from the validation imagery.

5.4 Mixed Pixel Modeling Remote sensing data is recorded for fixed-size spatial units (pixels) and therefore pixels that contain a mixture of land cover types are inherently present in these data sets. This occurs even with Landsat, which is one of the highest resolution publicly available data sources (Landsat's 30m pixel size roughly translates to an area of 0.22 acres on the ground). Naturally, the occurrence of mixed pixels increases significantly for coarser scale data sets such as MODIS (250m) and SPOT (1km). However, most land cover classifiers are trained with pure classes because mixed pixels have tremendous heterogeneity and it is infeasible to generate a representative set of training samples that adequately captures all kinds of mixed pixels. Therefore, the ability to tag mixed pixels (whether they appear on boundaries or in clusters) as belonging to a mixed *class* using output from classifiers that were trained for only pure classes represents a significant advance in land cover mapping.

Our model for land cover class label sequences assumes that the pixels are either $\mathcal{W}$, $\mathcal{V}$ or $\mathcal{U}$. For example, Figure 3a shows the validation imagery for a spatial region around an urban area. Running DLRclass trained on the set of target classes ($\mathcal{W}$, $\mathcal{V}$ and $\mathcal{U}$), a



(a) Validation image for August 2008. **(b)** Classification map using DLRclass for August 2008. **(c)** Classification map from Z^1 for August 2008. **(d)** Confidence score map, black implies low score. **(e)** Classification map Z^2 after including mixed class in model for August 2008. **(f)** Zoomed-in image for the area in C6 in August 2008.

Figure 3: These figures show the classification maps for a region of study in Belo Horizonte between 2003 and 2011. (Note that a larger version appears as Figure 5 in Section S-3 in [10] due to space constraints.)

classification map for this region is shown in Figure 3b. DLRclass is typically able to identify the water, urban and vegetation classes, but we see that it has misclassified several pixels as vegetation in an urban area inside circle C6 (as it appears from validation imagery). As was previously discussed in Section 3, confusion between classes is one of the reasons for this; Figure 3c which shows the new classification map has significantly reduced the number of pixels classified as vegetation in that region. However, we see some of the pixels inside C6 are still misclassified.

To investigate this issue, we computed the confidence score for each pixel of this region. Recalling the discussion from Section 3.3.3, this score measures the conformity of a given sequence to the model parameters and assumptions. Figure 3d shows a map of confidence scores for the pixels in the region. This map clearly shows that C6 and some other subregions in the region of study have very low probability compared to the rest of the image. We examined these low probability regions and found that these pixels have a significant (and almost equal) proportion of both vegetation and urban labels in their sequences. The validation imagery for these regions also reveals that these pixels are neither completely vegetation nor urban, but rather a mixture of the two classes. If our model attempts to assign these pixels to either the $\mathcal{V}$ or $\mathcal{U}$ class, the confidence score would be very low, as they have much higher confusion than expected by the HMM model.

Next, we extend our approach to include the concept of a mixed class ($\mathcal{M}$), which is a latent class in our model that is equally biased towards observing the $\mathcal{V}$ or $\mathcal{U}$ class labels in its sequence. Figure 3e shows the classification output (Z^2) after including $\mathcal{M}$. We see that several pixels in the low probability areas are now assigned to $\mathcal{M}$. This novel, mixed class replaces most of the inaccurate vegetation class labels, while the actual vegetated areas are classified as vegetation. Moreover, we see that pixels at the boundaries of two land cover classes are also assigned to $\mathcal{M}$. This is not surprising as boundary pixels are typically expected to have a proportion of both classes present in them and also confirms that the latent mixed class in our approach actually captures the pixels that physically consist of a mixture of two classes on the ground. The above experiment demonstrates the capability of the extended model to identify novel, latent classes from sequences of pure class labels by using the temporal context.

We also found that many spurious changes in classification maps tend to occur with mixed pixels, and modeling a latent mixed class avoids identification of these spurious changes to a large extent. We return to the change detection experiment (Section 5.3) to illustrate that mixed class modeling can actually reduce some of the spurious changes occurring at the boundaries of land cover classes and in regions of mixed pixels. Figure 2f shows the change map produced after re-classification with mixed class modeling Z^2. The $\mathcal{M} \rightarrow \mathcal{M}$ transition is prominently visible (in deep purple color) along the boundaries of the $\mathcal{V}$-$\mathcal{U}$ regions. Figure 2c shows the confidence score for the pixels and we can see that the boundaries between land cover types are assigned a low confidence score. Moreover, the region inside C6 has a dense concentration of low probability pixels. Validation imagery in Figure 2b confirms that these pixels are similar to the mixed pixels seen in Figure 3a. The new change map from Z^2 correctly assigns these pixels to a $\mathcal{M} \rightarrow \mathcal{M}$ transition, most of which were earlier inaccurately identified as changes in Figure 2e.

5.5 Synthetic Data We designed a set of experiments with synthetic data in order to quantitatively evaluate the proposed framework. Details of the experimental setup and discussion of results is presented in S-4 in supplement [10]. In particular, we evaluated classification performance in terms of the number of confusions, imputation of missing data, and change detection precision and recall and found the results on synthetic data to be consistent with our case studies on the real data. We also evaluated the proposed *change score* and found that is desirable to use it for ranking pixels when

the end-user wants to analyze a high precision subset of changes identified.

6 Concluding Remarks

In this paper, we applied data mining methods to advance the state-of-art in land cover change mapping by improving the existing classification products. In particular, we proposed an HMM-based generative model for land cover class label sequences and used it to infer new, more accurate land cover state sequences. Case studies on real data demonstrate that the proposed generative model is able to leverage the temporal context of class labels to improve classification. Furthermore, we used the probabilistic model to compute useful statistics, namely the confidence and change scores, which can be leveraged while analyzing change detection queries.

The goal of a semi-supervised global-scale land cover change detection system involves many challenges and this work is a step towards its realization. We explored the use of temporal context for improving land cover classification using an HMM-based model. HMM restricts the duration in same state to a geometric distribution. Future work will explore use of models such as Hidden Semi-Markov model that allow the explicit modeling of duration in the state before next transition. Another direction for research is to develop models that also use the spatial context of labels to correct classification errors. Moreover, the proposed approach was used for analyzing small regions of the size of a single city. Learning of transition and confusion probabilities is impacted by the size of selected region and future work will investigate the sensitivity to this parameter. Finally, for some regions multiple classification products are available, each with its own strengths and weaknesses. The model can be extended to integrate these multiple sources of class information in a principled framework to achieve better land cover classification than what can be achieved using a single source.

Acknowledgements

This research was supported in part by the National Science Foundation under Grants IIS-1029711 and IIS-0905581, as well as the Planetary Skin Institute. Access to computing facilities was provided by the University of Minnesota Supercomputing Institute.

References

[1] M. Alberti. The Effects of Urban Patterns on Ecosystem Function. *IRSR*, 2005.

[2] J. Barros. *Urban Growth in Latin American Cities*. PhD thesis, University College London, 2004.

[3] E. Bartholomé and A. Belward. Glc2000: a new approach to global land cover mapping from earth observation data. *IJRS*, 2005.

[4] CIESIN. Global rural-urban mapping project (GRUMP). *CIESIN, Columbia University*, 2004.

[5] C. Elvidge et al. Global distribution and density of constructed impervious surfaces. *Sensors*, 2007.

[6] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson. Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294–300, 2006.

[7] C. Huang, L. S. Davis, and J. R. G. Townshend. An assessment of support vector machines for land cover classification. *IJRS*, 23(4):725–749, 2002.

[8] E. Jaser et al. Temporal post-processing of decision tree outputs for sports video categorisation. *SSSPR*, 2004.

[9] N. Keshava and J. Mustard. Spectral unmixing. *Signal Processing Magazine, IEEE*, 19(1):44–57, 2002.

[10] V. Mithal, A. Khandelwal, S. Boriah, K. Steinhaeuser, and V. Kumar. Supplement for change detection from temporal sequences of class labels: Application to land cover change mapping. Technical Report UMN-CSE-13-003, Department of Computer Science, University of Minnesota, Minneapolis, Minnesota, January 2013.

[11] F. Nava, A. Nava, J. Lamolda, and M. Redondo. Change detection in remote sensing images with graph cuts. In *Proceedings of SPIE,*, pages 59820Q–1. Society of Photo-Optical Instrumentation Engineers, 2005.

[12] D. Potere et al. Mapping urban areas on a global scale: which of the eight maps now available is more accurate? *IJRS*, 2009.

[13] L. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Readings in speech recognition*, 1989.

[14] A. Salberg and O. Trier. Temporal analysis of forest cover using hidden markov models. In *Proceeding of IGARSS*, pages 2322–2325. IEEE, 2011.

[15] A. Schneider et al. A new map of global urban extent from MODIS satellite data. *Environmental Research Letters*, 2009.

[16] K. C. Seto et al. Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *PNAS*, 2012.

[17] J. Suutala et al. Discriminative temporal smoothing for activity recognition from wearable sensors. *Ubiquitous Computing Systems*, 2007.

[18] H. Taubenböck et al. Monitoring urbanization in mega cities from space. *RSE*, 2012.

[19] U.N. Department of Economic and Social Affairs. World urbanization prospects. `http://esa.un.org/unpd/wup/`.

[20] United Nations Environment Programme (UNEP). *Keeping Track of Our Changing Environment: From Rio to Rio+20 (1992-2012)*. 2011.

[21] U.S. Geological Survey. Landsat missions. `http://landsat.usgs.gov/`.

[22] N. Viovy and G. Saint. Hidden markov models applied to vegetation dynamics analysis using satellite remote sensing. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4):906–917, 1994.