2014

Inter-annual stability of land cover classification: explorations and improvements

Abercrombie, Stewart Parker

http://hdl.handle.net/2144/14299

Boston University
BOSTON UNIVERSITY
GRADUATE SCHOOL OF ARTS AND SCIENCES

Thesis

INTER-ANNUAL STABILITY OF LAND COVER CLASSIFICATION: EXPLORATIONS AND IMPROVEMENTS

by

STEWART PARKER ABERCROMBIE
B.S., University of California, Santa Barbara, 2003

Submitted in partial fulfillment of the requirements for the degree of Master of Arts
2014
Approved by

First Reader

______________________________
Mark A. Friedl, PhD
Professor of Earth and Environment

Second Reader

______________________________
Curtis Woodcock, PhD
Professor of Earth and Environment

Third Reader

______________________________
Luis Carvalho, PhD
Assistant Professor of Mathematics and Statistics
Acknowledgments

This work would not have been possible without the help of many people. I would like to thank my advisor, Mark Friedl, for his support and encouragement throughout this project. Thanks to Curtis Woodcock, Luis Carvalho, and Robert Kennedy for their invaluable advice on technical issues in land cover classification and statistical methods. Thanks to my colleagues and fellow graduate students at Boston University for their friendship and encouragement, especially Josh Gray, Eli Melaas, Mary Farina, Jon Wang, Dan Gionatti, Jian Bi, Chris Holden, and Hunter Glanz. I would like to thank Damien Sulla-Menashe for helping me get oriented to the MODIS Land Cover product, and Xioman Huang for providing the Xingu deforestation data set. Thanks to Michael Toomey for first suggesting that I consider Boston University for graduate school, and for hosting me during my first week in Boston. Finally I would like to thank Rachel Weber for her love and support.
Land cover information is a key input to many earth system models, and thus accurate and consistent land cover maps are critically important to global change science. However, existing global land cover products show unrealistically high levels of year-to-year change. This thesis explores methods to improve accuracies for global land cover classifications, with a focus on reducing spurious year-to-year variation in results derived from MODIS data. In the first part of this thesis I use clustering to identify spectrally distinct sub-groupings within defined land cover classes, and assess the spectral separability of the resulting sub-classes. Many of the sub-classes are difficult to separate due to a high degree of overlap in spectral space.

In the second part of this thesis, I examine two methods to reduce year-to-year variation in classification labels. First, I evaluate a technique to construct training data for a per-pixel supervised classification algorithm by combining multiple years of spectral measurements. The resulting classifier achieves higher accuracy and lower levels of year-to-year change than a reference classifier trained using a single year of data. Second, I use a spatio-temporal Markov Random Field (MRF) model to post-process the predictions of a per-pixel classifier. The MRF framework reduces spurious label change to a level comparable to that achieved by a post-hoc heuristic stabilization technique. The timing of label change in the MRF processed maps better matched disturbance events in a reference data, whereas the heuristic stabilization results in label changes that lag several years behind disturbance events.
Contents

1 Introduction ................................................. 1

2 Background .................................................. 5
   2.1 Global Land Cover Classification ......................... 5
   2.2 Stability of Land Cover Classification ..................... 7

3 Methods ...................................................... 9
   3.1 Data .................................................. 9
      3.1.1 MODIS Land Cover Type ............................. 9
      3.1.2 IGBP Classification Scheme ......................... 10
      3.1.3 Classification Features ............................... 14
      3.1.4 Parameter Estimation Data ......................... 14
      3.1.5 Validation data .................................. 16
   3.2 Algorithms ............................................ 16
      3.2.1 Decision Trees ................................... 16
      3.2.2 Markov Random Fields ............................ 19
   3.3 Analysis .............................................. 32
      3.3.1 Classification Accuracy ........................... 32
      3.3.2 Land Cover Sub-classes .......................... 34
      3.3.3 Multi-Year Training Data ......................... 38
      3.3.4 Markov Random Field Post-processing ............... 39

4 Results ..................................................... 43
   4.1 Land Cover Sub-classes ................................ 43
## List of Tables

3.1 IGBP land cover class definitions. (Source: Friedl et al., 2002) . . . . 11

4.1 Landscape homogeneity of conterminous United States . . . . . . . . 58
4.2 Classification accuracy of STEP training pixels . . . . . . . . . . . . . 61
4.3 Confusion matrix (Markov Random Field) . . . . . . . . . . . . . . . 62
4.4 Confusion matrix (Decision tree) . . . . . . . . . . . . . . . . . . . . 62
4.5 Confusion matrix (MLCT Collection 5) . . . . . . . . . . . . . . . . . 63
4.6 Error rate by time window . . . . . . . . . . . . . . . . . . . . . . . . 63
4.7 Commission errors in deforestation sites . . . . . . . . . . . . . . . . 67
# List of Figures

3-1 Land cover class distribution in STEP database ........................................... 12
3-2 Location of training sites ................................................................................. 13
3-3 Logic to aggregate 30-meter NLCD pixels to 500-meter IGBP pixels ............. 16
3-4 National Land Cover Database 2006 land cover map ..................................... 17
3-5 Location of deforestation reference sites .......................................................... 18
3-6 Graph structure of a grid of pixels ................................................................. 22
3-7 Temporal neighborhood system .................................................................... 29

4-1 Label stability vs. classification accuracy ....................................................... 45
4-2 Sub-classes of open shrubland land cover class ............................................ 46
4-3 Sub-classes of grassland land cover class ...................................................... 47
4-4 Sub-class confusion matrix ............................................................................. 49
4-5 Principal components of low-latitude shrubland and grassland sites ............ 50
4-6 Percent year-to-year label change using multi-year training data ................. 52
4-7 Accuracy of multi-year training set vs. single year training set ..................... 53
4-8 Random field energy as a function of ICM iteration ........................................ 55
4-9 Landscape homogeneity for contiguous United States ..................................... 57
4-10 Landscape homogeneity of MODIS tile h8v5 ................................................. 57
4-11 Effect of spatial smoothing on MODIS tile h8v5 ........................................... 60
4-12 Example classification of a deforested pixel .................................................. 65
4-13 Error rate for deforestation sites ................................................................... 66

A-1 Sub-classes of evergreen needleleaf forest class .............................................. 78
A-2 Sub-classes of evergreen broadleaf forest class

A-3 Sub-classes of deciduous needleleaf forest class

A-4 Sub-classes of deciduous broadleaf forest class

A-5 Sub-classes of mixed forest class

A-6 Sub-classes of closed shrubland class

A-7 Sub-classes of open shrubland class

A-8 Sub-classes of woody savanna class

A-9 Sub-classes of savanna class

A-10 Sub-classes of grassland class

A-11 Sub-classes of permanent wetland class

A-12 Sub-classes of cropland class

A-13 Sub-classes of cropland/natural vegetation mosaic class

A-14 Sub-classes of snow/ice class

A-15 Sub-classes of barren/sparsely vegetated class

A-16 Sub-classes of water class
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVHRR</td>
<td>Advanced Very High Resolution Radiometer</td>
</tr>
<tr>
<td>C5</td>
<td>Collection 5</td>
</tr>
<tr>
<td>EVI</td>
<td>Enhanced Vegetation Index</td>
</tr>
<tr>
<td>ICM</td>
<td>Iterated Conditional Modes</td>
</tr>
<tr>
<td>IGBP</td>
<td>International Geosphere-Biosphere Programme</td>
</tr>
<tr>
<td>LST</td>
<td>Land Surface Temperature</td>
</tr>
<tr>
<td>MERIS</td>
<td>Medium Resolution Imaging Spectrometer</td>
</tr>
<tr>
<td>MLCT</td>
<td>MODIS Land Cover Type</td>
</tr>
<tr>
<td>MODIS</td>
<td>Moderate Resolution Imaging Spectroradiometer</td>
</tr>
<tr>
<td>MRF</td>
<td>Markov Random Field</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NDVI</td>
<td>Normalized Difference Vegetation Index</td>
</tr>
<tr>
<td>NLCD</td>
<td>National Land Cover Database</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PRODES</td>
<td>Programa de Cálculo do Desflorestamento da Amazônia</td>
</tr>
<tr>
<td>STEP</td>
<td>System for Terrestrial Ecosystem Parameterization</td>
</tr>
</tbody>
</table>
Section 1

Introduction

We live in an era in which nearly every part of the Earth’s surface has been effected by humans (Sanderson et al., 2002). As the Earth’s population continues to grow global land cover will continue change as a result of human activity. Change in land cover drives change in regional and global ecosystems, and has serious implications to the ecosystem services that provide human society with food, water, and clean air (Foley et al., 2005). Land cover information is also a key input to many Earth system models (Bonan et al., 2002; Ek et al., 2003; Running and Coughlan, 1988; Sellers et al., 1997; Sterling and Ducharne, 2008), and is therefore important to studies of climate change (Myhre et al., 2005), ecosystem modeling (Wiedinmyer et al., 2011), and hydrology (Gerten et al., 2004; Zhang et al., 2010). Reliable maps of global land cover are therefore critical to our ability to understand and adapt to a changing planet.

Over the last two decades, remote sensing has become the primary source of global land cover information. To support global change science, land cover classification from remote sensing must be not only accurate and repeatable, but also internally consistent. Change in land cover maps should indicate real change on the ground, and not classification errors or ambiguous samples flipping between spectrally similar classes (for example, an area being classified as savanna one year and woody savanna the next). Unfortunately, current data and methodologies are not available that produce land cover maps that are both accurate and internally consistent across time.
Several global land cover products are available (Friedl et al., 2010; Loveland et al., 2000; Hansen et al., 2000; Bartholome and Belward, 2005). However, comparison between these products shows much disagreement, especially in transition areas between biomes (Hansen and Reed, 2000; Giri et al., 2005; McCallum et al., 2006; Herold et al., 2008). Each of these land cover maps is produced using different methodologies and data sources, and some of the disagreement is certainly due to differences in method and data. Herold et al. (2008) note that the classes that show the most disagreement are also mapped with low accuracy, which suggests that some of the disagreement is simply due to uncertainty in classification results. Land cover classification schemes are designed to capture vegetation categories that are meaningful to maps users, but these categories are not necessarily easy to resolve using satellite imagery. Thus, many land cover classes are difficult to separate using satellite measurements. This is especially true at coarse spatial resolution where many (if not most!) pixels include a mix of land cover types (Friedl et al., 2010). This problem is especially evident in ecosystem transition zones. These complications make the production of large scale land cover maps especially challenging.

In addition to presenting challenges for maps produced for a single date, the issues described above also lead to inconsistencies in maps produced using the same method and data sources across different years of observation. An excellent case study of this phenomenon is the Land Cover Type (MLCT) product, which is produced using the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument. The MLTC has been produced annually since 2001, and the classification exhibits about 30% change from one year to the next (Friedl et al., 2010). To address this, the MLCT algorithm includes steps to stabilize the classification across years, but despite these measures, the MLCT product exhibits about 10% annual change, which significantly exceeds the amount of actual global land cover change in a single year (Friedl et al.,
With these issues in mind, the primary goal of this thesis is to explore methods to improve classification accuracies for global land cover classifications from MODIS, with a specific focus on reducing spurious year-to-year variation in classification results. To achieve this, I investigate how the composition and definition of land cover classes affect classification results and evaluate two methodological approaches to improve the stability of multi-temporal classifications. In the first part of this thesis I conduct an exploratory data analysis to improve understanding of how sub-class groupings in land cover classes affect classification results. Specifically, many land cover classes exhibit multimodal frequency distributions at global scale (Friedl et al., 2002). Guided by this observation, I use clustering to identify sub-groupings within the defined land cover classes that may aid in understanding the spectral separability of the classes.

In the second part of this thesis I examine methods to reduce year-to-year variation in classification labels that arise from classifier uncertainty. Many land cover products are produced using pixel-by-pixel classification algorithms that do not consider spatial or temporal correlation structure in land cover classes. Contextual classification algorithms such as Markov Random Fields have proven effective at improving classification results at higher spatial resolutions (i.e., 1-30m) by incorporating spatial and temporal context (e.g. Lu and Weng, 2007; Cortijo and De La Blanca, 1998; Magnussen et al., 2004; Moser et al., 2013). Recent work by Cai et al. (in press) suggests that these methods may prove helpful at coarse resolution as well. To test this hypothesis, I evaluate two techniques of incorporating spatial and temporal context into the MLCT classification algorithm. First, I evaluate a method that utilizes multiple years of observation to train a per-pixel land cover classification algorithm. My hypothesis in this experiment is that the use of multiple years of training data acts to buffer year-to-year variability in the training samples, and therefore results in more
consistent land cover maps. Second, I evaluate a more formal technique of incorporating spatial and temporal context into the classification using a Markov Random Field model. Results from this work provide insights into the challenges involved in coarse spatial resolution land cover classification, and possible methodological solutions to those challenges.
Section 2

Background

2.1 Global Land Cover Classification

Prior to the 1990’s global land cover maps were created from mostly ground-derived data sets (Matthews, 1983; Olson, 1982; Wilson and Henderson-Sellers, 1985). These maps were compiled from different data sources, at different spatial scales, and using different classification schemes, making it difficult to compare maps, or repeat the classification to study change. In the 1990’s, the Advanced Very High Resolution Radiometer (AVHRR) made it possible to create global land cover maps from remotely sensed data. The first global AVHRR map was produced at 1° spatial resolution using a maximum likelihood classification of monthly Normalized Difference Vegetation Index (NDVI) measurements (DeFries and Townshend, 1994). Later maps were produced at 8-km resolution using decision tree classification algorithms (DeFries et al., 1998), and at 1-km resolution using both supervised (Hansen et al., 2000) and unsupervised (Loveland et al., 2000) classification approaches.

The next generation of coarse resolution sensors, such as NASA’s MODIS and Envisat’s Medium Resolution Imaging Spectrometer (MERIS), offered enhanced spatial, spectral, radiometric, and geometric quality compared to AVHRR. In addition, these instruments view the entire Earth every 1-2 days (depending on latitude), making them well suited to monitoring seasonal changes in vegetation. Several global land cover products have been produced using these coarse resolution sensors, using a variety of classification techniques. Data from MODIS is used to create the MODIS Land
Cover Type product (MLCT), which is produced annually using supervised decision tree classification. The Global Land Cover 2000 product was produced using unsupervised classification of data from the SPOT VEGETATION sensor (Bartholome and Belward, 2005), and the GlobCover product has now been produced for two time periods (circa 2005 and 2009) using unsupervised classification of MERIS data (Arino et al., 2008).

Despite decades of progress and research, producing accurate and consistent land cover classifications at global scale remains a significant challenge. As I describe above, several global land cover data sets are now available, but numerous studies have identified large amounts of disagreement between these maps, especially in transition regions between major biomes. A separate, but equally important problem is that repeated classifications using consistent algorithms applied to different data from different years show similar patterns of disagreement across years.

The research described in this thesis focuses on the MLCT product, which is produced using supervised classification algorithms (decision trees) applied to data from MODIS. The MLCT is unique among the global land cover data sets in that it is the only product that has been produced each year since 2001. All land cover maps face the challenge of accurately mapping land cover. But, as I have previously described, repeated land cover mapping faces the additional challenge of creating maps that are consistent across years. This issue is especially important because MODIS now provides a time series of land cover observations that spans nearly 13 years, a period encompassing substantial change in global land cover conditions. Realistic characterization of changes in the MLCT product is therefore critically important.
2.2 Stability of Land Cover Classification

Multi-temporal land cover maps must not only be accurate when produced, but also consistent across time with the other maps in the series. However, existing land cover products show great inconsistency from year to year; the raw output of the MLCT algorithm exhibits about 30% change from one year to the next (Friedl et al., 2010). To address this, the MLCT algorithm applies post-hoc heuristic solutions designed to stabilize classification results for each pixel by post-processing the raw classifier output. If the classifier assigns a different label than in the previous year, the label is changed only if the probability assigned to the new classification is greater than in the previous classification. This process is applied over a three year window to avoid propagating an incorrect classification. However, even after this “stabilization”, the MLCT product exhibits 10% annual change, which far exceeds the amount of actual global land cover change in a single year (Friedl et al., 2010).

Comparison of classification results across different global land cover maps shows similar patterns of inconsistency, but the inconsistencies are not distributed evenly across geographic regions and land cover classes. The maps show very good agreement in some classes, and poor agreement in others. Herold et al. (2008) compared IGBP DISCover (Loveland et al., 2000), the University of Maryland land cover dataset (Hansen et al., 2000), MLCT, and Global Land Cover 2000 (Bartholome and Belward, 2005) and found almost complete agreement in classification of tropical rain forests, large deserts, and the Greenland ice sheet. However, there is great disagreement in mixed forests, shrubs, and herbaceous vegetation, especially in biome transition areas. Not coincidentally, classes showing the most disagreement also have low accuracies, which suggests that much of the disagreement is simply due to uncertainty in classification results.

An additional source of uncertainty is the classification scheme used to map land
cover. Many classification schemes have been developed to meet the needs of the scientific community. However, these classification schemes are generally based on user needs, which cannot necessarily be accurately identified from satellite data (Heinl et al., 2009; Friedl et al., 2010). Consequently, many classes included in these classification systems are inherently difficult to separate using satellite data (for example, savanna vs. woody savanna). In addition, because of the relatively coarse spatial resolution provided by sensors such as MODIS, many pixels contain mixtures of land cover types, especially in biome transition zones.

Thus, a key challenge of global land cover classification methods is to consistently assign a set of broad, and not always well separated, land cover labels to the Earth’s entire land surface. The global scope of the problem often necessitates the use of coarse resolution data, resulting in many pixels that contain a mix of land cover types. As a result, current classification methods result in levels of year-to-year change that are unrealistically high. A key challenge of the research described in this thesis is to develop methods that address this challenge by both improving overall classification accuracies and reducing year-to-year variability in classification results that are unrelated to change.
Section 3

Methods

3.1 Data

The data used for this study exploits previously compiled land cover and training data that are used to produce Collection 5 of the MLCT product. By using this data set I am able to leverage both the context and the large body of existing knowledge related to land cover classification using MODIS data. Further, by framing this work in the context of the MLCT product, the results of this study are directly relevant to a widely used land cover data product.

3.1.1 MODIS Land Cover Type

The data used in this study was collected by the MODIS instrument onboard NASA’s Terra and Aqua satellites. MODIS provides global multispectral imagery at a 1-2 day repeat frequency and collects reflectance measurements in 36 spectral bands, seven of which are explicitly designed to study land processes. MODIS captures daily images of the Earth poleward of 30 degrees, making it particularly well suited to capturing seasonal vegetation patterns.

The MODIS land cover type (MLCT) product is a widely used global land cover data set produced using data collected by MODIS. The MLCT product is designed to provide scientific investigators with information on the current state of global land cover, and has been produced each year since 2001 (Friedl et al., 2002). The MLCT is the only annually produced global land cover data set available at the time of this
The MLCT product is produced using a boosted ensemble of decision trees, that are estimated using the C4.5 algorithm Quinlan (1993). Results of the ensemble classifier are post-processed to correct for biases inherent in the algorithm and training data set, and to leverage prior knowledge about the geographic distribution of land cover types (Friedl et al., 2010). The use of boosting allows the algorithm to assign class-conditional probabilities to each pixel, rather than a single classification label (Friedman et al., 2000; McIver et al., 2001). These class-conditional and prior probabilities are then used to estimate the posterior probability of each class at each pixel using Bayes’ rule.

Because the results from this general approach tend to be unstable from one year to the next, the posterior probabilities are also leveraged to stabilize classification results across years. For each year, a new land cover label is assigned to a pixel only if the probability of the new label is greater than the probability of the previous label. By doing so, the total proportion of changed pixels reduces from over 30% to less than 10%. The final MLCT product includes a primary land cover label for each 500-meter pixel, supplemented by several common land cover classification legends.

3.1.2 IGBP Classification Scheme

The primary land cover label assigned by the MLTC uses the International Geosphere-Biosphere Programme (IGBP) classification system (Loveland and Belward, 1997). This scheme defines seventeen land cover classes, described in Table 3.1. The IGBP land cover classes are non-overlapping, yet broad enough to include all of the Earth’s land cover types.

The IGBP scheme includes a class for urban and built-up environments (class 13). For this work, however, this class is omitted from the training data. Urban areas are complex and heterogeneous landscapes that present unique classification
<table>
<thead>
<tr>
<th></th>
<th>Land Cover Class</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Evergreen needleleaf forest</td>
<td>Needleleaf woody vegetation with a percent cover &gt;60% and height exceeding 2 m. Almost all trees remain green all year.</td>
</tr>
<tr>
<td>2</td>
<td>Evergreen broadleaf forest</td>
<td>Broadleaf woody vegetation with a percent cover &gt;60% and height exceeding 2 m. Almost all trees remain &gt;60% green year round.</td>
</tr>
<tr>
<td>3</td>
<td>Deciduous needleleaf forest</td>
<td>Woody vegetation with a percent cover &gt;60% and height exceeding 2 m. Consists of seasonal needleleaf tree communities with an annual cycle of leaf-on and leaf-off periods.</td>
</tr>
<tr>
<td>4</td>
<td>Deciduous broadleaf forest</td>
<td>Woody vegetation with a percent cover &gt;60% and height exceeding 2 m. Consists of broadleaf tree communities with an annual cycle of leaf-on and leaf-off periods.</td>
</tr>
<tr>
<td>5</td>
<td>Mixed forest</td>
<td>Percent tree cover &gt;60% and height exceeding 2 m. Consists of tree communities with interspersed mixtures or mosaics of the other four forest types. None of the forest types exceeds 60% of landscape.</td>
</tr>
<tr>
<td>6</td>
<td>Closed shrubland</td>
<td>Woody vegetation less than 2 m tall and with shrub canopy cover &gt;60%. The shrub foliage can be either evergreen or deciduous.</td>
</tr>
<tr>
<td>7</td>
<td>Open shrubland</td>
<td>Woody vegetation less than 2 m tall and with shrub canopy cover between 10% and 60%. The shrub foliage can be either evergreen or deciduous.</td>
</tr>
<tr>
<td>8</td>
<td>Woody savanna</td>
<td>Herbaceous and other understory systems, and with forest canopy cover between 30% and 60%. The forest cover height exceeds 2 m.</td>
</tr>
<tr>
<td>9</td>
<td>Savanna</td>
<td>Herbaceous and other understory systems, and with forest canopy cover between 10% and 30%. The forest cover height exceeds 2 m.</td>
</tr>
<tr>
<td>10</td>
<td>Grassland</td>
<td>Herbaceous types of cover. Tree and shrub cover is &lt;10%.</td>
</tr>
<tr>
<td>11</td>
<td>Wetland</td>
<td>Permanent mixture of water and herbaceous or woody vegetation. The vegetation can be present either in salt, brackish, or fresh water.</td>
</tr>
<tr>
<td>12</td>
<td>Cropland</td>
<td>Temporary crops followed by harvest and a bare soil period. Does not include perennial woody crops.</td>
</tr>
<tr>
<td>13</td>
<td>Urban and built-up lands</td>
<td>Buildings and other man-made structures.</td>
</tr>
<tr>
<td>14</td>
<td>Cropland/natural vegetation mosaic</td>
<td>Mosaic of croplands, forests, shrubland, and grasslands in which no one component comprises more than 60% of the landscape.</td>
</tr>
<tr>
<td>15</td>
<td>Snow and ice</td>
<td>Snow/ice cover throughout the year.</td>
</tr>
<tr>
<td>16</td>
<td>Barren</td>
<td>Exposed soil, sand, rocks, or snow. Never more than 10% vegetated cover during any time of the year.</td>
</tr>
<tr>
<td>17</td>
<td>Water bodies</td>
<td>Oceans, seas, lakes, reservoirs, and rivers.</td>
</tr>
</tbody>
</table>

Table 3.1: IGBP land cover class definitions. (Source: Friedl et al., 2002)
Figure 3-1: Land cover class distribution in STEP training sites. The distribution of land cover classes in the database does not represent the distribution of land cover in reality. In particular, the agricultural class (class 12) is heavily oversampled.

challenges. As a result, the MLCT algorithm treats urban areas separately from non-urban areas, which significantly complicates any methods that operate on urban land areas in the MLCT algorithm (Friedl et al., 2010; Schneider et al., 2009). Hence, this work addresses only classification of non-urban land cover. Note, however, that urban labels may be added to the post-processed land cover maps using the same method as is currently used to map urban areas in the MLCT product.

The MLCT algorithm is trained using time series of MODIS spectral reflectances for training sites included in the System for Terrestrial Ecosystem Parameterization (STEP) database (Muchoney et al., 1999). The STEP database contains more than 2000 globally distributed sites, and is continuously updated and expanded. Each training site is defined by a polygon that delineates the site border, a land cover label, and physical and biological characteristics of the site land cover. Site labels and characteristics are determined by analysts using high resolution aerial imagery. The sites range in size, from a single 500-meter MODIS pixel ($\sim 0.2$ km$^2$) to 376 pixels ($\sim 80$ km$^2$). However, most sites fall in the low end of this range, with a median size
Figure 3·2: Locations of STEP training sites.

of 16 pixels (Friedl et al., 2010). Figure 3·1 shows the frequency distribution of land cover classes in the STEP database, and Figure 3·2 shows the geographic locations of the training sites. The training sites are fairly evenly distributed across the globe. However, the distribution of sites among land cover classes is not even. The agriculture class in particular is sampled much more heavily than other classes. Decision tree classifiers are sensitive to the distribution of training samples, so this imbalance has implications for classification accuracies. The MLTC algorithm includes a step to correct for bias in the classification due to uneven training set size.
3.1.3 Classification Features

The MLCT classification algorithm is trained using one year of spectral measurements for each pixel in the training data set. These measurements include the MODIS “land bands” (bands 1-7), the Enhanced Vegetation Index (EVI; Huete et al., 2002), and the MODIS Collection 5 Land Surface Temperature (LST; Wan et al., 2002). The spectral measurements are adjusted to a nadir-viewing angle to correct for variation in viewing geometry and directional reflectance of the land surface (Schaaf et al., 2002), and these adjusted values are used to compute the EVI. Eight day composite measurements are aggregated into 12 sets of 32 day averages in order to create a monthly time series in each band. In addition, the classification input includes the annual maximum, minimum, and mean value for each spectral band, EVI, and Land Surface Temperature. In total, there are 135 classification features. These are the same classification features used to produce Collection 5 of the MLCT, and are described in more detail by Friedl et al. (2010).

3.1.4 Parameter Estimation Data

The parameters of the Markov Random Field energy function (described in Section 3.2.2) capture a prior belief about the likelihood that neighboring pixels will share the same land cover label. Estimating these parameters requires a land cover data set that is as accurate as possible, yet also encompasses a wide enough range of climatic regimes to produce generally useful conclusions about land cover distribution. The National Land Cover Database (NLCD; Fry et al., 2011) is produced for the conterminous United States at 30-meter spatial resolution, using data from the Landsat Enhanced Thematic Mapper+. The NLCD dataset represents a more accurate characterization of land cover in the United States than is possible to achieve using coarse resolution imagery, and as such the NLCD is a valuable resource for estimating land
cover characteristics.

The NLCD is produced at a much finer scale than the MLCT, and in a different map projection. Several pre-processing steps are required to compare the NLCD to MODIS derived land cover maps. To do this, the 30-meter NLCD pixels were reprojected to the MODIS Sinusoidal map projection and aggregated to 500-meter scale to create a comparable land cover map. In addition, the NLCD does not use the IGBP classification scheme, so the NLCD land cover labels were translated into IGBP labels.

The NLCD includes a land cover class, the percent developed imperviousness, and the percent tree canopy for each 30-meter pixel. The percent tree canopy and percent impervious layers were used to estimate the amount of tree cover and impervious surface within each low resolution pixel, and determine the appropriate IGBP class for each 500-meter MODIS pixel. This thesis uses the 2006 version of the NLCD land cover map and impervious surface layer, and the tree canopy layer from NLCD 2001 (NLCD 2006 does not include tree canopy data). Using tree cover data from 2001 results in slightly incorrect estimation of tree canopy in parts of the continent due to growth and deforestation, but these effects were deemed insignificant for the purpose of determining the Markov Random Field parameters.

Figure 3·3 shows the logic used to aggregate 30-meter NLCD pixels into 500-meter IGBP pixels. Note that the NLCD classification scheme does not distinguish needle leaf and broad leaf forest, or open and closed shrubland. Therefore, the IGBP open and closed shrubland classes and needleleaf and broadleaf forest classes were collapsed in the resulting 500-meter NLCD map. That is, all evergreen forest pixels were labeled as simply evergreen forest, etc. Figure 3·4 shows the NLCD 2006 land cover data set aggregated to 500-meter resolution and labeled using the simplified IGBP scheme.
3.1.5 Validation data

The PRODES (Programa de Cálculo do Desflorestamento da Amazônia) data set was used to assess the ability of the MRF post-processing model to eliminate spurious label changes without smoothing over real change. This product is produced annually by Brazil’s National Institute for Space Research (INPE, 2012) and gives polygon based maps of deforestation derived from Landsat data. The 30-meter pixels were used to derive an annual sub-pixel fraction of deforestation for 2001 to 2010 in each of 1,050 MODIS pixels located in the Xingu region of Brazil (Figure 3-5).

3.2 Algorithms

3.2.1 Decision Trees

This work focuses on land cover classification using a decision tree classifier. Decision trees are non-parametric supervised classification tools that have been shown to perform well with remote sensing data (Friedl and Brodley, 1997; DeFries et al., 1998;
Figure 3.4: NLCD 2006 land cover map aggregated to 500-meter spatial resolution, and cross-walked to the IGBP classification scheme.
Figure 3.5: Locations of deforestation reference sites along the Xingu river basin in Brazil.

Hansen et al., 2000; Friedl et al., 2010). A training data set is required for pixels of known land cover. The decision tree algorithm recursively partitions the training data, forming a tree of decision points, each of which tests one input feature against a threshold. The tree grows until every training point is correctly classified, and then the lower levels of the tree are pruned back to avoid over fitting the training data. The decision logic captured by the final tree is then used to classify unseen pixels. There are many algorithms for constructing decision trees. The algorithm used here is C4.5 (Quinlan, 1993).

Decision trees have several properties that make them attractive for remote sensing applications. The classifier is non-parametric, meaning that it does not make any assumptions about the statistical distribution of the input features. In addition, the C4.5 algorithm can handle missing values in both the data used to build the tree, and in unseen data to be classified (Quinlan, 1993). Missing data due to cloud cover
or low illumination is an ever-present challenge in remote sensing, especially in the cloudy tropics. Once estimated, a decision tree is simply a series of if-then tests, so the classification criteria are easy to interpret.

For many applications, an ensemble of several decision trees can provide better results than a single tree. Each of the trees in the ensemble makes a prediction for a data point, and the predictions are combined using a voting strategy. A common method of building an ensemble classifier is to train the trees in sequence and weight the training data to correct mistakes made by previous trees. Data points that are misclassified by one tree are assigned a heavier weight when training the next tree. This technique is known as boosting (Freund, 1995), and has been shown to improve the accuracy of decision tree land cover classification (Friedl et al., 1999; McIver et al., 2001). In addition to improving accuracy, boosting makes it possible to compute probabilities of class membership rather than an single prediction, (Friedman et al., 2000). This is extremely useful, as it allows the decision tree output to be manipulated using the rules of probability, and provides a measure of confidence in the prediction.

3.2.2 Markov Random Fields

Markov Random Fields were originally developed in the field of stochastic physics, and have been widely applied for image processing (Li, 2009; Besag, 1974, 1986; Geman and Geman, 1984). MRFs provide a mathematically sound model for incorporating contextual information into an image classification task. In the context of land cover classification, an MRF can be employed to explicitly capture the correlation between a pixel and its neighbors. An MRF allows us to treat the problem of land cover classification as that of finding the best labeling of a scene, rather than the best label for each pixel independently. This section briefly introduces the theoretical basis of an MRF, and its application in land cover classification.
Land Cover Classification as a Labeling Problem

The objective of pixel-based land cover classification is to assign a label to each pixel in an image. The image is made up of a rectangular grid of pixels. We will identify the set of labels as $\mathcal{L} = \{\omega_1, \omega_2, \ldots, \omega_k\}$, and the assignment of labels to pixels as $X = \{x_1, x_2, \ldots, x_n\}$. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be the spectral measurements of each pixel. $X$ is a set of random variables that represent the labels assigned to the pixels of the image, and $Y$ is a set of observed variables representing the spectral measurements of those pixels. $x_i$ represents that label assigned to pixel $i$, and $y_i$ represents the spectral measurements of pixel $i$.

A non-contextual classifier, such as the decision tree classifier discussed above, treats each pixel independently of the other pixels. In this case the classifier produces a probability that each pixel takes a given land cover label, given the spectral measurements of that pixel. We represent the probability that pixel $i$ takes label $\omega$ as:

$$p(x_i = \omega | y_i)$$  \hspace{1cm} (3.1)

When each pixel is treated independently, the most likely label for each pixel is simply the label that maximizes this conditional probability. The optimal label for pixel $x_i$ is given by:

$$\omega^*_i = \arg\max_{\omega} p(x_i = \omega | y_i)$$  \hspace{1cm} (3.2)

However, pixels in a satellite image are not truly independent. Nearby pixels are more likely to share the same label than distance pixels. Our goal is not to find the best label for each pixel individually, but to find the most likely labeling of the entire image. Many methods have been developed to incorporate context into
image classification. The *Markov Random Field* (MRF) (Kindermann and Snell, 1980) is one of the most frequently used contextual classification models (Lu and Weng, 2007). Before considering random field models, we must first formalize the spatial relationship between the pixels in the image though a neighborhood system.

**Neighborhood Systems**

The set of random variables $X$ introduced above is an unordered set without any spatial relationship. However, pixels in an image have spatial relationships, which are captured by a *neighborhood system*. A neighborhood system is a formal definition of how the random variables relate to each other in space. Let $\mathcal{N}$ represent the neighborhood system, and $\mathcal{N}_i$ represent the neighbors of pixel $x_i$. A neighborhood system must satisfy the following constraints:

1. A pixel is not a neighbor of itself: $x_i \notin \mathcal{N}_i$.

2. The neighborhood relationship is symmetric: $x_i \in \mathcal{N}_j \implies x_j \in \mathcal{N}_i$.

This thesis will use a simple neighborhood system in which the neighbors of each pixel are the pixels directly above, below, to the left, and to the right. This neighborhood system is illustrated in Figure 3·6. While not as powerful as higher order neighborhood systems, the simple neighborhood described here is often used in practice.

The set $X$ of random variables and the neighborhood relation $\mathcal{N}$ form an undirected graph, $\mathcal{G}$, in which the nodes of the graph are the pixels in $X$, and the edges connect pixels that are neighbors. A *clique* is a subset of the graph $\mathcal{G}$ in which all the nodes in the clique are neighbors. Under the neighborhood system described here, the cliques in the graph are simply single pixels, and pairs of pixels connected by an edge.
Random Fields Models

A collection of random variables is called a random field. A Markov Random Field is a random field in which the random variables obey two constraints. First, the probability of each possible labeling must be strictly positive, and second, the probability of the label of each pixel can depend only on the neighbors of that pixel. Each pixel is independent of all pixels that are not its neighbors.

\[ p(x) > 0 \]  \hspace{1cm} \text{(Positivity)}

\[ p(x_i | X_{-i}) = p(x_i | \mathcal{N}_i) \]  \hspace{1cm} \text{(Markovinity)}

The notation \( X_{-i} \) denotes the set of variables except for \( x_i \), and \( \mathcal{N}_i \) denotes the neighbors of \( x_i \). Thus, the Markovinity constraint states that the conditional probability of a pixel’s label given the labels of all other pixels in the image is equal to the conditional probability given only the labels of the pixel’s neighbors.

Another type of random field is the Gibbs random field (GRF). A random field is a Gibbs random field if and only if the joint probability of the field can be expressed by the Gibbs distribution:
\[ P(X) = \frac{1}{Z} \exp \left( -\frac{1}{T} U(f) \right) \]  
(3.3)

\[ Z = \sum_{f \in F} \exp \left( -\frac{1}{T} U(f) \right) \]  
(3.4)

\[ U(f) = \sum_{c \in C} V_c(f) \]  
(3.5)

\( Z \) is a normalizing constant known as the *partition function*, and ensures that the distribution sums to one. \( U \) is called the energy function, and \( T \) is a temperature parameter which controls the sharpness of the distribution. \( (T \) is typically set to one for convenience.) The energy function \( U(f) \) is expressed as a sum of potential functions over the cliques in the graph. Thus, the Gibbs distribution expresses the joint probability of the random field as functions on the cliques in the graph. Note that the computation of the partition function \( Z \) requires a sum of the full state space of the distribution (every possible labeling of an image). For general graphs, is not computationally feasible to calculate \( Z \) exactly.

The Hammersley-Clifford theorem (Hammersley and Clifford, 1971) states that any MRF is also a GRF, and any GRF is also an MRF. This equivalence allows us to express the joint probability of an MRF using the Gibbs distribution. Expressed in this way, the probability of a labeling may be computed as a function of local clique potentials. These clique potentials may be designed to capture the desired contextual information for an image processing task. For example, one of the early applications of MRFs to image processing (Besag, 1986) was removing noise from corrupted images. In this case, the clique potential functions balance the original value of each pixel in the noisy image against the prior knowledge that neighboring pixels typically are similar in color.

The probability of assigning labels to the pixels of \( X \) given observations \( Y \) is given
by some probability distribution function $p(X|Y)$. The task of finding the optimal assignment of labels is equivalent to the finding the maximum a posteriori (MAP) assignment in a Bayesian framework. Using the Gibbs distribution to express the posterior probability, the MAP assignment is given by the labeling that minimizes the energy of the distribution:

$$X^* = \arg\max_X P(X|Y)$$

$$= \arg\max_X \frac{1}{Z} \exp \left[ -\frac{1}{T} U(X|Y) \right]$$

$$= \arg\min_X U(X|Y)$$

For example, a simple energy function for an image classification task (e.g. Li, 2009; Solberg et al., 1996; Moser et al., 2013) is:

$$U(X|Y) = \sum_{x_i \in X} \left[ -\log(x_i|y_i) + \beta \sum_{i \in N_i} I(x_i, x_j) \right]$$

$$I(x_i, x_j) = \begin{cases} 0 & \text{if } x_i = x_j \\ 1 & \text{otherwise} \end{cases}$$

The energy function above is computed as a sum of local energies. Each local energy is composed of two terms: a pixel-wise term computed as a negative log-likelihood, and a pairwise term that favors similar labels in neighboring pixels. Given this function, we may evaluate the energy contributed by each pixel, for each possible label of that pixel. The optimal labeling will be that which minimizes energy over all pixels. A low conditional probability will push the energy away from optimal. Likewise, each neighboring pixel of a different label will incur a penalty that pushes
the energy away from the optimum.

Finding the optimal label assignment is equivalent to minimizing the energy function $U(X|Y)$ over all pixels in the image. This optimization problem is nontrivial, and many algorithms have been developed to find approximate solutions. This thesis uses a simple greedy search algorithm called *Iterated Conditional Modes* (ICM; Besag, 1986) to find an approximate solution to the energy minimization problem. The ICM algorithm is not guaranteed to find the global solution to the energy optimization problem, and the algorithm is sensitive to initial conditions. However, ICM is often used in practice due to its conceptual and computational simplicity.

**Application to Remote Sensing**

MRF models have found many applications in remote sensing. Random field models have been applied to super-resolution mapping (Kasetkasem et al., 2005), change detection (Bruzzone and Prieto, 2000; Chen and Cao, 2013), segmentation (Sarkar et al., 2002), and multi-source classification (Solberg et al., 1996; Tso and Mather, 1999; Nishii, 2003). One of the most common applications of MRFs is to incorporate spatial context into image classification (e.g. Magnussen et al., 2004; Moser et al., 2013).

MRFs have also been used to model temporal context in multi-date image classification. Jeon and Landgrebe (1992) developed a spatio-temporal Markov Random Field classification model based on the “cascade” approach proposed by Swain (1978). In this model, each image in a temporal series is classified, and the resulting map is used as context for the next image in the series. Jeon and Landgrebe evaluated their approach on a pair of Landsat images and found that the spatio-temporal classifier improved classification accuracy of crop types (corn, soy, wheat, and alfalfa/oat) compared to a per-pixel maximum likelihood classification. Solberg et al. (1996) developed a framework for fusion of multiple sensors, ancillary data, and spatio-temporal con-
text based on the MRF model. The authors included a temporal term in the MRF energy function to represent the probability of a pixel transitioning from one class to another class between time steps. Solberg et al. evaluated their approach through two experiments in fusing Landsat imagery with Synthetic Aperture Radar (SAR) and data from a geographic information system for land-use and crop classification. The authors found that the MRF model improved classification accuracy and accuracy of change detection compared to a per-pixel reference model.

Bruzzone and Prieto (2000, 2002) incorporated an MRF model into a framework for unsupervised change detection. The authors evaluated their technique using multi-date Landsat imagery, and found that the incorporation of spatial context through the MRF model improved accuracy of burned area detection. In contrast to the land cover classification works discussed above, Bruzzone and Prieto applied the MRF model to a spectral difference image, rather than to a sequence of land cover classification maps.

Melgani and Serpico (2003) proposed a modification to the cascade scheme employed by earlier studies. Rather than classify images in series, Melgani and Serpico propose a “mutual” scheme which allows for bidirectional exchange of temporal information. The authors evaluated their framework using a multi-temporal data set of Landsat and SAR imagery, and found that the mutual approach yielded improved accuracy over a reference cascade style MRF classifier. In particular, the mutual approach prevented classification errors from propagating into later images.

Most recently, Cai et al. (in press) successfully employed a MRF model to reduce the number of “illogical transitions” that appear in the MLCT product. The authors designed the MRF energy function to penalize land cover transitions that are not ecologically feasible in a single year, such as a grassland transitioning to a forest, while also incorporating a probability of land cover change based on spectral measurement of each pixel. Cai et al. found that the MRF framework successfully reduced the number
of illogical land cover transitions observed in the MLCT product. In addition, Cai et al. evaluated the accuracy of the MRF processed maps using the MLCT Collection 5 training data set, and found that MRF processing improved classification accuracy of the training pixels. While this data is designed for training the per-pixel classifier and is not an independent validation set, the result is encouraging. This study builds on previous research by applying the well established MRF framework to a global scale multi-temporal classification task. Most past MRF applications in remote sensing have focused on the 1-30 meter spatial scale. Results from post-processing the MLCT product are encouraging (e.g. Cai et al., in press), but more work is needed to assess the potential of these models for coarse resolution applications.

The energy function used in this study is designed to encourage the types of land cover homogeneity observed in real landscapes at 500-meter scale. Many MRF studies (e.g. Solberg et al., 1996; Moser et al., 2013) have employed a spatial energy function that simply optimizes for neighboring pixels to share the same land cover label. While attractively simple, this model ignores that fact that some land cover classes are more likely to co-occur than others. For example, savanna and woody savanna are relatively likely to occur in neighboring pixels, while agriculture and permanent snow/ice are not. The energy function used in this thesis includes a class-specific spatial smoothing term that takes into account the fact that some labels are more likely to co-occur than others. The parameters of this term are estimated using the NLCD land cover product for the contiguous United States.

Most multi-temporal MRF studies have utilized either a transition probability matrix (e.g. Jeon and Landgrebe, 1992; Solberg et al., 1996) or a zero-one penalty matrix (e.g. Melgani and Serpico, 2003; Cai et al., 2013) to smooth labels in the temporal dimension. This thesis evaluates the potential of a general purpose indicator function based on ancillary information to inform temporal smoothing. This model
has the advantage that it does not require the estimation of a transition probability change matrix, or expert knowledge to define an allowed transition matrix.

**Representing time**

The random field model discussed thus far represents only the spatial relationship among pixels in an image, and does not account for the temporal relationship between images in a time series. We will assume that multi-date images are co-registered. At each time point, we represent the land cover map as a rectangular grid of pixels with a first order spatial neighborhood as discussed above. In the case considered here, each map corresponds to land cover within a calendar year (e.g., global land cover in 2008), but the framework may be generalized to handle other multi-date classification tasks. Extending the notation introduced above, we will identify the observation of pixel $i$ in year $t$ as $y_i^{(t)}$, and the label assigned to this pixel as $x_i^{(t)}$. To reduce notational clutter, we omit the $(t)$ superscript unless explicitly dealing with the temporal dimension.

We will extend the spatial neighborhood introduced above into the temporal dimension. Thus each pixel has four spatial neighbors and two temporal neighbors: the pixel in the previous year ($x_i^{(t-1)}$) and the subsequent year ($x_i^{(t+1)}$). Pixels in the first and last years of the series have only a single temporal neighbor. This neighborhood system is illustrated in Figure 3.7.

**Optimization algorithm**

The optimal labeling of an image is given by the MAP assignment of the joint probability of all pixels in the image. However, computation of the partition function $(Z)$ is typically intractable, as it requires summing over all possible permutations of labels, so a traditional maximum likelihood solution is not possible. Instead, a variety of approximation algorithms have been developed to efficiently solve this optimization problem.
The algorithm employed here is a simple greedy search algorithm called Iterated Conditional Modes (ICM; Besag, 1986). The algorithm is initialized with a labeling for all the pixels (in this case each pixel is labeled with the maximum likelihood class predicted by the decision tree classifier). It then iterates over the pixels, and for each pixel updates the class to the label that minimizes the local energy of that pixel, given the labels of the neighbors. In this way, the total energy of the graph must decrease at each step. The process continues until the change in energy from one iteration to the next decreases below some threshold.

To handle multi-temporal classification the ICM algorithm must be adapted slightly. To do this, one iteration of ICM is run over each image in the sequence, holding the images before and after fixed. Then, one iteration of ICM is run on the next image in the sequence, and so on. After all images have gone through one iteration of ICM, we return to the first image in the sequence and run a second iteration of ICM, then a second iteration on the second image, etc.

Note that this algorithm allows influence to flow both forward and backward in
time. This is in contrast to “cascade” approaches to multi-temporal classification in which the first image in the scene is classified, and then held fixed while the next image is classified. Melgani and Serpico describe optimization over the whole sequence of images as a “mutual” approach to multi-temporal classification. Taking temporal context from images both backward and forward in time prevents incorrect labels from propagating forward, and allows more context to influence the classification results. The vast majority of pixels will not change from year to year, and including influence from observations forward in time encourages consistent labeling of stable pixels.

Energy function

The energy function is designed to capture the available prior information for the labeling of each pair of pixels. Intuitively, the energy function captures the “compatibility” of the labeling of two neighboring pixels given the available prior information. This prior information includes the predictions of the decision tree classifier, our prior expectation about the land cover in the part of the world that contains these pixels, our knowledge of landscape heterogeneity, and observations of the pixels at other points in time.

We shall divide the energy function into three pieces corresponding to a pixel’s observation, and its relation with spatial neighbors and temporal neighbors.

\[
U(x_i) = U_D(x_i | y_i) + U_S(x_i | N_i) + U_T(x_i^{(t)} | x_i^{(t-1)}, x_i^{(t+1)})
\] (3.11)

The data term \((U_D)\) represents the probability of a pixel taking a label given only the observation of that pixel. This component of the energy is provided by the posterior probability of a pixel’s label, predicted by the decision tree classifier. Boosting allows us to compute a class conditional probability from the decision tree output, and from this we can compute a posterior probability by applying a prior using Bayes’ rule.
The determination and application of this prior is discussed in Friedl et al. (2010) and McIver and Friedl (2002). Note that this posterior probability could be provided by any probabilistic per-pixel classification model. The data portion of the energy function is given by:

\[
U_D(x_i|y_i) = -\log p(x_i|y_i)p(x_i) \tag{3.12}
\]

The spatial portion of the energy function \(U_S\) encodes the interaction between a pixel and its neighbors (the pixels above, below, to the left, and to the right). This term represents our expectations of landscape heterogeneity. The spatial energy function used here encourages neighboring pixels to take labels that are observed to co-occur in real landscapes at 500-meter scale. The energy function is calculated as the probability of observing two land cover classes in neighboring pixels, and a weighting factor to control the influence of the spatial energy. The probability of observing two land cover classes at neighboring pixels is estimated using data from the National Land Cover Database, and the spatial energy function is given by:

\[
U_S(x_i|N_i) = \beta_s \sum_{j \in N_i} p(x_i|x_j) \tag{3.13}
\]

Similarly, the temporal portion of the energy function represents the interaction of a pixel with its temporal neighbors (observations of the pixel in prior and subsequent years). This energy will be expressed using a change indicator function, and a weighting coefficient. The indicator function determines if it is likely that a pixel has changed between two time points. The indicator function could use the spectral classification features to determine change likelihood, or some other ancillary information source.
\[
U_T(x_i) = \beta_t \left[ I_{\Delta} \left( x_i^{(t-1)}, x_i^{(t)} \right) + I_{\Delta} \left( x_i^{(t)}, x_i^{(t+1)} \right) \right] 
\]

\[
I_{\Delta}(x^{(t_1)}, x^{(t_2)}) = \begin{cases} 
1 & \text{if } x^{(t_1)} \neq x^{(t_2)} \text{ and ancillary data does not indicate change} \\
0 & \text{otherwise} 
\end{cases} 
\]

Previous multi-temporal MRF studies have utilized energy functions that attempt to model the probability of a pixel transitioning from one land cover class to another (e.g. Jeon and Landgrebe, 1992; Solberg et al., 1996), or by penalizing transitions that are considered infeasible (e.g. Melgani and Serpico, 2003; Cai et al., 2013). These strategies have proven quite powerful, but require either extensive training data, or an expert to specify the transition probability matrix. This thesis will evaluate a simpler scheme of incorporating temporal context using an indicator function to determine the likelihood of change between images in a series.

### 3.3 Analysis

#### 3.3.1 Classification Accuracy

The following sections describe techniques designed to improve the accuracy of large scale land cover maps. The accuracy of pixel based land cover classification can be measured in several ways. The most straightforward measurement is simply the percentage of pixels that are correctly classified (the land cover label assigned by the classification matches the label in the validation data). Overall accuracy treats all errors equally. A more nuanced view of accuracy is formed by considering errors omission and commission separately. The **producer’s accuracy** measures errors of omission, and the **user’s accuracy** measures errors of commission.

The producer’s accuracy for class \( C \) (also called “recall” in the machine learning...
literature) is the proportion of data points labeled as class $C$ in the validation data that are correctly labeled by the classifier. For example, the producer’s accuracy of the grassland land cover class is the percentage of pixels that are grassland in reality that are labeled as grassland in the classification map.

$$\text{Producer's accuracy} = \frac{tp}{tp + fn}$$  \hspace{1cm} (3.16)

$tp = \text{True positives}$

$fn = \text{False negatives}$

User’s accuracy (also called “precision”) is the probability a data point labeled as class $C$ by the classifier is actually of class $C$. The user’s accuracy for the grassland class is the probability that a pixel labeled as grassland is actually grassland. User’s Accuracy is calculated as:

$$\text{User's accuracy} = \frac{tp}{tp + fp}$$  \hspace{1cm} (3.17)

$tp = \text{True positives}$

$fp = \text{False positives}$

An ideal classification would have both high user’s and high producer’s accuracy. However, in practice one must balance between the two. In this thesis user’s and producer’s accuracy are considered equally important.

It is rarely feasible to check the accuracy of every pixel, so the accuracy may be estimated using design based sampling (e.g. Stehman and Czaplewski, 1998). However it is often impractical to perform a design based validation of coarse resolution global land cover. There is simply not enough high resolution validation imagery available to perform design based sampling for the entire globe. Instead, the accuracy of global
classification may be estimated using the data used to train the classifier.

Cross validation is a widely used method of evaluating classifier performance when training data is limited (Hastie et al., 2009). In cross validation, the training data set is partitioned into a number of subsets, called “folds”. The technique can be applied to any number of folds, but 10 is a common choice. The classifier is trained using 90% of the data, and used to predict the withheld 10%. Then the classifier is trained on a different 90% and tested on a different 10%. This continues for 10 runs, after which every data point has been used as both training and test data. The average accuracy of the cross validation rounds gives an estimate of the classifier performance on unseen data.

3.3.2 Land Cover Sub-classes

In this section, I perform an exploratory analysis of the STEP training data set in order to gain insight into the composition and definition of the IGBP land cover classes. Specifically, this analysis aims to answer the following questions:

1. Is there a relationship between classification accuracy and the temporal consistency of land cover labels?

2. Are sub-classes present within the IGBP land cover classes?

3. If sub-classes are present, do these groups help explain the spectral separability (or lack thereof) of the IGBP classes?

Accuracy and Consistency of Land Cover Labels

In order to investigate the relationship between classification accuracy and classification consistency, it is necessary to define a measure of “stability” between multi-date land cover maps. Given two co-registered class maps, we might simply compare each pair of pixels and compute the percentage of pixels which agree in label. However, in
order to investigate change within each class we need a measure that compares only
the pixels mapped as the class of interest. Given two land cover classifications, we
may express the “stability” of a class by comparing the number and location of pixels
mapped as that class in the two images. Specifically, we may compute the number of
pixels mapped as class $C$ in both images divided by the number of pixels mapped as
class $C$ in either image:

$$\text{Stability}(X^{(1)}, X^{(2)}, C) = \frac{|\{i | x_i^{(1)} = x_i^{(2)} = C\}|}{|\{i | x_i^{(1)} = C \lor x_i^{(2)} = C\}|}$$  \hspace{1cm} (3.18)$$

The stability index ranges from zero to one. A stability of zero indicates that
no class $C$ pixels in Image 1 are mapped as class $C$ in Image 2, and visa versa. A
stability of one indicates that all of the class $C$ pixels in Image 1 are also mapped as
class $C$ in Image 2, and visa versa.

Using this definition of stability I perform a linear regression using the producer’s
accuracy of each land cover class as the predictor variable, and the stability of the
land cover class as the response variable. The producer’s accuracy is determined by
cross validation of the training data set, and the stability is determined by computing
the stability index defined above, averaged over ten years of classification maps. My
hypothesis is that there is a positive correlation between classification accuracy and
stability. That is, classes of higher accuracy are more stable through time.

**Clustering**

My second objective in this analysis is to identify sub-groups within the IGBP land
cover classes. To achieve this goal, I use the K-means clustering algorithm to partition
each IGBP class into clusters of spectrally similar data points. K-means is a simple
but widely used clustering algorithm that partitions data points into clusters such
that the Euclidean distance from each point in a cluster to the cluster mean point is
minimized. The K-means algorithm includes a random initialization step, so different runs of the algorithm are not guaranteed to yield the same results. For each land cover class I ran 25 iterations of K-means with random initializations, and selected the best scoring result.

The K-means algorithm requires the user to specify the number of clusters to create. The data mining literature gives no generally applicable rules for determining the number of clusters in a data set, and this task is typically addressed by trying different numbers of clusters and selecting the best result based on some application-specific criteria (Jain, 2010). I selected the number of clusters by running the clustering with different values of $K$ and examining the resulting clusters using scatter plots of principal components, plots of the clustered points on a geographic map, and plots of the time series trajectory of the cluster centers, and I selected clusterings that yielded interpretable clusters.

To simplify computation and interpretation of the data clustering, pixels within each training site were aggregated to a single data point by taking the mean value of the spectral measurements of the individual pixels. The training data set also contains missing values due to cloud cover or poor illumination. C4.5 decision trees are able to handle missing values, but in order to cluster and visualize the data set it was necessary to impute the missing values. For purposes of exploratory analysis, the training data was preprocessed in the following manner:

1. For sites in which some pixels are observed and some are missing, fill the missing values using the mean of the observed values for the site.

2. For sites in which no pixels are observed, linearly interpolate using the previous and next observations for each pixel. The time series represents a full year of data, so it is treated as a cycle. That is, a missing value for January is filled by interpolating between December and February.
3. Discard pixels that are still missing data after the previous imputation steps.

4. Aggregate remaining pixels by mean value in each site.

Data Visualization

I used three visualization techniques to explore the clusters identified by K-means. First, I plotted the geographic locations of the training sites within each IGBP class on a world map, and used color to indicate cluster assignment. Second, I plotted the time series of each cluster centroid. The resulting plot gives a time series view of the average spectral trajectory of each cluster. This visualization is inspired by a technique used by Steinbach et al. (2001) to identify patterns in clustered climatology data.

The third visualization is a scatter plot of data points in spectral space, colored to indicate cluster assignment. I used principal component analysis (PCA) to approximate the data set in two dimensions, which can be directly plotted as a scatterplot. Principal component analysis is a mathematical technique that constructs a set of orthogonal vectors, called principal components (PC’s), that are aligned with the directions of maximal variance a data set. The vectors are constructed in such a way that the first principal component is aligned in the direction of largest variance in the data set, the second principal component describes the dimension of second largest variance, etc. We can project a data point onto the first two principal components in order to approximate that data point in two dimensions.

Hybrid Classification

The majority of error in MLCT classification is caused by confusion between similar land cover types, such as grasslands, shrublands, and savanna. To gain further insight into the confusion between the IGBP classes I created a version of the training data set in which each point is labeled not with its IGBP label, but with an identifier for the
cluster into which the point had been placed by K-means. Using this clustered data set, I trained a decision tree classifier and performed 10-fold cross validation to assess classification accuracy. The objective of this hybrid classification is to gain insight into the relationship of the sub-classes and the spectral separability of the IGBP classes (i.e. are the classification errors evenly distributed among the sub-classes, or are some sub-classes more difficult to separate than others?).

### 3.3.3 Multi-Year Training Data

Each annual MLCT map is produced using a single year of training data. For example, reflectance measurements captured by MODIS in 2008 are used to produce the land cover map for 2008. This thesis evaluates two techniques for incorporating temporal context in land cover classification. The first of these techniques is to construct the decision tree training data set by pooling multiple years of observation. For example, the training data set might be all the training points for 2007, 2008, and 2009, rather than only 2008. Each training pixel is represented by not one but several data points, corresponding to observation of the pixel in different years. A land cover map for 2008 may then be produced by processing the 2008 spectral measurements using the multi-year classifier.

Using a multi-year training set, I estimated an ensemble decision tree classifier, and evaluated the accuracy and consistency of the resulting land cover maps. Specifically, I constructed a data set that includes observations of the STEP training sites from 2007 to 2012. To reduce the computational burden of the larger training set, the pixels in each site were averaged to produce one data point per site per year. Each site is represented by six data points, one for each year. I used this data set to train a decision tree classifier and perform a global land cover classification for each year from 2007 to 2012, and compared these results to results produced using trees trained on single years of data. The resulting maps were evaluated by measuring the percentage
of pixels that changed labels between subsequent years. In addition, I performed cross-validation on the training data set in order to assess the effect of the multi-year training set on accuracy.

My hypothesis in this experiment is that including multiple years of observation as separate training points captures the behavior of each land cover type under a wider range of climatic conditions than is possible using a single year of training data. For example, consider a hypothetical grassland site. In a wet year the grassland will green up more than average, whereas in a dry year the site will green up less than average. These two years of observation will show different spectral profiles, but both represent the profile of a grassland pixel. By including training data for both years, this single pixel contributes more information about the distribution of grassland spectra than either observation contributes on its own.

3.3.4 Markov Random Field Post-processing

My analysis of the MRF post-processing model evaluates the potential for context to improve classification accuracy in areas of stable, homogenous, land cover, and the potential for context to reduce spurious change in land cover maps. As part of this analysis I used a collection of deforested sites to assess the model’s ability to reduce spurious change while retaining real change, and compared this method to the heuristic stabilization used in the MLCT algorithm. I also investigated the sensitivity of the MRF model to the model parameters.

Model Parameters

The MRF energy function requires several parameters to be specified. The pairwise spatial energy term requires a potential matrix that specifies the likelihood that two land cover types will appear in neighboring pixels, and a weighting coefficient that determines how much influence the spatial prior has on the outcome of the classi-
fication. This parameter matrix was estimated using the NLCD data set described in Section 3.1.4. I analyzed the NLCD land cover map to determine the probability that each combination of land cover types will be observed in neighboring pixels at 500-meter scale. To achieve this, I created a co-occurrence matrix by counting the number of times that each pair of land cover classes appeared in neighboring pixels. Specifically, let \( N \) be a \( k \times k \) matrix where \( k \) is the number of land cover classes (17 in the case of the IGBP scheme). Entry \( N_{m,n} \) represents the number of times that land cover class \( m \) was observed neighboring land cover class \( n \):

\[
N_{m,n} = N_{n,m} = \frac{1}{2} \sum_i \sum_{j \in N_i} I(x_i, x_j) \quad (3.19)
\]

\[
I(x_i, x_j) = \begin{cases} 
1 & \text{if } x_i = x_j \\
0 & \text{otherwise}
\end{cases} \quad (3.20)
\]

Note that the neighborhood relationship is symmetric, so \( N \) is a symmetric matrix. The co-occurrence matrix was then transformed to a probability matrix by iteratively normalizing the rows and columns until the matrix converged to a doubly stochastic matrix (each row and column sums to one).

The spatial energy function also requires a weighting coefficient (\( \beta_s \)) that defines how strongly the spatial energy influences the total energy. A high value of \( \beta_s \) will cause the model to favor more homogenous land cover maps. I evaluated the MRF model using a range of values for \( \beta_s \) in order to assess the sensitivity of this parameter. For reference, I used the NLCD data set to estimate the homogeneity of real landscapes at 500-meter scale. For this purpose, we represent the homogeneity of a land cover map as the number of neighboring pixels that have the same land cover label, divided by the total number of pixels:
\[ H(X) = \frac{1}{2|X|} \sum_i \sum_{j \in \mathcal{N}_i} I(x_i, x_j) \]  

\[ I(x_i, x_j) = \begin{cases} 
1 & \text{if } x_i = x_j \\
0 & \text{otherwise} 
\end{cases} \]

Using this definition, I estimated the homogeneity of real landscapes by evaluating \( H(\cdot) \) on the NLCD parameter estimation dataset.

The MRF energy function also contains a parameter to control the influence of the temporal energy term (\( \beta_t \)). High values of \( \beta_t \) lead to consistent land cover labels in subsequent images. A key challenge of determining this parameter is finding a value that minimizes spurious change without smoothing over real change. In order to determine the sensitivity of this parameter, I evaluated the MRF model using a range of values for \( \beta_t \) and a fixed value of \( \beta_s \). I measured the number of changed pixels from year to year in the resulting maps, and also the ability of the model to detect change due to deforestation.

**Evaluation of Unchanged Sites**

I evaluated the accuracy of the classification of the STEP training pixels in the MRF post-processed maps, and compared the results to the decision tree output and to the “stabilized” Collection 5 MLCT maps. The STEP training sites are selected to be good training samples for the decision tree classifier. As such, they are homogenous, and stable through time. In addition, I expect the accuracy of these pixels to be higher than average, as they form the training set used to build the decision tree. Evaluation using this data set represents a “best case scenario” for the contextual classification model, and assesses the potential for the model to correct errors in the decision tree output using spatial and temporal context.
Evaluation of Changed Sites

A data set of deforested sites in Brazil’s Xingu river basin was used to assess the ability of the MRF model to decrease spurious change while retaining real change. This data set, described in Section 3.1.5, gives the percentage deforestation within each of 1,050 500-meter pixels. I evaluated the degree to which the post-processed land cover maps agree with the deforestation statistics in this reference data set by counting errors of omission and commission within the ten year time series.

In this analysis, an error of omission was defined as a change in deforestation of at least 10% that does not coincide with a land cover change consistent with deforestation within a three year window. An error of commission was defined as a change in land cover that implies deforestation (for example, forest to savanna, or woody savanna to savanna) that does not coincide with a change in deforestation within a three year window. For purposes of this analysis, changes in land cover label that do not imply deforestation were not considered.
Section 4

Results

4.1 Land Cover Sub-classes

4.1.1 Accuracy and Consistency of Land Cover Labels

To investigate the relationship between classification accuracy and the temporal consistency of land cover labels, I performed a linear regression using the producer’s accuracy for each IGBP land cover class (estimated using 10-fold cross-validation) as the predictor variable, and label stability as the response variable. Label stability is measured on a scale of zero to one, where zero indicates that many pixels change label from year to year, and one indicates that no pixels change label (Equation 3.18). Figure 4-1 shows a plot of the stability of each class versus the producer’s accuracy for that class. The regression shows that accuracy explains only about 34% of the variance in stability, but that the relationship is statistically significant ($p < 0.05$). Although clearly not the only factor involved, classes with low accuracy are less stable, while classes with higher accuracies are more stable. This result supports my hypothesis that samples that are difficult to classify will exhibit higher levels of year-to-year “change” in land cover maps.

Broadleaf evergreen forest and barren land (classes 2 and 16) appear to be outliers in Figure 4-1. This might be explained by large, homogenous regions of forest and desert in global land cover, such as the Amazon rainforest and the Sahara desert. Although I do not test this factor here, it is likely that landscape heterogeneity also affects label stability. Intuitively, more heterogenous landscapes will contain more
mixed pixels, and thus include more pixels that are difficult to classify correctly. Conversely, very homogenous landscapes will contain relatively few mixed pixels, and I would expect less year-to-year label change.

These results demonstrate that there is a correlation between classification accuracy and year-to-year change in mapped land cover. The MLCT decision tree classifier achieves low accuracy for some land cover classes, and this low classification accuracy contributes to the problem of inconsistency in year-to-year land cover maps. The remainder of this section explores the nature of the classification training data, and investigates the spectral separability of land cover classes.

4.1.2 Clustering and Data Visualization

A key objective of this exploratory analysis is to identify whether sub-classes are present within the IGBP land cover classes that may explain separability of the classes. To explore this, I used K-means clustering to find groups of similar pixels within the training data for each class. The resulting clusters are presented in three visualizations: as a scatter plot of the first two principal components, as points on a world map, and as a time series plot of the cluster centroids. The resulting plots suggest that many of the IGBP classes do contain sub-groups, and that these sub-groups correspond to distinct land cover types in different parts of the world.

For example, Figure 4·2a shows the data points in the open shrubland class plotted in the principal component space, and also the geographic locations of the training sites. Figure 4·3a shows the same plots for the grassland class. (Clustering results for all IGBP classes are given in Appendix A.) The shrubland and grassland classes are each partitioned into three sub-classes. An interesting spatial pattern occurs when the training sites are plotted on a world map and labeled with the clusters identified by K-means. The clusters separate into high and low latitude groups, divided by a line at about 45°N latitude. That is, the points are clustered not only in the data
Figure 4.1: Year-to-year stability (Eq. 3.18) of each IGBP land cover class vs. cross-validated Producer’s Accuracy. A low stability score indicates that many pixels change label from year to year, while a high score indicates that few pixels change label from year to year.
(a) Left: Training sites plotted using first two principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure 4-2: Sub-classes of open shrubland class (class 7).
(a) Left: Training sites plotted using first two principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure 4-3: Sub-classes of grassland class (class 10)
space, but also in geographic space. It is worth noting that the features used to assign cluster labels are not based on the location of the training sites, so the spatial clustering is caused by differences in spectral behavior of training sites in different parts of the world.

To understand the spectral difference in high and low-latitude sites I examined the spectral time series of each cluster. Plotting the time series values for the cluster centroid shows the average spectral behavior of points in each cluster. Figures 4·2b and 4·3b show the time series trajectories of the shrubland and grassland clusters. The high-latitude clusters show much higher reflectance in bands 1-4 during the early and late months of the year. This high reflectance is likely due to snow cover, or bare soil exposed during the winter dormant season. In contrast, the low-latitude clusters shows a more level trajectory through the year in all bands. Thus, the first principal component captures the seasonality of the training sites. Note that these clusterings capture only broad patterns in the data. It is possible that more detailed analysis would reveal more subtle patterns within the general clusters identified here. For example, it is likely that the low-latitude clusters could be further divided into Northern and Southern Hemisphere groups,

The other land cover classes show similar sub-class patterns (Appendix A). The high-latitude clusters exhibit much higher reflectance in the early and late months of the year, and the low-latitude clusters show a comparatively level spectral trajectory. These results suggest that the largest source of variance in the data is seasonality and snow cover. The presence of identifiable sub-classes confirms that these land cover types exhibit multi-modal behavior in their spectral time series. The next section investigates the relationship between the sub-classes and mis-classified samples.
Figure 4.4: Confusion matrix for training sites labeled by sub-class. The outer axis label shows the IGBP classes, and the inner label shows the sub-classes of each IGBP class. The table shows that classification error between IGBP classes is often caused by confusion between certain sub-classes. For example, nearly all of the error between open shrubland and grassland is due to confusion between sub-classes 14 and 24.
4.1.3 Hybrid Classification

A large source of error in MLTC land cover classification is confusion between similar land cover types, such as grasslands, shrublands, and savanna. However, the previous section demonstrates that these classes exhibit different spectral behavior in different parts of the world, and in particular, that the spectral signature of high-latitude sites is dominated by seasonality. This section delves deeper into the confusion between these classes, and attempts to determine if the spectral confusion is especially problematic between certain sub-classes.

In order to gain insight into the relationships among the sub-classes identified by clustering and the spectral separability of land cover classes I created a version of the training data set in which each point is labeled not with its IGBP label, but with an identifier for the cluster into which the point had been placed by K-means.
I then trained a decision tree classifier using this new data set, and performed 10-fold cross validation to assess classification accuracy. The confusion matrix for the sub-class cross validation (Figure 4·4) shows that the confusion is not uniform among the clusters. For example, open shrublands (class 7) and grasslands (class 10) are each divided into three sub-classes. However, the classification error between these two classes is due mostly to confusion between only two to the sub-classes: cluster 14 (low-latitude open shrubland) and cluster 24 (low-latitude grassland). Indeed, confusion between low-latitude open shrubland and low-latitude grassland represents the single largest source of error in the confusion matrix. Figure 4·5 shows the first two principal components of the training samples in these sub-classes. The two groups are largely overlapped the principal component space, which explains the difficulty in classifying these points correctly.

This exploratory analysis shows that the largest source of variance in the vegetated class data is due to strong seasonality and snow cover of high-latitude training sites. Most of the classification error is due to confusion between low-latitude data points. Low-latitude sites of similar land cover (for example, grassland and open shrubland) are highly overlapped in the first two principal components, which explains the high degree of confusion between these classes. Based on these results, the available spectral and temporal classification features are not sufficient to separate some classes with high accuracy.

4.2 Multi-Year training data

To assess the potential for multiple years of training data to improve classification performance, I pooled the STEP training data for 2007 to 2012, and estimated an ensemble decision tree classifier using this multi-year data set. My hypothesis in this experiment is that multiple years of training data will act to buffer inter-annual
Figure 4-6: Percentage of changed pixels in land cover maps produced by decision trees trained on single years of data compared to trees trained on six years of data variability, and result in more consistent and accurate land cover maps. To assess the effectiveness of the multi-year data I measured the number of pixels that changed their label from year-to-year in the maps generated from multiple years of data versus single years of training data (Figure 4-6). The land cover maps produced using the multi-year data set consistently show fewer pixels changing class from year to year. A comparison of land cover maps for 2008 through 2012 shows on average about 40% change from year to year when using trees produced using single years of training data, and 32-33% change using trees trained using the 2007-2012 training set. The numbers presented here reflect changed pixels in the unmodified decision tree output, without further post-processing.

To assess how the multi-year data set affects classification accuracy I performed 10-fold cross validation using validation folds partitioned by the training sites, not the individual data points. Partitioning the data by site rather than by data point avoids spatial and temporal auto-correlation between observations of the same site in
Figure 4.7: Comparison of classification results using training data from 2007-2012 compared to year-by-year classification trees. Error bars indicate +/- one standard deviation.
different years from biasing the cross validation results (Friedl et al., 2001). Figure 4.7 shows the average accuracy of the multi-year classification and single year classifications. The mixed forest class (5) shows a slight drop in accuracy when using multiple years of training data, and barren (16) and water (17) show a small drop in producer’s accuracy. However, all other classes perform as well or better when using six years of training data, with several classes showing a notable improvement including closed shrubland (6), grasslands (10), and agricultural mosaic (14). However, the variance of the estimated accuracy also increases under the multi-year training set. The cause of this increased variance is not yet well understood, and will be the subject of future investigation.

The results from this experiment demonstrate that using multiple years of training data produces land cover maps with higher cross validation accuracy and fewer label changes from year to year. As previously discussed, levels of annual change of 30-40% are unrealistically high. However, the results presented here demonstrate that pooling multiple years of data has a measurable impact on the temporal consistency of the land cover maps. Thus, this experiment supports my hypothesis that the use of multiple years of training data can buffer inter-annual variability in the training samples.

4.3 Markov Random Field Post-processing

This section discusses results from post-processing per-pixel land cover classifications using a Markov Random Field model. I performed three experiments to assess the performance of this model. First, I investigated the effect of the MRF spatial smoothing parameter on the homogeneity of the land cover maps and compared this to the homogeneity of the reference NLCD data set. Second, I evaluated the accuracy of the STEP training pixels after MRF processing to assess the model’s performance in re-
Figure 4.8: Random field energy decreases with each iteration of ICM. Convergence is achieved within six iterations.

regions of stable, homogenous land cover. Third, I evaluated omission and commission errors for a set of deforested pixels in order to assess the model’s ability to reduce spurious label change while retaining real change.

4.3.1 Performance

The optimal labeling in the MRF model corresponds to the minimum of the random field energy function. This minimum was found using the Iterated Conditional Modes algorithm, described in Section 3.2.2. ICM is a simple greedy search in which each pixel, in raster-scan order, is assigned an optimal label given the current labels of neighboring pixels. The process iterates until the random field energy converges. Empirical results show that the energy converges within six iterations (Figure 4.8).

The ICM algorithm was implemented using the C programming language, and configured to process the output of the MLCT decision tree classifier. The random field grid is constructed as a time series of images corresponding to one tile in the
MODIS sinusoidal tile grid. Each tile is 2400 pixels wide and 2400 pixels tall, and the time series is 12 years in length (2001-2012). Processing one tile sequence takes approximately 4.75 minutes of computation time on a 2.5 GHz Linux computer.

4.3.2 Spatial Smoothing Parameter

A key challenge of applying the MRF model is determining the value of the spatial smoothing parameter \( \beta_s \). This parameter controls how strongly the model favors regions of homogeneous land cover over heterogeneous land cover. The homogeneity of a land cover map may be quantified as the probability that a pair of neighboring pixels will share the same label (Equation 3.22). Ideally, the MRF processed map would reproduce levels of homogeneity similar to the reference NLCD data set. To assess the potential benefit of MRF smoothing, I computed the average homogeneity of the NLCD data set, of the MLCT decision tree output, and the stabilized MLCT maps (Table 4.1). Contrary to my expectation, the NLCD map is less homogenous, on average, than the decision tree output. That is, a randomly selected pair of neighboring pixels in the NLCD map is less likely to share the same label than a pair selected from the decision tree maps.

However, the homogeneity of the NLCD map varies considerably across the North American continent. Figure 4.9 shows landscape homogeneity of the NLCD map averaged within a sliding 100-km window. Some parts of the map are very homogeneous (such as the deserts in the southwestern U.S., which consist primarily of shrubland), and others are very heterogeneous (such as the southeastern U.S., which is a mix of forest types, agriculture, and agriculture mosaic). In parts of the country, the probability of neighboring pixels sharing the same label is less than 0.5. This suggests that a single spatial smoothing parameter is insufficient to capture the variability of landscapes at continental scale.

One of the more homogenous regions of the NLCD map is the southwestern United
Figure 4.9: NLCD Landscape homogeneity (Eq. 3.22) averaged by 100-km block.

Figure 4.10: Homogeneity (Eq. 3.22) of land cover in MODIS tile h8v5 (Southwestern U.S.) as a function of the spatial smoothing parameter ($\beta_s$). Horizontal lines indicate the homogeneity of this tile in the decision tree output and the NLCD reference map. ($\beta_t = 1.25$)
Table 4.1: Average landscape homogeneity (Eq. 3.22) of MODIS tiles in the conterminous United States. NLCD = National Land Cover Database at 500-meter scale and cross-walked to IGBP; DT = Decision tree output, after priors; C5 = MLCT Collection 5 stabilized maps.

<table>
<thead>
<tr>
<th>Tile</th>
<th>NLCD</th>
<th>DT</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>h08v04</td>
<td>0.720</td>
<td>0.749</td>
<td>0.764</td>
</tr>
<tr>
<td>h08v05</td>
<td>0.844</td>
<td>0.795</td>
<td>0.815</td>
</tr>
<tr>
<td>h08v06</td>
<td>0.966</td>
<td>0.949</td>
<td>0.963</td>
</tr>
<tr>
<td>h09v04</td>
<td>0.755</td>
<td>0.802</td>
<td>0.821</td>
</tr>
<tr>
<td>h09v05</td>
<td>0.730</td>
<td>0.792</td>
<td>0.806</td>
</tr>
<tr>
<td>h09v06</td>
<td>0.678</td>
<td>0.691</td>
<td>0.758</td>
</tr>
<tr>
<td>h10v04</td>
<td>0.719</td>
<td>0.876</td>
<td>0.900</td>
</tr>
<tr>
<td>h10v05</td>
<td>0.547</td>
<td>0.700</td>
<td>0.734</td>
</tr>
<tr>
<td>h10v06</td>
<td>0.707</td>
<td>0.649</td>
<td>0.613</td>
</tr>
<tr>
<td>h11v04</td>
<td>0.719</td>
<td>0.876</td>
<td>0.900</td>
</tr>
<tr>
<td>h11v05</td>
<td>0.635</td>
<td>0.829</td>
<td>0.856</td>
</tr>
<tr>
<td>h12v04</td>
<td>0.543</td>
<td>0.721</td>
<td>0.750</td>
</tr>
<tr>
<td>h12v05</td>
<td>0.594</td>
<td>0.760</td>
<td>0.758</td>
</tr>
<tr>
<td>h13v04</td>
<td>0.581</td>
<td>0.818</td>
<td>0.869</td>
</tr>
<tr>
<td><strong>Avg</strong></td>
<td><strong>0.678</strong></td>
<td><strong>0.793</strong></td>
<td><strong>0.817</strong></td>
</tr>
</tbody>
</table>
States. To assess the effect of \( \beta_s \) in a fairly homogenous landscape, I processed MODIS tile h8v5 (Southwestern U.S.) using a fixed value of \( \beta_t \) and a range of values for \( \beta_s \). Figure 4·11 shows a subset of the land cover map for MODIS tile h8v5 processed using a range of values for \( \beta_s \). For low values of \( \beta_s \) the map shows many isolated pixels, or small regions of land cover different than the surrounding region. As \( \beta_s \) is increased, this high-frequency change is smoothed out, and the image tends toward larger, homogenous regions of the same label.

For each annual map of h8v5, I computed the homogeneity, and plotted the results as a function of the smoothing parameter (Figure 4·10). The homogeneity increases steadily as \( \beta_s \) increases, approaching and asymptote around 0.91. The horizontal lines in the plot show the homogeneity of h8v5 in the decision tree output and the NLCD maps. The decision tree map is slightly less homogenous than the NLCD map, which suggests that the map may be improved by spatial context. A parameter value of \( \beta_s = 0.25 \) produces a map of nearly equal homogeneity to the reference map for this tile.

### 4.3.3 Evaluation of Unchanged Sites

In order to assess the ability of the MRF model to correct errors in decision tree classification results I evaluated the accuracy of the decision tree training pixels after MRF processing. The training sites were selected to be homogenous and stable through time. Because these pixels represent the decision tree training set, the classification accuracy is quite high. Hence, this experiment is a “best case scenario” for the MRF, and evaluates the potential for the model to improve classification in areas that are homogenous and stable through time.

I compared the true label of each training pixel to the label assigned in the MRF processed maps, the stabilized MLCT Collection 5 maps, and the maximum likelihood class in the decision tree output (adjusted with prior probabilities). Table 4.2 shows
Figure 4.11: Land cover maps for a subset of MODIS tile h8v5 for a range of values of the spatial smoothing parameter. ($\beta_s = 1.25$)
Table 4.2: Classification accuracy of STEP training pixels. MRF = Markov Random Field ($\beta_s = 0.5$, $\beta_t = 1.25$), C5 = MLCT Collection 5, DT = Decision Tree output

<table>
<thead>
<tr>
<th>Year</th>
<th>User’s Accuracy</th>
<th>Producer’s Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DT</td>
<td>MRF</td>
</tr>
<tr>
<td>2001</td>
<td>94.1</td>
<td>95.7</td>
</tr>
<tr>
<td>2002</td>
<td>94.3</td>
<td>95.9</td>
</tr>
<tr>
<td>2003</td>
<td>94.2</td>
<td>96.1</td>
</tr>
<tr>
<td>2004</td>
<td>94.3</td>
<td>95.9</td>
</tr>
<tr>
<td>2005</td>
<td>94.1</td>
<td>96.0</td>
</tr>
<tr>
<td>2006</td>
<td>94.1</td>
<td>96.0</td>
</tr>
<tr>
<td>2007</td>
<td>94.2</td>
<td>96.1</td>
</tr>
<tr>
<td>2008</td>
<td>94.0</td>
<td>95.9</td>
</tr>
<tr>
<td>2009</td>
<td>94.1</td>
<td>96.0</td>
</tr>
<tr>
<td>2010</td>
<td>94.0</td>
<td>96.1</td>
</tr>
<tr>
<td>2011</td>
<td>94.5</td>
<td>96.3</td>
</tr>
<tr>
<td>2012</td>
<td>92.5</td>
<td>95.2</td>
</tr>
</tbody>
</table>

Table 4.2: Classification accuracy of STEP training pixels. MRF = Markov Random Field ($\beta_s = 0.5$, $\beta_t = 1.25$), C5 = MLCT Collection 5, DT = Decision Tree output

The resulting user and producer accuracies. The MRF accuracy scores are slightly higher than the decision tree scores, but very similar to the Collection 5 stabilized accuracies. Tables 4.3-4.5 show confusion matrices for the MRF maps, decision tree output, and Collection 5 stabilized maps for the year 2012. The patterns of error are similar, and are largely among land cover types with poor spectral separability (vegetation classes, agriculture vs. agriculture mosaic, etc.). These results show that the context provided by the MRF model successfully corrects some of the decision tree classification error. However, the MRF accuracy scores are not, in general, superior to the Collection 5 scores. It is therefore likely that most of the accuracy improvement is due to the temporal context, rather than the spatial context.
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRF Prediction (IGBP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>976</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2281</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>10</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>144</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1397</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>25</td>
<td>1388</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>171</td>
<td>17</td>
<td>514</td>
<td>0</td>
<td>2</td>
<td>16</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2374</td>
<td>48</td>
<td>2</td>
<td>69</td>
<td>1</td>
<td>1</td>
<td>11</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>19</td>
<td>6</td>
<td>32</td>
<td>15</td>
<td>100</td>
<td>0</td>
<td>12</td>
<td>1713</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>65</td>
<td>32</td>
<td>1373</td>
<td>7</td>
<td>0</td>
<td>26</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>0</td>
<td>8</td>
<td>2914</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>11</td>
<td>21</td>
<td>67</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>19</td>
<td>15</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2876</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>28</td>
<td>117</td>
<td>0</td>
<td>8203</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>150</td>
<td>1173</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>589</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>120</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1463</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>55</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1315</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision Tree Prediction (IGBP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>973</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2254</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>12</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>136</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>1339</td>
<td>89</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>0</td>
<td>27</td>
<td>29</td>
<td>13068</td>
<td>0</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>170</td>
<td>18</td>
<td>514</td>
<td>0</td>
<td>2</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>2303</td>
<td>48</td>
<td>44</td>
<td>85</td>
<td>3</td>
<td>5</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>44</td>
<td>6</td>
<td>35</td>
<td>13</td>
<td>127</td>
<td>0</td>
<td>12</td>
<td>1649</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>16</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>24</td>
<td>3</td>
<td>77</td>
<td>39</td>
<td>1278</td>
<td>43</td>
<td>0</td>
<td>27</td>
<td>23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>71</td>
<td>7</td>
<td>2</td>
<td>2855</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>54</td>
<td>110</td>
<td>2</td>
<td>0</td>
<td>13</td>
<td>0</td>
<td>22</td>
<td>36</td>
<td>21</td>
<td>9</td>
<td>2635</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>1</td>
<td>36</td>
<td>28</td>
<td>75</td>
<td>261</td>
<td>0</td>
<td>7795</td>
<td>120</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>141</td>
<td>1154</td>
<td>0</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>370</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>120</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1461</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1370</td>
</tr>
</tbody>
</table>

Table 4.4: Confusion matrix for training pixels in 2012 in decision tree output. Vertical: true labels. Horizontal: predicted labels.
Table 4.5: Confusion matrix for training pixels in 2012 in MLCT C5 stabilized maps. Vertical: true labels. Horizontal: predicted labels.

<table>
<thead>
<tr>
<th>Window (years)</th>
<th>Omission Rate</th>
<th>Comission Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MRF</td>
<td>C5</td>
</tr>
<tr>
<td>1</td>
<td>0.476</td>
<td>1.03</td>
</tr>
<tr>
<td>2</td>
<td>0.364</td>
<td>0.79</td>
</tr>
<tr>
<td>3</td>
<td>0.336</td>
<td>0.293</td>
</tr>
<tr>
<td>4</td>
<td>0.33</td>
<td>0.271</td>
</tr>
</tbody>
</table>

Table 4.6: Error rate for different acceptance windows. For example, using a window of two years, an omission error is a deforestation event that does not coincide with land cover label change within the following two years. A commission error is a label change that does not coincide with deforestation within the previous two years.

MRF = MRF post-processed maps ($\beta_s = 0.25$, $\beta_t = 1.25$), C5 = MLCT Collection 5, DT = Decision Tree output
4.3.4 Evaluation of Changed Sites

To assess the sensitivity of the MRF model to the temporal smoothing parameter I measured omission and commission error rates for a selection of deforested sites in Brazil. The fraction sub-pixel deforestation within each 500-meter MODIS pixel from 2001 to 2010 was estimated using the 30-meter PRODES data set described in Section 3.1.5. In this analysis, an error of omission was defined as a change in sub-pixel deforestation of more than 10% in the reference data set that does not coincide with a change in land cover label within a three year window. An error of commission was considered a change in land cover label that implies deforestation (e.g., forest to savanna) that does not coincide with deforestation in the reference data set within the preceding three years. Label changes that do not imply deforestation (e.g., grassland to agriculture) are not considered here.

The use of a three year window to define commission and omission error was based on experimentation with different time windows. Deforestation events do not always coincide neatly with the calendar year, so change in the land cover maps does not always match exactly with the year of deforestation in the reference data. The problem is further complicated by the stabilization employed the MLCT algorithm. As discussed earlier, the MLCT algorithm only changes a pixel’s label if the probability of the new label exceeds that of the previous label. This process is applied over a three year window to avoid propagating incorrect labels. A deforestation event involves forest (in the Amazon, typically broadleaf evergreen forest) transitioning to non-forest. Broadleaf evergreen forest is well separated in spectral space, and thus forest labels are typically associated with high confidence. The posterior probabilities assigned to non-forest pixels tends to be lower simply because the non-forest vegetation classes largely overlap in spectral space. When the forest is cut down and replaced with agriculture, the posterior probability of the new classification is lower,
Figure 4.12: A close look at one deforested pixel in the validation data set. (a) EVI time series from 2003 to 2012. The vertical red line indicates the deforestation event recorded in the reference data set. (b) Class-conditional probabilities output by decision tree classifier. Numbers indicate the IGBP label of the maximum likelihood class. (c) Land cover label predicted by decision tree classifier. (d) Land cover label in stabilized MLCT Collection 5 maps. (e) Land cover label in MRF post-processed maps. Pixel centroid: 54°34′48.1326 W, 11°20′22.4982″ S

and the stabilization algorithm retains the forest label until the high confidence forest label moves out of the three year stabilization window. This effect causes label change in the MLCT Collection 5 maps to lag deforestation events by a couple years. This phenomenon is illustrated for a sample pixel in Figure 4.12.

Table 4.6 shows commission and omission error rates for different sizes of temporal window. A window of one year means that an omission error is a deforestation event that does not coincide with a label change in the following year, and a commission error is a label change that does not coincide with deforestation in the previous year. For a window of less than three years, the Collection 5 maps show high rates of
omission error, due to the stabilization step discussed above. In order to provide a fairer comparison with the Collection 5 stabilization, this analysis uses a window of three years to determine omission and commission error.

To investigate the relationship between error and the MRF temporal smoothing parameter ($\beta_t$) I processed land cover maps of the Amazon using a range of values for $\beta_t$, and a fixed value for $\beta_s$. For each value of $\beta_t$ I measured the omission and commission error, and the number of pixels that changed labels between subsequent years (Figure 4-13). Error rates are also shown for the decision tree output (after prior
Table 4.7: Number of commission errors in deforestation sites. The numbers indicate the count of commission errors observed in 1,050 reference pixels from 2001 to 2012. MRF = Markov Random Field post-processed maps \((\beta_s = 0.25, \beta_t = 1.25)\); C5 = MLCT Collection 5; DT = Decision Tree output.

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>MRF</th>
<th>C5</th>
<th>DT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woody savanna</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Savanna</td>
<td></td>
<td>30</td>
<td>9</td>
<td>46</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td>140</td>
<td>90</td>
<td>172</td>
</tr>
<tr>
<td>Wetland</td>
<td></td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>3</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Evergreen</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>broadleaf forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Savanna</td>
<td></td>
<td>0</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Mixed forest</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Woody savanna</td>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Savanna</td>
<td></td>
<td>6</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Savanna</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td>18</td>
<td>28</td>
<td>44</td>
</tr>
<tr>
<td>Wetland</td>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Woody savanna</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wetland</td>
<td></td>
<td>3</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Savanna</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Grassland</td>
<td></td>
<td>7</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>Wetland</td>
<td></td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Agriculture</td>
<td></td>
<td>40</td>
<td>35</td>
<td>49</td>
</tr>
<tr>
<td>Cropland mosaic</td>
<td></td>
<td>62</td>
<td>77</td>
<td>148</td>
</tr>
</tbody>
</table>
probabilities are applied), and the stabilized MLCT Collection 5 (C5) maps. About
25% of pixels change label from year-to-year in the decision tree output. Both the
MRF and C5 stabilization reduce the year-to-year label change to about 10%. Thus
the MRF model achieves stabilization levels similar to the heuristic stabilization.

The MRF maps and the decision tree output show nearly identical rates of omission
error, and this rate hardly changes as the smoothing parameter is increased.
This means that the MRF processing does not smooth over deforestation change in
the decision tree output, even for very high settings of $\beta_t$. Interestingly, the C5 stabi-
lized maps show slightly lower rates of omission error than the decision tree output.
This difference is not explained by the stabilization heuristic, which would tend to
eliminate change in the maps. A possible explanation is that the C5 maps are post-
processed to correct known errors in the classification (such as the overestimation of
wetlands) (Friedl et al., 2010), and this post-processing may slightly change the error
rate.

The decision tree maps show high rates of commission due to the high level of
year-to-year label change. For low values $\beta_t$ the MRF commission is also high, and
decreases as $\beta_t$ increases. After $\beta_t = 1.25$ the MRF commission error levels off slightly
higher than the C5 error rate. The commission errors are detailed in Table 4.7. The
patterns of error are similar between the MRF maps and the C5 maps, with the error
rate slightly higher in the MRF maps, and much higher in the decision tree maps.

Both the decision tree maps and the MRF maps show several commission errors
due to transition from forest to wetland. The C5 maps do not show these errors,
due to a post-processing step that specifically corrects for over-prediction of wetland
(Friedl et al., 2010). The MRF maps were not treated with this post-processing step,
and a visual comparison of the land cover map shows more wetland area predicted
in the MRF map than in the C5 map. However, even without explicit correction for
wetlands, the MRF eliminates much, although not all, of the wetland commission error present in the decision tree output. These cases are typically due to a relatively high confidence wetland label in one or two years of the time series, which is corrected by temporal context.
Section 5

Discussion

The objective of this thesis was to investigate the causes of year-to-year label change in coarse resolution land cover maps, and to evaluate techniques of reducing spurious “change”. The MLCT product is an excellent example of both the challenges of consistent land cover classification, and the need for improved techniques in this area. The Markov Random Field stabilization method presented here achieved results that are comparable to those achieved by the heuristic stabilization employed in MLCT Collection 5. Although the MRF model does not outperform the heuristic, the development of this model has contributed important insights into the challenges of consistent 500-meter land cover mapping. One of the key challenges is to consistently identify classes that are not well separated in spectral space from pixels that often contain a mixture of land cover types.

As demonstrated in Section 4.1.1, year-to-year label variability is partly caused by low accuracy pixels flipping between similar land cover classes. Most of the error in the MLCT product is caused by confusion between ecologically similar classes, such as savannas and woody savannas, and open shrublands and closed shrublands (Friedl et al., 2010). However, my exploratory analysis of the STEP training data showed that many of these classes include distinct sub-classes at global scale, and that the confusion is not uniform among these sub-classes. My analysis showed that many of the sub-classes are not well separated in spectral space, and that separation is particularly poor among the low-latitude vegetation sub-classes. This lack of sepa-
rability implies that there is an upper limit to the classification accuracy attainable using the available data. To improve classification accuracy, future work should focus on creating classification features that will help separate the difficult sub-classes. A possible strategy might be to compute discriminant functions designed to separate some of the especially problematic cases. For example, Section 4.1.3 identified low-latitude grassland and low-latitude shrubland as difficult sub-classes to separate. A linear or quadratic discriminant function (Hastie et al., 2009) could be computed from the training data for these sub-classes, and this discriminant value could be included as an input feature for the decision tree classifier. The new feature would be a transformation of the existing classification features designed to provide additional discriminatory power for grassland and shrubland pixels.

The presence of sub-classes in the STEP training data also has important implications for parametric statistical models. For example, a multivariate Gaussian distribution estimated from the shrubland training data would likely be a poor approximation of the true distribution. A better approximation would treat the high and low latitude sub-classes separately.

The second part of this thesis investigated two techniques for incorporating context into the classification with the aim of improving the accuracy and stability of the resulting land cover maps. Contrary to my expectation, spatial context seemed to be of limited use at 500-meter scale. Temporal context, however, was demonstrated to improve classification stability both by building the training data set from multiple years of data, and by incorporating context into a random field model. My experiment with multiple years of pooled training data produced land cover maps with higher cross-validation accuracy and lower levels of year-to-year change than maps produced from single years of training data. This implies that a single year of data does not fully capture the variance in land cover spectra, and has implications beyond decision
tree classification. Specifically, it suggests that many statistical models of global land cover might be improved by using several years of pooled data to estimate the model. It is perhaps surprising that including observations from 2010 and 2011 improved accuracy of the 2012 land cover map. However, the phenomenon may be explained by viewing the training data as a sample from a statistical distribution. Each land cover class follows some (unknown) probability distribution, and each year of data represents a different sample from this distribution. Although the samples are not independent, combining multiple samples (years of data) provides a slightly better approximation of the distribution than a single sample alone.

One disadvantage of a larger training data set is that more computation time is required to estimate the decision tree classifier. Also, the complexity of the resulting trees tends to increase as the size of the training set increases, and thus the classification computation time also increases. In practice, such computational disadvantages have to be weighed against the performance gains that come with multi-year training samples. With this tradeoff in mind, a better approach may be to pool multiple years of data for certain classes and not for others. The training data could be constructed by combining multiple years of observation for classes that achieve low accuracies or that are underrepresented in the training data (e.g., closed shrubland and deciduous needleleaf forest), with a single year of observation of classes that achieve high accuracies. Such an approach may provide much of the benefit of a multi-year data set, while reducing the computational burden.

The simple Markov Random Field model presented here was successful at reducing year-to-year label changes, while retaining actual changes. The MRF model produced error rates comparable with, although not superior to, those achieved by the MLCT stabilization algorithm. However, the nature of the MLCT stabilization results in label changes that lag several years behind deforestation events, even when the change
is very clear in the spectral time series. The MRF model was much more successful at capturing change close to the time of deforestation. The energy function evaluated here is a simple indicator that favors consistent labels in adjacent years. This energy function is attractive due to its simplicity. However, better results might be obtained by adjusting the sensitivity of the model based on an analysis of each pixel’s spectral time-series. Incorporating more sophisticated spectral change detection algorithms into the MRF framework is an area worthy of further investigation.

To apply an MRF model at global scale, a robust method of determining the model parameters will be required. This thesis investigated the effects and sensitivity of the spatial and temporal parameters individually, but did not address the joint optimization of these parameters. Many methods have been proposed for determining optimal parameter settings using a labeled training data set (Li, 2009). However, it seems unlikely that a single set of parameters can fully capture land cover variation and change at global scale. A possible strategy for the spatial smoothing parameter ($\beta_s$) is to vary the parameter for different regions of the globe. The parameter could be scaled by the average landscape homogeneity measured within a spatial window, such as shown in Figure 4-9. In this work, the optimal labeling for the MRF was determined using the Iterated Conditional Modes algorithm. While this algorithm is simple and computationally efficient, it is sensitive to initialization and is not guaranteed to find the global optimum to the minimization problem. Future work should evaluate the potential for more sophisticated optimization algorithms to improve the MRF solution.

Most of the benefit of the Markov Random Field model in this study appears to be due to the use of temporal, rather than spatial, context. Based on analysis of the NLCD data set, spatial homogeneity of land cover at 500-meter scale varies greatly across the North American continent (Figure 4-9). This suggests that the spatial
smoothing provided by a MRF model will be more effective for some landscapes than others, and that the model parameters must take this spatial variation into account. For very heterogeneous landscapes (e.g., the Southeastern United States), spatial context at 500-meter scale may provide little, if any, benefit. In heterogeneous landscapes, a model that deals only with the temporal dimension may perform as well as a spatio-temporal model. For example, the Markov Random Field model proposed in this thesis could be adapted to a one dimensional Markov Chain that captures label change through time without considering spatial neighborhoods. Such a model would be conceptually and computationally simpler than a random field, while retaining the most successful elements of the spatio-temporal model. Future work in this area should attempt to determine the characteristics of landscapes in which the benefit of spatial context outweighs the disadvantages of increased model complexity.
Section 6

Conclusions

This thesis has explored methods to improve the accuracy of coarse resolution land cover classification, with a focus on reducing spurious variation in annual land cover maps. Variation in year-to-year maps is caused by a variety of factors, including poor spectral separability between land cover classes and mixed pixels at coarse spatial resolution. The use of temporal context was demonstrated to be useful for reducing the amount of year-to-year variation both by using multiple years of observation to construct the training set for a per-pixel classifier, and by incorporating context using a Markov Random Field.

The Markov Random Field model presented here successfully reduced the amount of spurious change in the MLCT maps, while retaining real changes due to deforestation. While the MRF model did not outperform the MLCT stabilization heuristic, the MRF was more successful at “changing” land cover labels close to the time of deforestation, whereas label change in the heuristically stabilized maps tends to lag behind the deforestation event. In addition, the MRF model provides a flexible framework for including ancillary information into the classification task. However, the composition of landscapes at 500-meter scale varies greatly, and some landscapes, such as the Southeastern United States, are very heterogeneous at this scale. In heterogenous landscapes the spatial smoothing provided by an MRF model is of limited use, and a model that considers only the temporal dimension may be a more appropriate choice for label stabilization.
Accurate and temporally consistent land cover maps are critically important to our ability to understand our changing planet. This thesis contributes to our understanding of global land cover classes, the spectral relationship between classes, and the underlying causes of year-to-year variation in land cover maps. In addition, this work provides insight into the benefits and limitations of spatial and temporal context for coarse resolution land cover mapping.
Appendix A

IGBP Land Cover Sub-class Plots

This appendix presents sub-classes identified within the IGBP land cover classes. Sub-classes were identified by clustering the training pixels in the STEP training data set. The data points in each IGBP class are visualized in three ways. First, the points are presented as a scatter plot in principal component space. Second, the locations of the pixels are plotted on a world map. Finally, each sub-class is visualized as a spectral time series measured at the cluster centroid. The methods used to identify and visualize sub-classes are described in detail in Section 3.3.2.
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A.1: Evergreen needleleaf forest (class 1)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A.2: Evergreen broadleaf forest (class 2)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A·3: Deciduous needleleaf forest (class 3)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-4: Deciduous broadleaf forest (class 4)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-5: Mixed forest (class 5)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-6: Closed shrubland (class 6)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-7: Open shrubland (class 7)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-8: Woody savanna (class 8)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-9: Savanna (class 9)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-10: Grassland (class 10)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-11: Permanent wetland (class 11)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A.12: Cropland (class 12)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A-13: Cropland/natural vegetation mosaics (class 14)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A·14: Snow/ice (class 15)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A.15: Barren/sparsely vegetated (class 16)
(a) Left: Principal components. Right: Geographic location of training sites

(b) Time series plots of cluster centroids. Thin lines indicate +/- standard error of the mean.

Figure A·16: Water (class 17)
References


Parker Abercrombie

parker@pabercrombie.com
+1.805.455-6871

SKILLS

Programming: Java, Python, R, C/C++, SQL, C#, .NET, OpenGL, and COM.

GIS: Experience with ESRI ArcGIS, ENVI, and NASA World Wind.

Databases: MySQL, PostgreSQL, SQLite, and Hypersonic.

Mobile platforms: Experience writing software for Android and Windows Mobile.

Web technologies: HTML, PHP, XML, XSD, and JavaScript.

EDUCATION

Master of Arts, Geography 1/14
Boston University
Course work: remote sensing • data mining • machine learning • statistics

Bachelor of Science, Creative Studies, emphasis in Computer Science 6/03
University of California, Santa Barbara
Course work: computer science • mathematics • environmental studies

PUBLICATIONS


EXPERIENCE

Research Assistant 9/12–present
Land Cover and Surface Climate Group, Boston University
Analyze and process satellite imagery for global land cover mapping.
• Develop software for processing large data sets on a computing cluster.
• Evaluate methods of improving accuracy of land cover classification.
• Tools: R, Python.

Software Engineer, NASA World Wind 9/10–9/12
Development of new features for the World Wind open source 3D GIS engine.
• Designed a system to render graphics defined by MIL-STD-2525 symbol set.
• Ported World Wind Java code to the Android platform.
• Added support for visualizing KML documents in World Wind.
• Integrated Internet Explorer to render HTML information balloons.
• Tools: Java, C++, OpenGL, COM

Software Engineer, The Okori Group 1/08–6/12
Help early stage start-up companies turn ideas into technology products.
• Developed a prototype open source voting machine for the Open Voting Consortium.
• Developed a plug-in for Internet Explorer that allows a security application to selectively redirect web requests to a proxy server.
• Tools: Python, C#, C++, COM
GIS Technician (Volunteer), Channel Islands National Park 6/11–6/12
Assist with collection and processing of geospatial data.
  • Development of Python scripts to automate data management in ArcGIS.

Software Engineer, Department of Energy 6/08–10/11
Special Technologies Laboratory
Development of a sensor command and control solution built on the NASA World Wind 3D GIS engine. The product allowed the DOE to collect data from remote sensors and display relevant information in a 3D user interface.
  • Designed a system for dynamically loading plug-ins packaged as Java JAR files.
  • Designed APIs to allow plug-ins to interact with the main application.
  • Tools: Java, Maven, Ant, Hibernate, OSGi, PostgreSQL

Co-Founder/Software Engineer, Zentopy 7/07–1/10
Started a software company with two partners. Developed a network storage solution to allow users to backup documents to storage space in the cloud using an interface built into Windows Explorer. The software automatically detected changes and synchronized with the remote server.
  • Designed and implemented a networked file storage system with intelligent client side caching.
  • Designed and implemented a network protocol to efficiently transfer partial files to a server.
  • Tools: C++, Amazon EC2 and S3, SQLite

Software Engineer, ZimRide.com 7/07–6/08
Developed and managed a Facebook application (Carpool). The application allows users to search for shared rides between cities.
  • Brought a new product to market in a start-up environment.
  • Supported a production system with a growing user base.
  • Tuned SQL queries to execute user searches efficiently.
  • Designed a flexible infrastructure to allow Carpool to grow beyond Facebook and be deployed on many social networking platforms using one code base.
  • Tools: PHP, MySQL, Facebook API, Google Maps API

Desktop Publishing and IT Technician, Oasis Design Press 6/04-6/06
  • Prepared books for publication. Layout using Adobe InDesign, graphics production using Photoshop and Illustrator.
  • Developed web scripts with Perl and PHP.

Software Engineer, EnvEnergy 11/03-1/04
  • Built a Java web application for building automation from the ground up using Struts and Hibernate.

PRESENTATIONS