2019

Enhancing Landsat time series through multi-sensor fusion and integration of meteorological data

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

**ENHANCING LANDSAT TIME SERIES THROUGH
MULTI-SENSOR FUSION AND INTEGRATION OF
METEOROLOGICAL DATA**

by

CHRISTOPHER E. HOLDEN

B.A., Boston University, 2011
M.A., Boston University, 2011

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2019

Approved by

First Reader

_____

Curtis E. Woodcock, PhD
Professor of Earth & Environment

Second Reader

_____

Mark Friedl
Professor of Earth & Environment

Third Reader

_____

Lucy Hutyra
Associate Professor of Earth & Environment

Let us not, however, flatter ourselves overmuch on account of our human victories over nature. For each such victory nature takes its revenge on us. Each victory, it is true, in the first place brings about the results we expected, but in the second and third places it has quite different, unforeseen effects which only too often cancel the first.

— Friedrich Engels, Dialectics of Nature (1883)

Acknowledgments

My time at Boston University working on my dissertation and other research has been benefited enormously by the many wonderful people in the community. First, I would like to thank my advisor, Curtis Woodcock, for his continued support and help from my first class in remote sensing during my undergraduate degree, through conferences and publications, to the work in this dissertation. I would also like to thank the faculty members who have taught and introduced me to so much research, especially my committee members Lucy Hutyra and Mark Friedl, and Suchi Gopal who originally led me down this path when she hired me as a research assistant. I have had the pleasure of working with a number of amazing scientists from beyond the department, including committee member Tom Loveland and his colleagues at the USGS EROS and the Landsat Science Team, and committee member Josef Kellndorfer who I have almost exclusively to thank for my knowledge of radar data. I'm also very grateful to Pontus Olofsson for being the chair of my committee and for his advice and help along the way.

I would also like to thank all the students, staff, and faculty of the Earth and Environment department and the Biogeoscience Program. It's been fantastic working with researchers across so many fields. In particular I would like to thank members of my research group who have helped me, especially Valerie Pasquarella, Eric Bullock, Paulo Arevalo Orduz, Damien Sulla-Menasche, Zhe Zhu, and Xiaojing Tang. I am also grateful for the advice and companionship of officemates and friends across these years, including Eli Melaas, Conor Gately, Margaret Hendrick, Ha Nguyen, Kira Sullivan-Wiley, Dan Short Gianotti, Anne Short Gianotti, Jon Wang, Drew Trlica, Marta Ribera, and Doug Bolton. Two of my chapters would not be possible without the hard work of sample interpreters Xianfei Shen and Yihao Liu. I also would like to thank the staff of Earth and Environment, especially Alissa Beideck and Frederick

George III, for their support and for helping me navigate the university procedures and deadlines. I am also very lucky in the past few years to have found comrades in graduate union organizing and beyond BU who have taught and supported me so much.

I wouldn't be anywhere without my family, especially my mother, Jill McCafferty, for her sacrifice and support, my brother Andrew for his friendship and advice, and my aunts, uncle, and cousins for their love and support. Last but most importantly, I want to thank my partner Diana Sibbald for her patience, love, and advice throughout this entire process.

ENHANCING LANDSAT TIME SERIES THROUGH MULTI-SENSOR FUSION AND INTEGRATION OF METEOROLOGICAL DATA

CHRISTOPHER E. HOLDEN

Boston University, Graduate School of Arts and Sciences, 2019

Major Professor: Curtis E. Woodcock, PhD
Professor of Earth & Environment

ABSTRACT

Over 50 years ago, the United States Secretary of the Interior, Stewart Udall, directed space agencies to gather "facts about the natural resources of the earth." Today global climate change and human modification make earth observations from all variety of sensors essential to understand and adapt to environmental change. The Landsat program has been an invaluable source for understanding the history of the land surface, with consistent observations from the Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors since 1982. This dissertation develops and explores methods for enhancing the TM/ETM+ record by fusing other data sources, specifically, Landsat 8 for future continuity, radar data for tropical forest monitoring, and meteorological data for semi-arid vegetation dynamics.

Landsat 8 data may be incorporated into existing time series of Landsat 4-7 data for applications like change detection, but vegetation trend analysis requires calibration, especially when using the near-infrared band. The improvements in radiometric quality and cloud masking provided by Landsat 8 data reduce noise compared to previous sensors.

Tropical forests are notoriously difficult to monitor with Landsat alone because of clouds. This dissertation developed and compared two approaches for fusing Synthetic Aperture Radar (SAR) data from the Advanced Land Observation Satellite (ALOS-1) with Landsat in Peru, and found that radar data increased accuracy of deforestation. Simulations indicate that the benefit of using radar data increased with higher cloud cover.

Time series analysis of vegetation indices from Landsat in semi-arid environments is complicated by the response of vegetation to high variability in timing and amount of precipitation. We found that quantifying dynamics in precipitation and drought index data improved land cover change detection performance compared to more traditional harmonic modeling for grasslands and shrublands in California.

This dissertation enhances the value of Landsat data by combining it with other data sources, including other optical sensors, SAR data, and meteorological data. The methods developed here show the potential for data fusion and are especially important in light of recent and upcoming missions, like Sentinel-1, Sentinel-2, and NASA-ISRO Synthetic Aperture Radar (NISAR).

Contents

List of Tables

List of Figures

xvii

xxv

List of Abbreviations

ALOS	......	Advanced Land Observation Satellite
ARD	......	Analysis Ready Data
AVHRR	......	Advanced Very High Resolution Radiometer
BFAST	......	Breaks For Additive Season and Trend
BIC	......	Bayesian Information Criterion
CCDC	......	Continuous Change Detection and Classification
CDR	......	Climate Data Record
EVI	......	Enhanced Vegetation Index
EWMACD	......	Exponentially Weighted Moving Average Change Detection
FWHM	......	Full-Width at Half Maximum
HH	......	Horizontal-horizontal polarization
HLS	......	Harmonized Landsat-8 Sentinel-2
HV	......	Horizontal-vertical polarization
JERS	......	Japanese Earth Resources Satellite
LCMAP	......	Land Change Monitoring, Analysis, and Projection
MODIS	......	Moderate Resolution Imaging Spectroradiometer
MSS	......	Multispectral Scanner
NAIP	......	National Agriculture Imagery Program
NBR	......	Normalized Burn Ratio
NBR2	......	Normalized Burn Ratio 2
NDVI	......	Normalized Difference Vegetation Index
NIR	......	Near Infrared
NLCD	......	National Land Cover Database
PALSAR	......	Phase Array type L-band Synthetic Aperture Radar
PDF	......	Probability Distribution Function
PRISM	......	Parameter elevation Regression on Independent Slopes Model
RMSE	......	Root Mean Squared Error
RTC	......	Radiometrically Terrain Corrected
SAR	......	Synthetic Aperture Radar
SPEI	......	Standardized Precipitation-Evapotranspiration Index

SPI	. . . . . .	Standardized Precipitation Index
STARFM	. . . . . .	Spatial and Temporal Adaptive Reflectance Fusion Model
SVM	. . . . . .	Support Vector Machine
SWIR	. . . . . .	Shortwave Infrared
USDA	. . . . . .	United States Department of Agriculture
USGS	. . . . . .	United States Geological Survey
UTM	. . . . . .	Universal Transverse Mercator
VCT	. . . . . .	Vegetation Continuous Tracker
VH	. . . . . .	Vertical-horizontal polarization
VV	. . . . . .	Vertical-vertical polarization
WRS-2	. . . . . .	World Reference System-2

Chapter 1

Introduction

1.1 Background

The Landsat Program, spanning over 40 years, provides the longest and most informative satellite record of our changing planet (Loveland and Dwyer, 2012; Roy et al., 2014). After spending most of its lifespan operating with a pay-for-access model, the Landsat archive was finally made freely available in 2008 and methods that take advantage of this data policy quickly proliferated (Wulder et al., 2012). These methods that work on time series of Landsat data have been used extensively to map forest cover change (Kennedy et al., 2007; Asner et al., 2009; Kennedy et al., 2010; Huang et al., 2010; Vogelmann et al., 2012; Zhu, Woodcock, and Olofsson, 2012; Verbesselt et al., 2012; Zhu and Woodcock, 2014; Brooks et al., 2014; Hansen et al., 2014), including a study that mapped global forest cover change between 2000 and 2012 at Landsat's 30 m resolution (Hansen et al., 2013). This critical information on the magnitude and spatial patterns of forest change has also been extended to 1990 (Kim et al., 2014) using Landsat data. While many recent Landsat time series applications have focused on identification of forest cover change, some methods have attempted to identify changes in all land covers (Zhu and Woodcock, 2014) or classify specific agricultural crops (Yan and Roy, 2015). Landsat time series also provide unique traits suitable for combination with lidar data for estimation of forest biomass as disturbance information from Landsat MSS through ETM+ have improved estimates of forest biomass (Pflugmacher et al., 2012). Consistent and continuous Landsat obser-

vations enable the transition from static snapshots of ecosystem properties to characterization of processes and dynamics of interest to ecologists (Kennedy et al., 2014), including phenology (Melaas et al., 2013) and insect infestations (Meigs et al., 2011). Refined information about dynamics within human systems is now also a target for Landsat based research because Landsat's long time record, locally relevant spatial scale (Hansen et al., 2013), and recent advances in mapping frequency enables the analysis of change in land cover within the context of socioeconomic or policy drivers. Such efforts include the monitoring of forest cover change to address policy questions in regions (Kennedy et al., 2012; Griffiths et al., 2012) or nationally using wall to wall maps (Hansen et al., 2014) and within a sampling framework (Masek et al., 2013), and comparing rates of urbanization across rapidly growing cities (Schneider, 2012).

1.2 Survey of Remote Sensing Time Series Change Detection Methods

The classic two-date direct classification of change approach for monitoring of land cover change infers land cover change based on the difference between the two images. Time series methods are an extension of this idea in that they compare observations through time against one another, but there are generally many more observations and the comparisons can also incorporate non-discrete changes, such as trends in spectral reflectance. Kennedy et al. (2007) was one of the first studies to perform time series analysis of relatively large stacks of multi-date time series data from Landsat, but more importantly they introduced the conceptual model of understanding land surface processes as trajectories and not discrete events that take place between images. A change in a riverbank from land to water over ten years is a discrete event in a two date change map, but ten years of spectral reflectance trajectories might relate to geomorphological processes. Kennedy et al. (2010) generalized the idea of mapping

against idealized trajectories to fitting "segments" that describe processes occurring between two points in the time series with the development of the LandTrendr algorithm and the associated interpretation tool TimeSync (Cohen et al., 2010). A segment might represent a period of forest regrowth, and the trend over time of the shortwave infrared band might be used as the basis of classification of segments as regrowth trajectories. This conceptual model of time series analysis in remote sensing is very useful and has seen adoption, especially for parts of the community using Landsat time series analysis for land monitoring. Algorithms like Breaks For Additive Season and Trend Monitor (BFAST Monitor; Verbesselt et al., 2012) or Continuous Classification and Change Detection (CCDC; Zhu and Woodcock, 2014) can produce results that can be described using similar terms, and the applications of the algorithms usually rely on interpreting segments. Kennedy et al. (2014) and Pasquarella, Holden, Kaufman, et al. (2016) more explicitly link segments estimated using time series of Landsat spectral reflectance to landscape and ecosystem processes and conditions.

While other studies have attempted to summarize and categorize the entire gamut of time series approaches used by the community (Hansen and Loveland, 2012; Zhu, 2017), many components in this dissertation are either directly related to or inspired by three main types of time series change detection algorithms: "offline" segment fitting, "online" anomaly monitoring, and "class membership" monitoring.

"Offline" segment fitting refers to approaches that analyze the entire time series at once and attempt to decide how many segments should be fit to best describe the time series without overfitting. Examples include LandTrendr (Kennedy et al., 2010), Break Detection for Additive Season and Trend (BFAST; Verbesselt, Hyndman, Zeileis, et al., 2010; Verbesselt, Hyndman, Newnham, et al., 2010).

"Online" residual monitoring, by contrast, iterates through the time series and

makes judgments about fitting segments based on new and past observations. A very simple online monitoring algorithm is the Cumulative Sum (CUSUM; Page, 1954) test run on the recursive residuals from a time series regression model (Ploberger and Krämer, 1992). The Vegetation Continuous Tracker (VCT; Huang et al., 2010) algorithm, which finds change by looking for ununsually large cumulated z-scores, is another example of an "online" change detection method. Break Detection for Additive Season and Trend Monitor (BFAST Monitor; Verbesselt et al., 2012) is based on the MOSUM test on recursive residuals. Exponentially Weighted Moving Average Change Detection (EWMACD; Brooks et al., 2014) is based on the EWMA test on recursive residuals. Continuous Change Detection and Classification (CCDC; Zhu and Woodcock, 2014) is based on a custom approach of monitoring for consecutive, abnormally large residuals, which has some basis in the literature for EWMA.

Although not as related, methods that rely heavily on time series of class membership are useful to describe within the framework of segments. As described previously, the VCT algorithm is an "online" change detection algorithm that monitors z-scores, specifically z-scores based on the mean and standard deviation of examples of forest land cover. The use of class probabilities or similar metrics is also the basis of the Hansen et al. (2013) approach to mapping global forest cover change. Hansen et al. (2013) use decision trees to generate predictions of tree cover on yearly time scales based on training data iteratively collected by interpreters. Bayesian updating has been used to monitor for changes in time series of forest probabilities using a threshold approach (Reiche, de Bruin, et al., 2015; Reiche et al., 2018), and Hidden Markov Models have been employed similarly in situations when there is knowledge of the likelihood of land cover transitions among classes (Solberg et al., 2008; Trier and Salberg, 2011; Salberg and Trier, 2011).

Common limitations and strengths

For classic two date change detection, the two dates of imagery do not need to be radiometrically matched or atmospherically corrected (Song et al., 2001) because differences in measurement due to the atmosphere can be accounted for as simple differences in stable classes. For example, if the NDVI for two dates of a perfectly stable forest are 0.6 and 0.5, the 0.1 difference might be attributable to atmospheric interference. A deforested example, however, would have NDVI differences much larger. By comparison, time series methodologies require either relative correction, as the trajectory based monitoring precursor to LandTrendr (Kennedy et al., 2007) and Hansen et al. (2013) used, or a physically based atmospheric correction routine to normalize for atmospheric influences to make observations directly comparable (Song et al., 2001). This limitation is not as large as it once was due to fairly robust algorithms for surface reflectance correction applied as standardized products (Masek et al., 2008).

Another common limitation for change detection methods is image availability. Two-date change detection approaches could work with several years of gaps in image availability, assuming that whatever change process occurred within the gap years did not yield reflectance that look exactly like the previous land cover. Methods such as LandTrendr, VCT, CCDC, and BFAST Monitor generally require at least one observation each year, with LandTrendr and VCT requiring this observation to be within an "anniversary date" window. In situations with missing observations, these values could be estimated using smoothing or interpolation, but the accuracy of the change detection might suffer. Because they fit complicated models and require many observations of change, CCDC and BFAST Monitor are especially sensitive to missing observations. Time series methods will not work in all places all the time if observational frequency is too low.

The basis for time series methods is the comparison of observations against themselves, which makes it particularly challenging when there is no basis for comparison as in the start and the end of the time series. Algorithms like CCDC and BFAST Monitor use many coefficients to predict reflectance, and so the initialization period, or the "historic" period, is highly affected by noise and there may be false positives detected if the initial models are fit poorly. Likewise, LandTrendr has difficulty with the start and end of the time series since it compares any given observation against the previous and future values (Kennedy et al., 2010). The VCT may not have such a difficult time understanding the beginning and end of the time series because it uses derived values (i.e., a forest z-score), but it still has difficulty in assessing change due to the usage of consecutive derived values.

One of the largest strengths of time series analysis for change detection is the ability to understand the process by which two dates of reflectances may differ. In the case of a regrowing forest, an image acquired in 2000 right after deforestation might look like a barren area or perhaps a shrubland. By 2010, the forest has matured and will probably look much more like a forest. If two date change detection classification were performed in this scenario, one might infer that there was some very quick landscaping that occurred which sprouted trees magically by 2010. A time series approach would be able to capture the real signal of a slowly regrowing forest and infer that there was no sudden land cover change which grew a forest, but that there was some measurable rate of growth. The trend through time in NDVI, for example, might be related to the accumulation of leaf area which could yield additional information about the dynamics of the system.

1.3 Structure of this Dissertation

My dissertation research explores and provides new methods of analysis for difficult problems in the land cover and land use monitoring community while attempting to confront the ever growing volume of earth observation data available to investigators. The change in policy that made the Landsat archive free and openly available (Wulder et al., 2012) has fundamentally enabled all of this work, and each chapter explores how to incorporate new information with this Landsat historical archive of TM and ETM+ data.

1.3.1 An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations

The Landsat Program's latest satellite, Landsat 8, was given the name the "Landsat Data Continuity Mission" during its development because of the recognition of the importance for data continuity. Being the first Landsat satellite launched in almost fifteen years, Landsat 8 had significant technological improvements over previous sensors, including narrowing some of the wavelengths of spectral bands, radiometric or geometric improvements from switching to a pushbroom sensor, and the addition of a new spectral band specifically designed for detecting aerosols. In this chapter, I use data from special Landsat 7 and Landsat 8 "underflight" orbits and almost two years of Landsat 8 time series data to quantify the benefits to time series analysis from these sensor improvements and answer if Landsat 8 can be seamlessly integrated into existing time series of Landsat data.

1.3.2 Forest Change Detection By Radar/Optical Fusion

The Landsat data archive is invaluable and has been applied for global applications including forest monitoring since 2000 (Hansen et al., 2013) or since 1990 (Kim et al., 2014), but the historical archive is far from globally consistent (Wulder et al., 2016).

There are places and time periods where data intensive time series analysis from Lansat alone will not be possible because there simply aren't enough observations historically. While observation density from optical data might not be as much of a problem with two active Landsat and Sentinel-2 sensors, high frequency data is important for near real time monitoring of the land surface and some places on Earth have near persistent cloud cover for months at a time. Fusion of radar data sources has been offered as a possible solution to these issues (Joshi et al., 2016; Reiche et al., 2016), but few studies have evaluated the benefit of the fusion (Joshi et al., 2016). This study attempts to fill that gap by developing two fusion algorithms and assessing the difference in performance when fusing optical and radar data for change monitoring.

1.3.3 Landsat Time Series Meteorological Data Fusion In Semi-Arid Ecosystems

Vegetation phenology and other dynamics related to abundance and variability are very important biogeochemical processes because leaves are the primary exchange surfaces for fluxes of carbon and water. While various communities have studied these questions that relate to the timing and variability of greenness, frequently by incorporating data describing environmental drivers of plant ecophysiology, the land cover change community has typically focused on mapping abrupt changes over more gradual or subtle variability in recent time series investigations. Indeed, many algorithms have developed routines to reduce the impact of year to year variability associated with changes to precipitation or heat. This final study of my dissertation attempts to bridge this divide by trying to determine if incorporating meteorology into time series change detection methods improves change detection performance in the semi-arid grasslands, oak savannas, shrublands, and coniferous forests of the Central Valley in California. I test to see if change detection performance, either

commission or omission, can be improved by using precipitation and drought index data, and analyze the estimated relationships between precipitation and vegetation greenness.

1.3.4 Summary

This dissertation provides new methods of integrating data from multiple sensors in order to better monitor Earth's natural resources, including how climate variability and direct human action enhances or degrades these resources. Monitoring of the land surface continues to grow in importance, and this dissertation offers new methodologies and advice for meeting those issues that take advantage of the simultaneously increasing data availability and computational power.

Chapter 2

An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations

2.1 Introduction

Technological advances since the development of Landsat 7 and the move from a whiskbroom to a pushbroom style sensor have greatly increased the measurement accuracy of Landsat 8 data. The signal to noise ratio (SNR) on Landsat 8's OLI is much higher than the mission specifications and an order of magnitude higher than the SNR on Landsat 7 (Knight and Kvaran, 2014; Morfitt et al., 2015) while the increase to 12-bit radiometric resolution extends the range of measurable radiances and reduces pixel saturation (Morfitt et al., 2015). The geometric accuracy on Landsat 8 is astounding and meets or exceeds all mission requirements (Storey et al., 2014). The absolute geolocation accuracy of Landsat 8, in particular, improves so much on the mission specifications that it will be used to improve the geometric accuracy of the entire Landsat archive by improving ground control points in the Global Land Survey (GLS) network. Improvements to onboard storage and ground transmission capacity enable Landsat 8 to acquire 60% more image scenes per day than Landsat 7 as of 2014 (Roy et al., 2014) and currently approximately 725 scenes are acquired per day. This increased image acquisition frequency will help further reduce differences in image availability due to geographic variation in observation conditions and

ground receiving station access (Kovalskyy and Roy, 2013). Landsat 8 continues the legacy of the Landsat program by providing much higher quality and quantity Earth observations vital to time series investigations.

Designed to continue this legacy, the spectral band specifications for Landsat 8's Operational Land Imager (OLI) sensor are, for the most part, very similar to previous sensors (see Table 2.1). Several spectral bands have been narrowed slightly to provide more precise measurements that are less correlated with other bands. The near infrared band on Landsat 8 (OLI Band 5) changed the most compared to other bands as it was narrowed substantially from 0.77 µm to 0.90 µm on ETM+ to 0.85 µm to 0.88 µm specifically to avoid an atmospheric water absorption feature at 0.85 µm. As such, we might expect near infrared measurements from Landsat 8 to be brighter than from previous Landsat sensors. The first shortwave infrared band (ETM+ band 5 and OLI band 6) also narrowed significantly and the second shortwave infrared band (ETM+ band 7 and OLI band 7), while it did not change as much, now excludes wavelengths (>2.3 µm) with the strongest spectral response function in ETM+ band 7 (Flood, 2014).

The changes in spectral bandpasses are an improvement over previous sensors, but one that may affect the comparability with previous sensors. Li et al. (2013) analyzed Landsat 7 and Landsat 8 surface reflectance and vegetation indices in four land cover types. They found differences in spectral bands between the two sensors that contributed to further differences in vegetation indices, especially when using indices calculated using the visible bands. Zhu, Fu, Woodcock, Olofsson, Vogelmann, Holden, et al. (2016) analyzed land cover change and dynamics within the rapidly urbanizing city of Ghuangzhou, China and found the NDVI to be positively biased when adding Landsat 8 into time series of Landsat 4, 5, and 7. Flood (2014) used 793 cloud free 8-day pairs of images over 123 Landsat scenes in Australia to assess the

difference between Landsat 8 and Landsat 7 top of atmosphere (TOA) and surface reflectance, including how the differences propagate in biophysical models. Differences in TOA reflectance were close to 1% for all bands except the near infrared and second shortwave infrared which were brighter by 6% and 3%, respectively. Surface reflectance correction using the methods of Flood et al. (2013) reduced these differences to within 1 to 2% for all bands except the blue band (ETM+ band 1 and OLI band 2). Unadjusted surface reflectance inputs to a fractional cover model created biased estimates relative to Landsat 7 with bare cover underestimated by 5% and vegetated cover overestimated by 7%. The differences in surface reflectance estimates were systematic and adjustable using a simple linear regression. Landsat 8 adjusted surface reflectance values produced fractional cover estimates equal to estimates from Landsat 7. Flood (2014), however, also noted that there was some variance across the 123 scenes in the relationship between Landsat 7 and Landsat 8 surface reflectances and that the corrections performed are not likely to be extendable beyond the study area. Mishra et al. (2014) used Landsat 7 and Landsat 8 data in combination with simultaneous or near simultaneous EO-1 Hyperion acquisitions to assess differences in cross-sensor calibration. EO-1 Hyperion was used to develop Spectral Band Adjustment Factors (SBAF) that corrected for the measurement differences due to different spectral bandpasses. Once the SBAF were applied to Landsat 8 data, TOA reflectance differences in a near simultaneously acquired Landsat 7 and Landsat 8 image pair were less than 2% for all analogous bands except for a 4% difference in the near infrared.

Landsat 8 includes a new band in the mid-infrared centered on 1.375 µm that is designed specifically to find cirrus clouds in Landsat OLI imagery (Loveland and Dwyer, 2012; Irons et al., 2012; Roy et al., 2014). Successfully used on the Moderate Resolution Imaging Spectrometer (MODIS) instrument, this band significantly improves

upon alternative methods of detecting cirrus clouds by at least an order of magnitude (Gao and Kaufman, 1995). Without such a tool, cirrus cloud detection in historical Landsat imagery was significantly more difficult. For example, an analysis of all Landsat images acquired during the first year over the conterminous United States found that 7% of the data were identified as low confidence cloud but high confidence cirrus clouds, suggesting that 7% of historic Landsat data may be contaminated with cirrus clouds (Kovalskyy and Roy, 2015). While the percent of undetected cirrus clouds may vary across the globe, this result highlights another difference between Landsat 8 and previous sensors that may be important when combining datasets.

2.1.1 Objectives

As previous studies have indicated (Li et al., 2013; Flood, 2014; Mishra et al., 2014), the change in Landsat 8's spectral bandpasses create differences in radiance and reflectance measurements relative to Landsat 7. While these studies have provided excellent evidence on the differences between sensor, they have not assessed sensor-specific differences in the Landsat Climate Data Record (CDR) community product. The Landsat CDR product is designed to improve on the Level 1 Product usability by providing atmospherically corrected Landsat data and uses different algorithms for Landsat TM/ETM+ and Landsat OLI data. Because any potential differences in the correction algorithms might enhance target specific differences due to spectral bandpasses, it is important to assess the relative consistency of the CDR products across sensors. Finally, it is still unclear how differences between Landsat 8 and previous sensors are manifested in time series analyses.

To assess the potential differences between Landsat 7 and Landsat 8 characteristics in the Landsat CDR product, we posed the following questions:

1. Do Landsat 7 and Landsat 8 spectral reflectances from CDR differ significantly?

If so, what is the nature of the differences?

2. Are time series of Landsat 7 and Landsat 8 data different? What is the effect of combining data from the two sensors?

3. What are the spectral characteristics of cloud and cirrus clouds masked by Landsat 8 but omitted in Landsat 7 data?

4. Are the cloud masking improvements in Landsat 8 apparent in time series?

2.2   Datasets

Atmospheric correction of Landsat 7 and Landsat 8 data within the Landsat CDR is performed using two different algorithms driven by different input ancillary data. Landsat 7 data are corrected using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006) while Landsat 8 data are corrected using "L8SR", a newly developed algorithm that takes advantage of some of Landsat 8's new sensor characteristics (U.S. Geological Survey, 2015). The LEDAPS algorithm retrieves surface pressure, water vapor, and air temperature inputs from the NOAA National Centers for Environmental Protection (NCEP) reanalysis dataset while L8SR calculates surface pressure based on the target's elevation and uses the Moderate Resolution Imaging Spectroradiometer (MODIS) Climate Modeling Grid - Aerosol (CMA) product for water vapor and air temperature estimates. Ozone data are retrieved from the Ozone Monitoring Instrument (OMI) or Total Ozone Mapping Spectrometer (TOMS) for LEDAPS and from the MODIS Climate Modeling Grid (CMG) coarse resolution ozone product in L8SR. Aerosol Optical Thickness (AOT) is estimated in LEDAPS using the dark, dense vegetation (DDV) method (Kaufman et al., 1997) while L8SR uses the MODIS CMA product. LEDAPS uses view zenith angle geometry from the image metadata while L8SR hard codes this angle to 0 degrees.

These differences in ancillary data may be another cause for potential disagreement in the surface reflectance estimates, but diagnosing the effect of different ancillary information on the output reflectance estimates is outside of the scope of this effort.

The L8SR data products are currently considered "provisional" because the algorithm, the implementation of the algorithm in software, and the subsequent output products have not been completely validated (U.S. Geological Survey, 2015). The algorithm implementation has been corrected to fix artifacts near coastal land and water boundaries, but visual artifacts remain among cloud edges and areas of high topographic variation in some images. The correction of these artifacts should not significantly alter the findings of this work that is based on the analysis of millions of pixels across many scenes. We refer the reader to the product changelog in the Landsat 8 CDR product guide (U.S. Geological Survey, 2015) to monitor changes to the product. In addition, the L8SR algorithm has a few caveats: the algorithm is not run on scenes with a solar zenith angle greater than 76 degrees and surface reflectance retrievals may be uncertain in hyper-arid, snow-covered, coastal regions with small amounts of land relative to water, or high cloud cover environments (U.S. Geological Survey, 2015).

We used two datasets to address the questions posed about the continuity of surface reflectance observations from Landsat 7 to Landsat 8: Landsat "underflight" data that provides near simultaneous observations from both sensors, and time series of Landsat data covering seven years of Landsat 7 and two years of Landsat 8.

Band	Landsat TM (µm)	Landsat ETM+ (µm)	Landsat OLI (µm)
Blue	Band 1: 0.45-0.52	Band 1: 0.45-0.52	Band 2: 0.45-0.51
Green	Band 2: 0.52-0.60	Band 2: 0.52-0.60	Band 3: 0.53-0.59
Red	Band 3: 0.63-0.69	Band 3: 0.63-0.69	Band 4: 0.64-0.67
NIR	Band 4: 0.76-0.90	Band 4: 0.77-0.90	Band 5: 0.85-0.88
SWIR1	Band 5: 1.55-1.75	Band 5: 1.55-1.75	Band 6: 1.57-1.65
SWIR 2	Band 7: 2.08-2.35	Band 7: 2.09-2.35	Band 7: 2.11-2.29
Thermal	Band 6: 10.4-12.5	Band 6: 10.4-12.5	Band 10: 10.6 - 11.19

Table 2.1: Spectral band specifications for Landsat sensors TM, ETM+, and OLI

2.2.1 Landsat 8 Underflight

On March 29-30, 2013, prior to the establishment of its final orbit, Landsat 8 flew in "underflight" mode in position with Landsat 7 allowing for almost simultaneous image acquisitions. The scene centers for Landsat 8 did not exactly match with Landsat 7 because it had not reached its operational Worldwide Reference System (WRS-2) orbit, but the images acquired provided substantial overlap for comparison. The time delay between the scene center acquisition for Landsat 7 and Landsat 8 was usually between two to five minutes (see Table 2.2).

The underflight Landsat 8 data were preprocessed to a Level 1 Terrain corrected (L1T) product and processed to surface reflectance using the L8SR algorithm by the USGS EROS Data Center. The data were processed and made available to the Landsat Science Team by the USGS, but can now be downloaded and analyzed as part of the "Pre-WRS-2" Landsat archive (the Landsat Scene Identifier for each image is available in Table 2.2). Cloud, cloud shadow, and snow masks were also provided for Landsat 8 data using CFmask, a C implementation of the Fmask algorithm (Zhu and Woodcock, 2012; Zhu, Wang, et al., 2015). Two CFmask images were generated for each Landsat 8 image: a mask that incorporated information from the new cirrus band on Landsat 8, and a mask that ignored the cirrus band and used the same algorithm as Fmask uses for Landsat TM/ETM+ data. The CFmask output used in this evaluation differed from the standard product because the cloud masks

that used the cirrus band identified cirrus clouds separate from other types of clouds. Landsat 7 data corresponding to each Landsat 8 underflight acquisition were acquired for comparison. The Landsat 7 data were preprocessed to L1T products and atmospherically corrected using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006). Cloud, cloud shadow, and snow masks were also generated for each Landsat 7 image using CFmask. We created layer stacked images containing corresponding image bands for each matching pair of Landsat 7 and Landsat 8 images. The new shorter wavelength blue band and the cirrus band on Landsat 8 were excluded as they had no direct comparison with the Landsat 7 data. To avoid confusion relating to the band numbering change between Landsat 7 and Landsat 8, we refer to spectral bands on ETM+ and OLI using by the common name of the spectral range they measure. Refer to Table 2.1 for the pairing between the spectral wavelength names, the sensor band numbering, and the spectral wavelengths measured.

WRS-2	Common Name	Landsat 8	Landsat 7
P013 R010	Qaasuitsup, Greenland	LC80130102013089LGN01 15:23:37Z	LE70130102013089EDC00 15:20:58Z
P013 R029	New Hampshire, United States	LC80130292013089LGN01 15:31:09Z	LE70130292013089EDC00 15:28:32Z
P029 R031	Nebraska, United States	LC80290312013089LGN01 17:10:31Z	LE70290312013089EDC00 17:08:13Z
P102 R076	Northern Territory, Australia	LC81020762013089LGN01 01:02:39Z	LE71020762013089ASA00 00:57:18Z
P134 R042	Arunachal Pradesh, India	LC81340422013089LGN01 04:06:18Z	LE71340422013089EDC00 04:01:31Z
P134 R052	Andaman and Nicobar Islands, India	LC81340522013089LGN01 04:10:16Z	LE71340522013089EDC00 04:05:30Z
P150 R042	Rajasthan, India	LE71500422013089PFS00 05:44:52Z	LC81500422013089LGN01 05:40:24Z
P198 R047	Tombouctou, Mali	LE71980472013089ASN00 10:42:34Z	LC81980472013089LGN01 10:39:03Z
P230 R084	Mendoza, Argentina	LE72300842013089CUB00 14:14:24Z	LC82300842013089LGN01 14:11:34Z
P230 R094	Santa Cruz, Argentina	LE72300942013089EDC00 14:18:23Z	LC82300942013089LGN01 14:15:34Z

Table 2.2: Scene center acquisition time for underflight data

2.2.2 Landsat time series

We acquired all available Level 1 Terrain corrected (L1T) Landsat ETM+ and OLI images with less than 20% cloud cover between 2008 and 2015 for Worldwide Reference System (WRS-2) path/rows 16/41, 23/37, 34/32, and 43/34 as Landsat CDR products. According to the 2011 National Land Cover Database (NLCD) (Homer et al., 2015), the land cover in scene in Florida, United States (P016R041) is primarily comprised of natural or cultivated herbaceous (36%), wetlands (30%), and developed (12%) categories. The scene in Mississippi, United States (P023R037) is marked by cultivated herbaceous (53%), wetlands (17%), and forest cover (16%). Our site in Colorado, United States (P034R032) intersects the Rocky Mountains and the Denver urban area and primarily contains forest (40%), natural and cultivated herbaceous (26%), and shrub (21%) covers. Finally, the site in California, United States

(P043R034) is centered on the Central Valley and is covered by natural or cultivated herbaceous vegetation (57%), forest (18%), and shrub (13%) cover.

Landsat images were atmospherically corrected using LEDAPS (Masek et al., 2006) for ETM+ data and L8SR for OLI data (U.S. Geological Survey, 2015). Atmospheric correction within the Landsat CDR product was performed by the USGS EROS Science Processing Architecture (ESPA) (LEDAPS version "LEDAPS_2.2.1" and L8SR version "l8_surface_reflectance_0.2.0") (Loveland and Dwyer, 2012). We transformed the surface reflectance values into the NDVI, EVI, NBR, and NDMI spectral indices. Thermal band data were converted into top of atmosphere brightness temperature.

Clouds, cloud shadows, and snow were identified using the Fmask algorithm (Zhu and Woodcock, 2012; Zhu, Wang, et al., 2015) with a cloud probability threshold of 12.5 and cloud, cloud shadow, and snow mask dilation of 5 pixels. The cloud probability threshold and mask dilation parameters were chosen to provide a conservative cloud mask result intended to limit omission of clouds. Errors of commission in cloud masks are less harmful than errors of omission in time series with hundreds of available images, so we used a custom run of Fmask instead of the available CFmask product. We also generated an additional Fmask image for Landsat 8 imagery using the same parameters that did not use information from the new cirrus band (as in Zhu, Wang, et al. (2015)) to assess the difference in noise reduction associated with having a band specifically designed to aid cloud detection.

Images within each path/row were layer stacked and aligned to a uniform image extent for further analysis. We excluded the new blue and cirrus spectral bands measured by Landsat 8 and only used one thermal band (Band 10) in Landsat 8 to facilitate comparison with Landsat 7. The number of total images available in each scene location varied between 164 and 183 with Landsat 8 contributing between 36

and 39 images (see Table 2.3).

WRS-2	Common Name	Landsat 7	Landsat 8	Total
P016R041	Florida, United States	144	39	183
P023R037	Mississippi, United States	128	38	166
P034R032	Colorado, United States	153	36	189
P043R034	California, United States	139	38	177

Table 2.3: Number of images by sensor in time series datasets

2.3 Methods

2.3.1 Landsat 8 Underflight

Spectral comparison

Clouds, cloud shadows, and snow in the Landsat 7 and Landsat 8 image pairs were masked from analysis using the CFmask band. The clouds and cloud shadows shifted position and shape considerably during the short time interval between the acquisitions by Landsat 7 and Landsat 8 (see Supplementary Figure 1). Because no solution exists to mask every cloud in a remote sensing image, artifacts remained in the underflight data after masking. This cloud contamination problem was amplified when comparing two images because missed clouds or cloud shadows in one image were unlikely to be masked in the other due to the shifts in the cloud and cloud shadow positions. Buffering clouds and cloud shadows in each individual Fmask image by 50 pixels cleaned up most of the noise present in both images and ensured that the remaining area contained only clear observations in both satellite acquisitions. The choice of 50 pixels was a relatively arbitrary decision, but it was the smallest buffer size tested that removed the majority of obvious data contamination.

The cloud screened reflectance observations were used to generate several vegetation indices for analysis, including the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), Enhanced Vegetation Index (EVI) (Huete et al., 2002), Normalized Burn Ratio (NBR) (Key and Benson, 2005), and Normalized Difference Mois-

ture Index (NDMI) indices. The Greenness and Wetness components of the Tasseled Cap transforms (Kauth and Thomas, 1976) were also calculated using Landsat TM surface reflectance transform coefficients from Crist (1985). We tested using the tasseled cap coefficients for ETM+ at-satellite reflectance from Huang et al. (2002), but the outcomes were not meaningfully different. Linear regressions were constructed to estimate the relationship between Landsat 7 and Landsat 8 reflectance values as a gain and bias. We were concerned with what a corrective equation might look like, so we did not restrict the functional form of the regression to only a gain or a bias. Regression 95% confidence intervals for the slope and intercept estimates were generated using 1,000 bootstrap samples of each dataset. The lower and upper confidence bounds were taken as the 2.5% and 97.5% percentiles of the sorted bootstrap parameter estimates.

Cirrus cloud characterization

We also used the underflight data to observe the spectral characteristics of cirrus cloud observations in Landsat data that would not be picked up without the new cirrus band on Landsat 8. We extracted pixels classified as clear land in Landsat 7 and Landsat 8 CFmask results as reference examples for clean observations. Pixels classified as clear land or clear water within the Landsat 7 CFmask image that were classified as cirrus clouds within the Landsat 8 CFmask image were also extracted. To account for errors in the CFmask images and movement of the clouds and cloud shadows between image acquisitions, we eroded the boundaries of CFmask labels by 20 pixels. This erosion helped ensure that the disagreement between the two CFmask images was the result of detection capacity and not random error or spatial mismatches. Not all scenes acquired during the underflight period contained cirrus clouds or contained enough disagreement about cirrus cloud coverage to perform our analysis, so we were limited to two scenes in Mendoza Province, Argentina (P230R084) and Santa Cruz Province,

Argentina (P230R094). Visual inspection of these images showed that cloud masks in both scenes shared similar patterns. Areas of dense cirrus cloud cover were identified as cloud cover in Landsat 7 masks and cirrus cloud cover in Landsat 8, but areas of thin cloud cover were misidentified as clear land in Landsat 7 and correctly classified as cirrus cloud in Landsat 8. While the cirrus cloud classification may suffer from false positives over bright targets in dry environments or over high elevation targets (Zhu, Wang, et al., 2015), the mean elevation of the two scenes was relatively low (440m and 675m for P230R084 and P230R094, respectively) and we are confident from visual inspection that the vast majority of pixels identified as clear land and cirrus cloud in Landsat 7 and Landsat 8 masks were omitted in the Landsat 7 cloud masks.

Observations labeled clear in one CFmask image but cirrus cloud in another represent difficult to detect noise within historical Landsat data that may now be routinely filtered using OLI's cirrus band. These data were analyzed using bivariate plots comparing the spectral reflectance and brightness temperature from Landsat 8 in combinations of spectral bands. Plotted points were colorized according to the classification in Landsat 8's CFmask image. We added contour lines to describe the overlap of the clear land and cirrus cloud observations by delineating the two dimensional space occupied by 95% of the observations in each population. We used a Gaussian kernel density estimate to approximate the distribution of each population in bivariate space (Silverman, 1986). The 95% percent contour lines were estimated as the 95th percentile of the estimated distribution from the two dimensional kernel density estimate.

2.3.2 Landsat time series

Landsat time series were analyzed using a Python implementation of the Continuous Change Detection and Classification (CCDC) algorithm (Zhu and Woodcock, 2014;

Zhu, Woodcock, Holden, et al., 2015). The CCDC algorithm attempts to detect abrupt changes in a time series using an "online" monitoring approach by sequentially comparing reflectance observations against a prediction. Reflectances are predicted using a simple Fourier style model with independent variables representing the overall reflectance (intercept), the change in reflectance over time (slope), and the intra-annual variation in reflectance due to phenology or sun-sensor geometry (harmonics):

$$\hat{\rho}_i = \beta_0 + \beta_1 x_t + \sum_{j \in N} [\beta_{2j} \cos(\frac{2\pi j}{T} x_t) + \beta_{2j+1} \sin(\frac{2\pi j}{T} x_t)] + \varepsilon_t \qquad (2.1)$$

where $\hat{\rho}_i$ is the predicted reflectance or temperature in each spectral band i, x_t is the ordinal date of each observation, N is a set of integers specifying the frequency, j, of the Fourier series harmonics, T is the number of days in a year (365.25), and ε_t is the residual error term for each observation.

This study used a pair of year and half year ($N = \{1, 2\}$) harmonics to be able to capture most seasonal vegetation patterns while remaining simpler than the full 8 coefficient model used in Zhu, Woodcock, Holden, et al. (2015). The CCDC algorithm employs the Least Absolute Shrinkage and Selection Operator (LASSO) regularization method to reduce overfitting by estimating coefficients that minimize the residual sum of squares while also penalizing the absolute magnitude of the coefficients (Tibshirani, 1996). LASSO can provide coefficient estimates of exactly zero, and this type of model selection allows CCDC to use a possibly overly specified but generic model for all time series.

To help provide a stable and robust initial estimate of land surface reflectance, the CCDC algorithm screens the training period using a multitemporal filtering algorithm designed to remove any cloud and cloud shadows that might have been missed by Fmask (Zhu and Woodcock, 2014). After the initial time series model has been fit during this training period, successive observations are compared against the forecast

prediction. A difference between the observation and prediction is significant if the l^2-norm, or Euclidean norm, of forecasted residuals scaled by model RMSE across the set of tested bands is above a specified threshold:

$$T_{crit} > \sqrt{\sum_{i \in B} (\frac{\hat{\rho}_i - \rho_i}{RMSE_i})^2} \qquad (2.2)$$

where T_{crit} is the significance critical threshold, B is a set of spectral bands used in the change detection, $\hat{\rho}_i$ and ρ_i are the predicted and observed reflectances or values in band i, and $RMSE_i$ is the time series model RMSE for band i.

If observed reflectances significantly differ from the forecasted reflectances for some number of consecutive observations, CCDC places a break in the time series and begins trying to fit another time series model after the disturbance. The breaks or disturbances in the time series create distinct segments through time that should correspond to a period of stable land cover or land surface condition. These segments in the time series can then be classified into land cover categories using attributes derived from the time series models, including the coefficients and Root Mean Squared Error (RMSE) of the time series models estimated for each Landsat band. The estimated time series model coefficients and RMSE estimates, and not the change detection, are the focus of this study. For further information and detail about the Continuous Change and Classification Detection algorithm, including how it detects change, the reader is referred to Zhu and Woodcock (2014) and to Zhu, Woodcock, Holden, et al. (2015) for algorithm changes and improvements.

To address the questions of the effect of spectral differences between Landsat 7 and Landsat 8 in time series, we fit time series using CCDC to five different post-launch scenarios. All five time series were initialized between 2008 and 2013 using only Landsat 7 data, but were continued after the launch of Landsat 8 with different sets of Landsat observations:

1. Only Landsat 7 data post-launch

2. Only Landsat 8 data post-launch

3. Combined Landsat 7 and Landsat 8 data post-launch

4. Only Landsat 8 data post-launch, excluding information from the cirrus band in the Fmask images for Landsat 8 data

5. Combined Landsat 7 and Landsat 8 data post-launch, excluding information from the cirrus band in the Fmask images for Landsat 8 data

The first scenario containing only Landsat 7 data was used as a normalizing factor to isolate the influence of including Landsat 8 data within the time series of scenarios two and three. Intercept and RMSE values for scenarios two and three were normalized by dividing these estimates by the corresponding intercepts and RMSE values from the Landsat 7 models (scenario one). A value of one for the normalized intercept and RMSE indicate that the inclusion of Landsat 8 data (scenario two and three) did not change the quantity of interest. Values above one indicate the quantity was increased relative to just using Landsat 7 and values below one signify a decrease in the quantity. The slope estimates in the CCDC models are usually either zero or very close to zero. To avoid computational issues of dividing by zero, we normalized the slope values by subtracting the slopes estimated using Landsat 8 data (scenarios two and three) by the slope estimated only using Landsat 7 (scenario one). Similarly, scenarios four and five were compared after being normalized by scenarios two and three to understand the effect of improved cloud detection possible with the cirrus band on Landsat 8.

We masked and ignored from the analysis any pixels that changed between 2010 and 2015 as found by the CCDC algorithm because time series following a disturbance can be erratic or noisy. Large variability in the observed target might have created

false differences among the continuation scenarios simply due to random chance rather than due to a systematic difference. We took a large random sample of all remaining, stable pixels, and extracted the intercept and slope coefficients and the Root Mean Squared Error (RMSE) of each time series model for comparison.

2.4 Results and Discussion

2.4.1 Spectral Comparison

Spectral comparison results for scenes over Nebraska, United States (WRS-2 path and row P029R031) and Mendoza Province, Argentina (P230R084) are shown in Figure 2·1 and 2·2, respectively. While only the results for these two Landsat scenes are shown in the main text of this article, the patterns observed in these scenes are characteristic of the responses seen in other scenes containing sufficient clear observations of land. Figures and tables showing the spectral responses for all other Landsat scenes acquired are available in the supplementary material (Supplementary Figures 2 - 8).

The reflectances in the visible wavelengths (blue, green, and red) of the Landsat 8 atmospherically corrected product are consistently darker than reflectances observed in the Landsat 7 data. The largest bias occurs in the blue band and decreases with increasing wavelength from green to red. While the wavelengths measured by the visible bands only changed slightly from Landsat 7 to Landsat 8 (Table 2.1), atmospheric interference is greatest in the blue band and decreases as wavelength increases. If differences in atmospheric correction are responsible for some of the difference between sensor measurements, it makes sense that the greatest difference is visible in the band that requires the greatest adjustment for scattering and absorption and the LEDAPS and L8SR algorithms use different inputs to parametrize atmospheric constituents. Some proportion of this bias is also likely caused by differences calibration or target
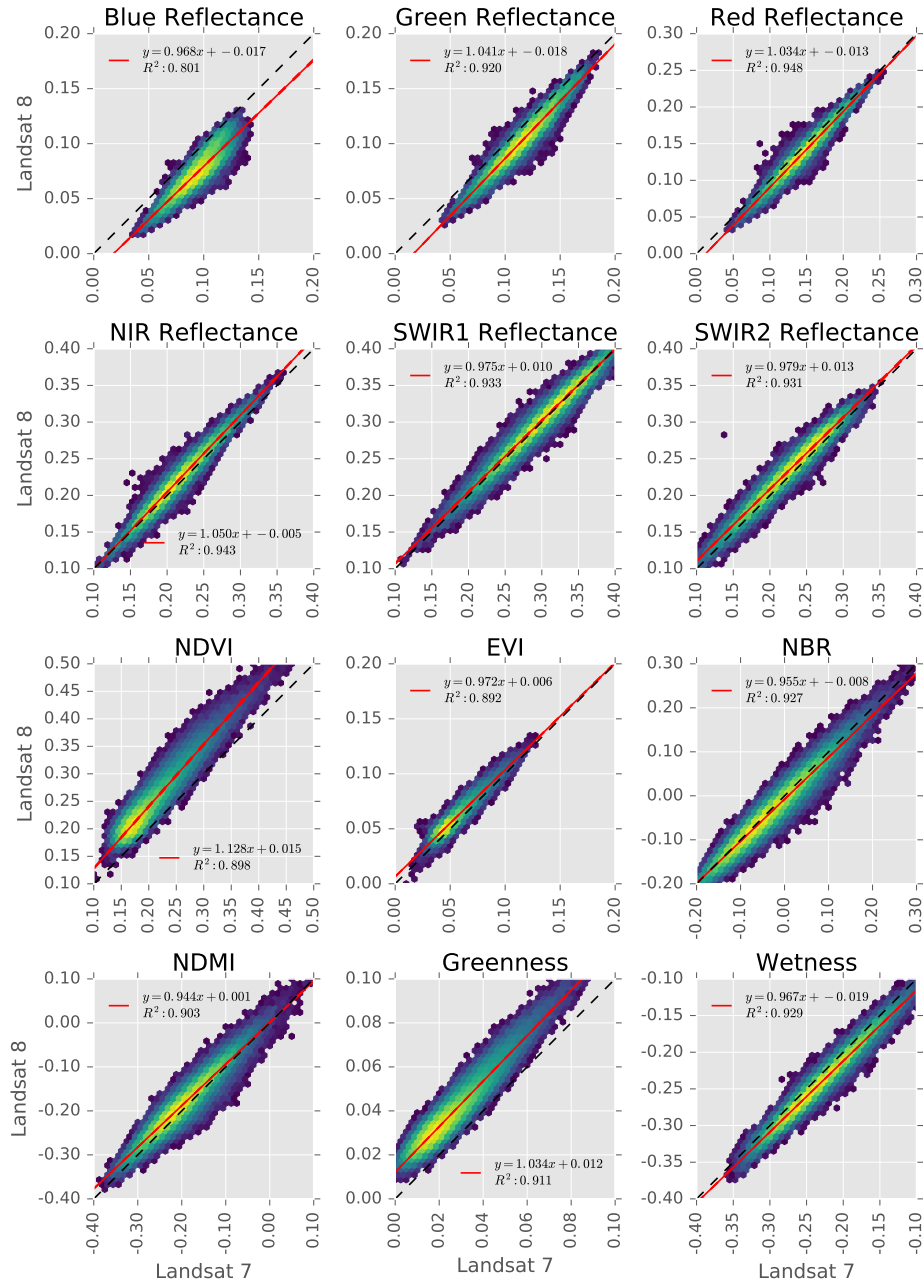
Figure 2·1: Landsat 7 and Landsat 8 underflight data comparison for Nebraska, United States (P029R031). Regression parameters and R^2 estimates describing the fitted line are provided in the top left of each plot. The 95% confidence interval range around the regression estimate is displayed as a dashed line around the regression line, but it may not be visible for small confidence intervals.
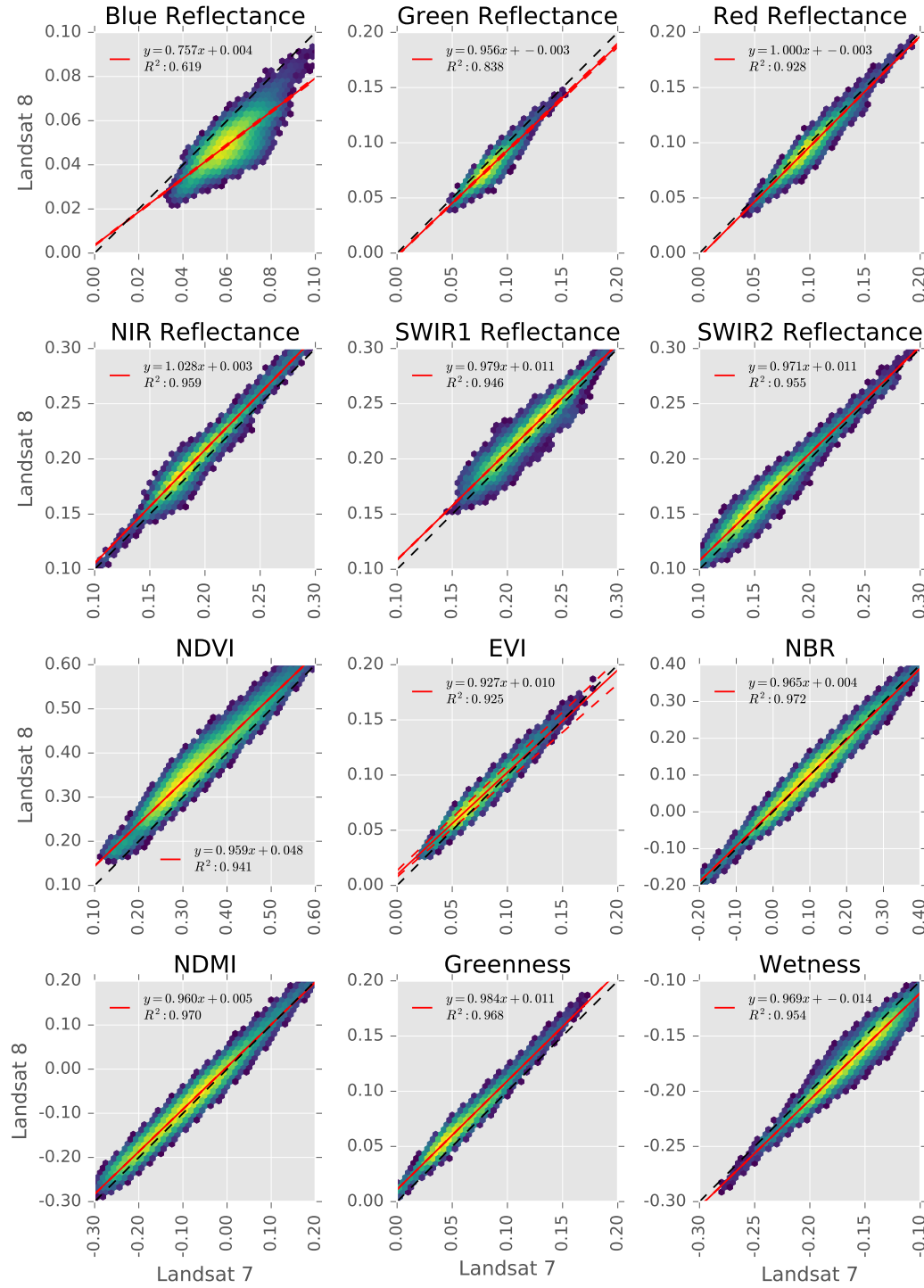
Figure 2·2: Landsat 7 and Landsat 8 underflight spectral comparison for Mendoza Province, Argentina (P230R084).

specific spectral responses, but estimating the relative contribution of these three causes is outside the scope of this analysis.

The near infrared band, with the largest physical difference in measured wavelengths, did not show considerable bias in Nebraska, United States (P029R031) (Figure 2·1). However, there was a large and consistent positive bias in the near infrared reflectance in Medoza, Argentina (P230R084) (Figure 2·2). Summer in Argentina during the month of March most likely has much higher atmospheric water vapor content than the United States during winter. The narrower wavelengths measured by Landsat 8 avoided an atmospheric water absorption feature, so it is understandable that the bias in near infrared reflectance is higher in an environment with more water vapor. The shortwave IR bands may be slightly brighter in Landsat 8 than in Landsat 7 due to the smaller spectral band widths, but the measurements have not changed as significantly as the other bands. The difference in bias across scenes observed in this study is similar to the findings of Flood (2014) that there are regional differences in the sensor bias.

Differences in individual bands between Landsat 7 and Landsat 8 are especially important when considering spectral or vegetation indices because these indices often rely on the contrast between two or more bands. The Normalized Differenced Vegetation Index (NDVI) relies on the contrasting relationship between the near infrared band and the red band. It appears that the near infrared band and red bands in Landsat 8 are brighter and darker than in Landsat 7, respectively, which enhances the contrast highlighted in NDVI. As such, we see a strong and consistent positive bias in NDVI with Landsat 8 having much higher NDVI. The Enhanced Vegetation Index (EVI) differs from NDVI by utilizing the blue band as an additional normalizing factor that corrects the red band for atmospheric influences. Either by coincidence or as the intention of Huete et al. (2002), it appears that the bias in the blue band

between Landsat 8 and Landsat 7 nullifies the bias in the red and near infrared band, resulting in a surprisingly similar EVI across sensors. The Normalized Difference Moisture Index (NDMI) and the Normalized Burn Ratio (NBR) both compare the shortwave infrared with the near infrared using either the first or second SWIR bands, respectively. These vegetation indices show little difference between Landsat 7 and Landsat 8. If there is a difference in the shortwave infrared bands between sensors, it appears to be in the same positive direction as the bias in the near infrared, resulting in a cancellation effect. The Tasseled Cap Greenness transform displays the same strong positive bias as the NDVI while the Wetness Tasseled Cap transform is darker in Landsat 8 than in Landsat 7. The Wetness transform for Landsat 7 places little weight on the red and near infrared bands in favor of contrasting the blue and green with the two shortwave infrared bands. Because the blue and green bands are darker in Landsat 8 than in Landsat 7 and the shortwave infrared bands might be brighter in Landsat 8, this contrast is enhanced and creates a small bias in the Wetness transform.

The effect of increased radiometric resolution is readily apparent when comparing reflectance estimates or spectral indices between sensors. Figures 2·1 and 2·2 both display a visual striping pattern distributed across the X axis that is not simply a visual artifact in the plot, but a result of the increased quantization possible with Landsat 8's 12-bit radiometric resolution. When the OLI instrument converts the continuous measurement of radiance by its detectors into digital numbers, the 12-bit radiometric resolution offers sixteen times the digitization detail compared to Landsat 7. The increased detail is carried through when converting from digital numbers back into radiance and the precision remains when converting to surface reflectance. The underflight spectral comparison figures for a Landsat scene over the coast of Greenland (P013R031), included in the supplementary material (Supplementary Fig-

ure 3), contains the best example of the effects of increased radiometric resolution. Reflectance estimates from Landsat 8 do not display any striping throughout the very low range of reflectance measured (2 - 8%) in the shortwave infrared while Landsat 7's measurements contain distinct discrete gaps representative of the lower quantization. Improvements to the radiometric resolution on OLI thus enhance detection capabilities by allowing the delineation of very similar, previously indistinguishable, targets.

2.4.2   Cirrus cloud characterization

We explored the spectral properties of clear land observations and those affected by cirrus clouds by plotting observations of both targets in bivariate combinations of spectral bands available in the Landsat 8 underflight data (Figure 2·3). Observations affected by cirrus clouds differ most from clear land observations in the visible wavelengths as seen in the high degree of separation within bivariate plots containing only the visible bands. The blue band, specifically, shows the largest difference between the two populations (subfigures A through E). Cirrus clouds, as expected (Gao et al., 1993; Gao et al., 1998; Gao and Li, 2000), increase the reflectance in visible wavelengths, but this increase is not very substantial. The centers of the cirrus cloud and clear land populations differ by approximately 4% reflectance in the blue band, about 3% in the green band, and by even less in the red band. It is interesting to note that even in the visible bands the values of observations affected by cirrus clouds are still within the range of unaffected observations, which helps explain why it is so hard to detect cirrus clouds in any automated fashion in the absence of a cirrus band like the one on Landsat 8. The near infrared and shortwave infrared bands provide no separation alone as the populations completely overlap when only using the infrared bands. The thermal band shows the largest separation between the two populations of approximately half a degree Celsius (subfigures F, K, and O). If our objective were

to find cirrus clouds in our dataset without the use of the cirrus band on Landsat 8, the shorter and thermal wavelength bands would be most valuable. The opposite perspective to this conclusion is that if we wish to avoid cirrus cloud contamination influencing our datasets from all Landsats prior to Landsat 8 we would exclude the visible bands and thermal bands from our analysis whenever possible and rely heavily on the NIR and SWIR bands as these appear to be least affected by cirrus clouds missed without the use of a cirrus band.

2.4.3 Landsat time series

Impact of Landsat 8 on time series of Landsat observations

Figure 2·4 shows an example time series for one pixel in an undisturbed evergreen forest stand in Colorado (P034R032) in red reflectance and NDVI. The regression models used are described by equation 2.1, but some model coefficients, such as the sinusoidal terms used to model seasonality, may be estimated as 0 due to the regularization performed by the LASSO method. As seen in the underflight data, the observations from Landsat 8 are darker in the red band, creating higher NDVI values. The time series models show a small difference in the intercept estimates and a much larger difference in the slope estimate that adjusts for the differences between sensors. The unexplained variation due simply to the sensor differences increases the Root Mean Squared Error (RMSE) in the models. Only observations in the visible bands and the NDVI were affected by sensor differences for this specific target. Time series models for the NIR, SWIR1, SWIR2, and thermal bands and for the EVI, NBR, and NDMI indices were not different between the Landsat 7 only (scenario one) and combined sensor (scenario two) datasets. This example is included to demonstrate the influence of observation differences on time series model parameter estimates. Results for millions of pixels across more study sites in all spectral bands and vegetation indices are given in this section.
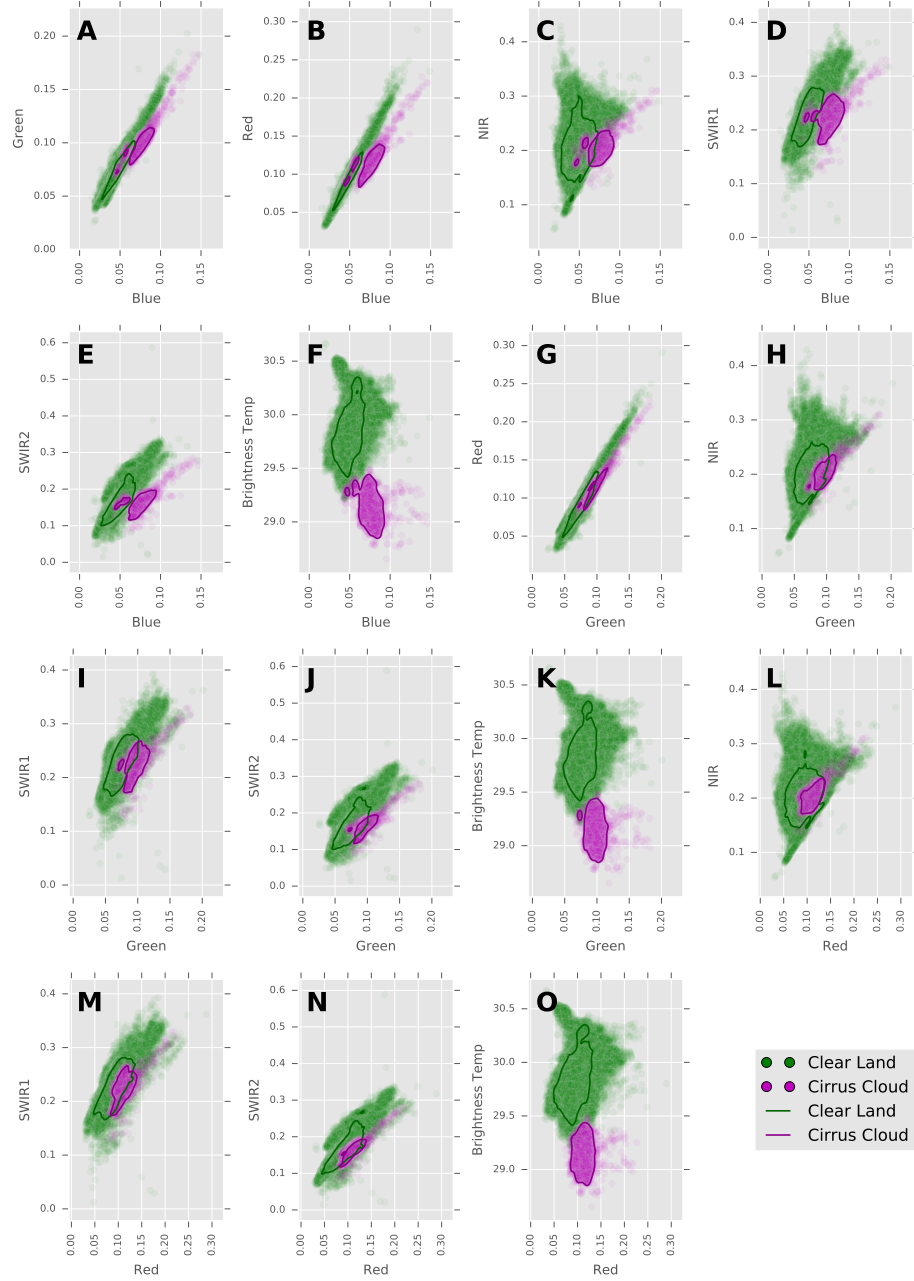
Figure 2·3: Clear land and cloud cirrus spectral properties for Mendoza Province, Argentina (P230R084). Contour lines contain 95% of the data in the bivariate distribution space for data labeled as clear land (green) and cirrus cloud (magenta) in Landsat 8 CFmask images.
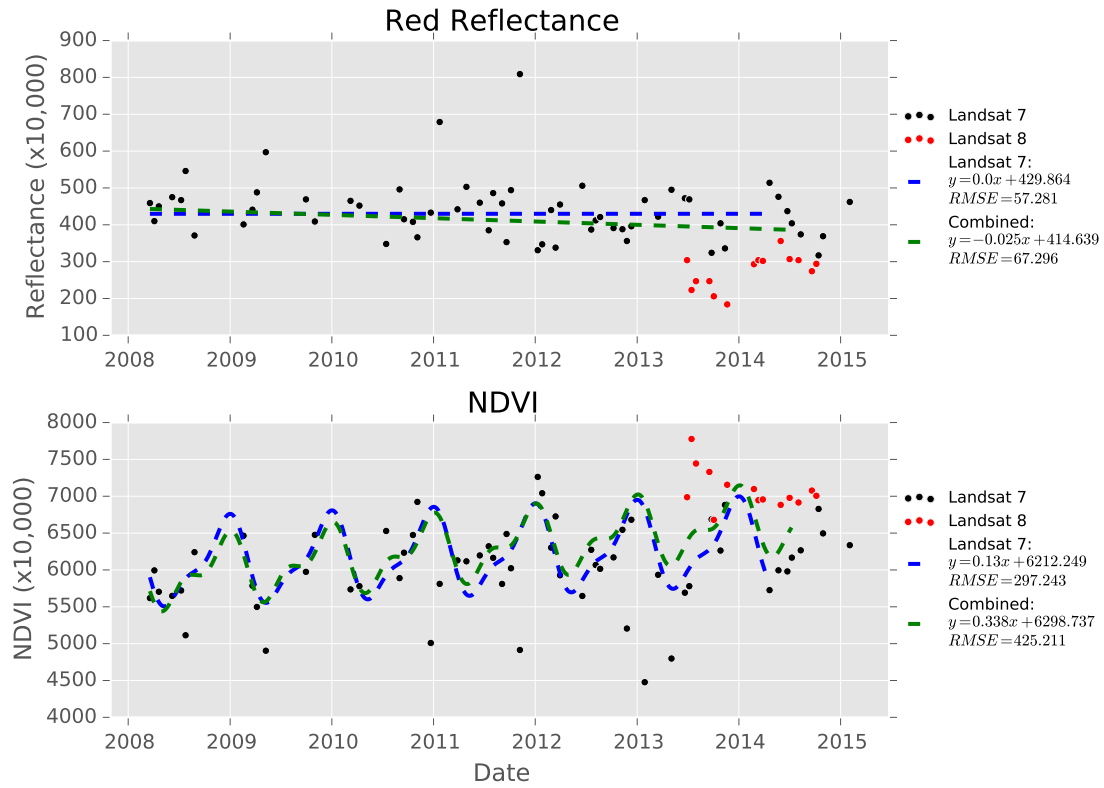
Figure 2·4: Example time series of an evergreen forest stand in Colorado, United States (P034R032). Observations from Landsat 7 and Landsat 8 are shown in black and red, respectively. CCDC model regressions for time series including only Landsat 7 data (scenario 1) and both sensors (scenario 2) are plotted in blue and green. Model slope, intercept, and RMSE estimates are included in the figure legend.

Three time series datasets covering four WRS-2 scenes initialized using Landsat 7 data and continued after the launch of Landsat 8 with either Landsat 7 (scenario one), Landsat 8 (scenario two), or both Landsat 7 and Landsat 8 data (scenario three) were analyzed with the Continuous Change Detection and Classification (CCDC) algorithm. Intercept, slope, and RMSE estimates for scenarios two and three containing data from Landsat 8 were normalized relative to Landsat 7 results as described in section 2.3.2. In essence, we are treating the use of only Landsat 7 as the baseline for comparisons with Landsat 8 or the combination of Landsats 7 and 8. A value of one for the normalized intercepts and RMSE values indicates no difference between scenarios, while values above or below one values indicate the attribute is brighter or darker when including Landsat 8 data. A value of zero in normalized slope estimates indicates no change between scenarios while positive or negative values indicate an increase or decrease in the time series slope relative to only using Landsat 7 data. If the introduction of Landsat 8 data into time series of Landsat 7 data caused no difference, histograms of these normalized parameter estimates from pixels within each scene should approximate a normal distribution centered on one for intercepts and RMSE or zero for slopes. Model parameters greatly affected by the inclusion of Landsat 8 observations should produce histograms that are offset from the "no change" normalized values of one or zero.

Time series model intercept, slope, and RMSE attributes for a scene in Colorado, United States (P034R032) are shown in Figures 2·5, 2·6, and 2·7, respectively and confirm the results previously found in the underflight data comparison. The blue, green, and red bands on Landsat 8 were consistently darker than the same measurements in Landsat 7 data in the underflight data, and this consistent offset impacts the modeling results in time series scenarios. Time series model intercepts for the blue, green, and red bands shown in figure 2·5 are unusually much darker due to the

inclusion of Landsat 8 data as shown by the shift in the median of the histograms from 1.0 to values between approximately 0.90 and 0.96. The median normalized intercept estimated for the blue band when continuing Landsat 7 time series with only Landsat 8 data (scenario two), for example, is 10% darker than the intercept that would have been estimated when using exclusively Landsat 7 data. Time series continued after the launch of Landsat 8 with both Landsat 7 and Landsat 8 (scenario 3) are less affected than with Landsat 8 data alone as seen in the reduction in the shift of the blue colored histogram (scenario three) relative to the red colored histogram (scenario two). However, the combined scenario still contains a bias in the visible band intercepts. The histograms of the normalized intercepts in the near infrared and shortwave infrared bands are centered approximately at one and are approximately normally distributed. Thus, the NIR and SWIR bands show little signs of bias relative to Landsat 7 and do not behave difference under the Landsat 8 or combined sensor scenarios. Among the vegetation indices, only NDVI is substantially affected and is characteristically brighter in Landsat 8 than in Landsat 7 as the histogram of normalized NDVI values is centered at approximately 1.03 to 1.04. While the thermal band is slightly darker for the time series in Colorado, United States, the pattern was not consistent in other Landsat scenes (Supplementary Figures 9, 12, and 15).

While the use of Landsat 8 in regressions might affect the overall mean reflectance, measured as the intercept of the regression, any differences between Landsat 7 and Landsat 8 are best explained using a slope estimate because the Landsat 8 data are placed at the end of the time series. Figure 2·6 shows that the Landsat 8 observations placed at the end of the Landsat 7 time series create a spurious time trends for many of the bands and indices analyzed. The darker observations in the visible bands in Landsat 8 create false positive decreasing time trends in both continuation scenarios, although the influence in the combined scenario is less pronounced. The median slope

response for the near infrared band is approximately zero indicating no bias in the slope coefficient. While the shortwave infrared bands appear to have a negative bias in the results for Colorado, United States (P034R032), this pattern was not present in the three other scenes analyzed (Supplementary Figures 10, 13, and 16). Repeating the patterns previously observed, NDVI is highly affected by differences in Landsat 8's reflectance in the visible bands but the effect does not carry over to EVI. The NBR and NDMI indices, despite not showing an effect in the time series intercept coefficient, show a slight positive time trend when using Landsat 8 data. The positive time trend in NBR and NDMI was also observed for two scenes over Florida (P016R041) and Mississippi (P023R037), but not in California (P043R034) (Supplementary Figures 10, 13, and 16). The distribution of thermal band slope coefficients for Colorado has a long negative tail, but this phenomenon is not observed in Florida or Mississippi and the slope distribution for California shows an opposite long positive tail. These results that show spurious time trends, especially in vegetation indices, can be caused by differences in reflectance across Landsat 7 and Landsat 8 reinforce the false positive slopes found in time series analysis over Canadian boreal forest using Landsat 5 and Landsat 7 (Sulla-Menashe et al., 2016).

The change in RMSE of time series models using Landsat 8 (scenarios two and three) relative to pure Landsat 7 time series models (scenario one) varies depending on how different a particular band or vegetation index looks in Landsat 8 and Landsat 7. For Colorado (P034R032), just as we see the largest difference in time series model intercepts and slopes in the visible bands and NDVI, we also see increases in the RMSE of time series models using these data (Figure 2·7). The other scenes in Florida (P016R041) and Mississippi (P023R037) share this pattern of increase in the unexplained variance in the visible and NDVI data (Supplementary Figures 11 and 14). The scene in California (P043R034) shows weaker evidence for increased noise

when using the combined data scenario (scenario three) and no sign of increasing noise when just using Landsat 8 data post-launch (scenario two) (Supplementary Figure 17). The estimated intercept and slopes for the near infrared and shortwave infrared bands did not differ among scenarios so we would expect little change in the RMSE of these models. The EVI, which has consistently shown the smallest smallest differences between Landsat 7 and Landsat 8, does not show increases in model RMSE relative to Landsat 7. While NBR and NDMI showed a positive time trend in Colorado, Florida, and Mississippi, this bias relative to Landsat 7 has only increased the RMSE of models in Mississippi. The influence of changes in the thermal band may affect the RMSE of models, but the magnitude and direction of this effect is not consistent among scenes analyzed.

A systematic difference between Landsat 8 and Landsat 7 observations affects time series analysis in three ways. First, a consistently darker or brighter sequence of Landsat 8 observations will eventually affect the mean reflectance of a time series. Variation in reflectance due to the dataset of origin rather than variation in physical attributes hampers inference. For example, if a forest were slightly degraded in 2013, the mean NDVI of a time series after the disturbance might be the same as the time series before the disturbance because the NDVI in Landsat 8 is brighter than the NDVI in Landsat 7. The effect of biases in Landsat 8 relative to Landsat 7 is especially concerning with regard to the time trend or slopes of time series models. Second, consistently darker or brighter observations from Landsat 8 placed at the end of an existing, stable time series of Landsat 7 data are shown to induce a false positive trend in reflectance through time. Time trends in reflectance or vegetation indices have been used as a proxy or indicators of changes in ecosystem function and structure, including vegetation greening, browning, recovery, or degradation. Inference of temporal trends will require relative calibration to ensure trends are not simply due
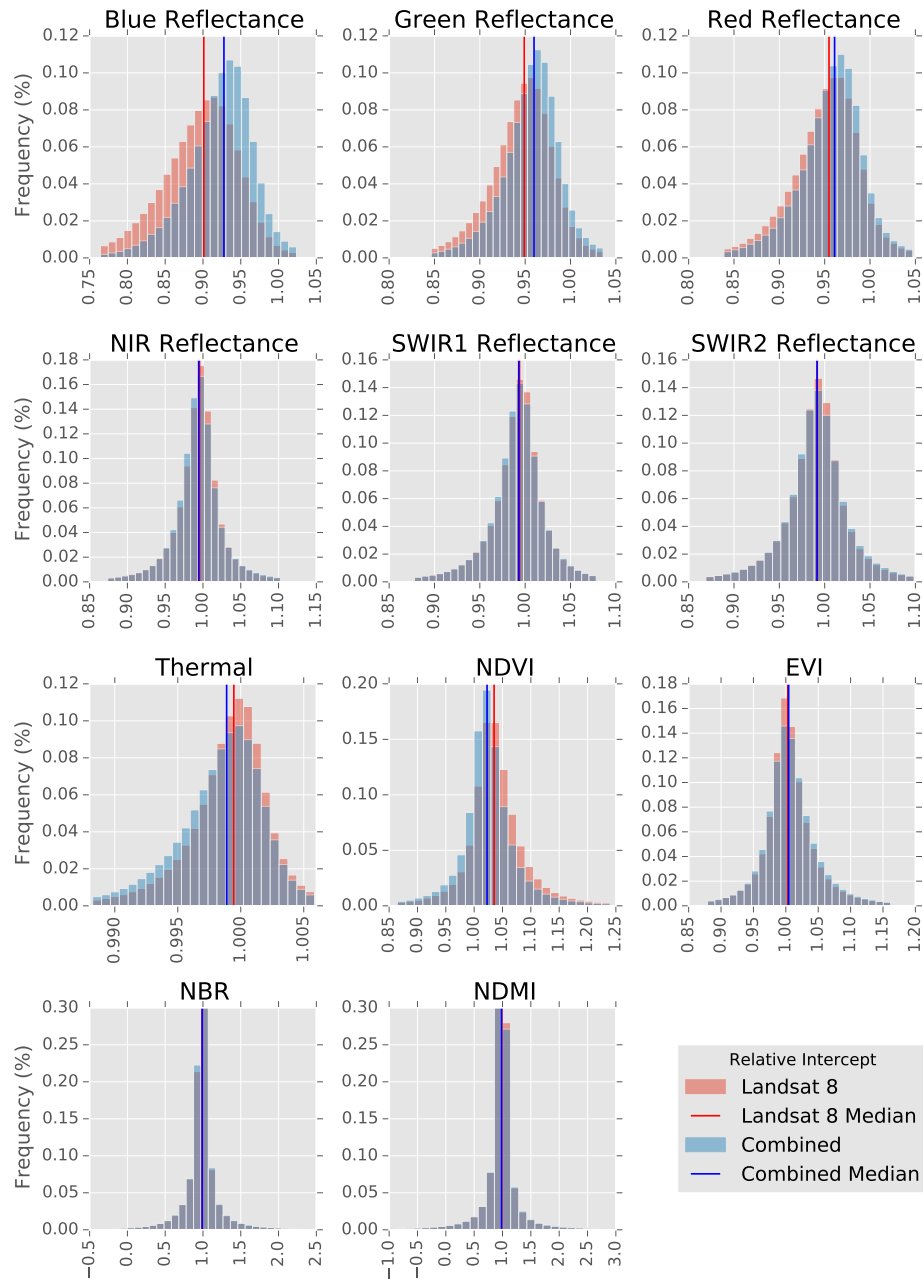
Figure 2·5: Normalized intercept estimates for time series models in Colorado, United States (P034R032) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 intercept estimates were normalized by dividing by the intercepts of time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
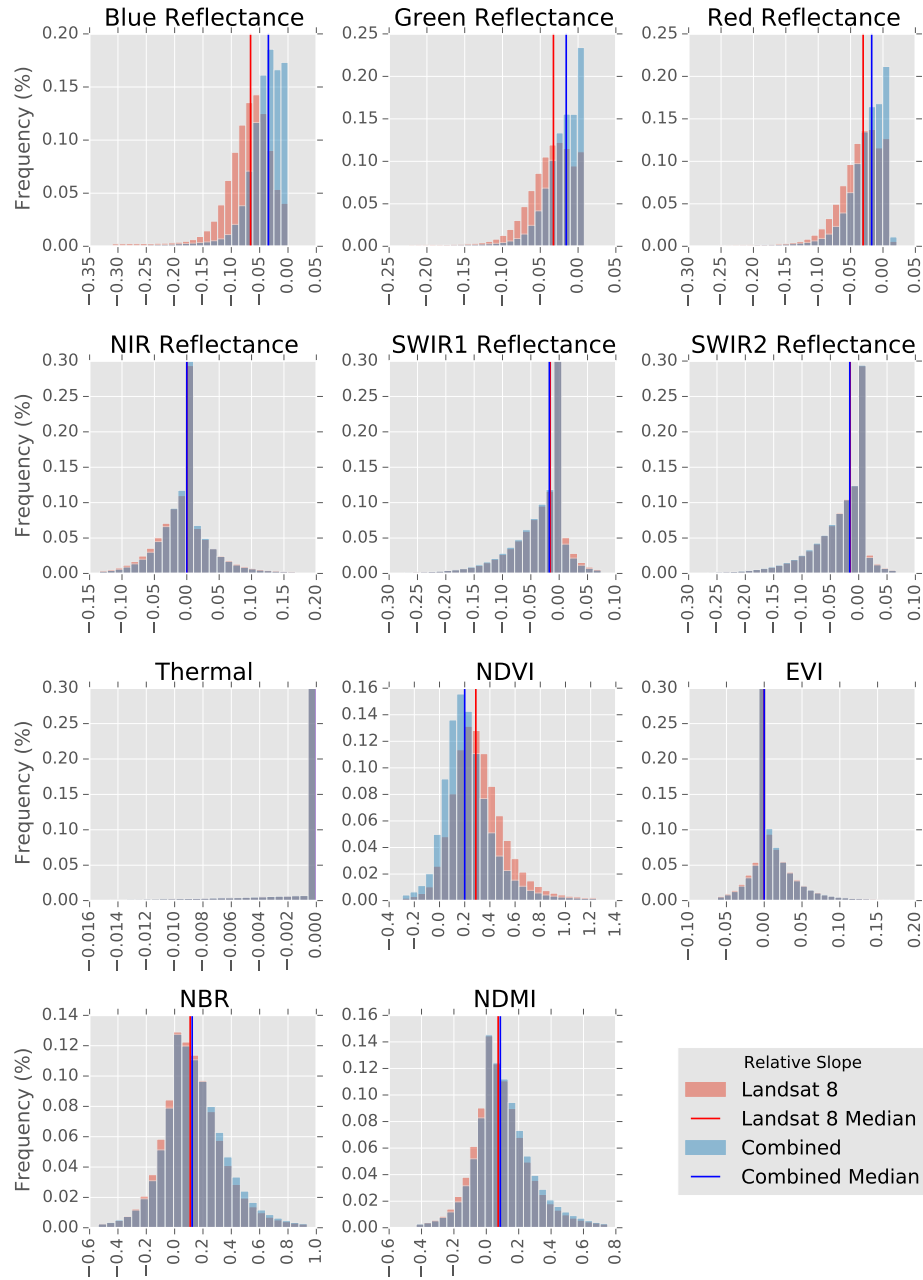
Figure 2·6: Normalized slope estimates for time series models in Colorado, United States (P034R032) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 slope estimates were normalized by subtracting the slopes from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
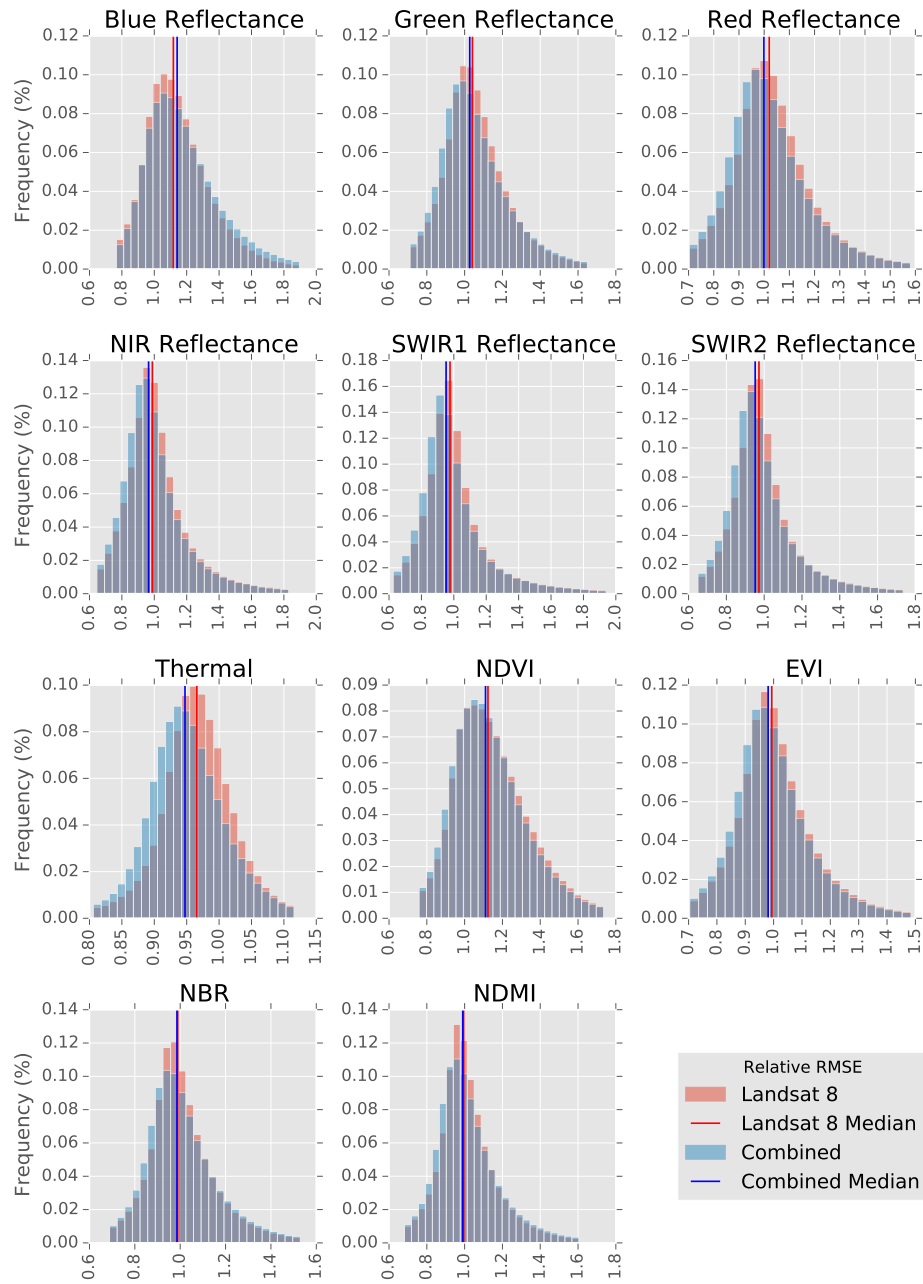
Figure 2·7: Normalized RMSE estimates for time series models in Colorado, United States (P034R032) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 RMSE estimates were normalized by dividing by the RMSE from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

to sensor differences. Lastly, bias in observed reflectance between the two sensors also hinders detection of land cover change in time series analysis. Change detection methods that use a regression framework to predict reflectance based on time series models (as in Verbesselt et al. (2012) and Zhu and Woodcock (2014)) identify a time series break occurs if predictions differ significantly from observations. Bias in reflectance values between sensors adds an unexplained source of variability to the time series that broadens the range of nonsignificant differences between observed and predicted values. A more familiar comparison is the decreased ability to infer separation between reflectances before and after a land cover change in a paired t-test if each population has larger within-group variance.

Time series algorithms that do not rely on surface reflectance observations or perform image normalization prior to analysis may not be as significantly affected by cross-sensor differences. For example, while the Vegetation Change Tracker (VCT) (Huang et al., 2010) uses surface reflectance, imagery also undergoes a normalization using examples of dense and dark forests prior to analysis. The LandTrendr algorithm (Kennedy et al., 2010) normalized all Landsat imagery to a reference image using the Multivariate Alteration Detection and Calibration (MADCAL) algorithm (Canty et al., 2004). Techniques which rely on modeled attributes, such as forest cover fraction, derived from Landsat imagery may also avoid image normalization if separate models are used for each sensor. Normalization might not be required, however, if the bias associated with a particular sensor can be modeled as a time series component. One extremely simple method that could account for the difference in reflectance associated with a categorical variable is the use of a "dummy" variable within each time series model. A "dummy", or treatment, variable acts as an additional intercept estimate specific to one category, or for Landsat 8 in this case. When the "dummy" variable is incorporated into the CCDC regression model, the regression model is

modified from equation 2.1 as:

$$\hat{\rho}_i = \beta_0 + \beta_1 x_t + \sum_{j \in N} [\beta_{2j} \cos(\frac{2\pi j}{T} x_t) + \beta_{2j+1} \sin(\frac{2\pi j}{T} x_t)] + \beta_{dummy} D_{sensor,t} + \varepsilon_t \quad (2.3)$$

where:

$$D_{sensor,t} = \begin{cases} 1, & \text{if Landsat 8} \\ 0, & \text{otherwise} \end{cases} \quad (2.4)$$

This "dummy" variable estimate would compensate for sensor specific differences allowing for a more efficient estimate of the time series model slope and intercept, provided that the difference in sensor behavior can be modeled as a fixed offset. Figure 2·8 shows the same time series over an undisturbed evergreen needleleaf stand as 2·4, but with an additional regression line estimated using a Landsat 8 "dummy" variable. "Dummy" variable coefficient estimates for Landsat 8 are negative in the red band to account for a darker measurement and positive for NDVI to account for a brighter measurement. The estimated slope in the red band does not match the slope estimated in the exclusively Landsat 7 data scenario, but the "dummy" variable accounts for the spurious time trend in the NDVI time series while greatly reducing the RMSE of both models. Much more work is required to understand the potential for these sensor specific intercepts to abate problems in time series due to sensor disagreements, but such a solution would be an easy and quick way of synthesizing the datasets without the need for more advanced cross-calibration.

Impact of cirrus cloud detection

The characterization of cirrus cloud spectral impacts from the Landsat 8 underflight data (section 2.4.2) shows that cirrus clouds increase the reflectance of the visible
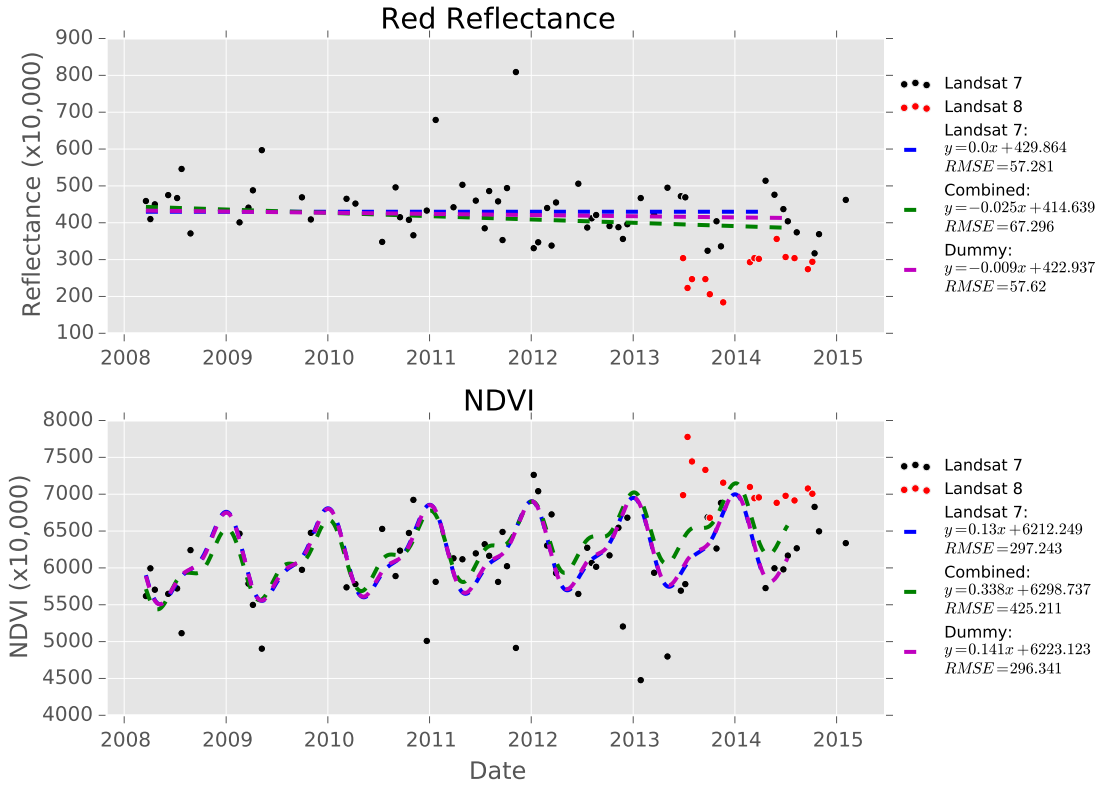
Figure 2·8: Example time series of an evergreen forest stand in Colorado, United States (P034R032). Observations from Landsat 7 and Landsat 8 are shown in black and red, respectively. CCDC model regressions for time series including only Landsat 7 data (scenario 1) and both sensors (scenario 2) are plotted in blue and green. The magenta regression line includes a dummy variable that captures the reflectance differences between Landsat 7 and Landsat 8 as an additional intercept. The dummy variable estimates were -94.3409 and 1169.12 for red and NDVI, respectively, which offset the bias in Landsat 8 measurements relative to Landsat 7.

bands but have negligible effects in the infrared bands. As a result, we expect time series containing cirrus clouds to positively bias the estimates of the intercept and slopes in the visible bands while increasing the RMSE of these models.

Figures 2·9, 2·10, and 2·11 show the time series model intercepts, slopes, and RMSE estimates for scenarios four and five which did not use cirrus band information in the Fmask cloud masks for a Landsat scene in Colorado, United States (P034R032). The intercepts and RMSE estimates were normalized by dividing these estimates by the same estimates from time series models using the cirrus band information (scenarios two and three). The slope coefficients were scaled by taking the difference between scenarios four and five and scenarios two and three, respectively.

The distribution of the relative intercepts for the visible bands have longer tails above the value of one, especially for the combined Landsat 7 and Landsat 8 data post-launch scenario (scenario 3). The near infrared and shortwave infrared bands appear to have less skew toward either side in the distribution. The intercepts in time series of brightness temperature appear colder without cloud screening using the cirrus band, demonstrated by the skew toward values less than one. Pixels affected by cirrus clouds appear to have reduced the intercept estimate of the NDVI, likely driven by the bias toward brighter values in the red band. The EVI, NDMI, and NBR vegetation indices do not appear to be have biased estimates for the time series model intercepts. The slopes of time series excluding the cirrus band relative to time series including the cirrus band show the same patterns as the intercept estimates (Figure 2·10) with large differences the visible bands and NDVI slope estimates. Eliminating cirrus clouds from time series images also contributes toward lowering the RMSE of time series models as shown in Figure 2·11. The skew in the distribution of relative RMSE estimates shows that many of the time series contain more noise without the cirrus band in all bands and indices, but especially in the visible bands and NDVI.
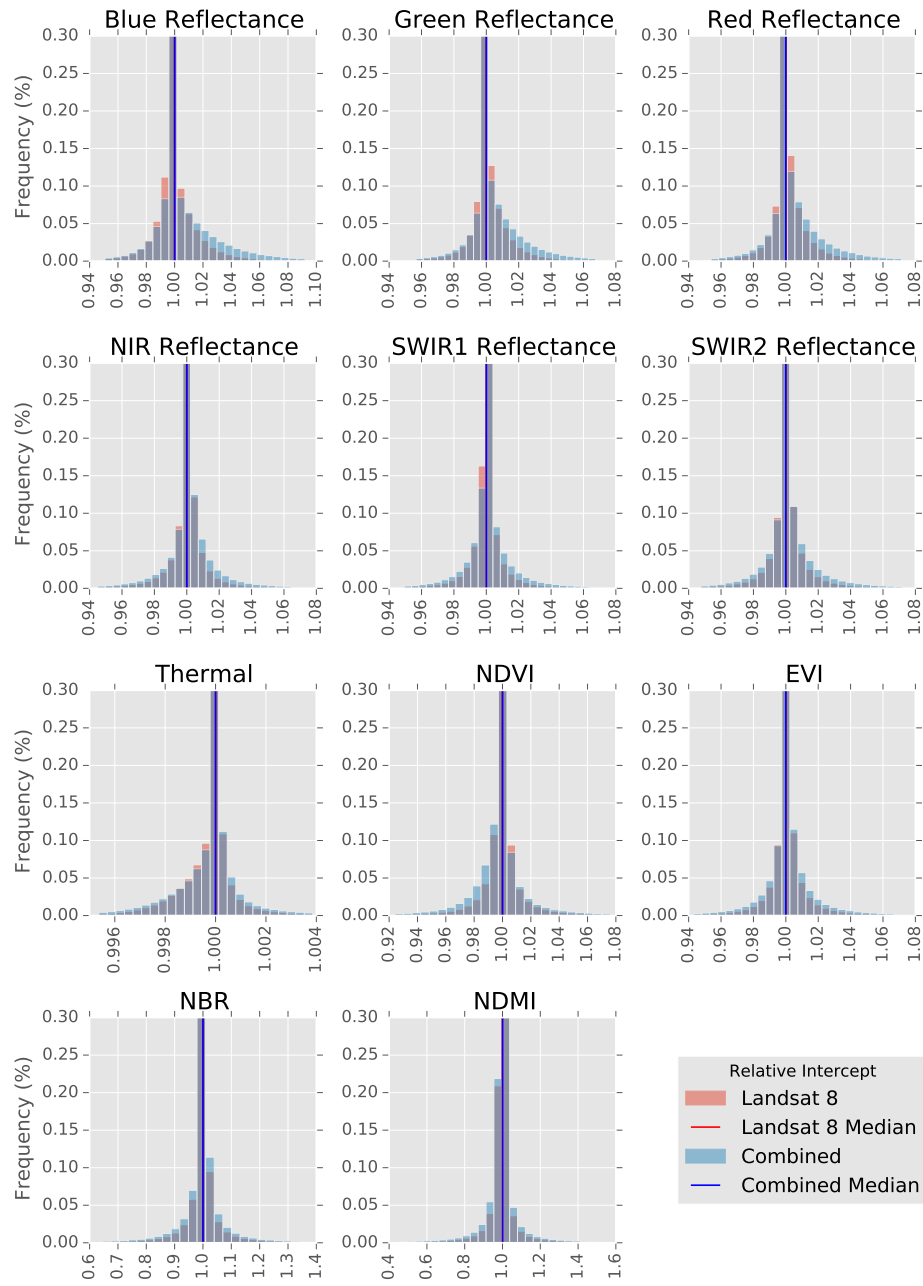
Figure 2·9: Normalized intercept estimates for time series models in Colorado, United States (P034R032) used to assess the influence of Landsat 8's cirrus band. Intercept estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by intercepts from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
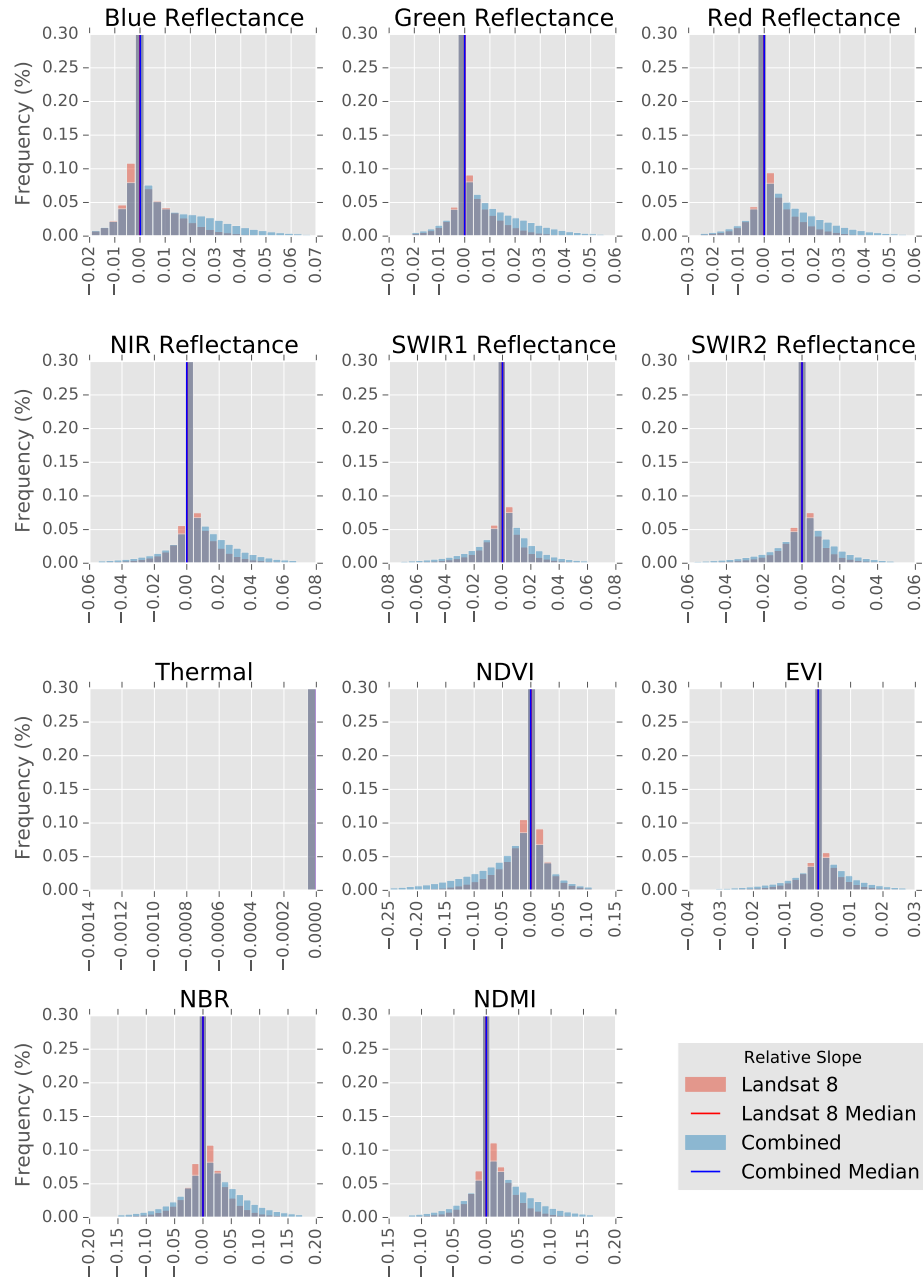
Figure 2·10: Normalized slope estimates for time series models in Colorado, United States (P034R032) used to assess the influence of Landsat 8's cirrus band. Slope estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by subtracting slopes from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
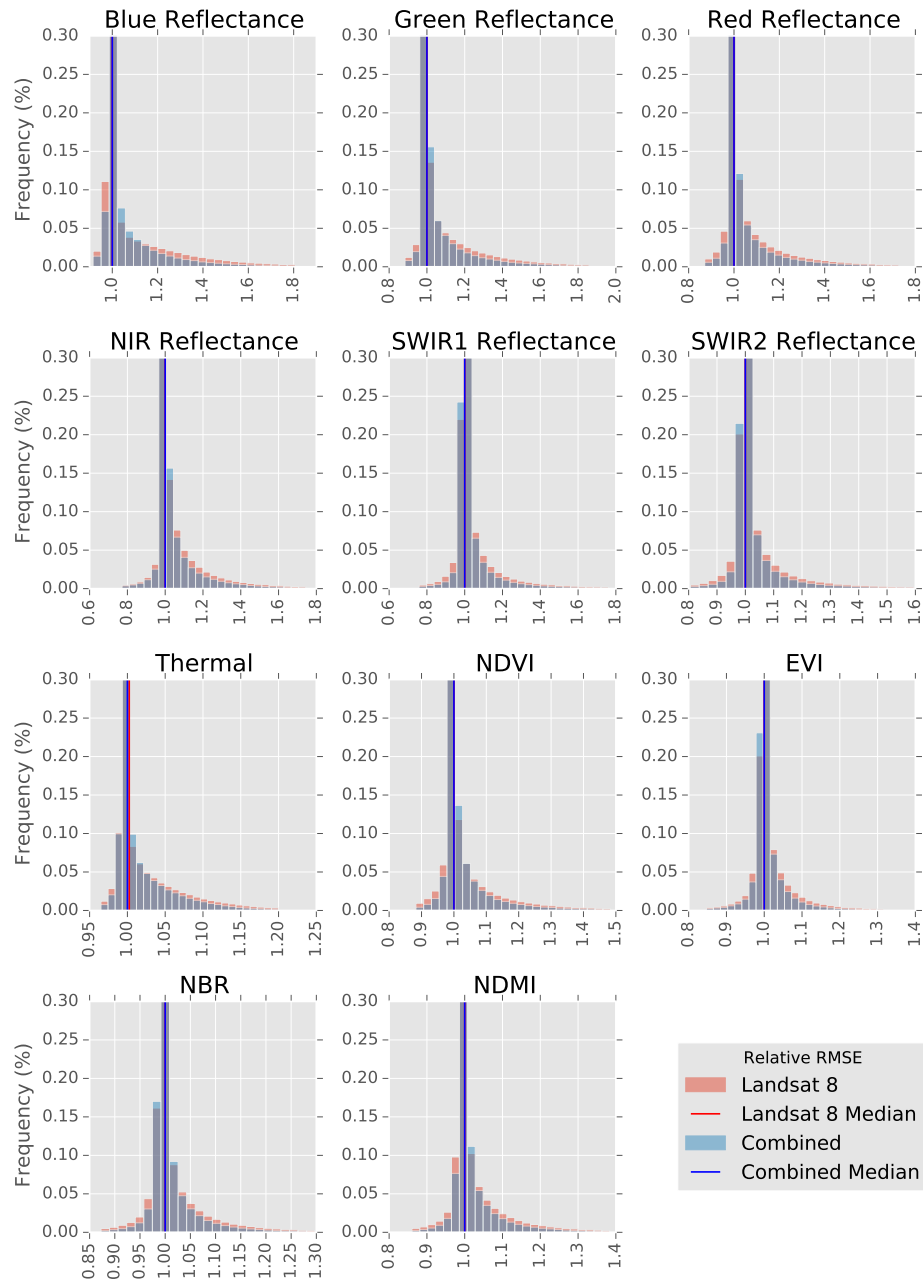
Figure 2·11: Normalized RMSE estimates for time series models in Colorado, United States (P034R032) used to assess the influence of Landsat 8's cirrus band. RMSE estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by RMSE from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

We did not specifically limit our analysis of time series attributes to pixels containing observations identified as cirrus clouds within the time series. Instead, we drew a random sample of pixels within each scene. As such, we do not expect there to be a systematic shift in slopes, intercepts, and RMSE estimates within the time series models not using cirrus information in the Fmask images relative to time series models that did mask cirrus clouds. The shifts in the tails of the relative distributions of these time series variables, positive in the visible bands and the negative in NDVI, are strong evidence to the benefits of the cirrus band. Landsat 8 observations will be much less noisy than observations from previous sensors and this difference might affect future analysis. Measurements in the visible bands from Landsat 8, integrated over enough time, are especially likely to be darker than the response in TM and ETM+, all else equal, simply due to this enhanced screening. The development of methods to screen historical Landsat TM and ETM+ observations of cirrus clouds, already an important task, should be an increased priority to bring these time series closer to parity with Landsat 8.

These results are reminders of previously established findings demonstrating how poor cloud and atmospheric contamination screening in time series can lead to spurious temporal trends or anomalies (Samanta et al., 2010; Samanta et al., 2012). Improved noise screening will also improve the consistency of surface characterization by removing artificial differences in spectral properties caused by the random chance of cirrus cloud contamination. Reductions in the amount of unexplained variance in reflectance caused by previously undetectable cirrus contamination also aids change detection of surface conditions or land cover because actual changes in surface reflectance will be more distinguishable if the overall variation decreases.

2.5 Conclusions

This study of near coincident "underflight" observations and time series analysis finds consistent differences in surface reflectance between Landsat 7 ETM+ and Landsat 8 OLI Climate Data Record (CDR) surface reflectance product in the visible wavelengths. While the near infrared and shortwave infrared bands appeared brighter in Landsat 8 within the "underflight" data, differences between sensors for these bands were small and inconsistent in the time series results. Vegetation indices using the visible bands, such as the NDVI and the Tasseled Cap Greenness, tend to be brighter in Landsat 8 and have larger differences than the individual reflectance bands because the biases in the visible and near infrared bands are in the same direction these indices attempt to contrast. Despite similarly contrasting the near infrared with the red wavelengths, the Enhanced Vegetation Index (EVI) is not very different between Landsat 7 and Landsat 8. The NBR and NDMI spectral indices, which contrast the shortwave infrared bands with the near infrared band, also do not vary across sensors. While these results were generally consistent across the Landsat footprints analyzed, the relationship between Landsat 7 and Landsat 8 varied too substantially to suggest that a global relative correction method could be applied successfully. As was also suggested by Flood (2014), a local, perhaps even per-scene, relative correction should be performed to incorporate the spatial variability of the sensor differences.

Systematic differences in surface reflectance between Landsat 7 and Landsat 8 are shown to have perilous consequences for inference of gradual and abrupt change in land surface conditions. The extension of Landsat 7 time series with Landsat 8 data produces spurious time trends in bands or indices affected by sensor specific differences. Until Landsat 8 surface reflectance can be operationally adjusted to match surface reflectance from Landsat 7, it is recommended to either estimate gradual changes over time using data from only one sensor, to perform relative correction

manually, or to avoid affected bands and indices. The sensor specific bias in Landsat 8 data also hampers detection of abrupt changes because time series models are artificially noisier if variation between sensors is left unexplained.

This study did not address the improvements in time series analysis that are due to the resumed 8-day repeat cycle possible with two Landsat spacecraft in operation. Given that much of Landsat 7's acquisition footprint is unobserved due to the failure of the Scan Line Corrector, the launch of Landsat 8 has more than doubled the observation frequency since the failure of Landsat 5. Increasing the sampling frequency of observations helps time series models capture rapid intra-annual transitions, such as the phenological transitions in deciduous forests at the start and end of the growing season or the number of cropping cycles in one year. Near real time monitoring of land cover change is significantly improved with two sensors in orbit because information on change can be retrieved in a more timely manner. Land cover change techniques that use annual image composites can now be more selective when selecting the best available pixel. Even the actual detection of land cover change might be more accurate as higher frequency observations make it more likely to capture the land surface during or immediately after the change event. Increased observation rates are especially useful in locations with pervasive cloud or snow cover or short growing seasons, including the incredibly important and changing tropical and boreal ecosystems.

Cirrus clouds detected by Fmask using Landsat 8's new cirrus band but not detected in Landsat 7 data are most distinct in the visible and thermal bands. Observations of cirrus clouds differ from clear land in visible wavelengths by less than 3% to 4% and show very little difference in the infrared wavelengths. An experiment comparing Landsat time series that do not use Landsat 8's cirrus band to screen cirrus clouds against time series that do remove cirrus clouds show that residual cirrus clouds create noticeable biases in the intercept, slope, and RMSE estimates of the

visible bands and in NDVI. These contamination induced biases mimic vegetation dynamics of interest, including browning or stress events, which highlights the value of Landsat 8's cirrus band. Efforts to enhance existing cloud mask products by improving detection of cirrus clouds within TM and ETM+ data may be aided by the characterization of cirrus clouds possible using Landsat 8 observations and would be a valuable addition.

Inconsistencies in surface reflectance between Landsat 7 and Landsat 8 observed in the Landsat 8 underflight dataset and in four Landsat scenes are due to a combination of three or more factors including physical changes to the spectral characteristics of Landsat 8, differences in relative calibration of sensor detectors, and differences in atmospheric correction ancillary datasets and methodology. This study did not seek to quantify the relative contribution of each factor to the differences observed, but instead focused on quantifying the observed effects and tried to offer practical advice for incorporating Landsat 8 information into existing time series methodologies. Image normalization procedures are recommended to equate the surface reflectance from Landsat 8 to Landsat 7, especially in the visible bands. The difference between sensors might also be modeled as a simple offset in reflectance which could be accomplished as an addition to a time series model specification. It is important to note that the Landsat CDR surface reflectance results described within are subject to change with alterations to sensor calibration or algorithm performance, but the techniques used and the lessons learned should be applicable when trying to extend the Landsat time series with observations from the upcoming Sentinel-2 satellite.

Chapter 3

Forest Change Detection By Radar/Optical Fusion

3.1 Introduction

Integration of optical data from medium resolution sensors like Landsat with medium resolution SAR data has been identified as a key research priority (Reiche et al., 2016) in the effort to better monitor our planet's forest resources. Radar satellites have historically been shorter lived than optical sensors, making data continuity difficult. Fortunately, several SAR missions have been launched or are planned that will produce consistent datasets suitable for time series analysis. Sentinel-1A and Sentinel-1B were launched in 2014 and 2016 respectively, and combined provide repeated observations in the C-band every 6 days at 10m spatial resolution (Torres et al., 2012). ALOS-2 was launched in 2014 in part to support global forest monitoring (Kankaku et al., 2014; Motohka et al., 2017), and has a potential 14 day gapless repeat time in ScanSAR mode with an acquisition strategy that aims for at least two acquisitions per year on all land except Antarctica, with potentially six per year for the tropics and some parts of the boreal zone (Rosenqvist et al., 2014). The National Aeronautics and Space Administration (NASA) and Indian Space Research Organization (ISRO) developed SAR, or NISAR, is expected to launch in late 2021 and will provide full polarimetric observations in L- and S-bands with a potential 12 day repeat depending on observation mode (Rosen, Hensley, et al., 2017; Rosen, Kim, et al., 2017; Sharma

et al., 2018). In historic analyses, data from past SAR sensors like the Japanese Earth Resources Satellite 1 (JERS-1) or the Phased Array type L-band Synthetic Aperture Radar (PALSAR) on the Advanced Land Observation Satellite (ALOS) platform, may also be useful in complementing optical time series, especially where there are gaps in the Landsat record.

A recent review of radar and optical fusion studies found that integration of SAR data with optical data increased result accuracy in 28 out of 32 cases (Joshi et al., 2016). Unfortunately, only 11 studies out of the 112 studies surveyed utilized optical and SAR data for land cover or land use change detection with the majority of studies focusing on single date classification. Joshi et al. (2016) also concluded that the vast majority of the experiments summarized were conducted over very small areas (300 - 3,000 km^2), did not have systematic or standardized means for comparing data fusion against single sensor results, and no study compared performance of any one data fusion method to another fusion method.

3.1.1 Background

One category of algorithm for fusing different sources of data are algorithms that predict one data source from another. A common example of this category is efforts to calibrate data from new versions of a satellite family with data from previous sensors. This type of fusion is frequently accomplished by fitting a linear regression model that corrects one source to another, as was the case in the Landsat 8 and Landsat 7 calibrating equations from Roy, Kovalskyy, et al. (2016) and Flood (2014). Increasing in complexity, the Harmonized Landsat Sentinel (HLS) product (Claverie et al., 2017) creates a fused time series by co-registration, bandpass adjustment, BRDF correction, and spatial resampling. Perhaps the best known example of this style of algorithm, the Spatial and Temporal Adaptive Reflectance Fusion Model (STARFM; Gao et al. (2006)) algorithm models the temporal and spatial relationship between

observations from MODIS and Landsat in order to create synthetic imagery with the temporal frequency of MODIS and with the spatial information from Landsat. This concept was also employed to improve near real time monitoring of deforestation by Xin et al. (2013) and Tang (2018). Predicting Landsat data from MODIS data is complicated by the vastly different observation footprints and real changes that may happen between Landsat observations, but the two sensors measure roughly the same portion of the electromagnetic spectrum and are cross-calibrated when possible. Despite the difficulties inherent in even fusing optical data from different sensors, Reiche, Verbesselt, et al. (2015) found that a weighted regression model could be used to fuse Landsat NDVI and ALOS-1 HH/HV ratio for deforestation mapping.

Rather than try to predict one from another, a second category of fusion algorithms combines data from multiple sensors by conversion to a new quantity. The most common example of this type of fusion are algorithms that fuse inputs by conversion to a probability quantifying membership in some group or class, typically derived from a supervised classification of land cover types. Although not developed for fusion purposes, the Vegetation Continuous Tracker (VCT) developed by Huang et al. (2010) can be described in these terms if one, for example, considers the similarities between the "Forest Z-Score" statistic and a forest class used in a Gaussian Naive Bayes classifier — both are parameterized by sample mean and standard deviation. Most importantly, conversion of individual observations to a Z-Score, or similar metric, creates a unitless variable that could come from any number of sources. This type of approach can take advantage of decades of research into classification problems in remote sensing, with approaches being distinguished among familiar contrasts of univariate to multivariate, parametric versus nonparametric, linear and non-linear, and more.

Based on this principle of "probability fusion", Solberg et al. (2008) combined

time series of optical and SAR data to monitor the progression of snow melt through a variety of stages. They modeled the distribution of optical data observations of snow cover using multiple Gaussian distributions and used a single Gaussian distribution to model the SAR data. They derived class probabilities for each observation from these distributions, and used a Hidden Markov Model (HMM) and transition probabilities that represented the directionality of the melting process to map snow conditions over time. In another set of analyses, Trier and Salberg (2011) used a HMM with time series of Landsat data for deforestation monitoring, and Salberg and Trier (2011) showed that their approach may also be used for mixing SAR and optical data. They estimated probabilities for forest, sparse forest, grass, and soil classes using multivariate Gaussian distributions, and used assumed values for the class transition probability matrix, though they noted that this could be estimated from the data. Reiche, Verbesselt, et al. (2015) used a "Bayesian updating" approach to fuse Landsat and ALOS-1 data for deforestation monitoring in a forest plantation in Fiji. They distinguished between forest and non-forest classes and used one variable from the optical (NDVI) and radar (HH/HV ratio) data sources each. Specifically, they parametrized the distribution of the NDVI and the HH/HV ratio using a single Gaussian distribution for non-forest and a single Weibull distribution for forest training data. The conditional probability of deforestation was estimated at each time step, and deforestation was recorded if this probability exceeded a user defined threshold. Reiche et al. (2018) also used this method in a seasonally dry topical forest to fuse Landsat NDVI, ALOS-2 PALSAR-2 L-HV, and Sentinel-1 C-VV data. They used spatial normalization (Hamunyela, Verbesselt, De Bruin, et al., 2016; Hamunyela, Verbesselt, and Herold, 2016; Hamunyela et al., 2017) to reduce the effects of seasonality, and used Gaussian distributions to model forest and nonforest classes for all data sources.

Algorithms like the Continuous Change Detection and Classification (CCDC; Zhu and Woodcock (2014)) or BFAST Monitor (Verbesselt et al., 2012) that rely on monitoring forecast residuals for change are widely used in the optical remote sensing community, but have not yet been used for fusion. In CCDC and BFAST Monitor, regressors are picked to represent the overall reflectance (intercept), the trend in reflectance over time (time trend), and seasonality (cosine and sine harmonics) with the goal of transforming the time series data into normally distributed forecast residuals with unit variance. Radar time series may require different independent variables to model, but this estimation of forecast residuals inherent to these methods presents an opportunity for fusion. The CCDC algorithm, for example, performs a very similar calculation to the "Forest Z-Score" from VCT when future observations are compared with model predictions and scaled by a measure of model uncertainty. While the forest class mean and standard deviations used in VCT create a unitless metric that describes how similar an observation is to a forest, the time series model prediction and RMSE from CCDC describe how similar an observation is to an adaptive running expectation. Assuming that changes to land cover create changes in both optical and radar signals, forecast residuals from either data source might be combined.

3.1.2 Objectives

In this chapter, we develop and test two fusion algorithms: a "Probability Fusion" algorithm based on time series of classification probabilities and a "Residual Fusion" algorithm that applies the concept of converting time series into forecast residuals. We attempted to answer the following three questions:

- Does adding observations from SAR into a Landsat time series improve the spatial and temporal accuracy of deforestation maps?

- How does the benefit of adding SAR observations change under different Land-

sat data availability condition?

- Which approach to optical and radar fusion is better: "Residual Fusion" or "Probability Fusion"?

3.2 Data

3.2.1 Study Site

We selected an area of evergreen tropical forest in the lowlands of the Peruvian Amazon near Yurimaguas, Loreto, Peru. Yurimaguas is 106m above sea level, and the study area is approximately centered on S6°2.633′ W76°6.5′ for this analysis. Burga Ríos (2016) attributed most deforestation in this region of Peru to conversion of forests to grow coca or as part of migratory agricultural practices that are used because of the relatively poor soil fertility in the Amazon. This study also identified mining, illegal logging, the construction of highways, and large scale industrial cultivation (usually of palm oil) as recent drivers of deforestation in Loreto, Peru. Recently, and especially around the border of Loreto and San Martín, deforestation has mostly been caused by large scale conversion of forests to oil palm plantations with increases of approximately 1,480 ha between 2009-2010 and 2,500 ha between 2010-2011 (Urquiza Muñoz and Burga Ríos, 2016; Cañote Amaya, 2018).

While large deforestation patterns can be observed from industrial scale production of palm oil, the Yurimaguas area also has a fairly complicated set of land uses related to agroforestry. Sotelo Montes and Weber (1997) interviewed farmers from Loreto, including from Yurimaguas, and found that they were familiar with and used a relatively large number of agroforestry species, and prioritized commodity crops like *Bactris gasipaes* (oil palm), fruit trees like *Inga edulis*, and rapidly growing hardwood tree species for lumber like *Cedrelinga cateniformis*, *Guazuma crinita*, and *Calycophyllum spruceanum.*

The vegetation, soil, and detritus around Yurimaguas contain a relatively large amount of carbon, even in successional, disturbed, or agroforestry cover types. Mature tropical forests surrounding Yurimaguas may reach 290 MgC ha^{-1} of live woody biomass after 40 years with only light extraction, while secondary forests may reach 184 MgC ha^{-1} after 15 years or 42 MgC ha^{-1} after 5 years (Guzmán and Arévalo, 2003; Viena Vela, 2011). Agroforestry systems in this area also store a significant amount of carbon, with palm oil storing approximately 148 MgC ha^{-1} after 16 years when planted in monoculture plantations and approximately 114 MgC ha^{-1} when planted with other species as part of a multi-level canopy (e.g., "shade grown") (Guzmán and Arévalo, 2003; Viena Vela, 2011).

Yurimaguas presents an excellent opportunity to study the effects of fusing optical and radar data for deforestation monitoring because deforestation in this area comes from complicated, small scale agroforestry and from the installation of large, commercial palm oil plantations. Most importantly for this study, a large amount of these changes occurred during the ALOS-1 PALSAR lifespan.

3.2.2 Passive Optical Data

We acquired all observations from the Landsat TM and ETM+ between 2003 and 2013 for WRS-2 Path 8 and Row 64, for a total of 183 images (75 Landsat 5 and 108 Landsat 7). We used the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm (Masek et al., 2006) to perform atmospheric correction to surface reflectance values. Cloud and cloud shadow objects were identified and removed from analysis using the FMask algorithm (Zhu and Woodcock, 2012). The atmospheric correction and cloud masking steps were performed by the United States Geological Survey (USGS) as part of the Climate Data Record (CDR) product produced by the USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) (U.S. Geological Survey, 2015).

3.2.3 Radar Data

We acquired all available acquisitions from the ALOS-1 PALSAR instrument from Frame 7060 to maximize overlap with the data from Landsat. We used the high resolution (12.5m pixel size) Radiometrically Terrain Corrected (RTC) product from the Alaska Satellite Facility Distributed Active Archive Center (ASF DAAC) (ASF DAAC, 2015). The ALOS-1 PALSAR RTC product from ASF undergoes a relatively high level of preprocessing, and are delivered in units of gamma nought power after undergoing co-registration with a DEM, radiometric terrain correction (Frey et al., 2013), and geocoding to 12.5m pixel spacing (Gens, 2015). Of the 17 acquisitions that were available, we had to discard 4 images because of ionospheric interference. Deforestation is easier to see in the crosspolarized L-HV channel than in the L-HH channel (Shimada et al., 2014; Reiche, de Bruin, et al., 2015), so we discarded an additional two single polarization (HH) images and were left with a total of 11 dual polarization (HH and HV) images.

To reduce speckle noise inherent to the radar signal, we applied the Quegan multitemporal speckle filter on each channel (HH and HV) of the radar time series (Quegan and Yu, 2001). The Quegan multitemporal filter uses information from the temporal domain of coregistered multidate SAR imagery to reduce the spatial filter window size needed for the noise reduction. Speckle filtering was performed on the original 12.5m spatial resolution imagery using a 7 pixel window. Finally, the ALOS data were resampled to 30m to match the Landsat data's grid using bilinear resampling. Bilinear resampling was chosen over nearest neighbor in order to further reduce noise. Once all preprocessing steps were complete, the ratio of HH to HV polarization was calculated in log space by subtracting the HV channel from HH.

3.3 Methods

We developed two new methods of fusing optical and radar data into a time series that may be monitored for changes. Sections 3.3.1 and 3.3.2 describe the two approaches for fusing the time series. Once the data are fused, both algorithms use the same method of change detection (see Section 3.3.3) based on the approach used by the CCDC (Zhu and Woodcock, 2014) algorithm.

3.3.1 Residual Fusion Algorithm

The "Residual Fusion" algorithm uses the forecast regression approach taken by algorithms like CCDC (Zhu and Woodcock, 2014) or BFAST (Verbesselt et al., 2012) to convert both the optical and radar time series into scaled forecast residuals. As shown at the top left of Figure 3·1, radar data from ALOS PALSAR is combined with a time series model specific to this data stream to produce "scaled forecast residuals". Through experimentation and analysis of sample time series, we found it appropriate to model the L-HV data using only an intercept and time trend (Equation 3.1) because this data is not affected by seasonality in our study location, though one could use more sophisticated models (e.g., in the case of shorter wavelength radar data that is affected by vegetation phenology).

$$L\text{-}HV_i \sim 1 + \beta Time_i + \epsilon_i \tag{3.1}$$

This time series model transforms radar backscatter into a series of scaled residuals with known statistical quantities (zero mean and unit variance). The optical data from Landsat undergoes the same process of combining a time series model with data to generate scaled forecast residuals, but we adapted the model to the character of the opitcal data. In the case of the Landsat data, we noticed some within-year variability, likely due to sun angle effects, and decided to pick a harmonic regression model that

would account for seasonal variability. Accordingly, the "Time Series Model" (see top right of Figure 3·1) chosen for the Landsat data is provided in Equation 3.2:

$$y_i \sim 1 + \beta Time_i + \delta_1 \cos \frac{2\pi}{365.25} Time_i + \delta_2 \sin \frac{2\pi}{365.25} Time_i + \epsilon_i \qquad (3.2)$$

The modeling of each time series produces coefficient estimates that have proven useful for visualization, and land cover classification (Zhu and Woodcock, 2014; Zhu, Woodcock, Holden, et al., 2015; Zhu, Fu, Woodcock, Olofsson, Vogelmann, Holden, et al., 2016; Pasquarella, Holden, and Woodcock, 2018), but we are interested in the error term for each observation, ϵ_i. These residuals should have zero mean if we're using an unbiased regression estimator, and we can use the estimate of the variance of the residual time series to scale the residuals to also have unit variance.

Once each data source has been converted into scaled forecast residual time series, the data may be combined by simply merging the time series together in chronological order because all data sources will have the same statistical properties (zero mean and unit variance). Despite the radar backscatter data and the optical reflectance or spectral index data originally coming in different units of measurement, both have been converted to a unit of measurement (scaled forecast residuals) that measures the departure at each time step from some data specific expectation.

Depending on the band or spectral index used for the Landsat data (y), the spectral signal of deforestation might have a different directionality than the typically decreasing backscatter observed during deforestation events. In order to align the signal from optical data that increase when deforestation occurs, a "directional alignment" step (see Figure 3·3) is performed by multiplying the scaled forecast residual time series by negative one to invert the sign. While the change in spectral characteristics of forest loss is reasonably well characterized (increase in visible and SWIR, but decreases in vegetation indices; decrease in L-HV), the more complicated spectral

signatures of other land cover type conversions may complicate the use of this algorithm for finding change. During algorithm development, we found that the SWIR1 channel on Landsat was the best single band to use for finding deforestation, though the algorithm could work with other bands or spectral indices.

After all input data streams (optical and radar) are converted to scaled forecast residuals, they may be merged together into one fused time series by simply combining and sorting the data by date. At this point, the time series have been fused through this "Residual Fusion" process and may be analyzed for change using a variety of algorithms (see 3.3.3).
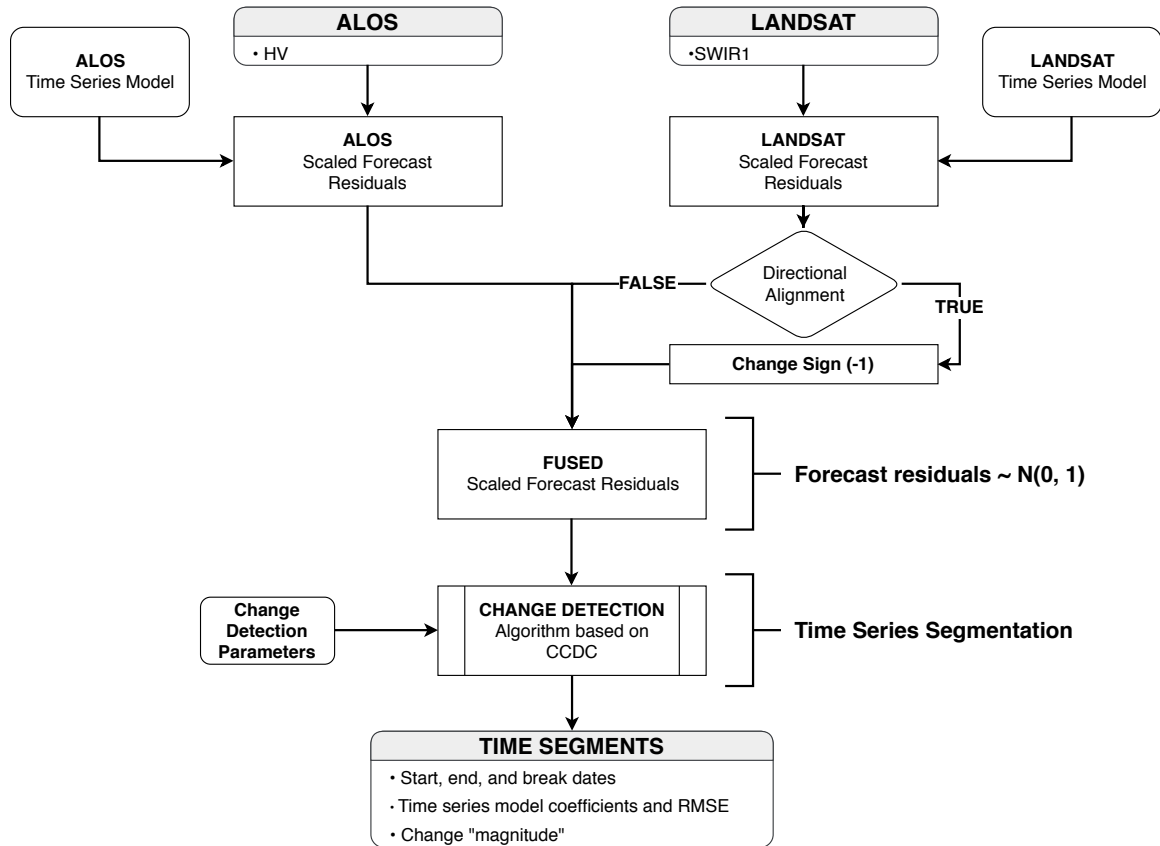
Figure 3·1: Residual fusion algorithm process diagram.

Figure 3·2 shows an example of the "Residual Fusion" algorithm steps for a pixel that experienced deforestation. The top panel shows a slight increase in the SWIR1
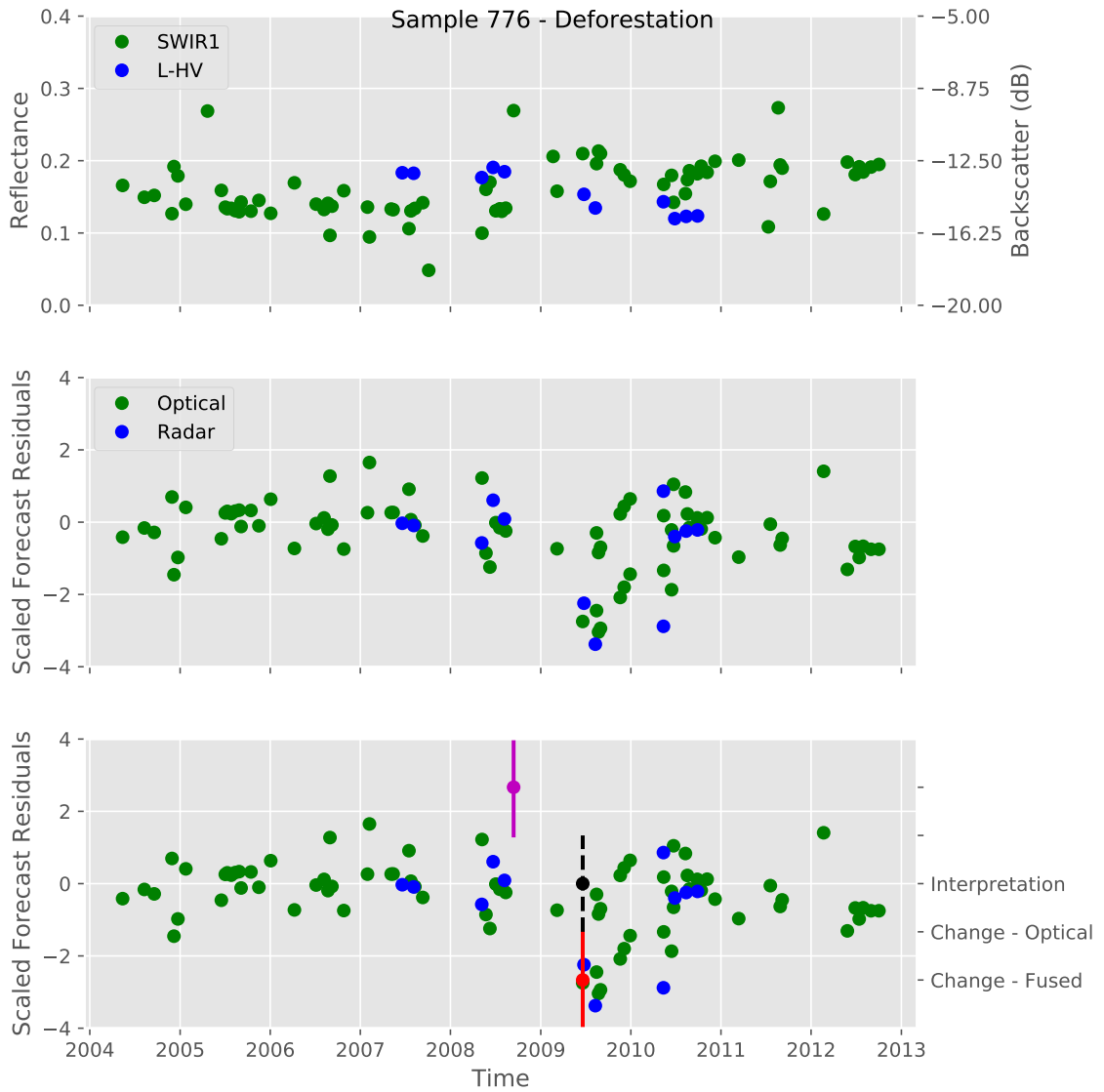
Figure 3·2: "Residual Fusion" results for sample 776 from our accuracy assessment. The top panel shows the original time series in SWIR1 and L-HV bands. The middle panel shows the fused time series of scaled forecast residuals estimated when fusing both data sources. Breaks detected by the "Residual Fusion" algorithm for both fused and non-fused results are shown as vertical lines on the bottom 66% of the graph, while the date of change given by the assessment interpreter is shown as the magenta line in the top 33% of the graph.

band and a slight decrease in the L-HV channel around 2009. The middle panel shows the time series of scaled forecast residuals for each data source. The forecast residual time series both have been transformed to have zero mean and unit variance, and so can be combined. We apply a change detection approach similar to CCDC (see Section 3.3.3) to the forecast residual time series to detect change. In this example, both the fused and non-fused runs detected change later than the interpretation. We are showing the radar time series in units of decibels, but the forecast model and change detection analysis is performed using linearly scaled units of power.

3.3.2 Probability Fusion Algorithm

The "Probability Fusion" algorithm is very similar to the approaches used by Solberg et al. (2008) and Reiche, de Bruin, et al. (2015) and Reiche et al. (2018). It depends on two main inputs: observations from optical and radar sensors and training data (see Figure 3·3). The training data are required to generate conditional class probabilities from a supervised classifier, and so collection of training data is a necessary and important step in this methodology. First, we decide the land cover categories we wanted to include for mapping deforestation, which included categories for stable forests, "herbaceous" (pastures and agriculture), impervious surfaces and barren land (development, river banks, mining), and water. To better distinguish deforestation as quickly as possible, we also included training data polygons that captured the spectral reflectance right after deforestation. Significant clouds and cloud shadows remained in the optical imagery, even after screening using FMask (see 3.2.2), so we also included training data targeting clouds and their shadows. Observations that had conditional probabilities for these "noise classes" larger than 0.5 were removed from the time series analysis before running the algorithm.

We collected the training data by drawing polygon "regions of interest" on Landsat and ALOS imagery and giving each polygon a reference class label. We checked
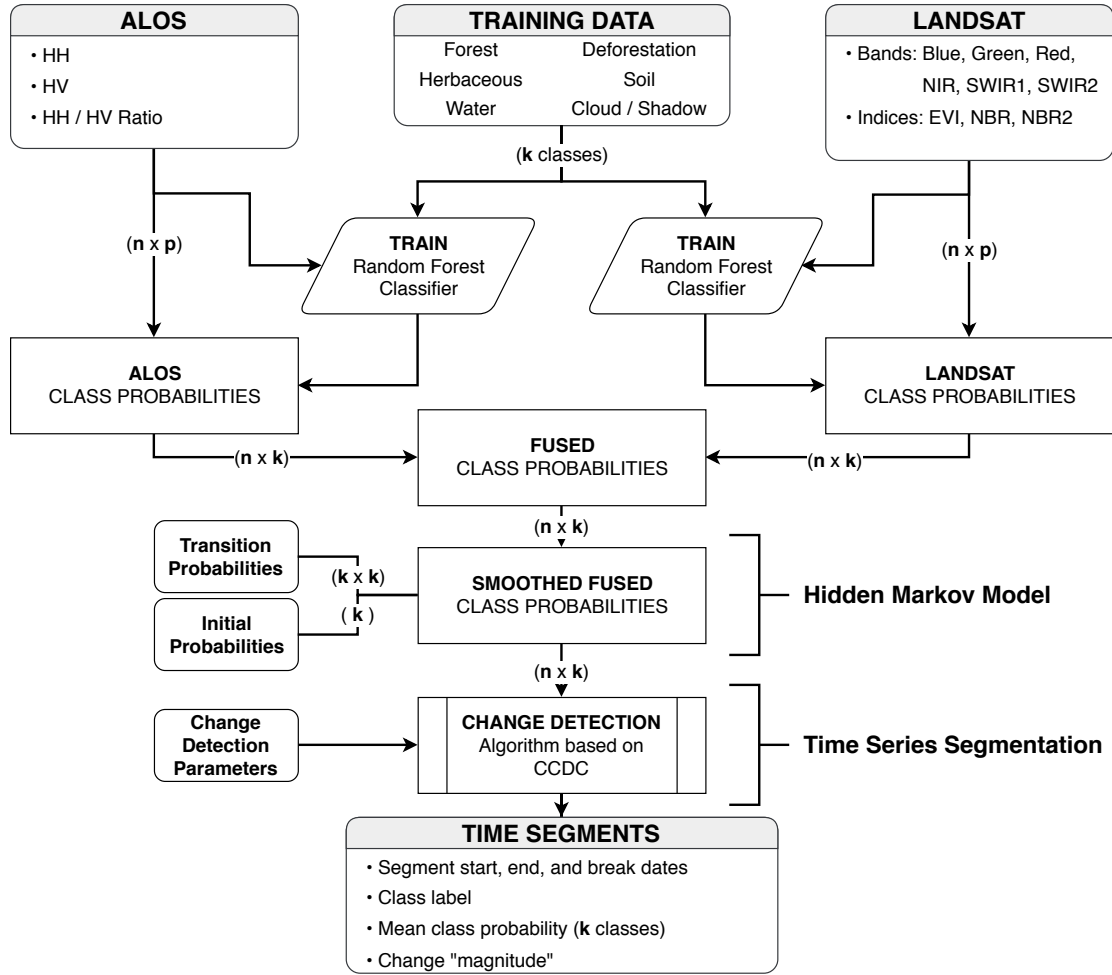
Figure 3·3: Probability fusion algorithm process diagram.

the time series of pixels within each polygon to ensure stability, and recorded what time periods the interpretation was valid for. Recording this timing information meant that training data "features", or independent variables, could be drawn from multiple dates of Landsat and ALOS imagery at once. Training data were collected in an iterative manner, by repeatedly assessing time series plots and single date images of the predicted class posterior probabilities and collecting additional training data where the classification posterior probabilities were poorly characterized. For example, we analyzed time series of forest class posterior probabilities alongside spectral reflectance time series to ensure that the forest probability was close to zero imme-

diately after a disturbance. In instances when this did not occur, we added training data from clear acquisitions before the disturbance for forest and data from after the disturbance to the "deforestation" class. We repeated this process several times to resolve misclassifiations from our trained classifiers.

The "features", or independent variables, used for the ALOS-1 data were only the original HH and HV bands and the ratio of these bands. The Landsat data have more difficulty in discriminating between forest and non-forested vegetation than L-HV, which was used globally to discriminate between these classes based on a regionally specific threshold in Shimada et al. (2014). To help increase separability using Landsat data, we used all of the reflectance data bands from Landsat and the EVI, NBR, and NBR2 vegetation indices.

We used the Random Forest algorithm (Breiman, 2001) implementation from the "scikit-learn" Python package (Pedregosa et al., 2011) in supervised classification mode to derive the conditional class probabilities for all time steps (n) and cover classes (k). We selected Random Forest because it has been widely applied, is non-parametric and can fit to non-linear distributions, usually has very good performance even without calibrating hyperparameters, and can perform well even with many, possibly redundant, features (Fernández-Delgado et al., 2014; Caruana and Niculescu-Mizil, 2006; Zhu, Woodcock, Rogan, et al., 2012). The conditional class probability was calculated for each class by dividing the total number of trees in the ensemble that voted for the class by the total number of trees.

Once both time series are converted into class probabilities, we apply two final steps before attempting to test for change. First, following Reiche, de Bruin, et al. (2015) and Reiche et al. (2018) we apply a clipping operation to limit the class probabilities to between 0.1 and 0.9 so as to avoid implying that we have very high or very low confidence as to the class membership using any single observation. Second,

we apply a first-order Hidden Markov Model (HMM) and use the "forward-backward" algorithm to smooth the time series of class probabilities. Hidden Markov Models are advantageous in this circumstance because they allow for information from the entire time series to influence the probability of class membership at any point in time. This is important because we know that land cover classes do not change rapidly from one observation to the next, but the time series of class probabilities may have sudden drops or spikes in class probability due to confusion from sources of noise in the data, like missed clouds or shadows.

HMMs accomplish this smoothing by conditioning the likelihood of class membership for any given observation on the likelihood of past observations and on the likelihood of transitioning from one class to another. Adopting the notation of Abercrombie and Friedl (2016), we can describe the process of estimating the joint probability of a pixel belonging to a given land cover label through the time series sequence as:

$$P(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n | \Omega) P(\Omega) = \prod_{t=1}^{N} P(\omega_t | \omega_{t-1}) P(\mathbf{x}_t | \omega_t) \tag{3.3}$$

where:

- $\mathbf{x}_t$ is the spectral measurement at time t

- Ω is the set of all possible classes, ω_i to ω_n

- ω_t is class at time t

- $P(\omega_t | \omega_{t-1})$ is the transition probability from the class at the last time step to the current time step

- $P(\mathbf{x}_t | \omega_t)$ is the likelihood of the spectral data at time t given the land cover label ω_t. These probabilities come from a supervised classification in our algorithm,

and are often referred to as "emission probabilities" in the context of HMMs.

As described by the arrow labels in Figure 3·3, we observe n measurements through time of p variables, where $p = 3$ for ALOS-1 and $p = 9$ for Landsat. Figure 3·4 describes these data as $X = x_1, x_2, \ldots, x_n$, which may come from either optical or radar data sources. The Random Forest classifiers convert the n observations of p features in X to $P(x_t | \omega_t)$, which describes the likelihood of observing the data given membership in label ω_t for all n observations and all k labels ($k = 6$).
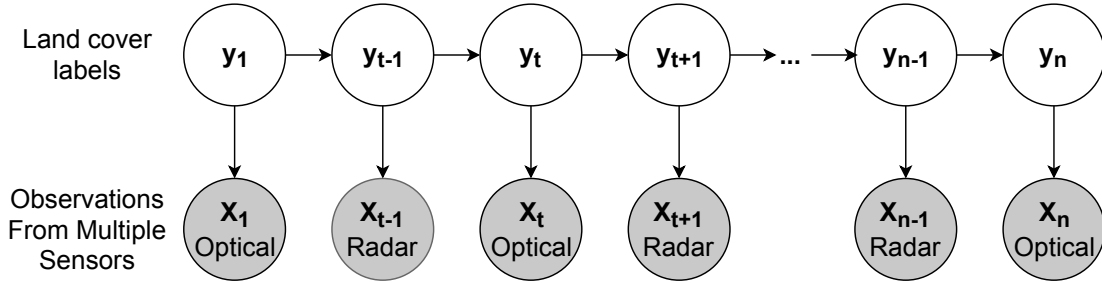
Figure 3·4: Diagram of the HMM used in this study. We directly observe spectral measurements x_t in p bands or channels for n observations, and use this information to infer the land cover label at each time y_t.

The calculation of the marginal probability for all observations according to Equation 3.3.2 is different depending on if you start with the beginning of the time series and move forward in time, or start with the end of the time series and move backward in time. As such, the "Forward-Backward" has been developed to address this inconsistency by running the Markov chain forward, backward, and then combines and normalizes both sets of information to calculate the marginal probability for each land cover label for each observation (Rabiner, 1989). We parameterized the HMM initial and transition probabilities using "naive" assumptions — all classes had equal initial probabilities and all transitions were assumed to occur with equal probability (0.05). In practice, these parameters may be either tuned for algorithm performance

or there may be good reasons for knowing the probabilities *a priori*, for example if one already had a map of the area from which to derive starting probabilities.

The marginal probability of n observations coming from each of the k classes that is calculated using the "Forward-Backward" algorithm is a smoother time series than the original probabilities output from a classifier because class transitions are parameterized to be unlikely events, which reduce the impact of noisy data that suggests frequent transitions. Consistency through time in the absence of real change is important because these data will be used to search for changes in the time series using a process similar to CCDC described in Section 3.3.3).

Figure 3·5 shows an example of the "Probability Fusion" algorithm steps for a pixel that experienced deforestation (also shown in Figure 3·2). The top panel shows the original data, and the middle panel shows the time series of forest land cover emission probabilities from Random Forest for each data source. These probabilities are "fused" together, and a HMM is applied to the fused emission probability time series to generate the "smoothed" time series shown in the bottom panel. The results of the change detection on this fused, "smoothed" time series detected change on the same date as the interpreter and the likelihood of forest for the remainder of the time series is very low, indicating a permanent transition.

3.3.3 Fused Time Series Segmentation

Both the "Probability Fusion" and "Residual Fusion" algorithms have been developed to identify change based on the approach used by the CCDC algorithm. The CCDC algorithm decides if a time series has changed if some number of consecutive observations in the time series are significantly different than the expected forecast from a time series model based on a user defined threshold. Multiple, consecutive observations are required in this approach in order to confirm that a change in one date is not caused by ephemeral conditions or noise. If the first of the next consecutive
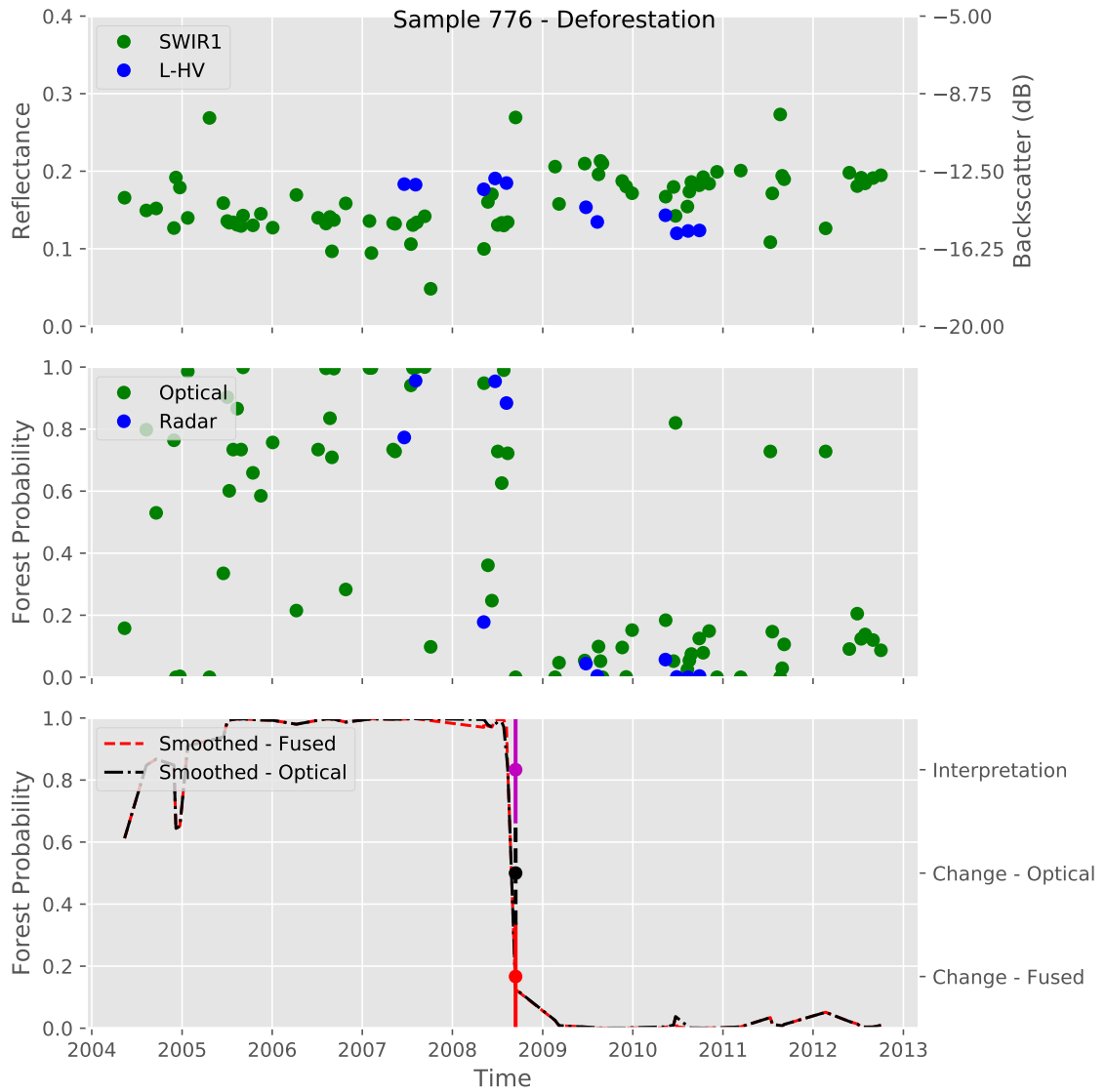
Figure 3·5: "Probability Fusion" results for sample 776 from our accuracy assessment. The top panel shows the original time series in SWIR1 and L-HV bands. The middle panel shows the fused forest emission probabilities from radar and optical data sources. The bottom panel shows time series of smoothed forest probabilities when using fusing both data sources (black) or when using Landsat data alone (red). Breaks detected by the "Residual Fusion" algorithm for both fused and non-fused results are shown as vertical lines on the bottom 66% of the graph, while the date of change given by the assessment interpreter is shown as the magenta line in the top 33% of the graph.

observations looks like a change occurred, but not all of the subsequent observations confirm this change, the CCDC change detection process will mask this first observation from further analysis. This process inside of CCDC addresses similar issues as robust modeling approaches that penalize outliers using observation weights. We found during algorithm development that these characteristics of the CCDC algorithm's change detection component were desirable for our purposes, but in theory any time series change detection method could be applied.

For the "Residual Fusion" algorithm, the data passed to the CCDC change detection process is already a normally distributed random variable with unit variance, centered on zero. We don't need to utilize a forecast model, and can simply scan through the time series looking for some number of consecutive observations that are larger than our desired threshold.

For the "Probability Fusion" algorithm, we have a matrix of n probabilities between 0 and 1 for k classes. We need to select which of the k classes we want to analyze for change and recenter and scale these data. We specify a very simple time series model, $P_{\omega,i} \sim 1 + \beta Time_i$, to fit an intercept and time trend. Just like the "Residual Fusion" algorithm and similar algorithms, we select some initial number of observations as the "training" or "historic" period to train our model. The probability modeling process produces time series for k classes, and we select which time series of probabilities to analyze by selecting the class ω with the maximum likelihood over all observations in the training period. We added a time trend to the model to account for situations like forest regrowth where the class probabilities may display a positive time trend. The goal of this modeling process is to recenter and scale future observations. Fitting only an intercept in the time series model would achieve the same results as if one just estimated the mean and standard deviation of the "training" period. While other approaches have determined a change has occurred based

on a threshold of a probability (Reiche, de Bruin, et al., 2015), this approach is more adaptive to the conditions of each time series because important detection parameters (mean and variance) are determined based on the "training" period.

3.3.4   Accuracy Assessment

We drew 800 random samples within the study area using a simple random sample to assess the accuracy of the deforestation detection of both algorithms when using data from just Landsat and from both radar and optical platforms. Each sample was interpreted primarily by analyzing plots and imagery from time series of Landsat and ALOS-1 data, with available high resolution true color image acquisitions from Google Earth used where possible to provide additional context. Interpreters recorded the land cover at the start of the analysis period (2004), whether the sample pixel experienced a land cover transition during the time series, and information on the land cover transition including the date of change and the land cover after the disturbance. Interpreters recorded the confidence in their overall interpretation using a three tiered qualitative scale, and the confidence in the date of change they provided in numbers of days. Interpretations were checked for accuracy by another interpreter, with special attention to samples marked with low confidence.

We assessed the accuracy of deforestation by comparing the answers from all four combinations of algorithms and observing conditions (optical data alone versus optical and radar data) to the reference interpretations. When deciding if a deforestation event observed in the reference data had been captured by a change detection algorithm, we required that the algorithms find the change within 1 year of the date recorded. To assess the commission errors of deforestation using reference interpretations of stable forest, we required that the algorithms not identify any change in the time series. This requirement is conservative, because it is possible that many of the false positive change detections could be removed through postprocessing. In

the case of the "Residual Fusion" approach, we could compare the magnitude of the change identified or compare the coefficient estimates for time series models before and after the disturbance. The "Probability Fusion" approach provides a classification probability for each time series segment, or stable time period identified, that could also be used to eliminate false positives. It is possible, for example, for the "Probability Fusion" algorithm to find an abrupt change in a time series due to noise despite there being high forest probabilities in the time periods before and after the break on average.

The most important hyperparameters for this change detection approach are the threshold to detect change and the number of consecutive observations that are outside the threshold needed to declare a change. While it was important to use the same algorithm for detecting change to isolate the comparison of algorithms to just the fusion process (Sections 3.3.1 and 3.3.2), we had to calibrate the threshold and number of consecutive observations to each fusion method. To help decide on the values of these parameters, we generated arrays of hyperparameter values (e.g., $threshold = [1.5, 2.0, 2.5, 3.0]$ and $consecutive = [3, 4, 5]$) and mapped deforestation over interesting subsets of our study site for all possible combinations of parameters. We visually analyzed these maps and chose parameters to balance finding deforestation events without having too many false positives. Based on this experience, we selected a threshold of 2.0 for both algorithms and required 3 and 5 consecutive observations of change for the "Residual Fusion" and "Probability Fusion" algorithms respectively.

3.3.5 Simulation Experiment

To test the performance of our algorithms in places with different acquisition histories, cloud cover regimes, or data availability, we analyzed how accurate and timely deforestation could be detected by randomly removing Landsat acquisitions from our

dataset before running the algorithms. We simulated 20 different Landsat data availability conditions ranging from from 100% of the data to 5% of the data. At each increment, we created 100 randomized subsets of the data by sampling without replacement the acquisition dates from the Landsat record. Because of its long history, free and open data policy, its commitment to data continuity, and its higher variability in the number of images acquired across the world, we consider the Landsat data to be the baseline dataset for this study. As such, we did not simulate reductions in the number of observations in the ALOS-1 PALSAR data. Using these 20 simulated observing conditions, we ran the "Residual Fusion" and "Probability Fusion" algorithms with both Landsat and ALOS-1 observations fused together and with Landsat observations alone. We calculated the omission and commission scores based on the 170 examples of deforestation from our reference data for all 100 trials across the 20 simulated conditions. Finally, we calculated the mean and standard deviation of the scores across all 100 trials, for both algorithms under optical and fused scenarios.

3.4 Results

3.4.1 Training Data

Figure 3·6 shows the training data in the EVI, NBR, and NBR2 vegetation indices from Landsat and the L-HH and L-HV channels from ALOS-1 PALAR as bivariate Gaussian kernel density estimates, with bandwidth calculated according to Silverman (1986). The probability distributions for some of the land cover classes could reasonably be characterized a single Gaussian distribution, but the data contain multiple concentrations or groupings that would warrant characterizing separately. The separation among the class labels is not trivial if using only one spectral index or band. For example, the PDFs of forests and herbaceous cover have the highest degree of separation in the EVI index but forests and cloud shadows (sampled usually over forest
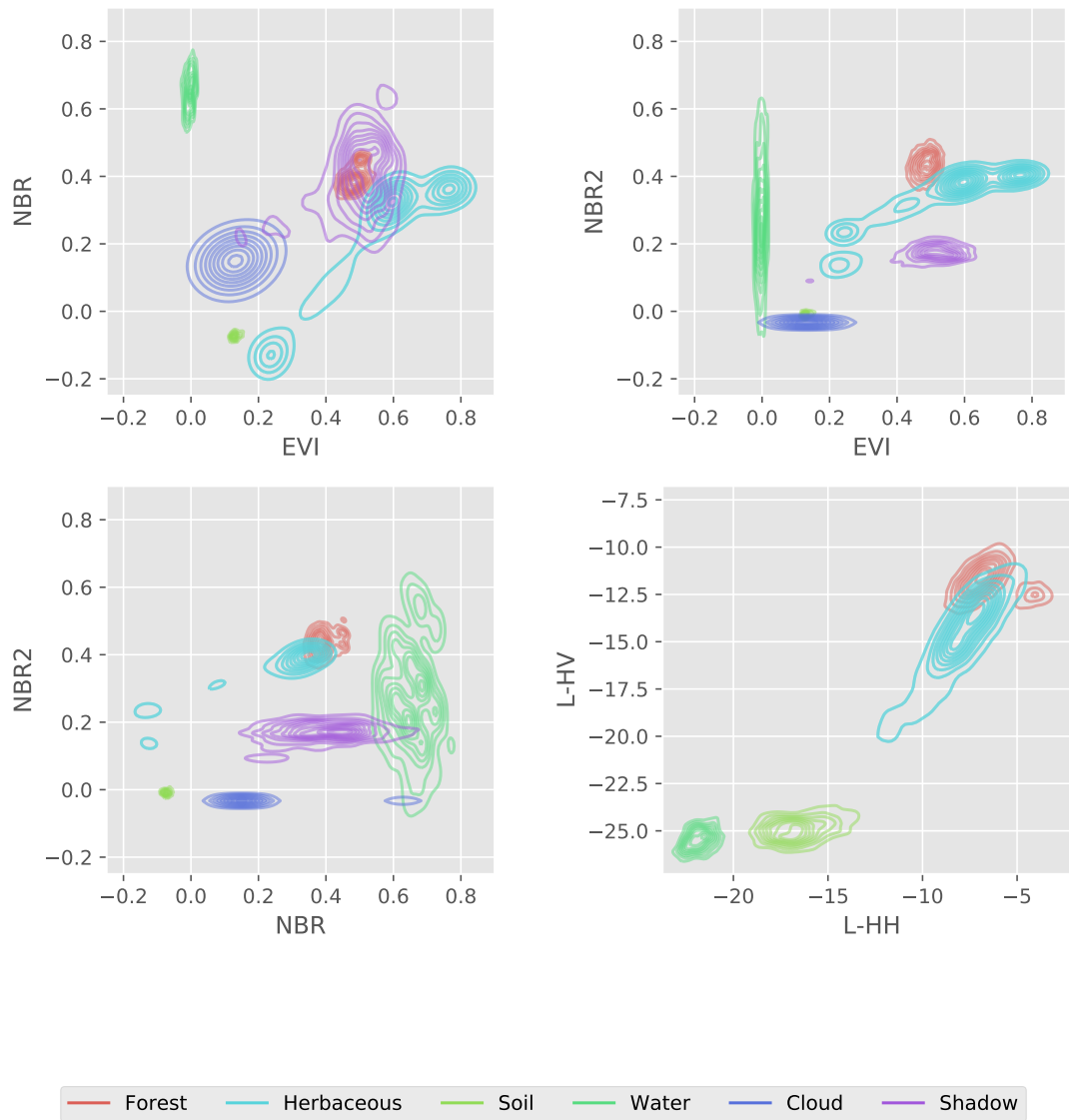
Figure 3·6: "Probability Fusion" algorithm training data visualized in the EVI, NBR, and NBR2 spectral indices from Landsat and L-HH and L-HV backscatter from ALOS-1 PALSAR. The distributions of the training data for all classes are estimated using a Gaussian kernel density estimate.

cover) look similar. Fortunately, cloud shadows and forests are very distinguishable in the NBR2 index. L-HH was the most important in separating water from soil, and there's some overlap in L-HH and L-HV for less dense forest samples and some

herbaceous examples.

3.4.2 Overall Accuracy
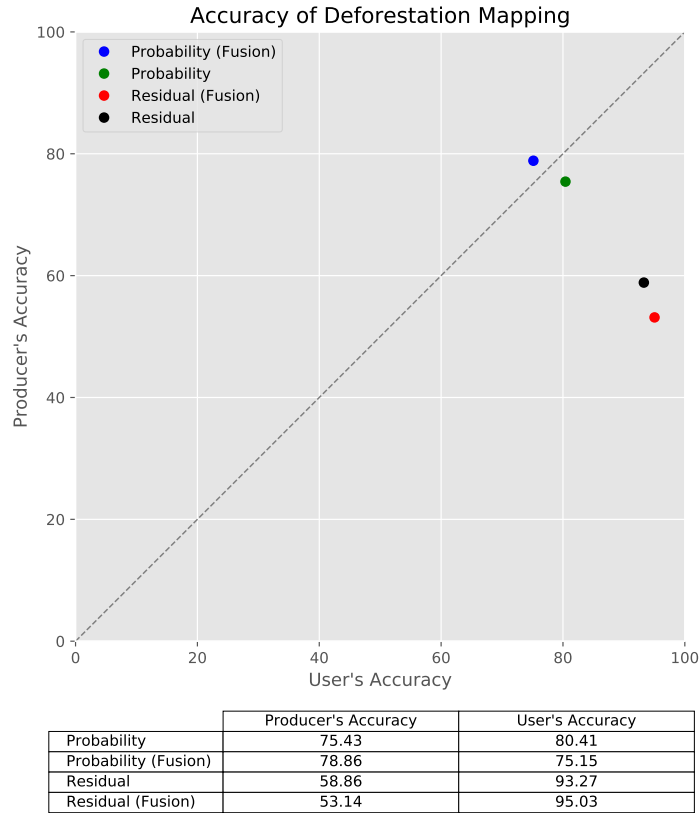
Figure 3·7: Accuracy of deforestation

The random sample of 800 units contained 170 samples that experienced defor-estation during the study time period and 346 samples of stable forest. The interpre-tations from these samples were used to generate omission and commission accuracies, which were plotted for each algorithm and data scenario in Figure 3·7. The "Prob-ability Fusion" algorithm had the highest Producer's Accuracy (1 - Omission Error) when radar data was incorporated, but did not detect more deforestation events than when using Landsat data alone. The User's Accuracy for the "Probability Fusion" algorithm was slightly higher when using just Landsat data. Overall the errors of

commission and omission were approximately balanced for this algorithm, which is important because the mapped areas are likely to fall within the confidence intervals of the accuracy adjusted, unbiased area estimates when doing area estimation (Olofsson et al., 2013). The "Residual Fusion" algorithm was significantly less accurate than the "Probability Fusion" algorithm, missing more than 40% of the deforestation events. This algorithm was slightly less accurate when incorporating radar data, and the algorithm was not balanced in omission and commission error, suggesting that alternative hyperparameters might detect more deforestation events at the cost of commission error.

3.4.3 Timing of deforestation
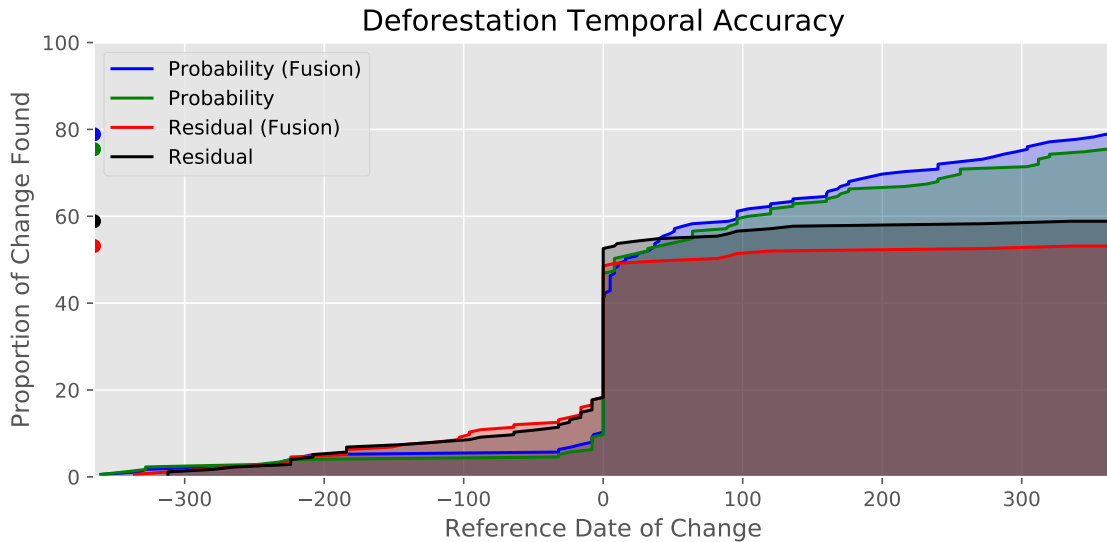
Figure 3·8: Time to detect deforestation by algorithm type.

For all samples of deforestation that were captured by our change maps, we calculated the temporal accuracy of the change detection by comparing the date of change in the reference interpretations against the algorithm results. Figure 3·8 shows the cumulative percent of deforestation events found by each algorithm and data sce-

nario, which was helpful in similar analyses by Tang (2018). As expected, the timing results for the "Probability Fusion" algorithm show that this algorithm finds more of the deforestation than the "Residual Fusion" algorithm, but this algorithm does not perform considerably differently when incorporating radar data into optical data time series. The "Residual Fusion" algorithm finds approximately 20% of the deforestation too early, and did not detect very many deforestation events after a few months have passed since the change occurred. The "Probability Fusion" algorithm detects approximately 10% of the sample too early, but is capable of finding approximately about 79 and 75% of the changes within 1 year when fusing data sources and when only using optical data. We did not consider a 2 year window to be sufficient for our accuracy analysis, but the "Probability Fusion" algorithm detected 90.2% of the deforestation samples within 2 years.

3.4.4 Observational Frequency Simulation
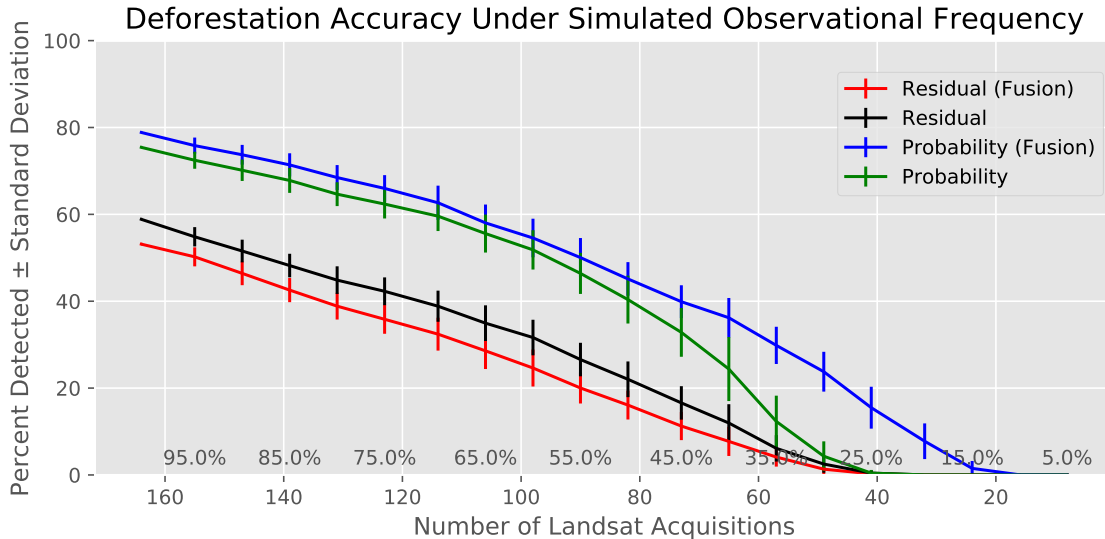
Figure 3·9: Accuracy of deforestation under simulated observational frequencies

The results of the Landsat data observation condition simulation experiment are

shown in Figure 3·9. The "Probability Fusion" algorithm only saw extra benefits larger benefits from utilizing radar data once the Landsat archive had been reduced to about 40% of its original size, or to about 7 observations per year. The "Residual Fusion" algorithm consistently performed better when only using Landsat data, and always worse than the "Probability Fusion" algorithm.

3.5 Discussion

3.5.1 Fusion Algorithms

The algorithms developed here may also be described in terms of their use of either univariate or multivariate data. The "Residual Fusion" algorithm currently operates using only one variable per time series data source. Many of the bands from optical sensors are highly correlated, but having to select one single index or band neglects useful information. Despite this, the of a single variable is common to BFAST Monitor Verbesselt et al. (2012), EMWACD Brooks et al. (2014), and LandTrendr Kennedy et al. (2010). A possible improvement for this algorithm is to incorporate additional spectral bands from optical and radar data sources by using the approach from CCDC, which combines scaled residuals across multiple bands by calculating the vector norm across all bands for each observation. By doing this, multiple bands per data source would be combined first, and the combined information from each data source would then be fused together. For example, we might want to monitor for change using the first three Tasseled Cap indices (Kauth and Thomas, 1976; Crist, 1985) with Landsat data and dual polarized data and texture metrics for SAR datasets.

While the "Probability Fusion" approach may only monitor for change in a univariate time series, the class probabilities that are the basis for the time series change detection can come from a classifier trained on multivariate data. The "Bayesian updating" approach in Reiche, de Bruin, et al. (2015) and Reiche et al. (2018) works

very similarly, but those studies only used univariate distributions of either NDVI, L-HV, or S-VV, though others have used multivariate Gaussian distributions (Salberg and Trier, 2011). As work toward identifying the cause of land cover change disturbances have demonstrated, methods of land cover change, even within a single type of transition, can show up differently, especially if analyzed in multiple spectral bands. While the "Probability Fusion" approach leverages multiple polarizations of radar data and multiple spectral bands and indices of optical data, there are many more transformations of these data that could be included when deriving the classification probabilities. Texture, for example, has been shown to be very useful in discriminating among land cover types (Lu and Weng, 2007; Zhu, Woodcock, Rogan, et al., 2012) and mapping forest cover, biomass, and stand age (Luckman et al., 1997; Simard et al., 2000; Kuplich et al., 2005; Lu, 2005; Walker et al., 2010; Cutler et al., 2012; Thapa et al., 2015). This study was interested in these algorithms in context of near real time monitoring, and so using as many observations as possible was a priority, but one could also produce temporal features to use in the classifier by resampling to a coarser temporal resolution. If change mapping at a yearly basis, as LandTrendr does for example, one could generate temporal metrics like the range of the observations or the mean value for each season that are useful features when classifying land cover, as in Abercrombie and Friedl (2016).

With the exception of Trier and Salberg (2011) and Salberg and Trier (2011) who mapped forest, sparse forest, grass, and soil classes, previous studies that used methods similar to the "Probability Fusion" algorithm have only attempted to identify forests separate from non-forest. Mapping forests apart from non-forests is a useful approach if users are interested in finding deforestation, but frequently users desire more information about what causes deforestation and identifying the land cover after disturbances is required for understanding what land cover or use transitions

occur. This information is very important for understanding historic carbon emissions (Houghton et al., 1999) because transitions from forest may cause different emissions directly and from changes in fluxes from the new land cover. Information on land cover change is also useful for estimating spatial relationships to drivers of land cover change or projecting land cover change through time (Thompson, Plisinski, Olofsson, Holden, et al., 2017).

We chose Random Forest primarily because it generally performs very well without needing to fine-tune hyperparameters (Fernández-Delgado et al., 2014) and based on past experiences using the algorithm, but there are a few other benefits and some tradeoffs when using a non-parametric classification algorithms over parametric ones. First, Random Forest is able to separate among classes that may be non-linearly separable and handle classes with multiple modes in some variable. The kernel density estimations of the probability distribution functions for each thematic class in Figure 3·6 shows that the herbaceous training data samples had several different groupings internal to the class. This phenomenon when collecting training data is commonly referred to as "many to one mapping", which references that there may be many sub-groupings that are representative of one larger class. In this example, we likely could have used a Gaussian distribution to model the herbaceous training data samples, but only after first isolating the individual groupings of data. By choosing a classifier that does not much such assumptions about the data, we avoid this complication. Random Forests, however, require a relatively larger amount of data to characterize the distribution of training data than when using parametric classifiers, which only need a representative sample so that estimates of model parameters (i.e., mean and variance) can be accurately characterized. Random Forest is also sensitive to the representation of each class in the training data, especially for imbalanced training data sets with few samples of rarer classes. We attempted to address this issue by

providing class weights for each training data sample that were inversely proportional to the number of samples for the class (Chen et al., 2004).

Random Forests along with other types of decision trees, Support Vector Machines (SVM), and naive Bayes classifiers have been shown to have difficulties generating unbiased class probabilities (Platt et al., 1999; Niculescu-Mizil and Caruana, 2005; Zadrozny and Elkan, 2001), although building larger decision trees may be less prone to this issue (Provost and Domingos, 2003). Algorithms like Random Forest or SVMs that emphasize the boundaries among classes tend have a sigmoidal shape to their classification predictions, avoiding predicting values close to 0 or 1, while approaches like naive Bayes that assume features are independent tend to push probabilities toward 0 and 1. We did not try to quantify or address this potential issue, but there are approaches to calibrate probability estimates that could be employed, including Platt or sigmoid calibration (Platt et al., 1999) and isotonic regression calibration (Niculescu-Mizil and Caruana, 2005).

The time series of SWIR1 from Landsat and of L-HV from ALOS-1 were not highly affected by seasonality for most examples of tropical forest. However, the Landsat data used had not been corrected for sun and sensor geometry, and so there was some amount of seasonality for all pixels in the Landsat data, especially along edges where projected shadow area can change greatly with sun angle. We attempted to characterize this apparent seasonality in the "Residual Fusion" algorithm by fitting a yearly harmonic, but there was no such characterization of seasonality in the classifications by the "Probability Fusion" algorithm. Reducing the temporal resolution of the analysis to a yearly scale would allow for characterization of seasonality by using annual profiles, but this would not be suitable for monitoring applications. Based on the approach used for Landsat data in Hamunyela, Verbesselt, and Herold (2016), Reiche et al. (2018) addressed seasonality in the time series signal by normalizing each

observation relative to neighboring pixels in a large moving window. This approach has proven very useful for forest monitoring applications, and Hamunyela, Verbesselt, and Herold (2016) found that the median value of the top 10% largest values in a vegetation index was robust to normalizing by disturbed pixels. Reiche et al. (2018) also demonstrated the utility of spatial normalization for reducing seasonality in C-band data from Sentinel-1.

3.5.2   Change Detection

During algorithm development, we informally tested and analyzed the performance of the Exponentially Weighted Moving Average (EWMA; Shewhart (1931) and Lucas and Saccucci (1990)), Cumulative Sum (CUSUM; Ploberger and Krämer (1992)), and Moving Sum (MOSUM; Chu et al. (1995)) statistical tests for change in a time series. In our experience, these statistical tests were overly prone to false positives, usually because of clouds and cloud shadows that were not detected in the FMask. Based on our experience running the CCDC algorithm (Zhu and Woodcock, 2014), we implemented the change detection test from CCDC which detects change if the scaled residuals for some number of consecutive observations are larger in magnitude than a given threshold. This method is similar to a "robust EMWA" approach described by Lucas and Saccucci (1990) that requires multiple, consecutive observations to be outside the EWMA threshold to detect a change. One modification, however, is that the change detection method based on CCDC will try to eliminate bad observations as it monitors the time series by removing an observation from the remainder of the analysis if its scaled residual exceeds the threshold but the subsequent observations are within the threshold. This process is performed during the monitoring phase of the algorithm, but attempts to address similar concerns as the Shewart X-Bar chart test performed prior to the historic time series model fitting step in the EWMACD (Brooks et al., 2014) change detection algorithm. We found that this noise screening

step that occurs while monitoring for change was very useful, but it may not be necessary in areas where clouds and cloud shadows are less of an issue. The exact algorithms that control how the change detection is performed on the fused time series may be interchangeable to some degree, and we consider the actual "time series fusion" part of the algorithms we developed here to come before this step.

One possible enhancement to the change detection algorithm to help take advantage of radar observations is to modify the number of consecutive observations needed if radar observations are being considered during the monitoring phase. Radar observations, for example, could count as two of the required consecutive observations to identify a break in the algorithm. Some other tests for change integrate information of change magnitude and persistence differently. The MOSUM (moving sums) test (Chu et al., 1995), for example, is calculated for time t by using past and future observations (lags and leads), and under this approach few large disturbances could add up to the same MOSUM score as many smaller disturbances. Under this framework, radar observations could be weighted more highly by scaling the radar radar forecast residual time series, potentially making residuals from radar time series count double or triple, thereby reducing the time needed for detection.

3.5.3 Accuracy

One of the common causes for errors in the "Probability Fusion" algorithm is due to inaccurate estimates of the classification emission probabilities, especially for examples that fall near the boundaries of another class. The forest emission probabilities from Random Forest were typically above 90% for examples of dense forest cover, typically one or more pixels away from an edge and with reflectance in the first Landsat SWIR band (1.55-1.75 μm) between 10-15% and backscatter in L-HV of approximately -10 to -12 dB (e.g., Sample 43). Many of these poorly characterized areas were also secondary or recently harvested, regrowing forests, and transitional states are difficult

to characterize with a discrete land cover label. Approaches that transform data into continuous variable, like estimation of percent tree cover or vegetation continuous fields (Sexton et al., 2013; Hansen et al., 2013) or spectral endmember percentage from a mixture model (Souza Jr et al., 2005), may do a better job of characterizing transitional classes than a discrete labeling algorithm. Conversions to oil palm plantation was identified very well with the "Probability Fusion" algorithm (38 of 40 examples for fused and non-fused), and much better than overall but lower than the "Probability Fusion" algorithm (31 and 27 of 40 examples for fused and non-fused).

Many of the poorly estimated forest emission probabilities from the radar dataset were near edges, but frequently had fairly stable L-HV time series that would otherwise suggest permanent forest cover. There was clear correlation of the L-HH and HH/HV ratio time series with forest probability estimates in these examples, suggesting that either the full variation in L-HH and the HH/HV ratio were insufficiently characterized in the training data or inherent variability that isn't as much connected to a notion of "stable forest". Backscatter received in the same plane that it was transmitted in (e.g., horizontal-horizontal or HH) does not need to interact repeatedly with target elements to alter the polarization, and the necessity of volume scattering for backscatter in cross-polarized channels makes L-HH inherently less stable than L-HV. Soil moisture, for example, increased the backscatter of L-HH more than L-HV (4.0 dB versus 2.5 dB) in Lucas et al. (2010), and this influence from soil moisture conditions was more sensitive for lower biomass forests. The large increase in backscatter at high soil moisture conditions (e.g., after a rainfall) for L-HH were attributed to the increased surface scattering, particularly increased double bounce or direct ground returns. Shimada et al. (2014) found that a regional thresholds on L-HV was sufficient for distinguishing forest cover globally, and a simple solution would be to drop the L-HH and ratio time series and only analyze L-HV. It should

also be possible, however, to retain the information from L-HH if the contributions to this signal from soil moisture were either better characterized in training data or input feature variables.

Another source of error in both algorithms comes from instability in estimates in the training period of the change detection step. If clouds or cloud shadows are present in the training period of the either change detection algorithm, the coefficients estimated for a time series forecast model can be very inaccurate, often predicting a large negative or positive time trend. This phenomenon seemed especially important for the "Probability Fusion" algorithm because one low class probability due to noise in the first few observations could weigh down adjacent observations due to the HMM smoothing process, which can induce a spuriously large coefficient estimate for the time trend and cause a spurious change shortly into the monitoring phase. Both fusion algorithms estimated time series models during the training period using iterately reweighted robust linear models to reduce the influence of such outliers on regression results, but did not implement more advanced methods of addressing noise in the training period that algorithms like CCDC implement. Instability in model parameter estimates is to be expected with noisy data and few observations, and it is possible to calibrate how long the training period has to be within both algorithms. As with CCDC, we required time series forecast models to have some minimum number of observations before the change detection algorithm will run. The "Residual Fusion" algorithm is at a relative disadvantage here compared to the "Probability Fusion" algorithm because the fusion of the time series happens after forecast models are run. As such, optical data and radar data each have separate minimum number of observations (16 and 3 here, respectively), limiting the number of observations that may be used for monitoring. By comparison, the "Probability Fusion" is able to train forecast models using data from radar and optical sources because fusion is

done before this step. While increasing the minimum number of observations may help with model stability, it can also cause errors if change occurs within the training period because a regression estimated with multiple land cover regimes is likely to also be spurious or unstable.

While the "Probability Fusion" algorithm outperformed the "Residual Fusion" algorithm, the "Probability Fusion" algorithm has a few relative tradeoffs when considering operational implementation. First, the "Probability Fusion" algorithm required significant time investment in gathering training data before the process can be run. Although there are hyperparameters that could be tuned for better performance locally, the "Residual Fusion" algorithm could be ran over large areas and could generate change maps without needing human input. The effort required to run the "Probability Fusion" algorithm could be reduced by using a parametric classifier that doesn't require as many training data samples, or by automating the generation of training data by collecting samples using other datasets or land cover maps as reference. While the need for training data requires time investment, it also provides the opportunity to refine the maps produced from the algorithm in a way not possible with the "Residual Fusion" time series. The fusion transformation in the "Residual Fusion" method is based on comparison with the expectation for each time series, which is not something one can calibrate or fix with human input. Classification probabilities, however, could be corrected by iteratively analyzing classification probability maps and adding more training data to correct the classification. Finally, the "Probability Fusion" algorithm took significantly longer to compute primarily because of the time needed to compute emission probabilities from Random Forest. Fortunately, these predictions are easily parallelized, and other classifiers (especially parametric ones) take much less time to compute class probabilities. There may be utility for both of these algorithms, since the "Residual Fusion" algorithm could quickly produce maps

of changes and anomalies that can be used to guide training data collection.

3.5.4 Radar Cost Benefit

This study only saw minor improvements in the spatial and temporal accuracy of deforestation from incorporating radar data into time series based on all available Landsat data when using the "Probability Fusion" algorithm. Radar contributed most strongly when using the "Probability Fusion" algorithm performed once the Landsat data archive had been simulated to have about 7 acquisitions per year on average. The study site we selected was located within the Peruvian Amazon and was within range of the Brazilian receiving station The observation reduction experiment highlights the potential of radar time series to benefit deforestation and land cover change analysis, especially when Landsat data is noisy or historically unavailable. The experiment design involved only changing the number of usable observations from Landsat and held all other variables constant, which potentially overestimates the omission error in practice because we would have adapted several hyperparameters to the reduced data density. For example, the number of consecutive observations required to detect a change for the "Probability Fusion" algorithm could likely be reduced from 5 to 3 and reduce omission error at the cost of an acceptable increase in commission error. For historic analyses, the number of available observations, even after consolidation of the archive, can vary greatly with lower densities in important forests in Africa and Southeast Asia (Wulder et al., 2016). Data from the Japanese Earth Resources Satellite (JERS-1) could help extend time series analysis back into the 1990s for places that were poorly covered by Landsat-5 relative to Landsat-7, or decrease uncertainties in conjunction with Landsat data, within the same workflows as ALOS PALSAR data. For example, Thomas et al. (2015) were able to utilize data L-band data from JERS-1 and ALOS-1 PALSAR during the 1990s and 2000s to map mangroves using the procedures. For more recent time periods and for ongoing

monitoring applications, the methods tested here could combine data from Landsat 8, ALOS-2, and Sentinel-1, as was done in Reiche et al. (2018).

3.6   Conclusion

Two algorithms were developed in this study that can fuse data from optical and radar sensors for mapping deforestation. The "Probability Fusion" algorithm caught almost 80% of deforestation within one year of the reference interpretation date, and over 90% within two years, and combining radar and optical data was slightly more accurate in finding deforestation than using optical data alone. The "Residual Fusion" algorithm was much less accurate, and despite individual examples where radar data improved detection this algorithm was overall more accurate using optical data alone (58.86% for optical data versus 53.14% when fused). The improvement in deforestation mapping from adding radar data for the "Probability Fusion" algorithm was much more pronounced as the Landsat data availability was decreased in our simulations. We saw small benefits to using radar data given the relatively good coverage from Landsat data, but there were only 11 usable dual polarimetric observations from ALOS-1 over our study site. The combination of L- and S-band observations from ALOS-2 and NISAR, along with the potential for integration of C-band time series from both Sentinel-1 satellites, would yield significantly denser radar time series than what was historically available.

Conversion of observations into class probabilities — particularly forest class probabilities — has consistently proved to be a useful method for time series fusion (Solberg et al., 2008; Salberg and Trier, 2011; Trier and Salberg, 2011; Reiche, de Bruin, et al., 2015; Reiche et al., 2018). Probabilities are an obvious method for combining data, and have been used in remote sensing context as prior probabilities to assist land cover mapping (Schneider et al., 2003; McIver and Friedl, 2002; Friedl et al.,

2002). Spatial information may also be introduced by modeling the spatial correlation among neighboring pixels, which has been done, for example, by using Markov Random Fields (Solberg et al., 1996). In addition to modeling the spatial relationship among neighboring pixels of the same sensor, it may be possible to use class probabilities to help fuse data from multiple resolutions. The STARFM model, for example, explicitly models the relationship between different observations of different spatial resolutions using weight functions. Information from multiple scales is frequently used to identify images in recent deep learning methods (Zhang et al., 2016; LeCun et al., 2015), and Zhao and Du (2016) provided a recent application of deep learning in remote sensing by developing a multiscale convolutional neural network (MCNN) that integrated information from three spatial scale representations of very-fine resolution imagery.

Research and development of frameworks that can leverage multiple streams of information to answer questions using remote sensing data should be of high priority given the increasing availability of free and consistent observations. Many applications in remote sensing have benefited from having access to the entire archive of Landsat data, and probability presents a useful framework for connecting to data that describes human (e.g., spatial development patterns, zoning, taxes) and natural systems (e.g., decadal climate variability, logging or fire frequencies) dynamics.

Chapter 4

Landsat Time Series Meteorological Data Fusion In Semi-Arid Ecosystems

4.1   Introduction

4.1.1   Background

Vegetation phenology, or the seasonal patterns of dormancy, growth, and senescence, is a first order control on many physiological processes of interest, including the annual budgets of carbon dioxide and water vapor exchanges. The phenological timings of plants are optimized for favorable growing conditions, to optimize seed dispersal, to avoid competition or predators, or to avoid hazardous environmental stresses (Cleland et al., 2007). Phenology in grassland ecosystems, especially in arid or semi-arid areas, is greatly affected by variation in precipitation and has been found to vary on the order of a month due to precipitation and temperature conditions (Lesica and Kittelson, 2010; Shen et al., 2011).

Satellite remote sensing has been very useful in estimating phenological patterns at large spatial scales, at coarse spatial resolutions using Advanced Very High Resolution Radiometer (AVHRR) (Moulin et al., 1997), Moderate Resolution Imaging Spectroradiometer (MODIS) (Zhang et al., 2003), Landsat (Melaas et al., 2013), and ground base sensors or combinations of all (Liu et al., 2017). Many of the remote sensing studies of vegetation phenology are more focused on mapping the timing of the transition than modeling or understanding the drivers, frequently deriving transi-

tion dates from logistic or smoothing functions fitted to a vegetation index. Mapping transition dates is very useful for further analysis, like investigation of trends in earlier spring or later autumn transitions. Others have used phenological date information for model inference, established that precipitation can help predict phenology in arid and semi-arid regions of Africa for monitoring applications (Zhang et al., 2005), or helping calibrate phenology models by providing estimates over large spatial areas (Xin et al., 2015).

The total vegetation abundance, whether mapped as a vegetation index or more physical quantities like percent cover or biomass, is also a domain in which remote sensing studies use climate data. Hausner et al. (2018) used the Google Earth Engine (Gorelick et al., 2017) to assess the benefit of riparian ecosystem restoration by comparing the relationship between yearly NDVI and precipitation totals. They found that the two variables were very highly correlated, and that restoration projects change the relationship between the variables, suggesting a potential for monitoring abrupt surface process changes (e.g., restoration or degradation).

Formica et al. (2017) analyzed the relationship between anomalies in annual precipitation and NDVI in the deserts of Central Asia using a separate model per pixel, and a pooled model of all observations that included the pixel location as a random effect. They found significant positive correlations between anomalies of NDVI and precipitation that showed spatial variation across gradients of arid and semi-arid environments. Birtwistle et al. (2016) used this relationship between precipitation and greenness to determine which ephemeral stream channels had seen flow from monsoon rains, and were able to explain 73% of the variance in monsoon rain totals with the difference in NDVI before and after the rains.

In addition to timing, remote sensing is also used to analyze the variability of vegetation associated with climate and environmental drivers. Seddon et al. (2016)

analyzed the sensitivity of vegetation to climate variables (air temperature, water availability, and cloud cover) by using these climate variables to explain standard anomalies of 14 years of MODIS EVI time series. Their models also included an autoregressive term to account for time series memory effects, and their study produced global maps of temperature, water, and cloud sensitivity that they combined into an overall vegetation sensitivity index. Also looking for vegetation response to variability in rainfall and flooding, Broich et al. (2018) used 26 years of Landsat time series over floodplains in Australia to determine the influence of rainfall amount, flooding condition, and structural breaks on EVI standard anomalies. They found that rainfall and flooding were strong predictors of EVI anomalies, and that estimating breakpoints frequently, though not always increased explanatory power, albeit less than when adding flooding alone. The authors used the same approach for estimating breakpoints as the Breaks For Additive Season and Trend (BFAST; Verbesselt, Hyndman, Newnham, et al., 2010; Verbesselt, Hyndman, Zeileis, et al., 2010) algorithm, which is known as the "Bai-Perron" test after Bai and Perron (1998). The breakpoints fit using this method were frequently associated with the onset of the South East Australian Millennium Drought, suggesting that their method could detect not just climate variability, but structural changes to the relationships between EVI and moisture variables that drive the variability. Chen et al. (2018) used time series of precipitation and a greenness index to infer irrigation from variability, relying on the idea that increases in greenness that were not coincident with precipitation were assumed to be caused by irrigation. Yang et al. (2017) used remote sensing derived time series of evapotranspiration to look at the response of vegetation across land cover type and across regrowth trajectories to drought. They finding that evapotranspiration decreased before NDVI during the drought and that the mature pine forest plantation stands were not as affected as younger stands, presumably because deeper

roots provide better access to water.

Many studies have investigated the links between variability in greenness and drought for the Amazon, with some studies focusing on the response, if any, to the 2005 drought (Huete et al., 2006; Samanta et al., 2010; Samanta et al., 2012), the 2011 drought (Xu et al., 2011), or to droughts and El Niño events since 2000 (Hilker et al., 2014). Others have focused on the boreal ecosystems of the world, looking for greening or browning trends as the vegetation cover in these regions experience some of the most rapid responses to temperature and moisture conditions due to global climate change (Myneni et al., 1997; Goetz et al., 2005; Ju and Masek, 2016; Sulla-Menashe et al., 2018). Many of the studies focusing on this topic used coarse resolution data from either AVHRR or MODIS data, but more recent studies that use medium resolution data from Landsat have found that disturbance and the legacy of past disturbance drives many of the observed trends (Ju and Masek, 2016; Sulla-Menashe et al., 2018).

Most existing methods of finding land cover change using time series analysis have not incorporated ancillary information (Kennedy et al., 2010; Huang et al., 2010; Verbesselt et al., 2012; Zhu and Woodcock, 2014), but some approaches have built-in methods for handling variability due to meteorological conditions. The Vegetation Continuous Tracker (VCT; Huang et al., 2010) uses a Z-score to measure distance away from examples of forests in an image to detect change, and the algorithm can take two steps to deal with effects of drought, particularly in semi-arid open woodlands in the Southeast United States. First, they increase the number of years they require of consecutive, abnormally high integrated Z-scores from two to three. Second, they perform the calculation of Z-scores twice, specifically targeting open forests as a separate population for the calculation of the mean and variance. The mapping scheme used by Huang et al. (2010) did not want drought impacts to be classified as

a disturbance, so these steps were crucial for map accuracy. LandTrendr (Kennedy et al., 2010) allows for single year disturbances within it's conceptual framework, but by ignoring meteorological data it misses the opportunity to distinguish if environmental conditions caused these temporary disturbances. Many of the effects of climatic variability, including onset and offset phenology or changes to the trajectory of greening, are able to be excluded immediately when using only one observation per year, typically at the height of the growing season, as many algorithms do (Kennedy et al., 2010; Huang et al., 2010). Using BFAST Monitor, Hamunyela, Verbesselt, De Bruin, et al. (2016) have tried to address the issue of seasonality in a similar way by normalizing forest pixels based on the mean calculated from forest pixels within a large window.

The BFAST Monitor (Verbesselt et al., 2012) and Continuous Change Detection and Classification (CCDC; Zhu and Woodcock, 2014) methods assume that greenness can be explained by decomposing the signal into linear combinations of a intercept, time trend, and seasonal terms (i.e., as a Seasonal-Trend Decomposition Cleveland et al. (1990)). One study attempted to enhance the performance of BFAST Monitor over tropical forests by incorporating the Standardized Precipitation Index (SPI) as an explanatory variable, but found that change detection performance decreased when adding this extra variable (Dutrieux et al., 2015). They found that using spatial normalization (Hamunyela, Verbesselt, De Bruin, et al., 2016) was a more effective method of removing the influence of climatic variability than SPI. They hypothesized that the forests in their study area might not be sufficiently water limited to be influenced by dry conditions or have non-linear responses to drought, or that SPI might not be ideal at representing the moisture conditions. The CCDC approach does not explicitly account for variability in vegetation due to environmental drivers, but the requirement for consecutive observations of change and the temporally depen-

dent forecast uncertainty calculation can help. Specifically, the calculation of Root Mean Squared Error for the forecast uncertainty in CCDC for any given observation is only based on predictions for similar days of the year, resulting in wider confidence intervals for periods of high variance (e.g., phenological transition periods or spring snow melt) than stable periods (e.g., peak growing season). Despite these factors, CCDC assumes greenness follows a deterministic pattern, irregardless of environmental drivers, which causes spurious changes to be identified that reduce the accuracy of land cover classification and complicates the inference of land cover change from the "structural breaks" found using CCDC.

Just as climate driven variability makes it harder to detect change, land cover change makes it harder to establish the impacts of natural variability because land cover change usually fundamentally alters the vegetation. Studies that calculate parameters over long time scales, like greening or browning trends or shifts in growing season, using low resolution data are particularly susceptible to this issue. Zhu, Fu, Woodcock, Olofsson, Vogelmann, Holden, et al. (2016) explored this by separating greenness changes caused by land cover change in Guangzhou, China from gradual greenness change from vegetation growth or regrowth (e.g., from a forest plantation). Both Ju and Masek (2016) and Sulla-Menashe et al. (2018) utilize disturbance information, either in the form of fire permiter databases or disturbance maps, to separate greening or browning trends in the boreal forests of North America due to climate drivers from trends that are simply caused by disturbance or land cover change. Melaas et al. (2016) explicitly accounted for land cover change when analyzing phenological patterns over eastern North America using maps from CCDC, but had to remove pixels with multiple changes or changes in the first half of the time series. Pasquarella, Holden, and Woodcock (2018) used phenological transition dates for each CCDC time series segment estimated according to Melaas et al. (2013), and

found that including phenological transition dates improved classification of forest species.

4.1.2 Objectives

This research effort seeks to incorporate environmental conditions into time series analysis of vegetation greenness. Such integration will allow for variability in phenology to be separated into variability due to natural variations in environmental drivers and variability in phenology due to anthropologically or naturally driven structural changes to grass, shrubland, and forest ecosystems. In order to guide this assessment, this effort seeks to answer the following:

1. What is the relationship between vegetation greenness and meteorological variables in time series analysis?

2. Does meteorological data meaningfully improve land cover change detection efforts?

We answer these questions by following, or at least being inspired by, the methodologies for integrating meteorological information into Landsat time series taken by Dutrieux et al. (2015), who included a drought index as an external regressor in an online change detection algorithm based on harmonic regression (BFAST-M), and Broich et al. (2018), who used rainfall, flood condition, and breaks to explain EVI anomalies. We chose to focus this study on a Landsat data footprint over the San Joaquin Valley and Sierra Nevada Mountains for its diverse land cover and disturbance regimes, and recent droughts that have affected the vegetation. By assessing an area that has experienced recent precipitation anomalies as a contrast with "baseline" or "historic" conditions, we hope to gain insight into responses to climate variability similar to previous studies (Friedl et al., 2014).

4.2 Data

4.2.1 Landsat Time Series

We used the newly available Landsat Analysis Ready Data (ARD) product from the USGS (Dwyer et al., 2018) from tile h03v09 in this analysis. Landsat ARD data are available to download for all of the Contiguous United States (CONUS), Alaska, and Hawaii in a highly preprocessed and easy to use form suitable for large scale analysis. Landsat ARD are preprocessed using the same Landsat Collection 1 algorithms and are atmospherically corrected using approaches familiar to Landsat data users: LEDAP for Landsat TM and ETM+ (Masek et al., 2006) and LaSRC for Landsat 8 (Vermote et al., 2016). Cloud and cloud shadow masks generated by the "CFMask" implementation of Fmask (Zhu and Woodcock, 2012) are also available from ARD.

While similar to Landsat time series datasets the community uses and was used in the previous studies of this dissertation, ARD offer a number of specific improvements to data quality and ease of use. For large area mapping, for example, ARD have better geometric accuracy than if one were to preprocess by reprojecting and tiling the scene-based, Collection 1 data. ARD are directly projected into Albers Equal Area (for CONUS) when going from swath to grid, avoiding resampling artifacts that occur when Landsat scenes are reprojected from UTM. The tiling system used by ARD facilitates analysis by combining data from multiple WRS-2 paths and rows within a single tile, increasing the data density in areas of overlap between adjacent WRS-2 footprints. This increase in density, however, is not uniform through space (only in the "overlap" regions) and ARD are not corrected for differences in view or solar geometry.

4.2.2   Meteorological Time Series

We used the Parameter elevation Regression on Independent Slopes Model (PRISM) gridded meteorological data product (Daly et al., 1997), which is available at 4 km spatial resolution for monthly time steps. While we had the temperature variables from this dataset available for analysis when interpreting time series data, we only used the precipitation variables from this dataset in quantitative analysis.

In order to capture a much wider range of moisture dynamics, we also used the Standardized Precipitation and Evapotranspiration Index (SPEI) developed by Vicente-Serrano et al. (2010) which takes into account moisture availability by modeling the difference between precipitation and potential evapotranspiration. The SPEI is standardized, which helps account for differences in moisture availability due to different hydrologic sources (e.g., rainwater versus snowmelt) over space. The SPEI is available integrated over timescales ranging from one month to four years, which also helps to quantify moisture conditions over time as droughts evolve. Vicente-Serrano et al. (2010) use the example that shorter SPEI integration periods are more likely to be related to short term fluxes like rainfall, while longer integration periods are likely related to fluxes from river, lake, or groundwater storage. Many aspects of ecosystem response to drought operate on different time scales, potentially making SPEI a good predictor variable for vegetation in semi-arid environments. We acquired the SPEI data for California from the West Wide Drought Tracker (Abatzoglou et al., 2017), which provides SPEI at monthly time steps and uses the PRISM dataset as inputs and provides SPEI integrated over one month to six years.

4.3 Methods

4.3.1 Preprocessing

We masked observations from the Landsat time series that were either saturated or were found to be contaminated with cloud, cloud shadow, or snow cover by Fmask. The total number of observations we used from ARD between 1986 and 2018 was 3120, although there is significant spatial variability.

In order to harmonize the meteorological time series data to the Landsat data, we first reprojected the PRISM and SPEI time series data from geographic to the ARD Albers Equal Area projection and upsampled from 4 km resolution to 30 m using nearest neighbor resampling. The spatial mismatch is obviously enormous, but using a gridded product, coarse as it was, captured gradients in our data, especially on either side of the Sierra Nevada Mountains. The PRISM data are also available at 800 m resolution, but only as a product for purchase which would hurt the reproducibility of this study, but may better capture gradients in this study area.

4.3.2 Time Series Model Selection

There are many ways of using pseudo-physical or empirical modeling techniques to predict phenology, but these approaches typically try to estimate observed transitions or assess drivers of phenology, not try to predict what it should be in a near real time forecast scenario. A complicated model that tries to model physiologically relevant processes using a variety of data sources, like that from Hufkens et al. (2016), would be difficult to estimate in an "online" change detection approach where initial forecasts are typically made with 10 to 20 observations. Methods like model selection using forward or backward step-wise regression or cross validation procedures could be useful, but these methods require some sort of objective function like a F-statistic or a likelihood value. These objective functions are typically derived from the data,

but in our study the objective is to fit the data while finding abrupt changes on the land surface, but might not lead to the most "optimal" model according to model performance metrics.

Inspired by Dutrieux et al. (2015) and Broich et al. (2018), we tested estimating models predicting EVI using rainfall and using a drought index (SPEI). We compared both of these approaches to a fairly standard harmonic regression model with six coefficients: an intercept, a time trend, and cosine and sine pairs for once and twice a year harmonics. The final models we used to test rainfall and SPEI both included a base model with four terms — intercept, time trend, and yearly harmonic cosine and sine terms — as a basis for explaining overall reflectance, trends through time in reflectance, and a basic representation of yearly cycles of solar geometry and vegetation phenology.

When constructing the model using rainfall, we began by testing the influence of rainfall within the same month and rainfall from previous months to attempt to represent memory effects. We fitted CCDC regression models using binned precipitation into groups of three months from up to a year prior to each observation as predictors in addition to the four term base model. We did not perform a robust, quantitative analysis of these models, but instead visually interpreted maps of coefficient estimates and plots of example time series fits to determine which features to select. We found little benefit to including terms other than the monthly rainfall and last 3 months of rainfall terms (Equation 4.1), which were typically significant and important to the estimation of EVI, especially in herbaceous grasslands. Monthly rainfall was almost always estimated to be more important than the rainfall over the last 3 months, but this latter term was likely important in capturing moisture from medium to longer term reservoirs than water from surface runoff likely represented by the monthly term.

The model equation used for testing precipitation is:

$$\hat{EVI}_i = \beta_0 + \beta_1 x_t + \beta_2 \cos(\frac{2\pi}{T} x_t) + \beta_3 \sin(\frac{2\pi}{T} x_t) + \gamma_1 PPT_i + \gamma_2 PPT_{-3:0\ mo} + \varepsilon_t \quad (4.1)$$

where,

T	The number of days in a year (365.25)
x_t	The ordinal date of each observation

We performed a similar process of qualitative interpretation of maps and figures to determine the features to include from the SPEI dataset, but chose only two features to match the two significant features from the precipitation model. We decided to include the 3 month and 24 month integration periods for SPEI to represent shorter term and longer, persistent droughts (Equation 4.3). We also chose to pick a six coefficient harmonic model (Equation 4.2) to use for comparison to keep the number of regressors equal, despite CCDC typically using 8 coefficient models.

The equations for the exclusively harmonic term model used for comparison is:

$$\hat{EVI}_i = \beta_0 + \beta_1 x_t + \sum_{j \in 1,2} [\beta_{2j} \cos(\frac{2\pi j}{T} x_t) + \beta_{2j+1} \sin(\frac{2\pi j}{T} x_t)] + \varepsilon_t \quad (4.2)$$

and the model combining harmonics and SPEI is:

$$\hat{EVI}_i = \beta_0 + \beta_1 x_t + \beta_2 \cos(\frac{2\pi}{T} x_t) + \beta_3 \sin(\frac{2\pi}{T} x_t) + \gamma_1 SPEI_{3mo} + \gamma_2 SPEI_{24mo} + \varepsilon_t \quad (4.3)$$

These three models will be referred to as "Harmonic" for the harmonic only model, "Harmonic + PPT" for the harmonic and precipitation model, and "Harmonic + SPEI" for the harmonic and SPEI model.

4.3.3   Change Detection Algorithms

We used the CCDC algorithm (Zhu and Woodcock, 2014) as the primary means of comparison among the harmonic (Equation 4.2), harmonic with precipitation (Equation 4.1), and harmonic with SPEI models (Equation 4.3) statistical models. In order to isolate the effect on greenness monitoring applications, we modified the CCDC algorithm to use only the EVI band to detect change instead of the default combination of the red, near infrared, and shortwave infrared bands. We also increased the number of consecutive observations required to find change to 7 to help account for the high data density of ARD.

We also evaluated model performance using the Bai-Perron (Bai and Perron, 1998) test for structural breaks in time series, which is the same method used by the BFAST (Verbesselt, Hyndman, Zeileis, et al., 2010; Verbesselt, Hyndman, Newnham, et al., 2010). In order to use this method, we resampled the EVI time series data to monthly time steps matching the precipitation and SPEI data by using a maximum value composite. The Bai-Perron evaluates the best number and location of structural breaks to introduce into a time series by testing all possible combinations of the number and locations of segments. This structural break detection method is a logical extension of the Chow Test (Chow, 1960), which uses an F-test to check if fitting a break at a known point in time, and thus estimating two models, significantly improves the performance over a single model with no break. Building on this idea, Andrews (1993) developed critical distributions for the *sup-F* test distribution that allow for testing the significance of fitting a breakpoint at an unknown location in time. The Bai-Perron takes this approach further by allowing for the estimation of multiple breaks at unknown points in time.

Using the example of a 30 year time series, if the minimum segment size was 2 years then a total of 15 models could be fit over this period. It is likely, however,

that absent real structural change that only a single model would be adequate for explaining the time series. Fitting 15 segment models to 30 years of data will obviously give better predictions than fitting just one model but uses 15 times the number of regressors (i.e., same number of coefficients per model, but 15 more models). The Bai-Perron test accounts for this effect by measuring performance according to the Bayesian Information Criterion (BIC), which weighs increases in model explanatory power with more regressors against the decrease in degrees of freedom. In addition to providing insight into how other change detection methods (BFAST) behave when using meteorological data as driving variables and comparing "online" (CCDC) and "offline" (BFAST) methods, we used the Bai-Perron test to quantify the potential increase in predictive power according to the overall model BIC. When running the Bai-Perron test, the two most influential hyperparameters are the maximum number of breaks allowed and the minimum segment size, which we set to 6 and 2 years respectively.

4.3.4   Sample Design

In order to assess the performance of our changes to the CCDC algorithm, namely changes to the data used in the forecast models, we drew a sample of the study area designed to highlight the differences. The number of changes found by each approach is central to the question of reduction in spurious noise while also being technically and conceptually simple to understand. Confusion matrices that show the difference in the number of changes detected by the "CCDC" and "CCDC PPT" models are shown in Figure 4·1, for the pixel counts, and Figure 4·2, for the percentage of changes found by the "Harmonic + PPT" based model for each number of changes found by the "Harmonic" based model.

The confusion matrices confirmed our suspicion that comparing the two mapping approaches on the basis of the number of changes detected would yield a diverse
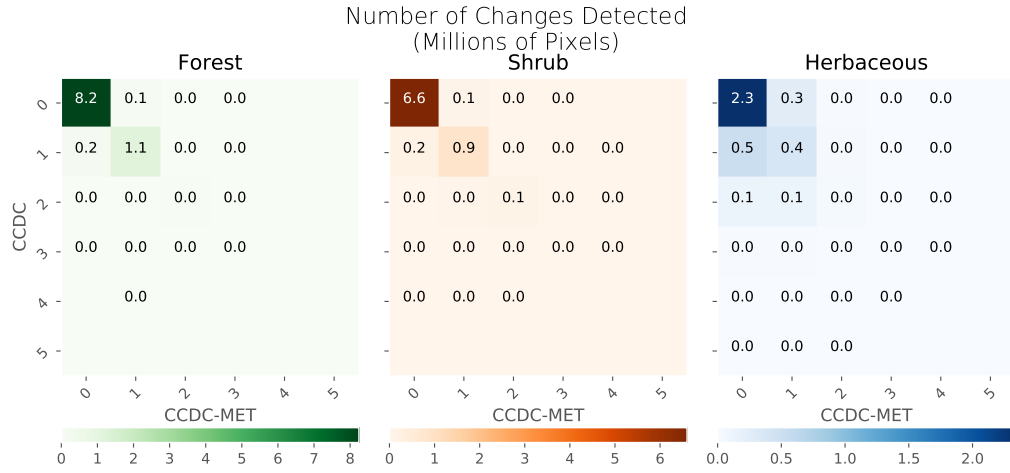
Figure 4·1: Confusion matrix comparing the number of changes detected when running CCDC on EVI data using models based on just harmonic variables ("Harmonic") on the Y axis versus harmonics and rainfall ("Harmonic + PPT") on the X axis. Numbers shown are pixel counts in the millions.

sample that could provide clarity about which approach was better.

4.3.5 Interpretation

We used a team of three interpreters to analyze each sample using a combination of Landsat time series data, PRISM meteorological data, historical high resolution imagery from Google Earth and the National Agricultural Imagery Program (NAIP), and the CAL-FIRE fire perimeters (California Department of Forestry and Fire Protection, 2018) and Great Basin fire (Welty et al., 2017) databases. We used the TSTools time series visualization QGIS plugin to help the interpretations because, similar to TimeSync (Cohen et al., 2010), it allows for simultaneous exploration of the temporal, spatial, and spectral dimensions.

Interpreters were randomly assigned points and tasked with recording information about the change processes between 1987 and 2017, including if there a permanent land cover conversion (e.g., forest to development), how many abrupt change (e.g.,
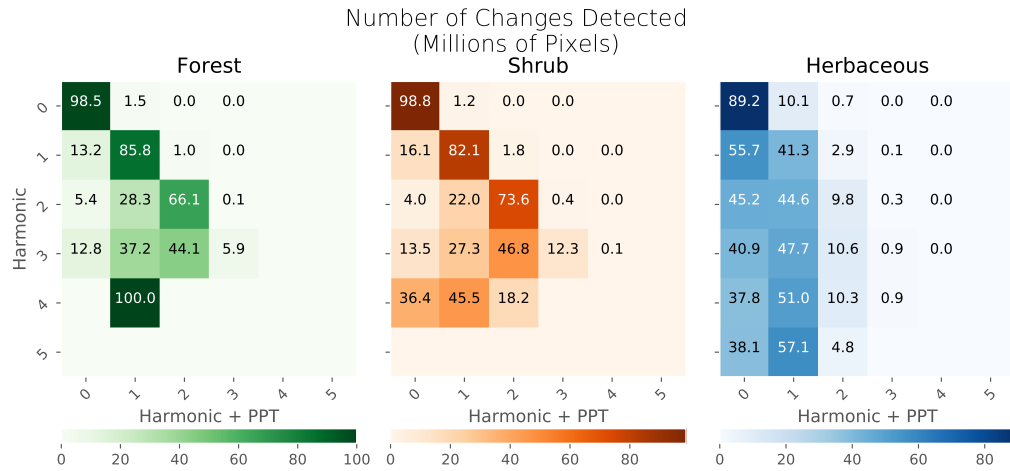
Figure 4·2: Confusion matrix comparing the number of changes detected when running CCDC on EVI data using models based on just harmonic variables ("Harmonic") on the Y axis versus harmonics and rainfall ("Harmonic + PPT") on the X axis. Numbers shown are percentages when summing across the Y axis, and indicate the percent of changes found by the "Harmonic + PPT" model for each number found by the "Harmonic" model.

cyclical forest harvest or fire in a shrubland that doesn't change land cover), when and what process caused each change, and how confident they were overall and about each change date. Because we limited the sample to pixels identified as forest, shrub, or herbaceous cover in 2011 by the NLCD, we could be reasonably assured that there was no permanent land cover change in our sample because most conversions in this area are unidirectional. A forest in the mountains might burn or be logged, thus changing land cover types of "shrub" or "herbaceous" under the NLCD definitions, but it is usually the case that the land cover will return to forest eventually. Central to this study is the idea that drought adapted vegetation may show variability purely driven by environmental conditions, which is distinct from abrupt changes from things like fire, mortality, or logging. As such, we instructed the interpreters to be predominantly looking for abrupt changes that were traceable to events relevant to the land cover,

including fire, flooding or landslides, logging or thinning, permanent conversion, and forest mortality, especially mortality related to the recent drought in the area.

4.3.6 Analysis Design

We assessed the accuracy of the change detections by comparing the "true" number and timing of abrupt changes, as measured by the sample interpretations, against the "predicted" abrupt changes found by the "Harmonic", "Harmonic + PPT", and "Harmonic + SPEI" models when fitted using both CCDC and BFAST. When comparing the "true" versus "predicted" abrupt changes, we required that the date of change be within 365 days of the "true" value to match. Using this method for agreement, we recorded the number of correctly identified abrupt change detections ("Detection"), the number of missed abrupt changes ("Omission"), and the number of "predicted" abrupt changes that did not correspond to a "true" abrupt change ("Commission"). In order to compare these scores across pixels, we normalized the tallied "Detection", "Omission", and "Commission" to be within the range of zero to one based on the number of changes in the data. We plotted the distribution of these accuracy metrics for all scenarios, and calculated the average "Detection", "Omission", "Commission" for each land cover class (herbaceous, shrub, and forest) and time series model ("Harmonic", "Harmonic + PPT", and "Harmonic + SPEI"). We also calculated paired, two-sided t-tests comparing the detection, omission, and commission rates across all approaches (e.g., "Harmonic" vs "Harmoinc + PPT", "Harmonic" vs "Harmonic + SPEI", and "Harmonic + PPT" vs "Harmonic + SPEI") for each change detection method.
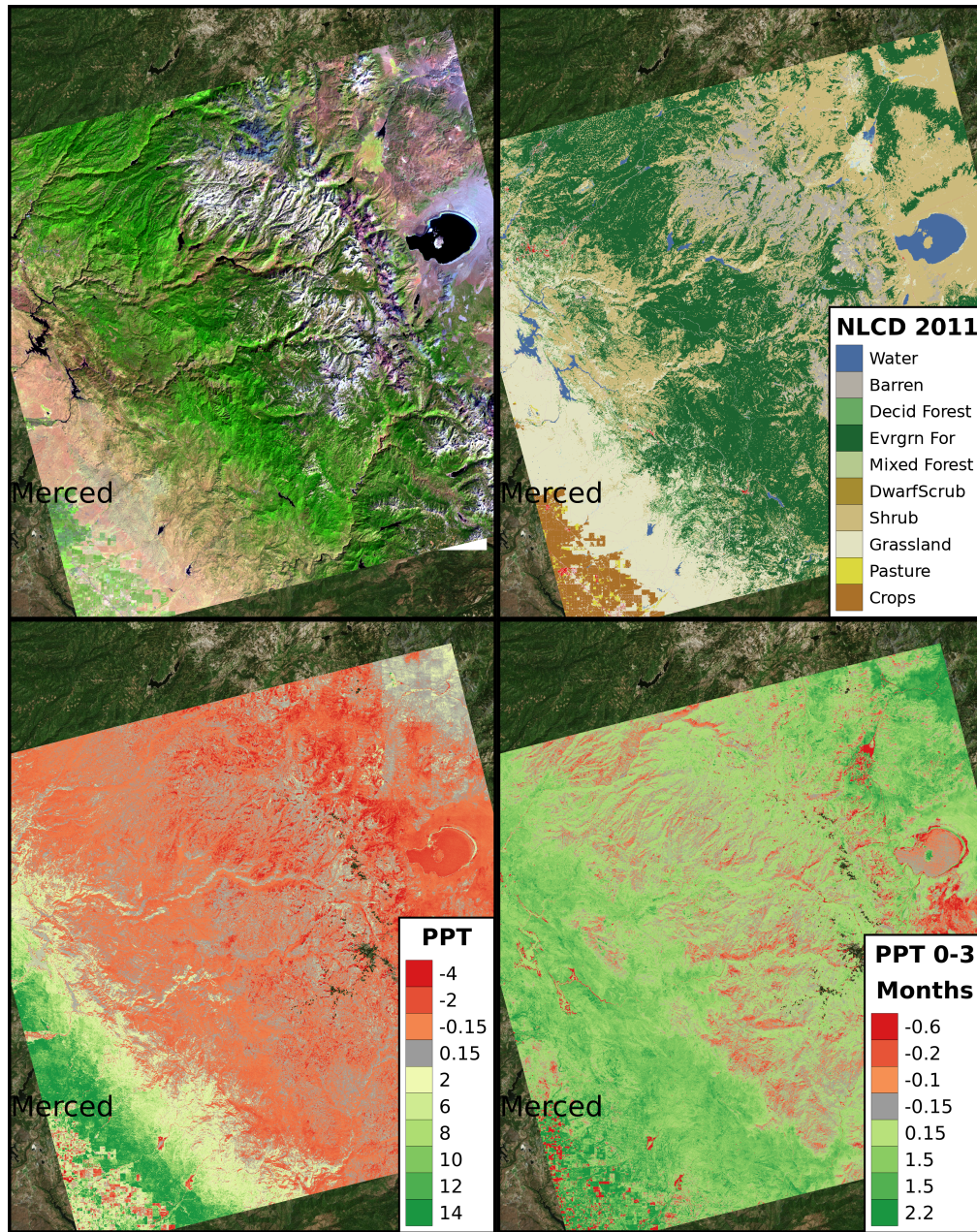
Figure 4·3: Time series model coefficient estimates for the precipitation based model ("Harmonic + PPT"). The top left panel shows the intercept coefficient estimate, normalized for segment slopes, for the SWIR, NIR, and Red bands as RGB. The top right shows the National Land Cover Database map for 2011. The bottom two panels show the coefficient estimates for monthly precipitation (left) and for total precipitation over the last 3 months (right) when estimating EVI.
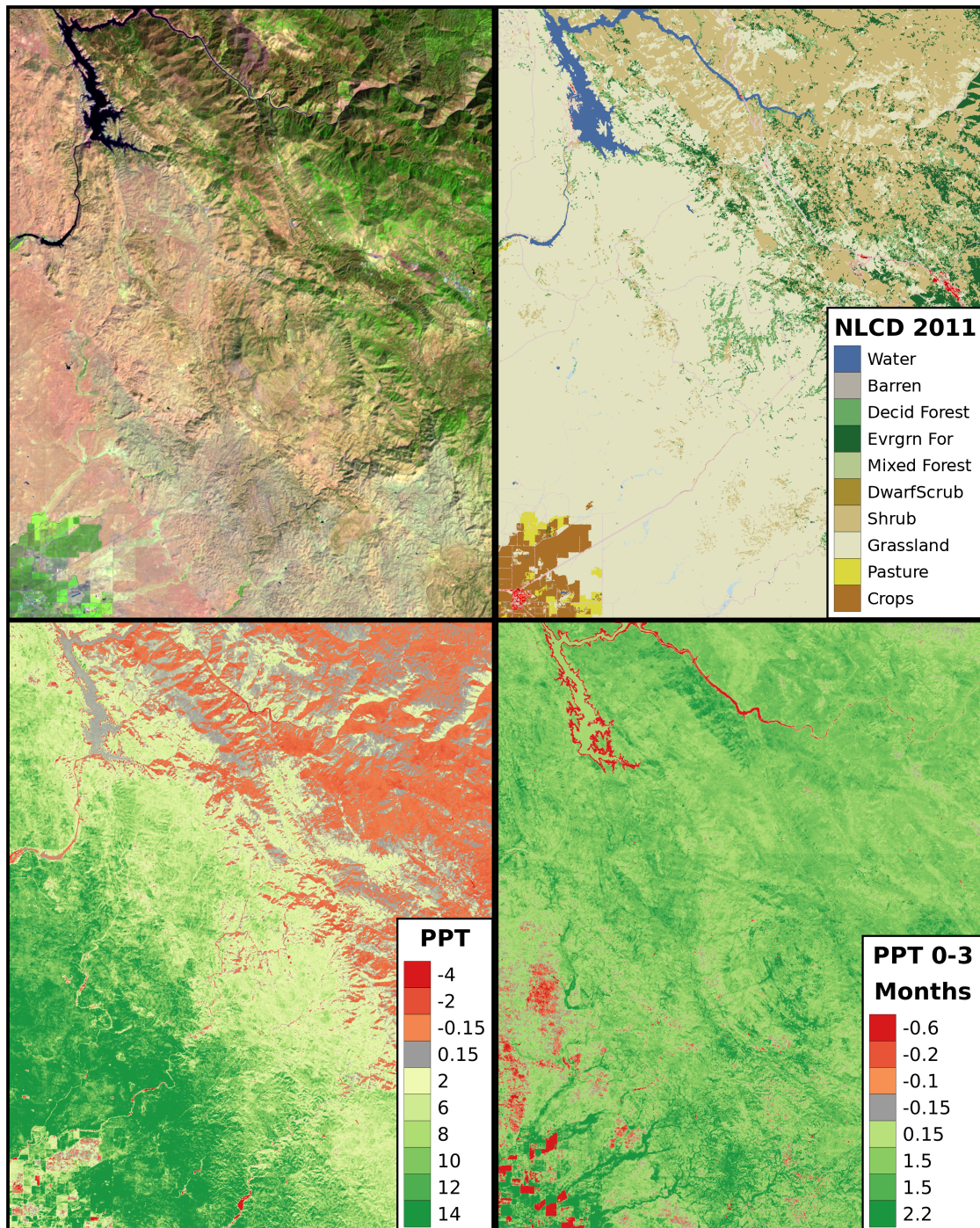
Figure 4·4: Time series model coefficient estimates for the precipitation based model ("Harmonic + PPT"). For interpretation, see Figure 4·3.

4.4 Results

4.4.1 Coefficient Analysis

We mapped the coefficients estimated from "Harmonic + PPT" models for segments intersecting July 1st, 2005 to show the spatial variability in rainfall sensitivity (Figure 4·3). A zoomed in version of these data is shown in Figure 4·4 for a transect of the agricultural fields of the San Joaquin Valley, the grasslands, shrublands, and oak savannas, and the coniferous forests of the Sierra Nevada Mountains.

There is a strong correlation between the location of grasslands (light yellow in the top left NLCD map) and high coefficient estimates for the monthly rainfall feature. Most of the areas mapped as shrub or forest by NLCD show no strong influence from either meteorological variable. Most land covers show positive correlation with monthly precipitation, except for areas in the Sierra Nevada that are typically covered by snow for most of the year or are in the shadow of mountains for all but the summer months. One of the exceptions, which is highlighted in the bottom left of the zoomed coefficient images in Figure 4·4 are the herbaceous crops and orchards (mapped as cropland in the NLCD, and brown on the map), which appear to be negligibly impacted by precipitation.

The maps of the coefficient on the total precipitation over the last 3 months shows much greater spatial variability. The bottom right of the zoomed map in Figure 4·4 shows some of the most extreme coefficient values in the entire map over row crops, including corn, and walnut or almond orchards. This effect in the cropland areas is likely due to a combination of differences in irrigation, or the timing of the planting and harvesting seasons for herbaceous crops. Many of the edges of rivers and lakes in the study site, as well as the mostly barren or permanent snow covered mountain peaks, show negative correlations with the three month precipitation total. The small, seasonal flood channels that extend east to west from the foothills to the valley are

visible in the three month precipitation coefficient image.

We summarized the data shown in Figure 4·3 by grouping the data according to the NLCD land cover class and plotting summary statistics using a boxplot. Figure 4·5 shows the distribution of the coefficient for monthly precipitation when predicting the Landsat optical bands and EVI; the latter of these is also shown in the bottom left corner of Figures 4·3 and 4·4. The monthly precipitation coefficient estimated when mapping herbaceous cover types has the most range and largest magnitude of median and minimum or maximum values. The trend groups as you would expect, with visible and shortwave infrared bands decreasing with precipitation for most of the area, and EVI showing positive correlation with precipitation. The deciduous forest class, which is predominantly comprised of sparse oak savannas with grass understories, is the next most affected by the monthly precipitation term, with approximately 75% of the pixels mapped as this cover in the 2011 NLCD showing positive correlations.

We also summarized the three month precipitation term as boxplots grouped by land cover in Figure 4·6. As expected from looking at the map of this coefficient, the majority of vegetation cover in this study site shows positive correlation between EVI and moisture. The herbaceous and deciduous forest cover types are again the most sensitive to moisture, and these classes show the same pattern of coefficient values for the Landsat spectral bands. Despite general agreement that the three month precipitation term increases EVI and the near infrared reflectance, the shrub, mixed forest, and evergreen forest cover show opposite patterns than the herbaceous and deciduous forest in the visible and shortwave infrared bands.

4.4.2 Change detection examples

CCDC Examples

Example time series and CCDC model results for examples of grassland, shrubland, and regrowing forest pixels are show in Figures 4·7, 4·8, and 4·9. These figures show
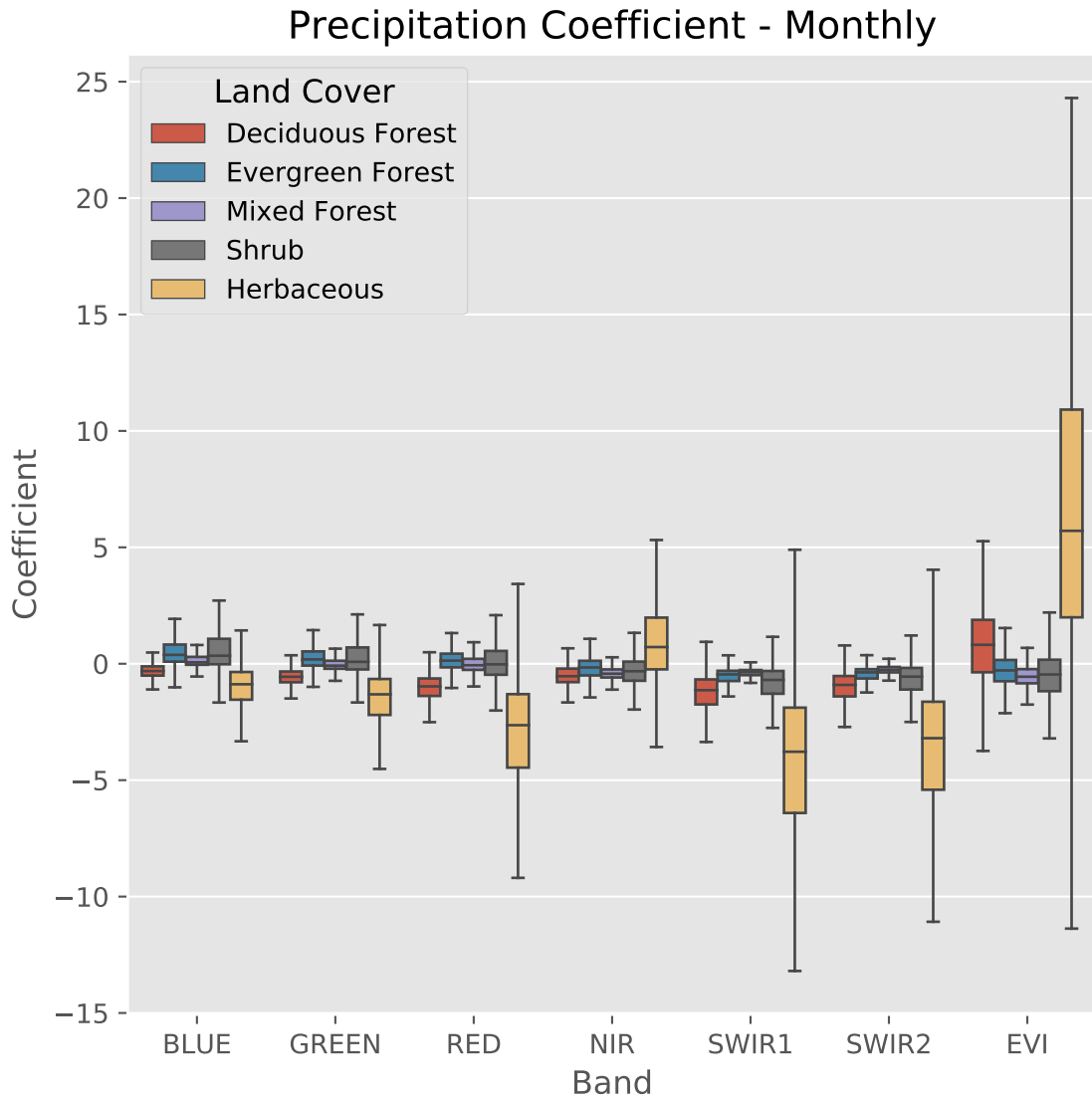
Figure 4·5: Coefficient estimates for the monthly precipitation feature in the "Harmonic + PPT" model, grouped by land cover and ordered by the band or index fitted.

the monthly precipitation values from PRISM on the top, and the EVI time series and CCDC model results for the "Harmonic", "Harmonic + PPT", and "Harmonic + SPEI" experiments on the bottom three subplots. We repeat these same three examples for the Bai-Perron change detection component of this study, and readers may find it helpful to refer to the 24 month integrated SPEI time series at the top
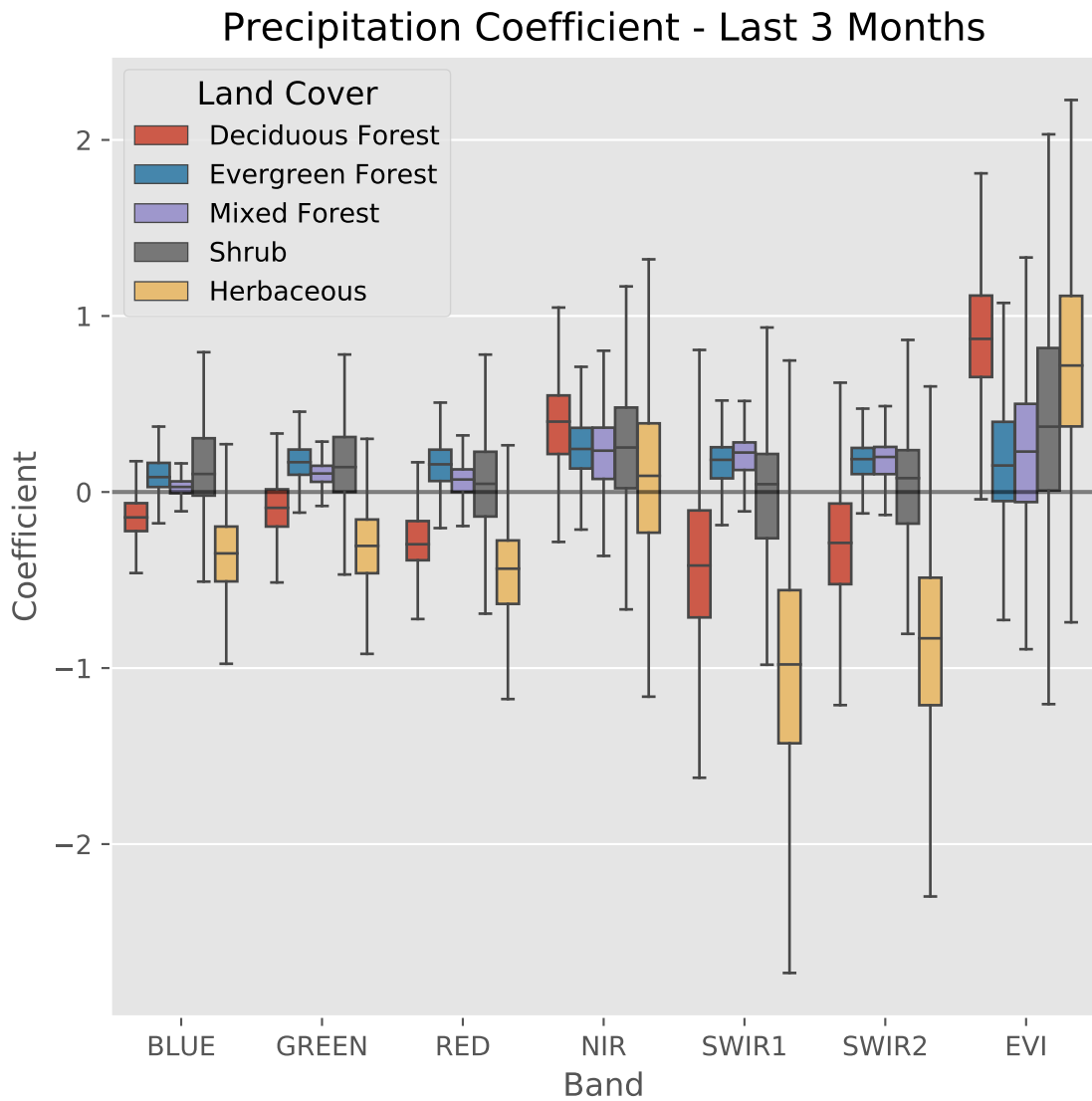
Figure 4·6: Coefficient estimates for the last 3 month precipitation total feature in the "Harmonic + PPT" model, grouped by land cover and ordered by the band or index fitted.

of those figures as another metric for wetness in addition to the precipitation values plotted here.

The example grassland pixel in Figure 4·7 did not have any abrupt changes found by the interpreter, so the breaks CCDC fitted to these time series in this example are considered errors of commission. As we expected, the "Harmonic" model does not
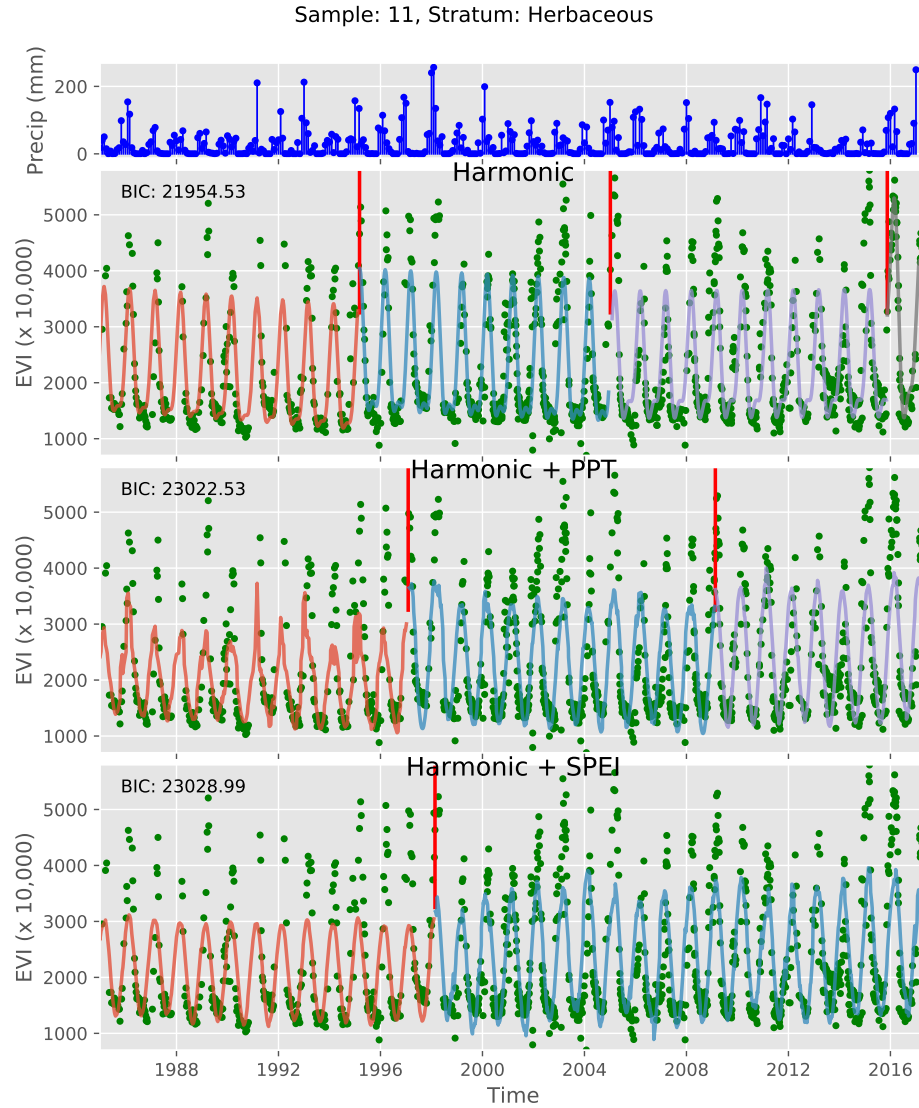
Figure 4·7: Example of model performance for CCDC on time series of Landsat EVI for a stable grassland pixel. The top panel shows the PRISM precipitation time series. The bottom three panels shows the monthly maximum value of EVI, with model diagnostics from the SPEI-based regression model and harmonic-only regression models. Time series model predicted values for each segment are shown as lines, breaks in the model as vertical red lines, and the overall BIC value for each time series is shown in the top left.

represent the dynamics of peak greenness observed during each year in the time series and finds spurious changes in the time series. Despite this, the "Harmonic" model has

a lower overall BIC score than the other models. The "Harmonic + PPT" model finds one fewer change than the "Harmonic" model, and the "Harmonic + SPEI" model finds the fewest number of breaks (n=1). Despite inclusion of the meteorological information in these last two models, the predicted values do not match the variability in peak greenness.
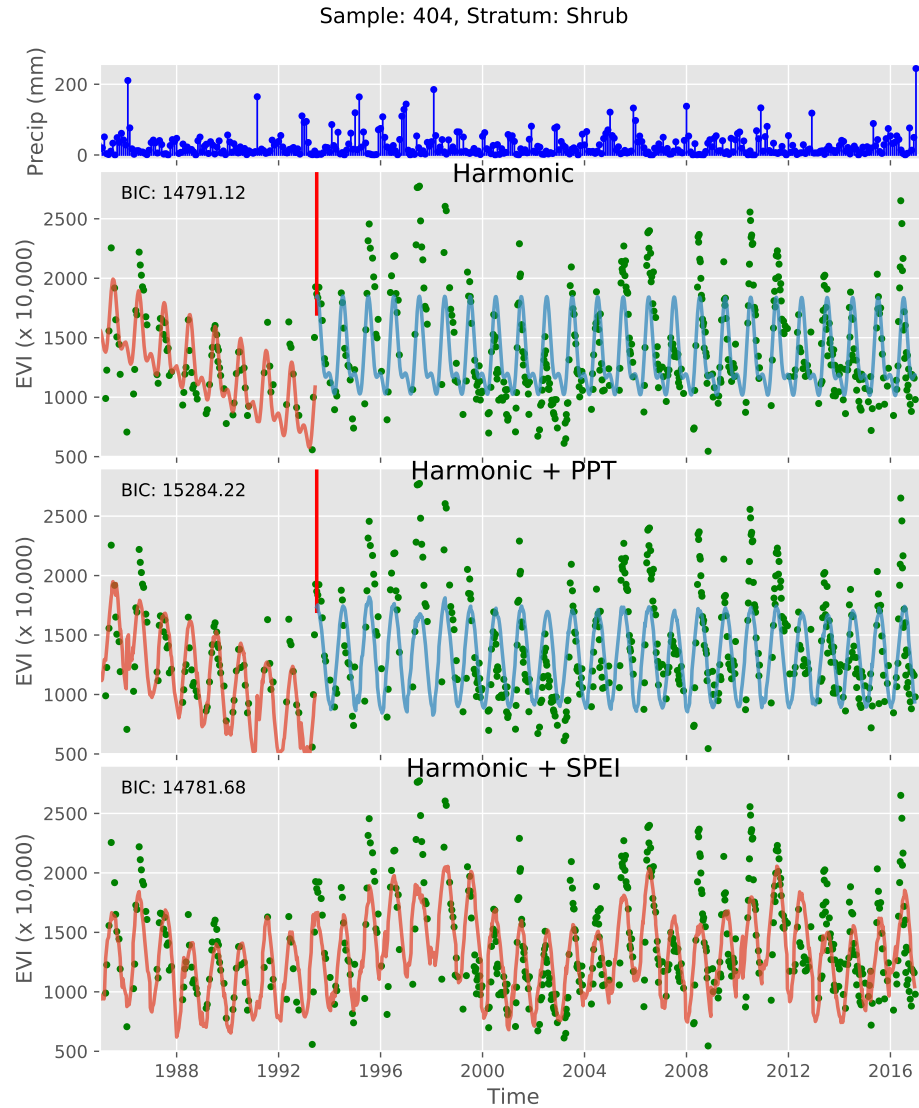
Figure 4·8: Example of model performance for CCDC on time series of Landsat EVI for a stable shrub pixel. For interpretation of this plot, refer to Figure 4·7.

The example included here for shrub cover types shows great range in EVI values over time (Figure 4·8), usually following the patterns of drought, although this variability can be difficult to see underneath the predicted segment lines. The interpreter of this sample noted that there were no abrupt changes in this time series, but that the shrub cover exhibited variability likely due to climate factors. Both the "Harmonic" and "Harmonic + PPT" models estimate abrupt changes early in the 1990s, which approximately coincide with the end of a drought that began in the late 1980s. The precipitation and SPEI based forecast models show some variability in the predicted values, with the SPEI based model seeming to fit the time series the best of all approaches (as indicated by model BIC and visual comparison with observed data).

Figure 4·9 shows a time series for a coniferous forest stand that burned in the Clavey and Paper fires of 1987 and the Rim fire in 2013. Between these dates the forest regrew, but exhibited year to year variability that looks correlated with climate variability. Using CCDC, all three models missed the first change in 1987, likely because the change occurred near the beginning of the time series and the minimum number of observations needed to begin each model was set at 24.

Bai-Perron Examples

Example time series and Bai-Perron model results for the examples of grassland, shrub, and regrowing coniferous forest pixels used previously are show in Figures 4·10, 4·11, and 4·12. Except for the model predictions and breakpoints, which differ because of the estimation method (CCDC vs Bai-Perron), the bottom three panels are the same as the previous figures (Figures 4·7, 4·8, and 4·9). The top panel, however, now shows one of the driving variable behind the "Harmonic + SPEI" models — the 24 month integrated SPEI anomaly. Dryer than normal anomalies are colored in red and wetter than usual anomalies are colored in green. The BIC values estimated for
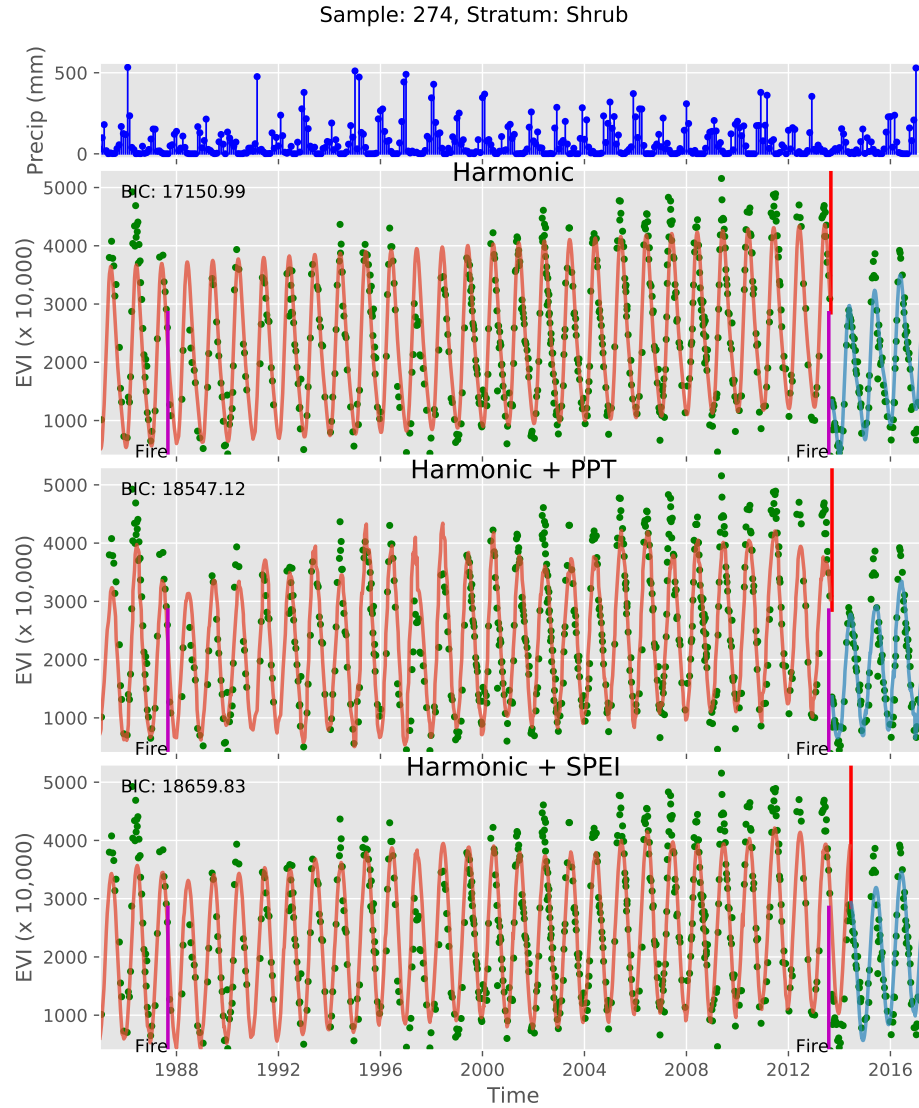
Figure 4·9: Example of model performance for CCDC on time series of Landsat EVI for a forest pixel that was burned in 1987, regrew, and was burned again in 2013. Note that this pixel was a regrowing forest in 2011, but NLCD considers regrowing forest to be within the shrub class. For interpretation of this plot, refer to Figure 4·7.

these models may be used to compare against other models fitted using the Bai-Perron method, but are not applicable to the CCDC model estimates because the underlying data source is different (all available observations versus monthly maximum value composites).
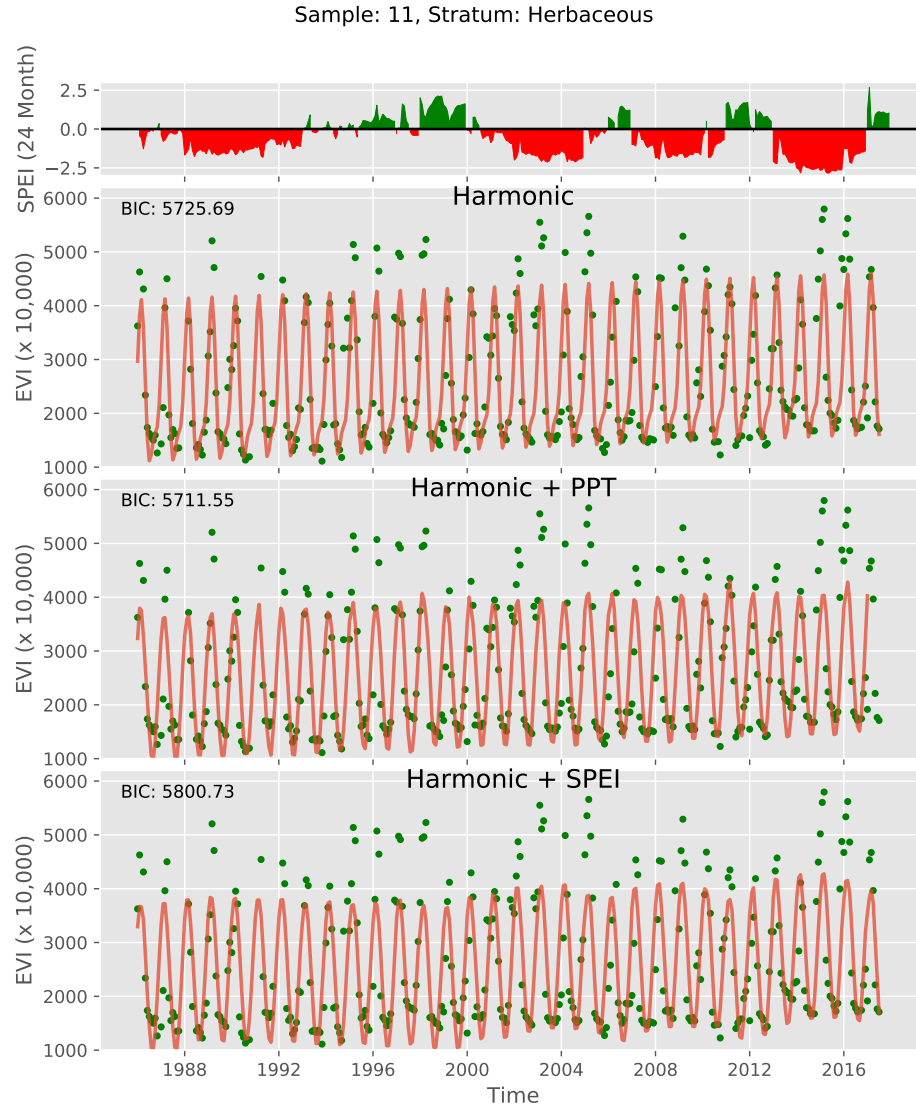
Figure 4·10: Bai-Perron change detection method for a stable grassland pixel. The top panel shows the 24-month integrated SPEI index over the time series record, with green periods indicating wetter than average conditions and red indicating dryer than average conditions. For interpretation of the bottom three panels of this figure, refer to Figure 4·7.

Figure 4·10 shows the performance for the Bai-Perron breakpoint detection method for a stable herbaceous grassland. Compared to the result using CCDC for this same example (Figure 4·7), the Bai-Perron method finds no spurious breaks. The size of

the coefficients estimated using the Bai-Perron method are also much larger than the coefficients estimated using CCDC, allowing all three models to better capture the peaks and troughs of greenness over the time series.
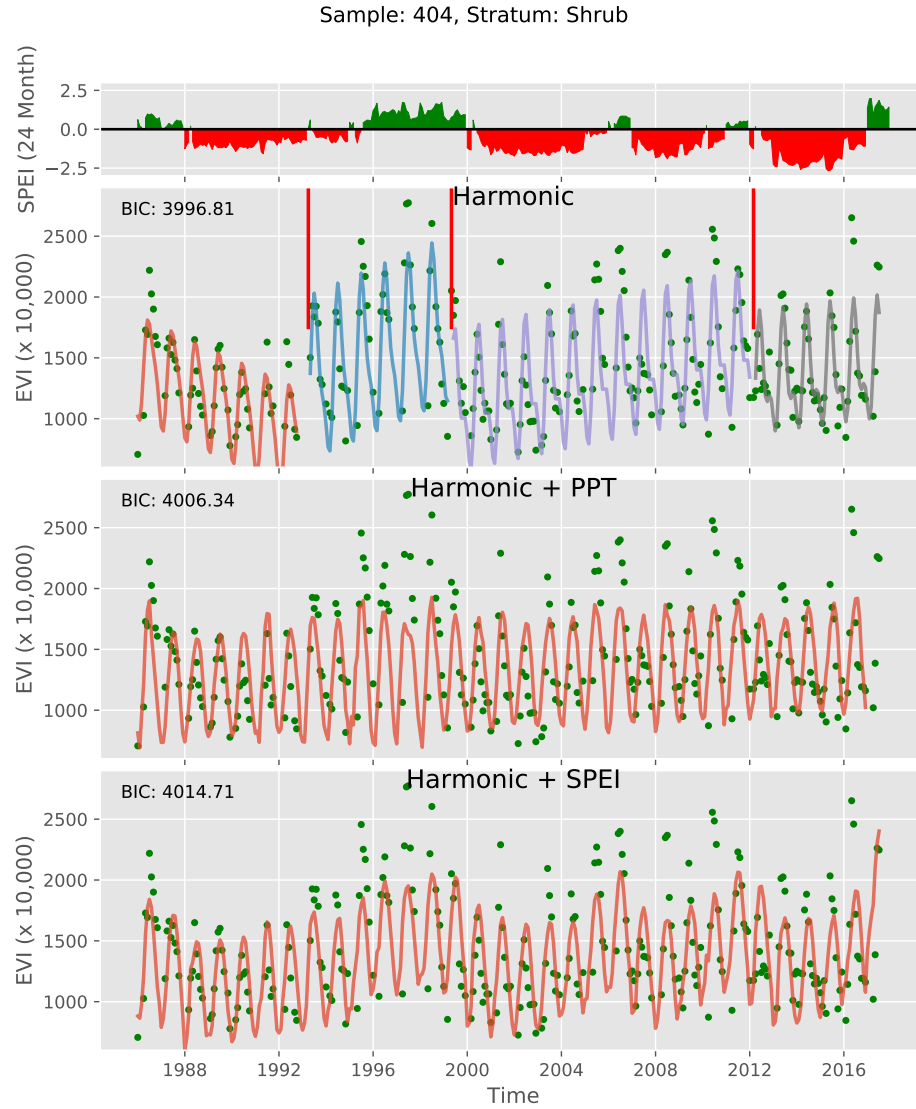
Figure 4·11: Example of model performance for Bai-Perron structural break detection on time series of Landsat EVI for a stable shrub pixel. For interpretation of this figure, refer to Figure 4·10.

The Bai-Perron results for the "Harmonic + PPT" and "Harmonic + SPEI" models do not estimate a break in the time series for the stable shrubland (Figure 4·11).

Lacking information on moisture dynamics, the "Harmonic" model fits a total of four segments. In place of capturing moisture dynamics, these models fit large time trends that seem to match moisture variability, but these models need to break whenever the moisture conditions change. By comparison, the precipitation and SPEI based models show almost no trend through time, instead varying stochastically with the meteorological data. Despite what visually looks like an increase in model performance, the harmonic regression model does achieve a lower BIC with multiple segments, even despite BIC penalizing the degrees of freedom removed with each newly estimated segment.

The Bai-Perron test parameterized using a minimum segment length of two years was able to capture the fire that burned the forest stand in the example pixel shown in Figure 4·12, while our parameterization of CCDC was not able to capture this break (see Figure 4·9). All three time series model approaches find this change late, likely immediately after the 24 minimum observations (or two years) were met. As with other examples, the model only using harmonic features estimates more breakpoints during the regrowth trajectory than the models using precipitation or SPEI. All three modeling approaches find the second fire event in 2013.

4.4.3   Change detection accuracy

Accuracy Assessment

We summarized the results of the accuracy assessment in Table 4.1 for the CCDC results and Table 4.2 for the Bai-Perron results. The "Detection" column shows the average across all sample pixels of the percentage of the changes that were identified in the reference data that were also found by the land cover change detection algorithms. The "Omission" column is calculated as the proportion of changes identified by interpreters that were not captured, and is the inverse of the detection rate. The "Commission" column is the average across all pixels of the proportion of false
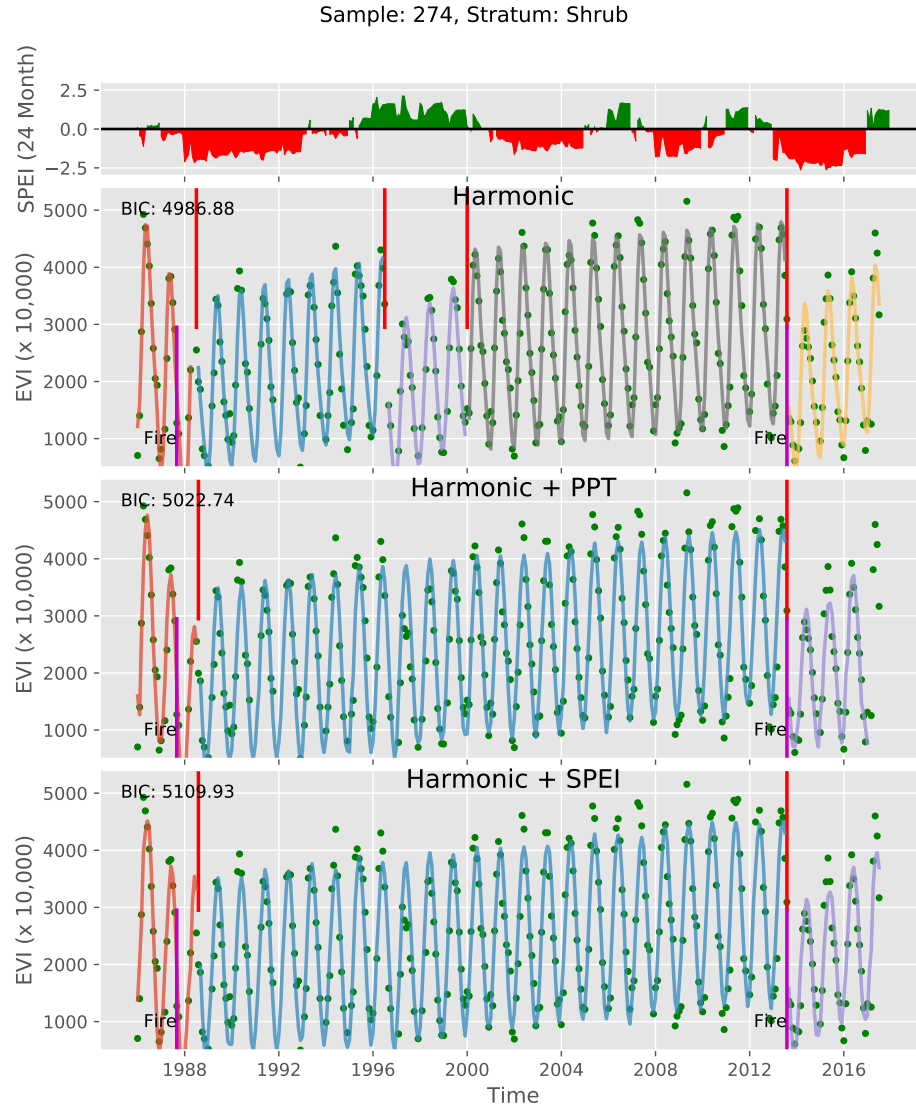
Figure 4·12: Example of model performance for Bai-Perron structural break detection on time series of Landsat EVI for a forest pixel that was burned in 1987, regrew, and was burned again in 2013. For interpretation of this figure, refer to Figure 4·10.

positives identified by the change detection methods. Violin plots showing the full distribution of these accuracy metrics for each time series model and land cover type are included in the Appendix (Figures B.1.1 and B.1.2).

Using the regrowing forest pixel example and the Bai-Perron results from Figure

4·12 as an example, there were two fire events found in the reference data. All three models found both of these fire events, so the "Detection" proportion would be 100% and the "Omission" proportion would be 0%. The precipitation and SPEI based models did not find any other changes in the time series that were not recorded in the reference data, so their "Commission" proportion would be 0%. The harmonic only model, however, fitted two breakpoints out of four total that are not confirmed in the reference data, so the "Commission" proportion would be 50%. We took the average of these calculations over all pixels in the sample and present them in Tables 4.1 and 4.2.

Model	Land Cover	Detection	Omission	Commission
Harmonic	Overall	85.58	14.89	62.91
	Forest	77.34	23.38	47.12
	Herbaceous	92.20	7.80	91.49
	Shrub	87.09	13.62	50.00
Harmonic + PPT	Overall	84.91	15.56	59.24*
	Forest	75.30	25.42	49.28
	Herbaceous	92.91	7.09	84.75**
	Shrub	86.38	14.32	43.66
Harmonic + SPEI	Overall	85.35	15.13	57.56**
	Forest	74.22*	26.50*	48.20
	Herbaceous	93.62	6.38	87.06
	Shrub	88.03	12.68	37.44***

Table 4.1: Average correct number of abrupt change detections, and omission and commission rates for sample using the CCDC algorithm. Asterisks indicate a significant difference between the harmonic only and other models (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$)

Overall, the CCDC method was more accurate at finding abrupt disturbances

Model	Land Cover	Detection	Omission	Commission
Harmonic	Overall	86.70	13.30	61.21
	Forest	77.82	22.18	64.95
	Herbaceous	93.98	6.02	53.57
	Shrub	88.44	11.56	64.91
Harmonic + PPT	Overall	86.21	13.79	46.03***
	Forest	76.10	23.90	58.76***
	Herbaceous	93.23	6.77	30.45***
	Shrub	89.42	10.58	48.53**
Harmonic + SPEI	Overall	84.85*	15.15*	45.55***
	Forest	72.79**	27.21**	61.27
	Herbaceous	92.48	7.52	28.42***
	Shrub	89.42	10.58	46.58***

Table 4.2: Average correct number of abrupt change detections, and omission and commission rates for sample using the Bai-Perron algorithm. Asterisks indicate a significant difference between the harmonic only and other models (* $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$)

within the forest and shrub classes than the Bai-Perron method, but less accurate when finding change in the grasslands. Omission rates were the highest for the forest class regardless of model or break detection approach, and the "Harmonic + SPEI" had significantly higher omission rates than the "Harmonic" model using either CCDC and Bai-Perron. Other omission rates were not significantly different when comparing against the harmonic only model. There was no significant difference between the performance of the "Harmonic + PPT" and the "Harmonic + SPEI" models.

The largest differences for the both the CCDC and Bai-Perron approaches occur when comparing the rates of commission for the shrub and grassland pixels under the harmonic versus the harmonic and meteorological data models. Incorporating

meteorological data, whether it is precipitation or SPEI, significantly reduced the commission rate overall, but especially for the herbaceous and shrub classes. Using CCDC, the precipitation based model significantly reduces commission error for the herbaceous compared to the "Harmonic" model, while the SPEI based approach significantly reduces the omission error for the shrub class. We saw a a greater impact on the commission error when using the Bai-Perron test, with the precipitation model significantly lowering commission error rates overall and for each land cover class individually. Meanwhile, the SPEI based model significantly reduced the commission error overall and for shrubs and herbaceous cover, but this model did not significantly affect the commission error for forest.

Model goodness of fit comparisons

While comparison against reference data gives information on whether or not including meteorological data affects the change detection performance, we also wanted to compare the difference in model fit performance between the typical harmonic regression model and those using meteorological data. The Bai-Perron structural break test uses the BIC to decide which possible breakpoint models is best, and this metric, or the related Akaike Information Criterion, is frequently employed when performing model selection in time series regression contexts. We calculated the differences in the BIC scores when using the "Harmonic" model compared to the "Harmonic + PPT" and "Harmonic + SPEI" models, and plotted the distribution of these differences in BIC as a function of the difference in abrupt changes detected in Figures 4·13 and 4·14.

The BIC weighs model complexity against model likelihood, and so models that are more likely (i.e., have better performance) will have a lower BIC score. As such, a positive difference between the harmonic only and harmonic with precipitation or SPEI models indicates superior performance when using meteorological data. Nega-
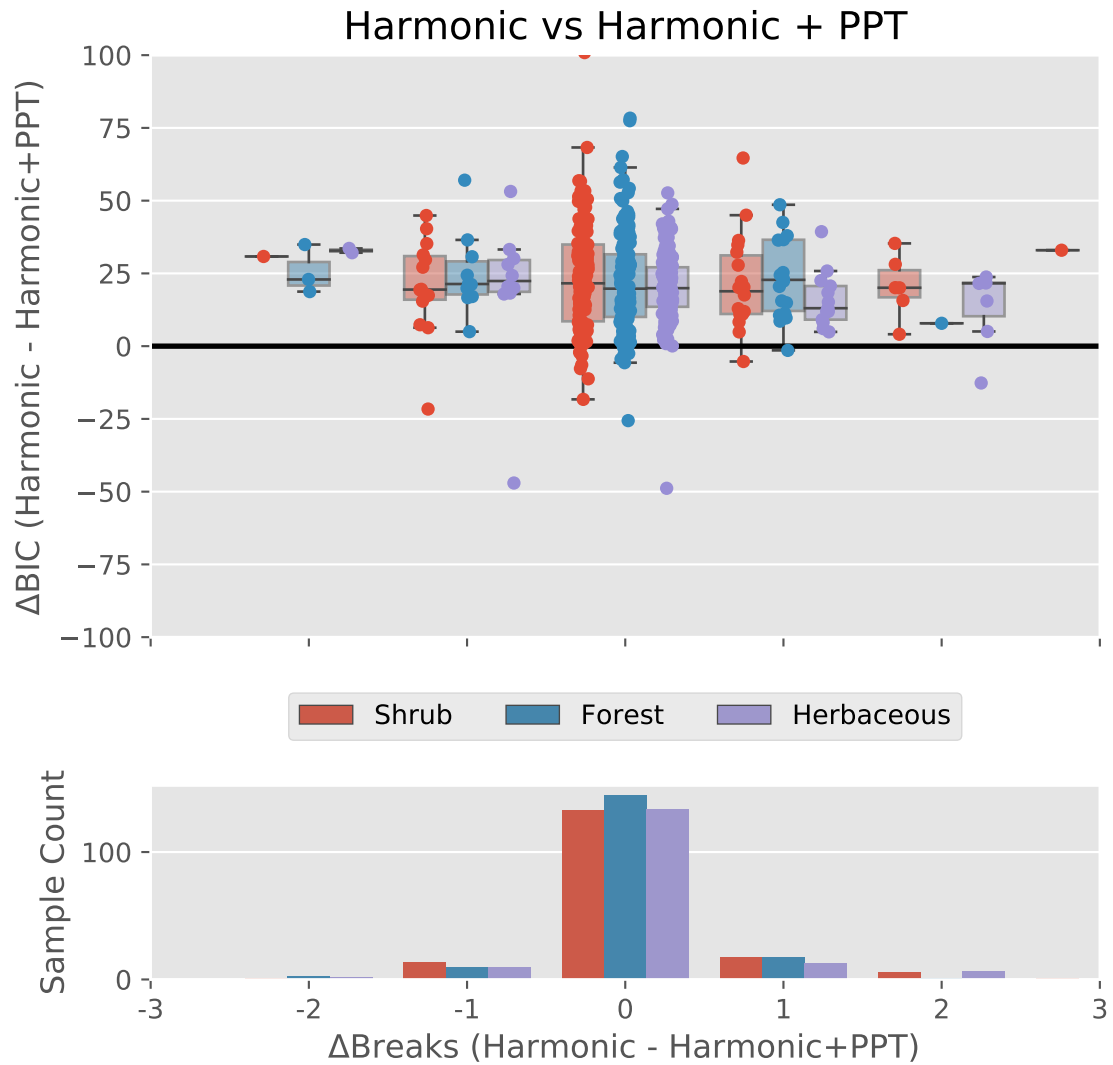
Figure 4·13: Model performance comparison for Bai-Perron break-point detection on time series, sorted according to the difference in the number of breaks between the harmonic only and harmonic and precipitation models. The top panel shows the difference in BIC between the two models, with positive values indicating superior performance for the harmonic and SPEI model over the harmonic-only model. The X axis has been truncated to show differences in abrupt changes fitted of less than 3 for readability.
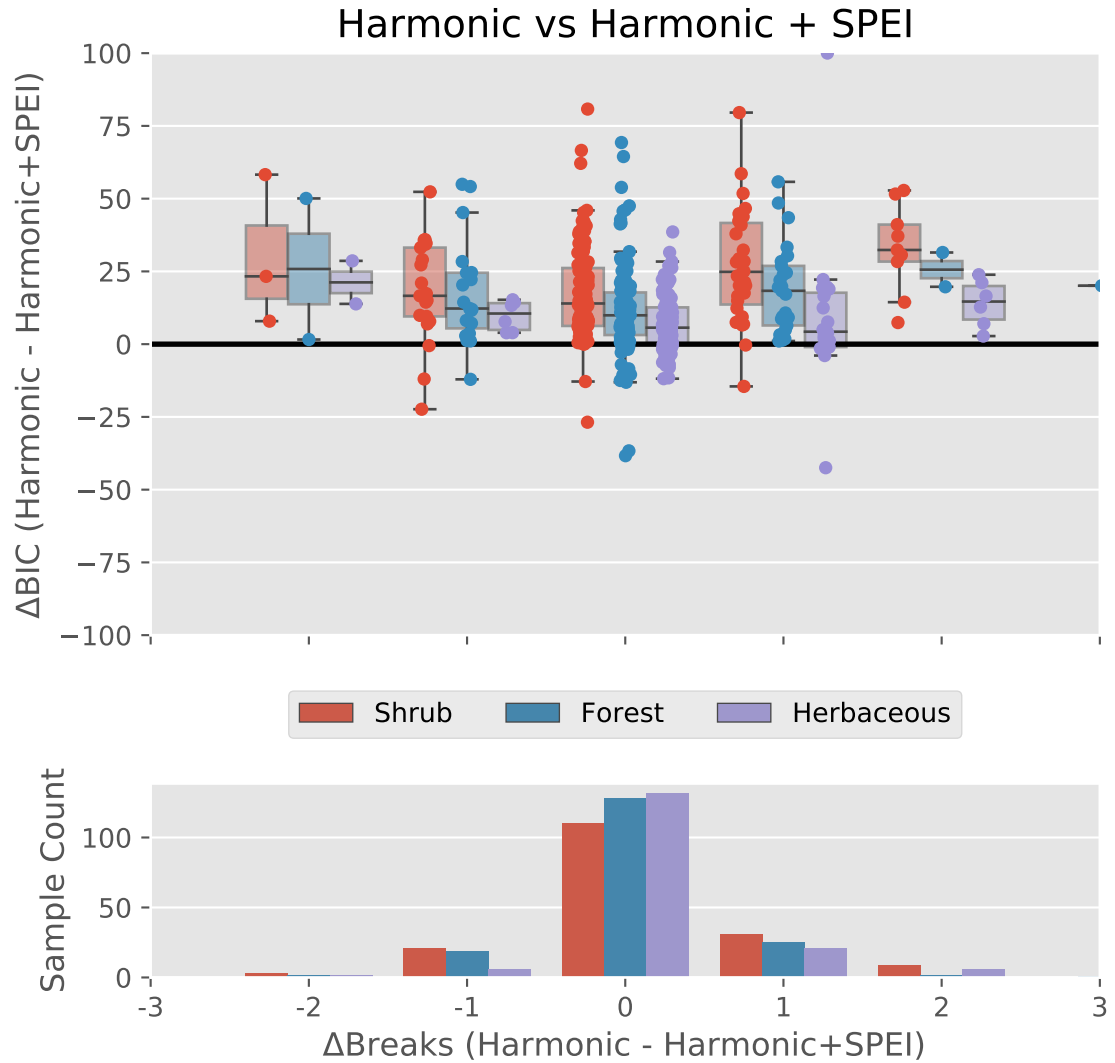
Figure 4·14: Model performance comparison for Bai-Perron break-point detection on time series, sorted according to the difference in the number of breaks between the harmonic only and harmonic and SPEI models. See Figure 4·13 for full interpretation details.

tive difference in BIC values indicate that the harmonic model was superior.

For the majority of samples, the precipitation and SPEI based models outperformed the harmonic only model for all three land cover types considered. The improvement when using meteorological data was especially pronounced for pixels

that estimated different numbers of breaks than the harmonic models. As seen in the bottom panel barchart that show sample count frequencies binned by the difference in the number of changes detected, the SPEI based model tends to estimate slightly fewer breakpoints than the harmonic only model.

4.5 Discussion

4.5.1 Regression coefficient analysis

Figure 4·3 shows patterns of precipitation driving greenness that varies across land use (e.g. irrigation and crop type) and land cover type (herbaceous, shrubs, oak savanna, and coniferous forests). The coefficient estimates for monthly precipitation are negatively related to greenness for the predominantly shrub and evergreen forest cover of the Sierra Nevada mountains and the shrublands of the Great Basin to the east. The relationship estimated here is likely because precipitation is acting as a proxy for other, omitted variables, like temperature and other physiological controls, or solar radiation geometric effects. A typical EVI profile for the shrublands or coniferous forest is fairly flat (Pasquarella, Holden, Kaufman, et al., 2016), and while there is down-regulation of photosynthesis, the influence of sun angle in this data, which hasn't been corrected for any direction effects, is likely the main driver. Precipitation in California typically arrives in the winter, just as the measured EVI decreases with decreasing solar angle, and this mechanism is likely driving these estimates. Artifact or alias for omitted variables, the spatial variability and grouping by land cover of these estimated coefficients suggest they would be useful for classification.

The monthly precipitation coefficient tends to be very negative around rivers and the edges of lakes, because the added rainfall is linked to expansions of these bodies of water, which reduces greenness. We also see large coefficient estimates for flood channel features in the grasslands along the San Joaquin Valley, suggesting that these

results may be useful for identifying "boom and bust" ephemeral grasslands fed by seasonal streams (Broich et al., 2018). The spatial patterns observed in the croplands and orchards of the San Joaquin Valley require further investigation, but they may be related to whether or not and for how long the field is irrigated.

Of interest to time series change detection methods, many of which rely on the shortwave infrared bands, Figure 4·5 shows that for many pixels there is a considerable negative influence of precipitation on the first and second SWIR channels for the herbaceous and deciduous forest classes. This effect is likely due to the reduction in brightness of the soil due to either wetter soil or herbaceous vegetation cover.

4.5.2 Change detection performance

The addition of meteorological information to the forecast models used by time series models generally reduced the commission error when compared to the "Harmonic" model. The only trade-off for using this information was seen for the SPEI model when mapping forest cover change, with significant increases in omission error for the forest class when using both CCDC and the Bai-Perron methods. The accuracy of the forest class was also the lowest of all land cover types, likely due to the use of the EVI over a shortwave infrared based index or band. We also observed examples where forest change was not detected in the "Harmonic + PPT" and "Harmonic + SPEI" models which appear to be caused by the forest cover change coinciding with a drought shortly after model initialization. Instead of finding a break in the time series, many of these models fit large, negative estimates to the meteorological data coefficient, allowing for the reduction in EVI that came with forest change to be explained using the moisture conditions instead of a land cover change.

Using the CCDC, the precipitation based model significantly reduced the commission error for the herbaceous cover while the SPEI significantly reduced the commission for the shrub class, which makes sense intuitively and ecologically. Grasslands

are much more tightly coupled with precipitation and previous results have noticed that precipitation can cause "boom and bust" cycles of greenness (Seddon et al., 2016; Broich et al., 2018). Shrubs, by contrast, persist year to year and likely have deeper roots than grasses that provide access to longer term water resources than runoff driven moisture in the first inches of soil. The shrubs in this area are drought adapted, and when we visited this site in the spring of 2017 we noticed that many shrubs had allowed portions of their branches to die off, likely in response to recent droughts. An index like SPEI that can be integrated over long time periods is a better predictor of this slower response to moisture conditions.

The BIC comparisons showed that inclusion of these environmental drivers almost always improved model performance, which has implications for the number and type of changes that are detectable. In the accuracy assessment sample data, we saw instances where the increased predictive performance when using SPEI or precipitation resulted in accurate predictions that avoided finding a false positive where the simple harmonic model fit a spurious change. In other cases, the increased predictive ability allowed the CCDC or Bai-Perron models to detect more breaks than the simple harmonic model. In these cases, the simple harmonic model tended to have very high RMSE values so only very large abrupt changes would fall outside the forecast confidence intervals. Analyzing the BIC of these models (Figures 4·13 and 4·14) and assessing model predictive power is important because the observed decrease in commission error could have been driven by increases in model error. If moisture dynamics were not important for time series predictions, model error would be higher than when using harmonics, making it more difficult to fit a break, real or spurious, using CCDC. The BIC analysis helps prove that the decrease in commission error was due to meaningful improvements in model predictive performance and not from increases in model error.

Similar to what was observed in (Broich et al., 2018), we also found that many of the changes detected coincided with periods of extreme drought, especially during the dry periods of the early 1990s and in the early 2010s. The example included in this analysis of a stable shrubland that is affected by drought (Figure 4·8) is a good example of changes that are found by the "Harmonic" model which may provide information, albeit confusing information, about vegetation response to drought. For example, it may be possible to predict a "change agent" that caused the falsely identified abrupt disturbance, and with training data perhaps classify this abrupt change as being caused by drought. It is also likely that the land cover classifications of the first and second segments would be the same (i.e., shrubland), which could help identify the abrupt change as an ephemeral change or some other kind of disturbance and not a land cover conversion.

While breaks caused by ephemeral processes or climate variability might not be of utility for land cover change applications, one might actually want to include these breaks in the time series. In fact, instead of monitoring for structural changes in the residuals, future studies could monitor for change in the coefficient estimate for rainfall or SPEI, as done in previous studies (Verbesselt, Hyndman, Newnham, et al., 2010; Broich et al., 2018). Monitoring for structural change in residuals was able to capture most of the abrupt changes, but it is possible for the relationship between predictor variables and EVI to change over time as the time series models incorporate more data. Monitoring the estimated coefficient values can help identify these changes in variable relationships (Zeileis et al., 2003), although it is likely that maps produced by monitoring of coefficients will differ than maps produced from residual monitoring. Just as there has been development of spectral indices for specific tasks in remote sensing, time series algorithms might develop specific routines or parameterizations for a particular question. There is already evidence that outputs from existing time

series methods contain unique information when compared against each other, even when using the same input data across that focus on forest change monitoring (Healey et al., 2018), and combining outputs from two time series algorithms has shown to be useful when classifying forest community composition (Pasquarella, Holden, and Woodcock, 2018).

Comparing the models estimated by CCDC and by the Bai-Perron test for the same pixels (Figures 4·8 and 4·11 for shrubs and Figures 4·7 and 4·10 for grassland), one immediately sees that the coefficients estimated by CCDC are reduced compared to those estimated by the Bai-Perron approach. This is because CCDC uses the Lasso (Tibshirani, 1996) regression method which applies a regularization penalty on the size of the size of the regressors as a balance to the model prediction performance (i.e., sum of squared residuals), while the Bai-Perron models are simple ordinary least squares. The use of the Lasso versus ordinary least squares is also likely partially responsible for the differences in the reduction of commission error when comparing the CCDC and Bai-Perron results (Tables 4.1 and 4.2). Regularization shrinks the impact of the meteorological data in the models, which reduces its potential for affecting results.

The Lasso regression method is useful for change detection, especially "online" change detection, because this shrinkage reduces variance at the cost of bias. Being a biased estimator, the BIC scores are always much larger for CCDC estimated coefficients. The use of regularization also limits the ability for the climate data features to influence the regression. Estimating the relationship between spectra and meteorological variables might be a good step to perform using OLS after finding changes using CCDC, so the change detection could have the relative stability of a regularized regression while unbiased estimation procedures could be used to estimate parameters for each segment. One might also estimate a robust regression using iteratively

reweighted least squares to provide unbiased estimates of the coefficient parameters while avoiding contamination from noise like missed clouds or shadows.

4.5.3 Time series interpretations

We found the CAL-FIRE dataset to be very useful in helping to quickly confirm signals in the Landsat time series plots, especially in assisting in finding the Landsat observations closest to the disturbance. Not all fires were captured, and the product is more prone to errors of commission than omission, as is typical with hand drawn polygons. When double checking the interpretation results, it was very common that the interpreters would identify the date of change for a fire in the early spring that actually happened in the late summer, even when there were available clear observations of the burn. In California, the winter months (October - February) are typically the rainy season and cloud cover reduces the availability of imagery, which contributes to the temporal inaccuracy of the interpretations. The CAL-FIRE database was at least a worthwhile dataset to screen interpretations for possible omissions or clerical errors, and might be useful in time series analysis to train or calibrate models or perhaps serve as a prior probability for finding change.

4.5.4 Landsat Analysis Ready Data

We are among the first studies to use Analysis Ready Data (ARD; Dwyer et al., 2018), and generally have a very positive impression of the product. During initial exploratory phases of this research, we assembled and preprocessed all Landsat WRS-2 path and row footprints ourselves for this part of the San Joaquin Valley to take advantage of overlap. The process of downloading the data, performing reprojection from Universal Transverse Mercator (UTM) it into a wide area projection suitable for large area mapping (which also allowed us to combine data from adjacent paths with different projections), and tiling the data to our destination grid required a lot of

computation that ARD completely eliminates. Worse, the result of these preprocessing steps would never have been as accurate as ARD because ARD is only resampled once while our approach resamples the data a second time.

The increased temporal density presents new opportunities, but also challenges due to the spatial heterogeneity in the density of observations and because of differences in observations from adjacent paths. While ARD allows for greatly increased data densities in the overlap regions, algorithms that do not use fixed intervals (e.g., 16-day, monthly, or yearly composites) may encounter issues, especially with algorithms that rely on a moving window, such as the the bandwidth of the Moving Sum (MOSUM; Chu et al., 1995; Zeileis et al., 2003; Zeileis, 2005) test for BFAST Monitor (Verbesselt et al., 2012), or the consecutive number of observations from CCDC. In our experience running CCDC, we needed to increase the number of consecutive observations required to find a change globally to avoid spurious change resulting from noise in high density ARD data, but corrections should be attempted on a per-pixel level to account for the spatial heterogeneity of the ARD.

Despite the advantages of increased observation density, combining observations from adjacent WRS-2 path and rows introduced visible noise into our time series, with apparent systematic differences likely due to bidirectional reflectance effects. Future efforts might attempt to characterize this uncertainty within time series models (Holden and Woodcock, 2016) or correct for it prior to analysis (Melaas et al., 2016; Sulla-Menashe et al., 2018). Progress toward a bidirectional reflectance corrected Landsat data (Roy, Zhang, et al., 2016) might allow for future versions of ARD to come corrected for angular effects, and ARD already include the solar and sensor geometry information for each pixel that would enable this research (Dwyer et al., 2018). During exploratory phases of this study, we were impressed with the performance of ARD in creating monthly maximum EVI composites because the in-

creased data densities meant a good value would be found more often. Composites of ARD ought to reduce the noise from directional effects while likely also eliminating the heterogeneity of observation density over space, and composite methods have long been a topic of investigation so some best practices have emerged (e.g., Griffiths et al. (2013) and White et al. (2014)).

4.5.5 Limitations and Future Work

Applicability to the CCDC Algorithm

One of the goals of this effort was to provide information that would be useful to the development of the CCDC algorithm, especially for implementation in the USGS Land Change Monitoring, Analysis, and Projection (LCMAP) project that is using CCDC to map land condition continuously over the last 30 years for the United States. Meanwhile, in order to narrow our experiment to consider the impact of climate data on greenness we altered the typical configuration of the CCDC algorithm to only detect changes using the EVI band. While using EVI alone showed relatively high agreement about the number of breaks estimated for our sample data, the CCDC tends to be much more performant when considering departures from forecasts across multiple spectral or index series, especially if those data use the shortwave infrared.

Our results indicate that using either precipitation data or drought indices like SPEI decrease commission error rates while improving detection rates for herbaceous or shrub cover (Table 4.1). Unfortunately, but perhaps unsurprisingly, inclusion of these moisture related data did not reduce commission errors and slightly hurt change detection accuracy for the forest class, which is primarily comprised of evergreen coniferous species. We also observed that precipitation does have a negative influence on the optical and shortwave infrared bands, especially for oak savanna and grassland cover types, suggesting that using meteorological data may even benefit more typical CCDC analyses that rely on the shortwave infrared band. Future efforts should be

devoted to applying the test framework developed here to the current, official version of CCDC as implemented by the USGS ("pyCCD") in order to better target the question of algorithm improvements.

We also saw that meteorological data improve the performance of change detection when using the Bai-Perron test used by BFAST (Table 4.2), particularly in reducing the false positive or commission error rates for the herbaceous and shrub cover types. We could not isolate the reasons why the commission error decreased substantially more for the Bai-Perron test method than for CCDC, but it is likely primarily influenced by the difference in change detection approaches. The CCDC is an "online" change detection method, which means that it iterates through the time series, typically in a forward direction, looking for anomalous values within a short forward window. While these "online" approaches have memory information about the time series (e.g., RMSE and past residual values), most of the information CCDC has when it decides on a change is limited to some number of consecutive observations ahead of the current iteration. By comparison, the BFAST algorithm is an "offline" algorithm which, similar to LandTrendr (Kennedy et al., 2010), makes decisions on where to place breakpoints based on knowledge of the entire time series. Indeed, the Bai-Perron test checks all possible combinations of breakpoint numbers and locations to determine the optimal location and number of breaks. The Bai-Perron test also differs because it was run on the monthly maximum EVI values instead of all available observations. This was necessary because the Bai-Perron test should be run on regularly spaced data, but the maximum EVI resampling process also removed a significant amount of noise from the time series, including random noise from cloud or cloud shadows or atmospheric interference and systematic noise from directional effects in the Landsat ARD data.

Predictive modeling framework

One of the major limitations to the predictive accuracy of the time series modeling was using the EVI data directly, instead of trying to work with a derived value like a standard anomaly (see Seddon et al. (2016) and Broich et al. (2018)). Conceptually, the EVI data are an integrated time series, such that the value of EVI at time t is dependent on the value at time $t - 1$. In other words, vegetation greenness grows, persists, or deteriorates over time while a quantity like monthly precipitation is an unintegrated random variable, even if it shows its own temporal patterns. If we were to normalize the EVI time series to be unintegrated, either by taking the first difference or including autoregressive terms in our model, or by calculating standardized anomalies, we would be able to relate incoming precipitation to changes in EVI. Preliminary investigation into modeling EVI as standard anomalies show much higher relative performance for precipitation (Equation 4.1) and SPEI (Equation 4.3) based models than the harmonic model (Equation 4.2).

We used the original EVI time series data as CCDC for historic reasons, since this is the predominant way these data are used for time series analysis in the land cover change community and we were already experimenting with the independent variables in the regression equations. Future work in this domain should explore how standard anomalies might be used for change detection instead of the original observations, including how these anomalies may be calculated given the potential existence of land cover change in the historic record.

Ancillary Datasets

There were several issues we noticed when using the NLCD as a source of categorical information used in stratification, visualizations, and groupings. First, there exist several categories that represent the land use than vegetative traits, like the inclusion

of walnut and almond orchards in with herbaceous croplands. We also observed that there was a large amount of variability within what was mapped as herbaceous cover, but would be more useful to this study if mapped as sparse oak savanna. Future investigations might investigate the relationship between meteorological data and greenness using data that better captures the plant functional type, like a continuous fields product, instead of broad land cover label that flatten such distinctions.

4.6   Conclusion

In this study, we investigated the relationship between vegetation greenness and meteorological variables, whether using rainfall information could help predict when vegetation greens up, and whether meteorological data would improve the performance of change detection. Our analysis of time series model coefficients for precipitation (Figures 4·3 and 4·4 for maps and Figures 4·5 and 4·6 for summary boxplots) show that precipitation for the current month is most important in estimating the greenness of herbaceous vegetation, while precipitation over the last 3 months is influential across a broader range of vegetation. These relationships exhibit large spatial variability that make sense in context of ecological and geophysical dynamics. Our analysis of the time series model goodness of fit using BIC (Figures 4·13 and 4·14) show that both precipitation and the SPEI drought indices are better predictors of greenness than the harmonic models alone. Many of the examples from our accuracy assessment show how the addition of meteorological data allow time series models to capture the stochastic variability in greenness associated with variability in climate (e.g., Figure 4·8), which should also help improve "synthetic" image generation from CCDC which has been useful instead of the original Landsat data in other change detection algorithms (Healey et al., 2018). Finally, our analysis of the mapped changes compared to reference interpretations (Tables 4.1 and 4.2) show that meteorological data can

significantly reduce the commission error of the CCDC and BFAST change detection methodologies for grasslands and shrublands.

Advancement in the understanding of how to integrate environmental conditions into time series analysis of spectral observation stands to benefit land cover monitoring projects, including the USGS Land Change Monitoring, Analysis, and Projection (LCMAP) project. While a relationship identified between greenness patterns and environmental drivers is of less importance to this study, such a relationship could be of interest to ecologists. In addition, the model parameters that describe such a relation might be useful for the identification of shrub and grassland species in the same way that parametrization of phenology transition dates is important for the discrimination of forest species communities.

Beyond showing that meteorological data can be useful for land cover monitoring, this study also helps enable further analysis that could further clarify relationships between greenness and climate variability. First, this study produced a substantial (n=540) number of interpretations of forest, shrub, and herbaceous time series over the last 30 years. Because the stratification of the sample that generated these interpretations, there are many examples of complicated change trajectories, including fire, logging, regrowth, succession, natural variability with multi-year to decadal climate variability, and forest mortality in both sparse oak woodlands and conifer forest. Starting without these sorts of data and wanting parsimony for the purposes of change detection, model selection in this study was performed by interpretation of coefficient maps and intuition. The accuracy assessment time series interpretations could be used in future analyses that seek to identify the best set of predictor variables to explain EVI, typically through some sort of cross validation exercise. Second, progress toward separating abrupt disturbances that fundamentally alter the state of the land surface (e.g., fire, logging, mortality) will help future studies separate vegetation re-

sponse from land cover change from more natural variability related to environmental conditions. Doing this work at the Landsat scale is important because of the spatial scale of its observations frequently matches that of change processes, and Landsat analyses may be useful in diagnosing uncertainty or improving estimates from sensors like MODIS that have better temporal resolution but worse spatial resolution.

Constructing a framework to integrate meteorological information into time series analysis of remote sensing data will also improve our understanding of radar time series that are sensitive to moisture. With the launch of both Sentinel-1 satellites, we now have very frequent access to C-band Synthetic Aperture Radar (SAR) measurements which are very sensitive to moisture. Accommodating for the influence of rainfall in these measurements will produce a more coherent understanding of the C-band time series dynamics.

Chapter 5

Conclusion

Global climate change and human modification make earth observations from all variety of sensors essential to understand and adapt to environmental change. Fortunately, earth observation was envisioned as a priority in the 1960s and early 1970s by people wanting to apply new technology to natural resources problems. The Landsat program was launched to meet these goals and has been an invaluable source for understanding the history of the land surface, with consistent observations from the Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors since 1982. While these sources of data have been tremendously valuable, new or upcoming satellites, sensor technologies, and ancillary datasets are promising sources of information to combine with the Landsat record. This dissertation develops and explores methods for enhancing the TM/ETM+ record by fusing other data sources, specifically, Landsat 8 for future continuity, radar data for tropical forest monitoring, and meteorological data for semi-arid vegetation dynamics.

5.1   Key Findings

- Landsat 8 data may be incorporated into existing time series of Landsat 4-7 data for applications like change detection, but vegetation trend analysis requires calibration, especially when using the near-infrared band. The improvements in radiometric quality and cloud masking provided by Landsat 8 data reduce noise compared to previous sensors.

- Tropical forests are notoriously difficult to monitor with Landsat alone because of clouds. This dissertation developed and compared two approaches for fusing Synthetic Aperture Radar (SAR) data from the Advanced Land Observation Satellite (ALOS-1) with Landsat in tropical forests of Peru, and found that radar data increased accuracy of deforestation. Simulations indicate that the benefit of using radar data increased with higher cloud cover.

- Time series analysis of vegetation indices from Landsat in semi-arid environments is complicated by the response of vegetation to high variability in timing and amount of precipitation. We found that quantifying dynamics in precipitation and drought index data improved land cover change detection performance compared to more traditional harmonic modeling for grasslands and shrublands in California. Specifically, models that incorporate moisture dynamics have significantly lower commission error without significantly affecting the omission error.

This dissertation enhances the value of Landsat data by combining it with other data sources, including other optical sensors, SAR data, and meteorological data. The methods developed here show the potential for data fusion and are especially important in light of recent and upcoming missions, like Sentinel-1, Sentinel-2, and NASA-ISRO Synthetic Aperture Radar (NISAR). Trends in both computing and community algorithm development are continuing to enable sophisticated analysis of increasingly large datasets, making fusion of multiple large time series datasets increasingly possible.

5.2 Recommendations for future work

5.2.1 Continuity among Landsat or Landsat-like sensors

While the investigations into using radar and meteorological data treated data fusion as an opportunity to add complimentary information, Chapter 2 of this dissertation attempted to determine if Landsat 8 provided the same or comparable information to the Enhanced Thematic Mapper Plus (ETM+) despite advances to technology and differences in design decisions. Creating long records of Landsat data that are comparable across time is incredibly important for efforts that characterize the history of the land surface, and recent efforts to extend this analysis as far back as the Landsat Multispectral Scanner (MSS) have enabled studies that span 40 years (Vogeler et al., 2018). However, the creation of a unified record will increasingly neglect possibly useful information as sensors grow more sophisticated, like the higher resolution visible and near-infrared bands on Sentinel-2 or the coastal blue band on Landsat 8. It is also much easier to degrade Landsat 8 data to look like Landsat Thematic Mapper (TM) and ETM+ data than it is for MSS data to be comparable to TM/ETM+ data. Even the harmonization of newer sensors is a difficult task. Efforts to create a Harmonized Landsat 8 Sentinel-2 (HLS) data product have needed to unify atmospheric correction routines, perform image registration, correct for BRDF effects, and apply bandpass adjustments (Claverie et al., 2017; Claverie et al., 2018). As the sensors used in medium to high resolution passive optical remote sensing become more sophisticated and grow in number, it may be beneficial to adopt the perspective taken when fusing radar or meteorological data sources by treating the data from each sensor more independently. There still may be benefits to using a unified product, but especially as fusion methods for combining other data sources advance it may be worthwhile to consider the Landsat archive as discontinuous. A historical analysis beginning with Landsat 1 and ending with data from the upcoming Landsat 9 will have different

uncertainties over time, and thinking of the addition of Landsat 8 and Landsat 9 as "fusion" rather than extending the time series might reduce overall uncertainties by taking better advantage of the features of these new sensors.

5.2.2 Optical and radar data time series fusion for land cover monitoring

Just as during the opening of the Landsat archive enabled researchers to go from using one or several Landsat images to eventually thousands (Wulder et al., 2012), changes to data policy for many sensors might transform the practice of using one satellite data source to many. The open data policies and vision for collaboration with other satellite products from the Sentinel program (Malenovskỳ et al., 2012; Torres et al., 2012), for example, help make the cost benefit of data fusion much more favorable. Community standards for data promoted by data distributors in the Landsat community, like the Climate Data Record (U.S. Geological Survey, 2015) or Analysis Ready Data (Dwyer et al., 2018), and from the European Space Agency's Sentinel missions (Torres et al., 2012), have made research more efficient and reproducible, and these efforts are a model for future data product releases. Software and tools to make geospatial analysis by making data available as collections to work with, such as the Open Geo Data Cube (Lewis et al., 2017) or the Google Earth Engine (Gorelick et al., 2017) are extremely valuable to the adoption of fusion algorithms because they eliminate specific data format, preprocessing, or file location details, making simply accessing the data much more streamlined.

The historical archive of Landsat data is far from globally consistent (Wulder et al., 2016) is far from consistent, with many places lacking data in the 1980s and some with few usable observations in the 1990s. Data from sensors like the Japanese Earth Resources Satellite (JERS-1) or the Advanced Land Observation Satellite (ALOS-1) could help fill in these gaps in the historic record to produce a more accurate accounting of the late 20th and early 21st centuries. Studies that follow the example of

(Reiche et al., 2018) and others by incorporating more than two time series (Landsat NDVI, Sentinel-1 C-band, and ALOS-2 L-band) should be encouraged, as the challenge of fusing three data sources should help create robust solutions that can scale to arbitrary number of input data sources.

Methods for fusing across data sources could also be helpful for making better used of radar data acquired from different incident angles, ascending or descending orbits, polarizations, and perhaps scanning modes. Using the example of fusion by class probability (i.e., Reiche et al. (2018) and Chapter 3), one might collect regions of interest to parameterize classification models for forest cover for all data sources, but train and apply classification models by polarizations or orbit characteristic. The investigation into probability fusion in Chapter 3 used data from the dual poliarmetric mode, and trained classification models on L-HH, L-HV, and HH/HV ratio features, but the L-HH and ratio features were sometimes a source of confusion when estimating forest probability. It might be more advantageous to use the L-HH time series separately from the L-HV data in future work because it could strengthen the L-HV results while adding additional observations in time when only single polarization is available. Most of the change detection algorithms employ some conception of noise in the time series (Zhu and Woodcock, 2014; Brooks et al., 2014; Verbesselt et al., 2012) that would show differences between L-HH and L-HV, and it might also be useful to weight observations from different polarizations in calculation of change detection statistics. For example, an unusual observation from L-HV might count for two consecutive observations of change in CCDC while an observation from L-HH would be treated with the same suspicion as Landsat observations. Currently, practitioners using Sentinel-1 time series separate the data based on ascending or descending orbit because the data are fundamentally different. Fusion algorithms could resolve this difference and further increase the density of observations from Sentinel-1.

Fusing across the spatial resolution using MODIS and Landsat data have been successful in increasing change detection performance (Tang, 2018; Xin et al., 2013), and algorithms that make fusing data easy, like the STARFM by Gao et al. (2006), have been used in a variety of applications. These approaches have required somewhat complicated modeling of the spatial overlap of the data based on the swath footprints (Xin et al., 2013; Tang, 2018) or by using weighting functions on gridded data (Gao et al., 2006). These approaches might be useful when applied to high and lower resolution radar scanning modes, like the Fine Beam and ScanSAR modes on the ALOS-1 PALSAR, to increase the temporal density of radar observations while retaining high spatial resolution.

One promising approach to data fusion is the monitoring of class probabilities, as used in Chapter 3 and in (Solberg et al., 2008; Salberg and Trier, 2011; Reiche, de Bruin, et al., 2015; Reiche et al., 2018). By first using classification procedures to convert disparate time series into class membership estimates, this approach was able to fuse information from multiple sensors and improve our understanding of the land surface. Most of the examples of this approach have focused on likelihood of a single class, or between two classes, but this approach may be able to generate robust and descriptive land cover class transitions that are essential to understanding what drives land cover conversion. Using spectral unmixing to transform a time series of Landsat data into a time series of fractional cover, Bullock et al. (2018) was able to capture landscape conversion process dynamics on a continuous scale among axes of spectral endmembers. It might be possible to incorporate data from other sources, including radar, into the endmember analysis, although the selection of endmembers would have to be tailored to the data source. For example, optical and radar data sources might share dense tree cover as an endmember, but it's unclear how a shadow endmember would apply to radar or how endmembers driven by moisture content

would align across the data types.

5.2.3 Landsat Time Series Meteorological Data Fusion In Semi-Arid Ecosystems

This study is among the first to incorporate environmental drivers into analysis of vegetation greenness dynamics and abrupt change at the Landsat scale. While many of the time series change detection methods using Landsat data developed since the opening of the Landsat archive attempt to work around or ignore the influence of climatic drivers on time series signals, this work found that change detection is more accurate when contextualizing observations within moisture dynamics. While change detection performance decreased by about 4% for forests when using a drought index in CCDC, the integration of moisture dynamics in time series forecast models greatly decreased the commission error of CCDC and the BFAST algorithms for grasslands, forests, and shrublands.

Advancement in the understanding of how to incorporate environmental conditions into time series analysis of spectral observation stands to benefit land cover monitoring projects, including the USGS Land Change Monitoring, Analysis, and Projection (LCMAP) project. While a relationship identified between greenness patterns and environmental drivers is of less importance to this study, such a relationship could be of interest to ecologists. Maps of estimated meteorological data coefficients also reveal spatial patterns that appear linked to surface hydrological processes and water availability for semi-arid, seasonal grasses that could help monitor water resources. In addition, the model parameters that describe these relations might be useful for the identification of shrub and grassland species in the same way that parametrization of phenology transition dates is important for the discrimination of forest species communities (Pasquarella, Holden, and Woodcock, 2018).

Most studies, including the work presented here, has tried to use vegetation in-

dices to quantify vegetation response to environmental conditions, but recent work (Yang et al., 2017; Yang et al., 2018) have used thermal remote sensing and a land-atmosphere exchange model to analyze time series of evapotranspiration (ET). ET changes in reaction to changes in environmental conditions more quickly than changes to photochemical properties. By comparing estimated ET against a reference ET, they were able to see downregulation of ET during drought and returns to higher rates of ET after rainfall. Yang et al. (2018) formalized this comparison against the expected ET of a well watered plant into an Evaporative Stress Index. Estimation of carbon exchange through solar-induced fluorescence (Meroni et al., 2009; Frankenberg et al., 2014) has many advantages to previous methods that use vegetation indices as proxies for productivity, as the measurement of fluorescence is much more directly connected to short term physiological processes than reflectance due to photochemicals. Utilizing approaches that relate more closely to the physical processes of vegetation is more difficult and requires modeling or ancillary data, but these data might prove more effective for assessing vegetation response to climate variability than using a broad greenness index.

Appendix A

Chapter 2: An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations

A.1 Underflight results

This section contains supplementary tables and figures for the "underflight" dataset spectral comparison.

Table A.1.1: Underflight comparison regression results for all available WRS-2 path and rows. The 95% confidence intervals around slope and intercept estimates are estimated using a percentile bootstrap.

WRS-2	Band	Slope (95% CI)	Intercept (95% CI)	R^2
P013R010	Blue	0.9811 (0.979 - 0.9832)	-0.0143 (-0.0159 - -0.0126)	0.9879
P013R010	Green	0.8303 (0.8285 - 0.8322)	-0.0203 (-0.0217 - -0.0188)	0.9875
P013R010	Red	0.8847 (0.883 - 0.8866)	-0.0038 (-0.0053 - -0.0025)	0.9877
P013R010	NIR	1.0038 (1.0023 - 1.0054)	-0.0147 (-0.0158 - -0.0136)	0.9919
P013R010	SWIR1	0.757 (0.7501 - 0.7634)	0.0012 (0.0008 - 0.0015)	0.9006
P013R010	SWIR2	0.8419 (0.8372 - 0.8465)	0.0125 (0.0124 - 0.0127)	0.8548
P013R010	NDVI	1.0166 (1.012 - 1.0215)	0.056 (0.0559 - 0.0562)	0.8876
P013R010	EVI	0.6533 (0.5828 - 0.7345)	0.2758 (0.2694 - 0.2833)	0.15
P013R010	NBR	0.8996 (0.8899 - 0.9088)	0.0724 (0.0641 - 0.081)	0.8881
P013R010	NDMI	0.9714 (0.9633 - 0.98)	0.0308 (0.0233 - 0.038)	0.9407
P013R010	Greenness	0.876 (0.8715 - 0.8803)	0.0594 (0.0589 - 0.0598)	0.83
P013R010	Wetness	0.913 (0.9104 - 0.9152)	-0.0149 (-0.016 - -0.0136)	0.9842
P013R029	Blue	1.1519 (1.1312 - 1.1736)	-0.0149 (-0.0159 - -0.014)	0.9286
P013R029	Green	1.0851 (1.0723 - 1.0987)	-0.0096 (-0.0104 - -0.0089)	0.9681
P013R029	Red	1.0574 (1.0469 - 1.069)	-0.0073 (-0.0081 - -0.0067)	0.9771
P013R029	NIR	1.096 (1.0788 - 1.1132)	-0.0118 (-0.0151 - -0.0085)	0.9517
P013R029	SWIR1	1.0201 (1.0109 - 1.0297)	0.0019 (0.0001 - 0.0036)	0.9814
P013R029	SWIR2	1.0241 (1.015 - 1.0334)	0.0045 (0.0034 - 0.0056)	0.981
P013R029	NDVI	1.0877 (1.076 - 1.0988)	-0.0075 (-0.0135 - -0.0015)	0.9686
P013R029	EVI	0.9677 (0.9437 - 0.9892)	0.007 (0.005 - 0.0094)	0.92
P013R029	NBR	0.9442 (0.9317 - 0.9564)	0.0016 (-0.002 - 0.0053)	0.9631
			Continued on next page	

Table A.1.1 – continued from previous page

WRS-2	Band	Slope (95% CI)	Intercept (95% CI)	R^2
P013R029	NDMI	0.9492 (0.9359 - 0.9611)	0.0037 (0.0018 - 0.0056)	0.9608
P013R029	Greenness	1.0832 (1.0594 - 1.1065)	0.0014 (-0.0005 - 0.0034)	0.9313
P013R029	Wetness	0.9934 (0.983 - 1.0034)	-0.01 (-0.0113 - -0.0088)	0.9782
P029R031	Blue	0.9654 (0.9622 - 0.9687)	-0.0171 (-0.0174 - -0.0168)	0.8956
P029R031	Green	1.0397 (1.038 - 1.0414)	-0.0173 (-0.0175 - -0.0171)	0.959
P029R031	Red	1.0313 (1.03 - 1.0326)	-0.0125 (-0.0127 - -0.0123)	0.9731
P029R031	NIR	1.0472 (1.0456 - 1.0487)	-0.0048 (-0.0051 - -0.0044)	0.97
P029R031	SWIR1	0.9715 (0.9693 - 0.9735)	0.0107 (0.0101 - 0.0114)	0.9642
P029R031	SWIR2	0.9758 (0.9737 - 0.9779)	0.0135 (0.0131 - 0.0141)	0.963
P029R031	NDVI	1.1236 (1.1192 - 1.1278)	0.016 (0.0152 - 0.0169)	0.9472
P029R031	EVI	0.9675 (0.9651 - 0.9699)	0.0068 (0.0067 - 0.0069)	0.9445
P029R031	NBR	0.9563 (0.9535 - 0.959)	-0.0076 (-0.0077 - -0.0074)	0.9627
P029R031	NDMI	0.9465 (0.9429 - 0.9499)	0.0013 (0.0006 - 0.0021)	0.9502
P029R031	Greenness	1.032 (1.03 - 1.034)	0.0122 (0.0122 - 0.0123)	0.9545
P029R031	Wetness	0.9639 (0.9616 - 0.966)	-0.0194 (-0.0199 - -0.0189)	0.9621
P102R076	Blue	0.7682 (0.7646 - 0.7723)	0.0045 (0.0043 - 0.0048)	0.7451
P102R076	Green	0.9195 (0.9169 - 0.9218)	-0.0007 (-0.001 - -0.0005)	0.8768
P102R076	Red	0.9845 (0.9829 - 0.9861)	-0.0029 (-0.0031 - -0.0027)	0.9344
P102R076	NIR	0.9943 (0.9927 - 0.9961)	0.0045 (0.0042 - 0.0048)	0.9233
P102R076	SWIR1	0.9578 (0.9561 - 0.9595)	0.0155 (0.015 - 0.0159)	0.9269
P102R076	SWIR2	0.9434 (0.9419 - 0.9449)	0.015 (0.0147 - 0.0154)	0.9355
P102R076	NDVI	1.0565 (1.0541 - 1.0589)	0.0181 (0.0177 - 0.0185)	0.8996
P102R076	EVI	0.9009 (0.8972 - 0.9048)	0.0078 (0.0077 - 0.008)	0.8424
P102R076	NBR	0.893 (0.891 - 0.8952)	-0.0058 (-0.006 - -0.0056)	0.9159
P102R076	NDMI	0.8753 (0.8727 - 0.878)	-0.025 (-0.0256 - -0.0245)	0.8674
P102R076	Greenness	0.9831 (0.9804 - 0.9857)	0.0088 (0.0087 - 0.0088)	0.9005
P102R076	Wetness	0.9347 (0.9332 - 0.9363)	-0.0217 (-0.022 - -0.0213)	0.9277
P134R042	Blue	0.912 (0.9065 - 0.9175)	-0.0071 (-0.0072 - -0.0069)	0.8094
P134R042	Green	0.986 (0.9837 - 0.9882)	-0.0037 (-0.0038 - -0.0036)	0.9363
P134R042	Red	0.9973 (0.9948 - 0.9996)	-0.0023 (-0.0023 - -0.0022)	0.9648
P134R042	NIR	1.0641 (1.0628 - 1.0655)	-0.0131 (-0.0135 - -0.0128)	0.9625
P134R042	SWIR1	1.0033 (1.0021 - 1.0044)	0.0011 (0.001 - 0.0013)	0.9791
P134R042	SWIR2	1.0032 (1.0015 - 1.0048)	0.0025 (0.0024 - 0.0026)	0.9819
P134R042	NDVI	1.0332 (1.0304 - 1.0359)	-0.0088 (-0.0109 - -0.0067)	0.9656
P134R042	EVI	0.9758 (0.9744 - 0.9772)	-0.001 (-0.0012 - -0.0007)	0.96
P134R042	NBR	0.9633 (0.9619 - 0.9647)	0.0113 (0.0104 - 0.0122)	0.973
P134R042	NDMI	0.9609 (0.9596 - 0.9621)	0.0111 (0.0106 - 0.0116)	0.9695
P134R042	Greenness	1.0561 (1.0548 - 1.0574)	-0.0034 (-0.0036 - -0.0032)	0.9645
P134R042	Wetness	0.9885 (0.9871 - 0.9899)	-0.0048 (-0.0049 - -0.0047)	0.9784
P134R052	Blue	0.1868 (0.1762 - 0.1972)	0.0467 (0.0462 - 0.0472)	0.1965
P134R052	Green	0.5533 (0.544 - 0.562)	0.0223 (0.022 - 0.0226)	0.6462
P134R052	Red	0.3972 (0.3906 - 0.4043)	0.0233 (0.0232 - 0.0235)	0.4346
P134R052	NIR	0.4272 (0.3851 - 0.466)	0.0168 (0.0159 - 0.0177)	0.4697
P134R052	SWIR1	0.2963 (0.2854 - 0.3079)	0.0121 (0.0119 - 0.0123)	0.3408
P134R052	SWIR2	0.2385 (0.234 - 0.2428)	0.0092 (0.0091 - 0.0092)	0.3262
P134R052	NDVI	0.1088 (0.1019 - 0.1166)	-0.1084 (-0.1087 - -0.108)	0.2385
P134R052	EVI	0.33 (0.282 - 0.3775)	-0.0074 (-0.0076 - -0.0073)	0.5135
P134R052	NBR	0.0865 (0.0848 - 0.0883)	0.3497 (0.3491 - 0.3502)	0.1939
P134R052	NDMI	0.1262 (0.1241 - 0.1281)	0.2353 (0.2348 - 0.2358)	0.2498
P134R052	Greenness	0.5405 (0.4789 - 0.6087)	-0.0088 (-0.0095 - -0.008)	0.652
P134R052	Wetness	0.3054 (0.2981 - 0.3131)	0.006 (0.0059 - 0.006)	0.3842

Table A.1.1 – continued from previous page

WRS-2	Band	Slope (95% CI)	Intercept (95% CI)	R^2
P150R042	Blue	0.9378 (0.9353 - 0.9401)	-0.0022 (-0.0025 - -0.0018)	0.9093
P150R042	Green	0.9915 (0.9891 - 0.9937)	-0.0029 (-0.0033 - -0.0024)	0.9545
P150R042	Red	0.991 (0.9889 - 0.9929)	0.0017 (0.0012 - 0.0022)	0.9636
P150R042	NIR	0.9385 (0.9354 - 0.9415)	0.023 (0.022 - 0.024)	0.951
P150R042	SWIR1	0.9237 (0.9201 - 0.9269)	0.0379 (0.0366 - 0.0392)	0.9497
P150R042	SWIR2	0.9045 (0.902 - 0.9072)	0.0402 (0.0392 - 0.041)	0.9519
P150R042	NDVI	1.0961 (1.0935 - 1.0988)	-0.0052 (-0.0055 - -0.0048)	0.9312
P150R042	EVI	1.0169 (1.0144 - 1.0197)	-0.0001 (-0.0002 - -0.0)	0.9159
P150R042	NBR	0.9291 (0.9265 - 0.9315)	-0.0082 (-0.0083 - -0.0081)	0.9327
P150R042	NDMI	0.9137 (0.9097 - 0.9174)	-0.0146 (-0.015 - -0.0143)	0.9134
P150R042	Greenness	1.0628 (1.0606 - 1.0648)	0.0034 (0.0034 - 0.0035)	0.9368
P150R042	Wetness	0.8743 (0.8712 - 0.8774)	-0.0469 (-0.0478 - -0.046)	0.9336
P198R047	Blue	0.9141 (0.9116 - 0.9163)	0.0137 (0.0133 - 0.0142)	0.8474
P198R047	Green	0.7662 (0.7631 - 0.7693)	0.0795 (0.0784 - 0.0806)	0.7214
P198R047	Red	0.6346 (0.6307 - 0.6382)	0.1854 (0.1835 - 0.1874)	0.5888
P198R047	NIR	0.6452 (0.6413 - 0.6493)	0.2078 (0.2053 - 0.2102)	0.5804
P198R047	SWIR1	0.6305 (0.6267 - 0.6346)	0.2736 (0.2708 - 0.2763)	0.603
P198R047	SWIR2	0.6337 (0.63 - 0.6373)	0.241 (0.2387 - 0.2435)	0.6485
P198R047	NDVI	0.2607 (0.2566 - 0.2647)	0.0578 (0.0575 - 0.0581)	0.3937
P198R047	EVI	0.3398 (0.3356 - 0.3438)	0.0181 (0.018 - 0.0183)	0.4588
P198R047	NBR	0.7336 (0.7306 - 0.7364)	-0.0174 (-0.0175 - -0.0173)	0.7296
P198R047	NDMI	0.8243 (0.8217 - 0.827)	-0.0305 (-0.0307 - -0.0303)	0.7921
P198R047	Greenness	0.4013 (0.3977 - 0.4049)	0.0093 (0.0092 - 0.0093)	0.4799
P198R047	Wetness	0.7395 (0.7365 - 0.7421)	-0.149 (-0.1506 - -0.1476)	0.7632
P230R084	Blue	0.755 (0.7468 - 0.761)	0.0037 (0.0033 - 0.0041)	0.7877
P230R084	Green	0.9574 (0.9526 - 0.961)	-0.0027 (-0.003 - -0.0023)	0.9169
P230R084	Red	0.998 (0.9936 - 1.0011)	-0.0032 (-0.0035 - -0.0028)	0.9627
P230R084	NIR	1.0259 (1.024 - 1.0274)	0.003 (0.0027 - 0.0034)	0.979
P230R084	SWIR1	0.9789 (0.9779 - 0.98)	0.0107 (0.0104 - 0.0109)	0.9736
P230R084	SWIR2	0.9706 (0.9696 - 0.9715)	0.0109 (0.0108 - 0.0111)	0.9774
P230R084	NDVI	0.9583 (0.957 - 0.9595)	0.048 (0.0476 - 0.0484)	0.9699
P230R084	EVI	0.874 (0.7474 - 0.9556)	0.0135 (0.0077 - 0.0225)	0.9336
P230R084	NBR	0.9647 (0.964 - 0.9655)	0.0034 (0.0033 - 0.0035)	0.9858
P230R084	NDMI	0.9596 (0.9586 - 0.9606)	0.0051 (0.005 - 0.0052)	0.9849
P230R084	Greenness	0.9831 (0.9814 - 0.9845)	0.011 (0.011 - 0.0111)	0.9831
P230R084	Wetness	0.9687 (0.9678 - 0.9696)	-0.0141 (-0.0143 - -0.014)	0.9773
P230R094	Blue	0.8924 (0.8906 - 0.8944)	-0.0013 (-0.0014 - -0.0011)	0.8802
P230R094	Green	0.9068 (0.9049 - 0.9087)	0.0026 (0.0024 - 0.0028)	0.9084
P230R094	Red	0.887 (0.8849 - 0.8891)	0.009 (0.0087 - 0.0093)	0.9053
P230R094	NIR	0.8614 (0.8591 - 0.8637)	0.0202 (0.0198 - 0.0206)	0.8773
P230R094	SWIR1	0.8032 (0.8008 - 0.8057)	0.0379 (0.0374 - 0.0384)	0.8577
P230R094	SWIR2	0.7746 (0.7721 - 0.7773)	0.0352 (0.0347 - 0.0356)	0.847
P230R094	NDVI	1.0569 (1.0508 - 1.0626)	-0.0002 (-0.0012 - 0.0009)	0.8747
P230R094	EVI	0.8522 (0.8483 - 0.8558)	0.0053 (0.0052 - 0.0054)	0.8639
P230R094	NBR	0.6417 (0.6357 - 0.6476)	-0.0006 (-0.0007 - -0.0006)	0.7563
P230R094	NDMI	0.6824 (0.6777 - 0.6871)	-0.0388 (-0.0393 - -0.0383)	0.7888
P230R094	Greenness	0.9408 (0.9372 - 0.9448)	0.0041 (0.004 - 0.0041)	0.8982
P230R094	Wetness	0.7392 (0.7364 - 0.7418)	-0.039 (-0.0394 - -0.0386)	0.821
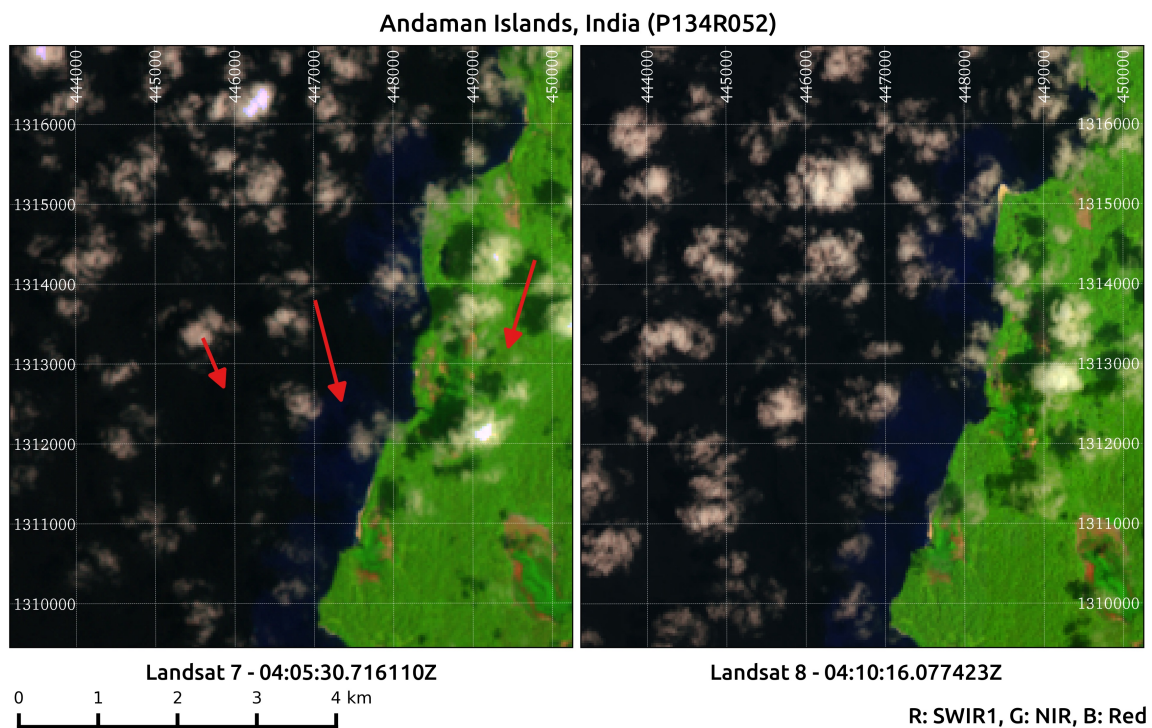
Figure A.1.1: Movement of clouds in underflight data over the Andaman and Nicobar Islands, India (P134R052) in the roughly 5 minute interval between acquisitions.

A.2   Underflight cirrus comparison

This section contains a figure from the "underflight" cirrus cloud characterization.
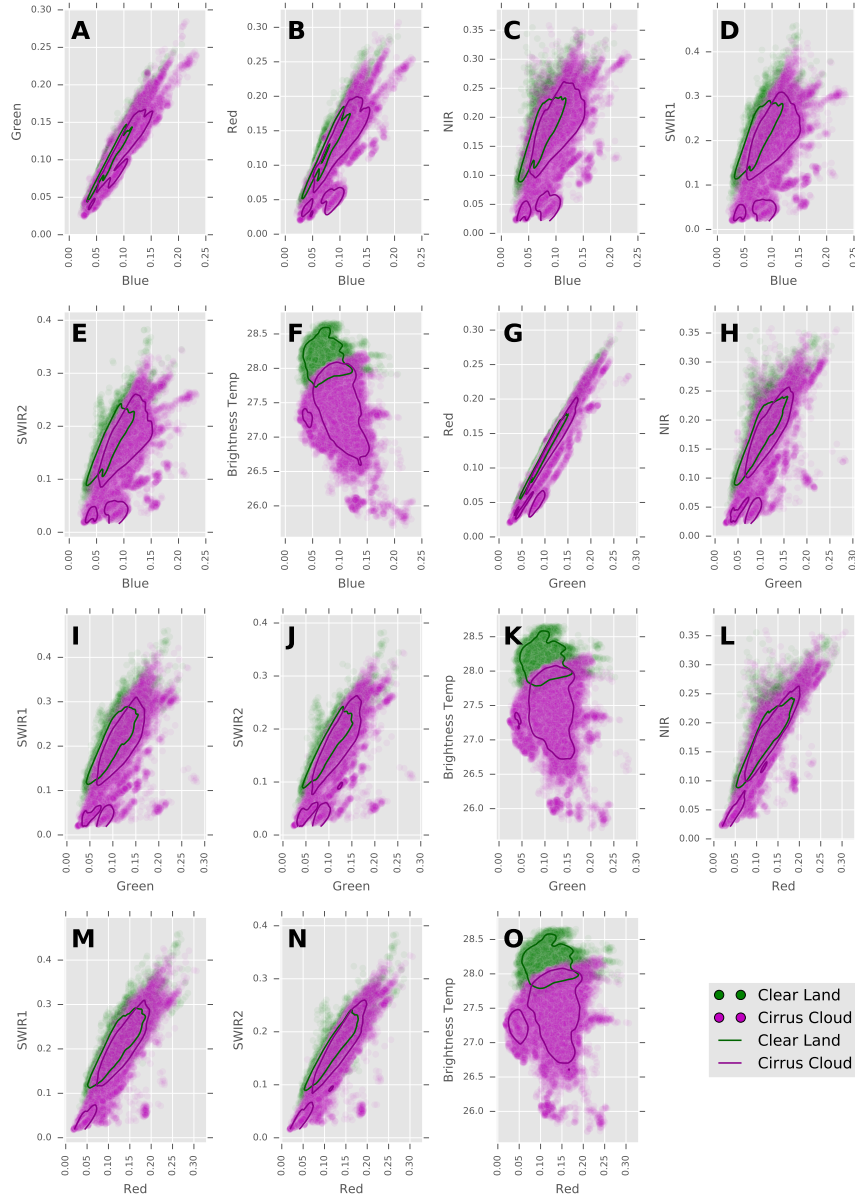
Figure A.2.1: Clear land and cloud cirrus spectral properties for Santa Cruz Province, Argentina (P230R094). Contour lines contain 95% of the data in the bivariate distribution space for data labeled as clear land (green) and cirrus cloud (magenta) in Landsat 8 CFmask images.

A.3   Time Series spectral results

This section contains supplementary figures for the Landsat time series spectral comparison.

Figure A.3.1: Normalized intercept estimates for time series models in Florida, United States (P016R041) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 intercept estimates were normalized by dividing by the intercepts of time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

Figure A.3.2: Normalized slope estimates for time series models in Florida, United States (P016R041) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 slope estimates were normalized by subtracting the slopes from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
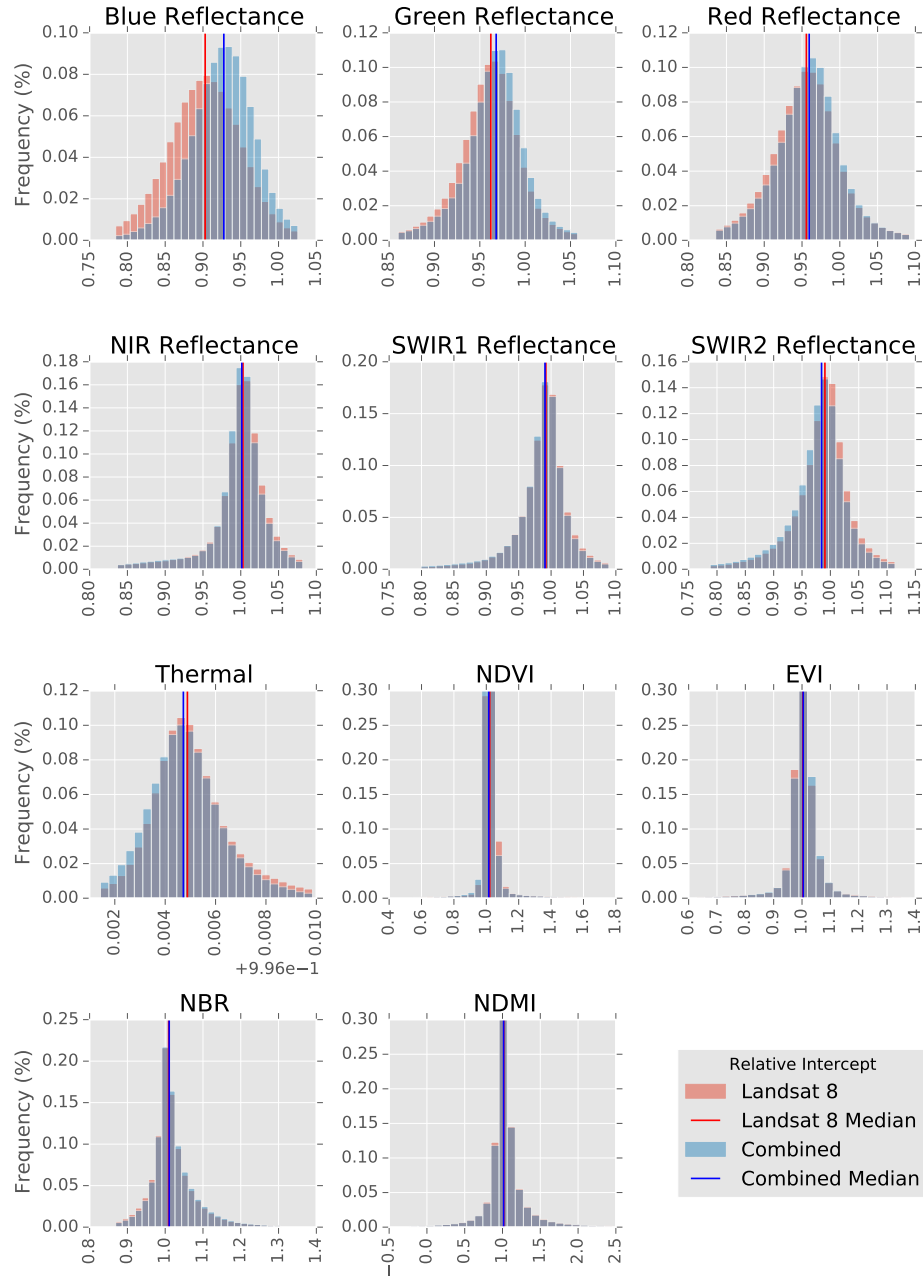
Figure A.3.3: Normalized RMSE estimates for time series models in Florida, United States (P016R041) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 RMSE estimates were normalized by dividing by the RMSE from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

Figure A.3.4: Normalized intercept estimates for time series models in Mississippi, United States (P023R037) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 intercept estimates were normalized by dividing by the intercepts of time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
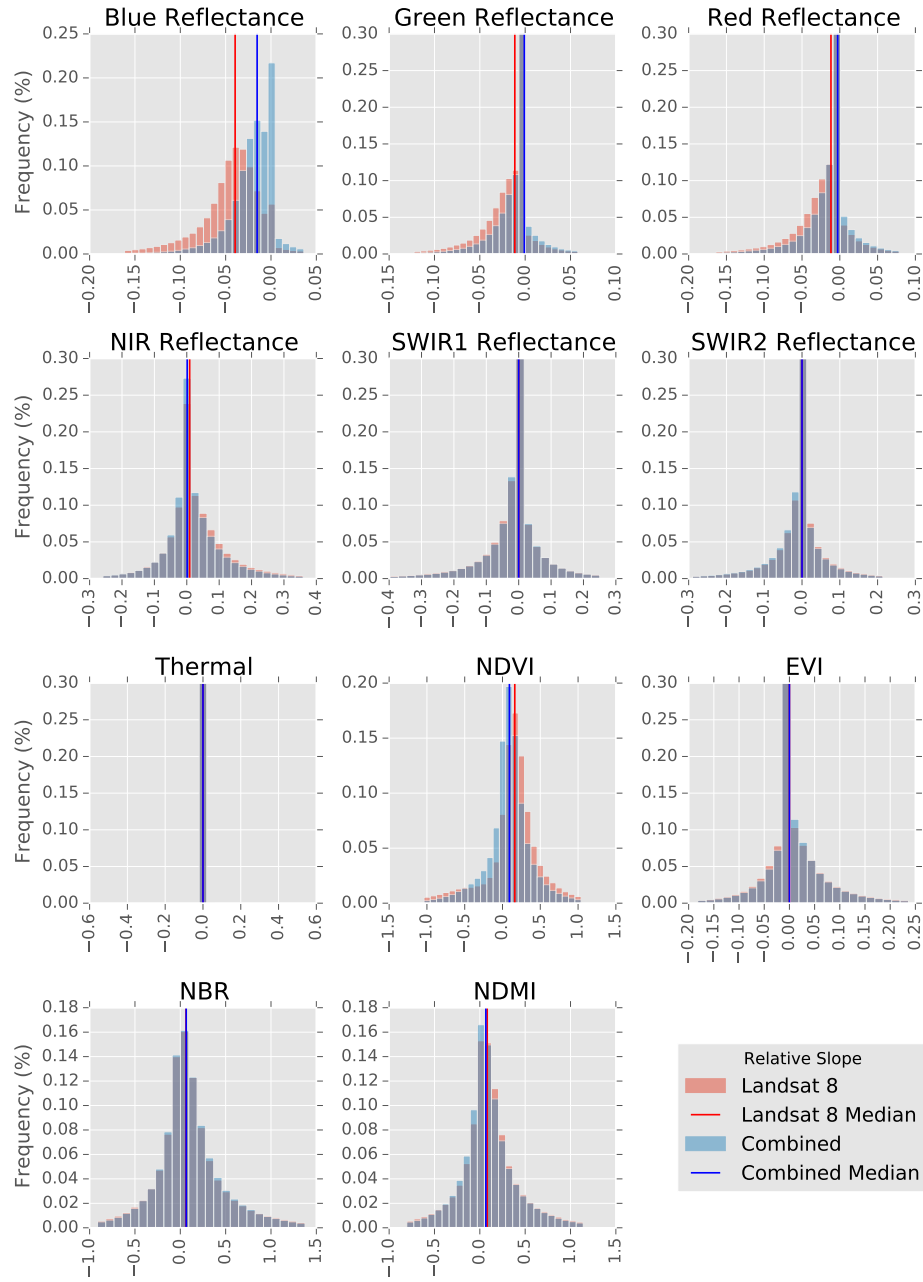
Figure A.3.5: Normalized slope estimates for time series models in Mississippi, United States (P023R037) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 slope estimates were normalized by subtracting the slopes from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
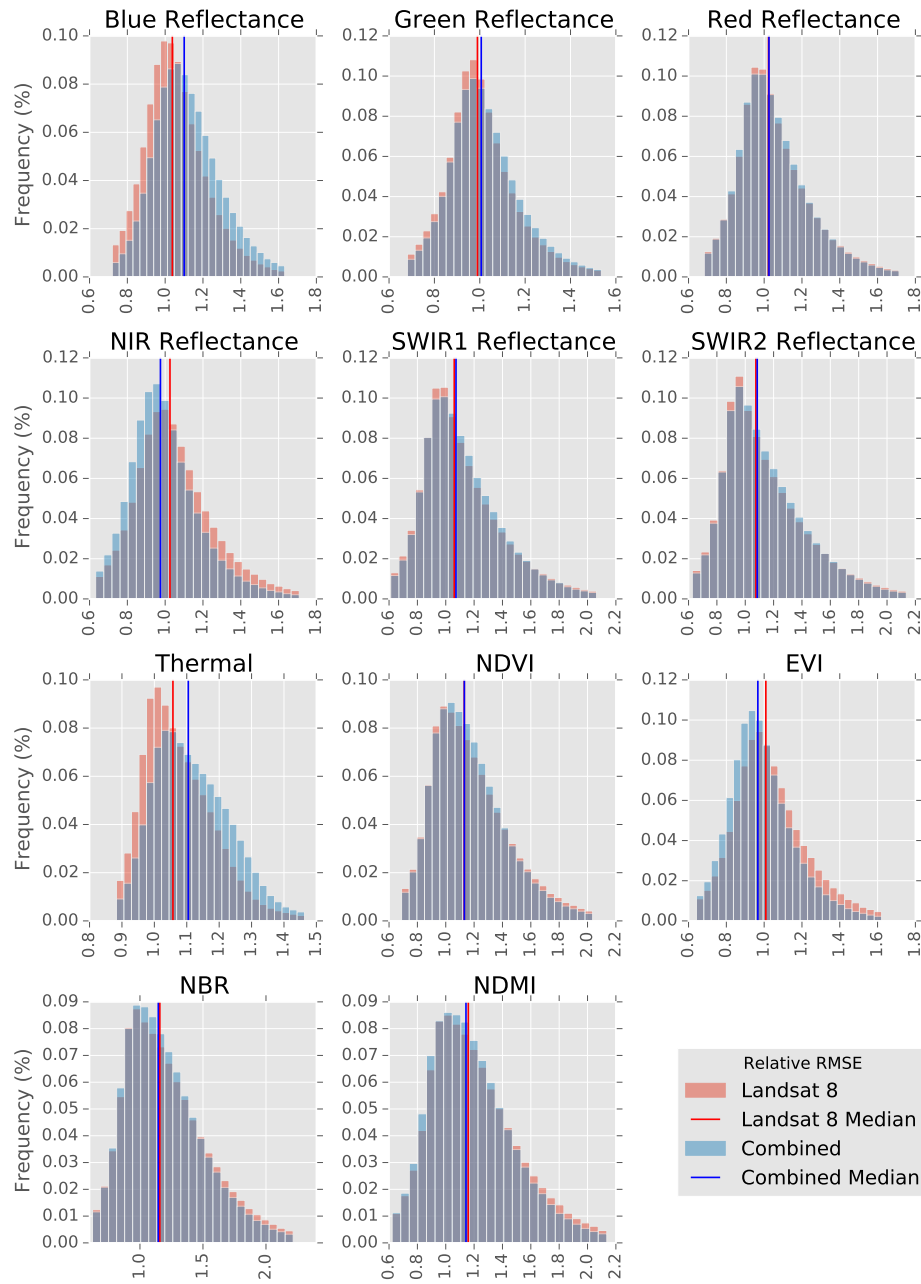
Figure A.3.6: Normalized RMSE estimates for time series models in Mississippi, United States (P023R037) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 RMSE estimates were normalized by dividing by the RMSE from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
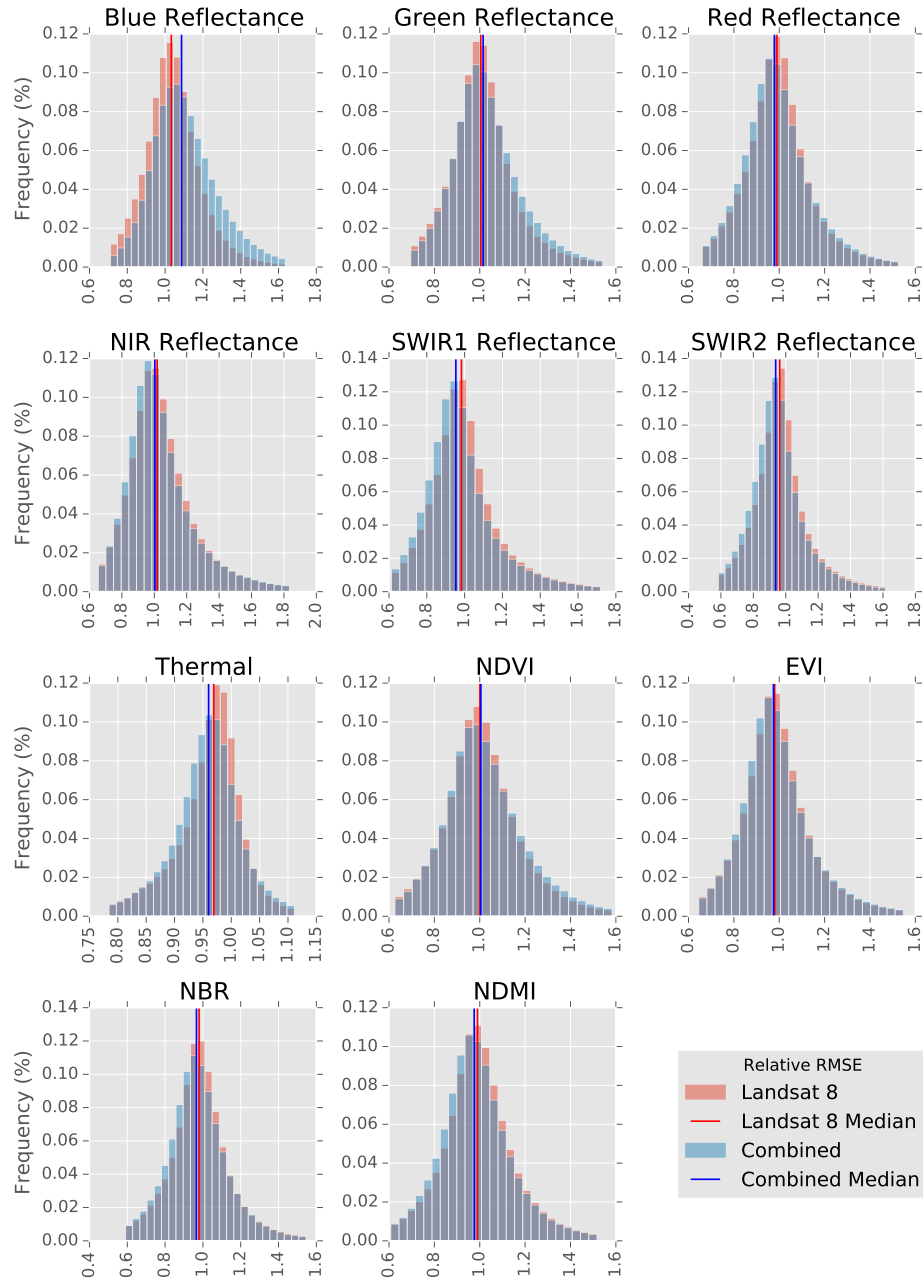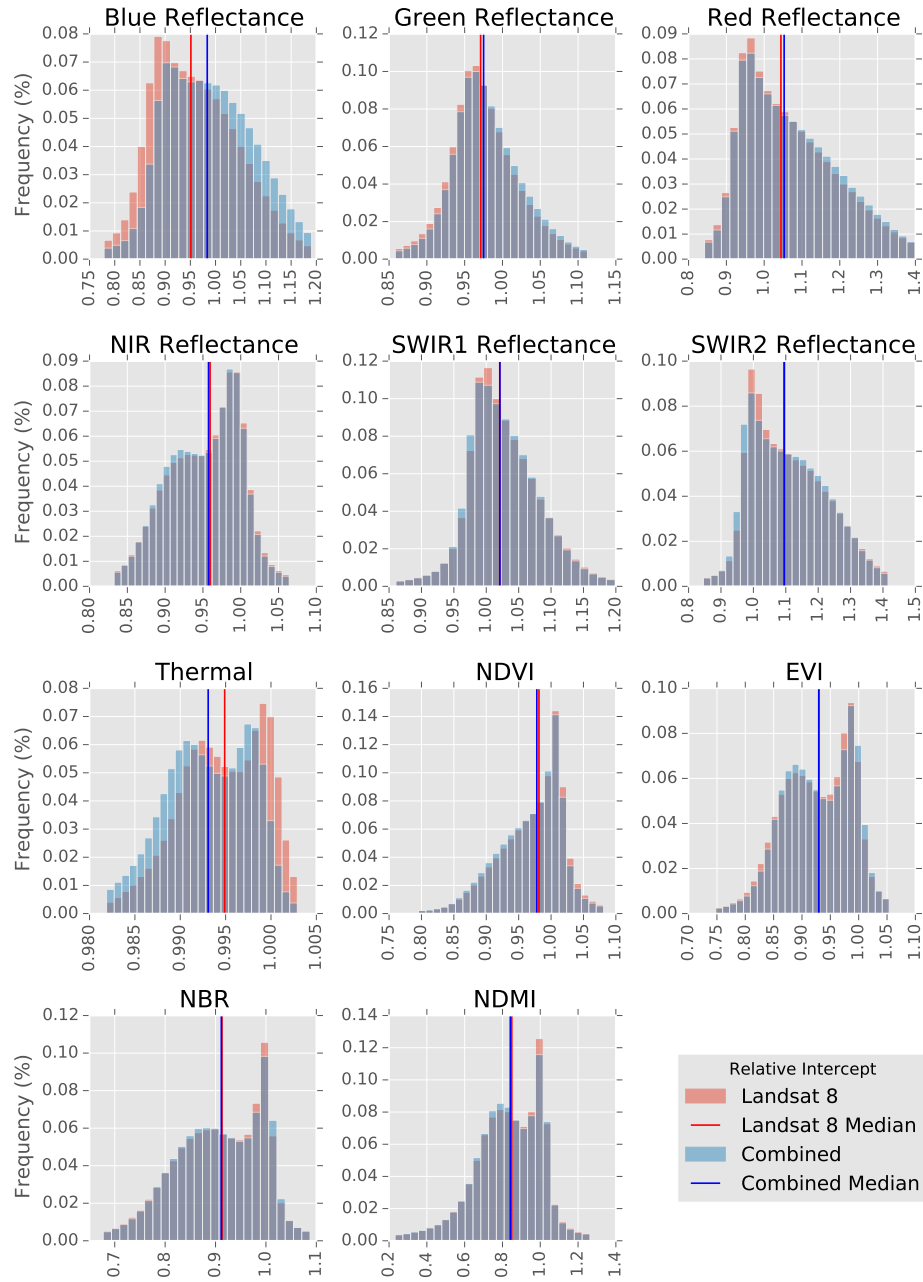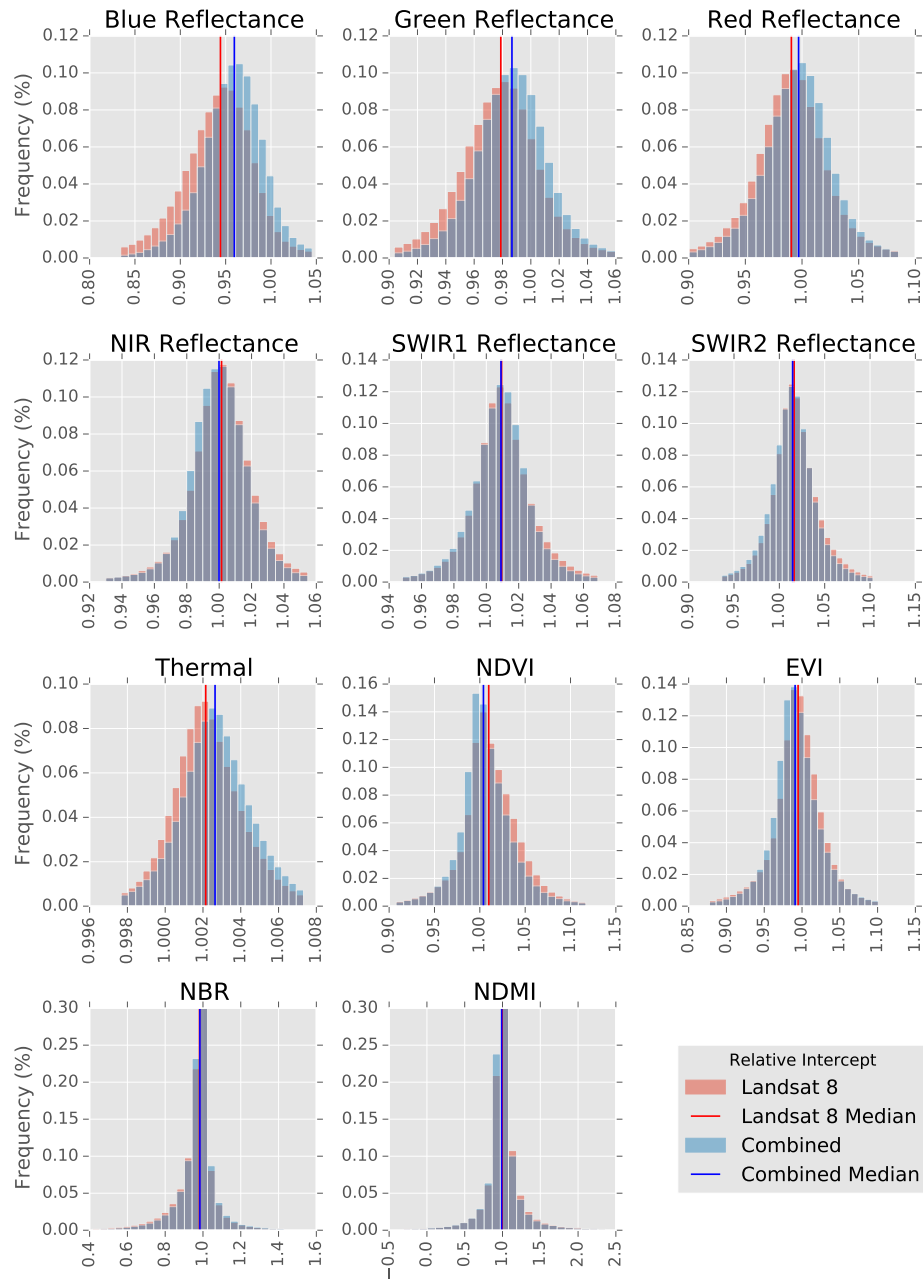
Figure A.3.7: Normalized intercept estimates for time series models in California, United States (P043R034) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 intercept estimates were normalized by dividing by the intercepts of time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
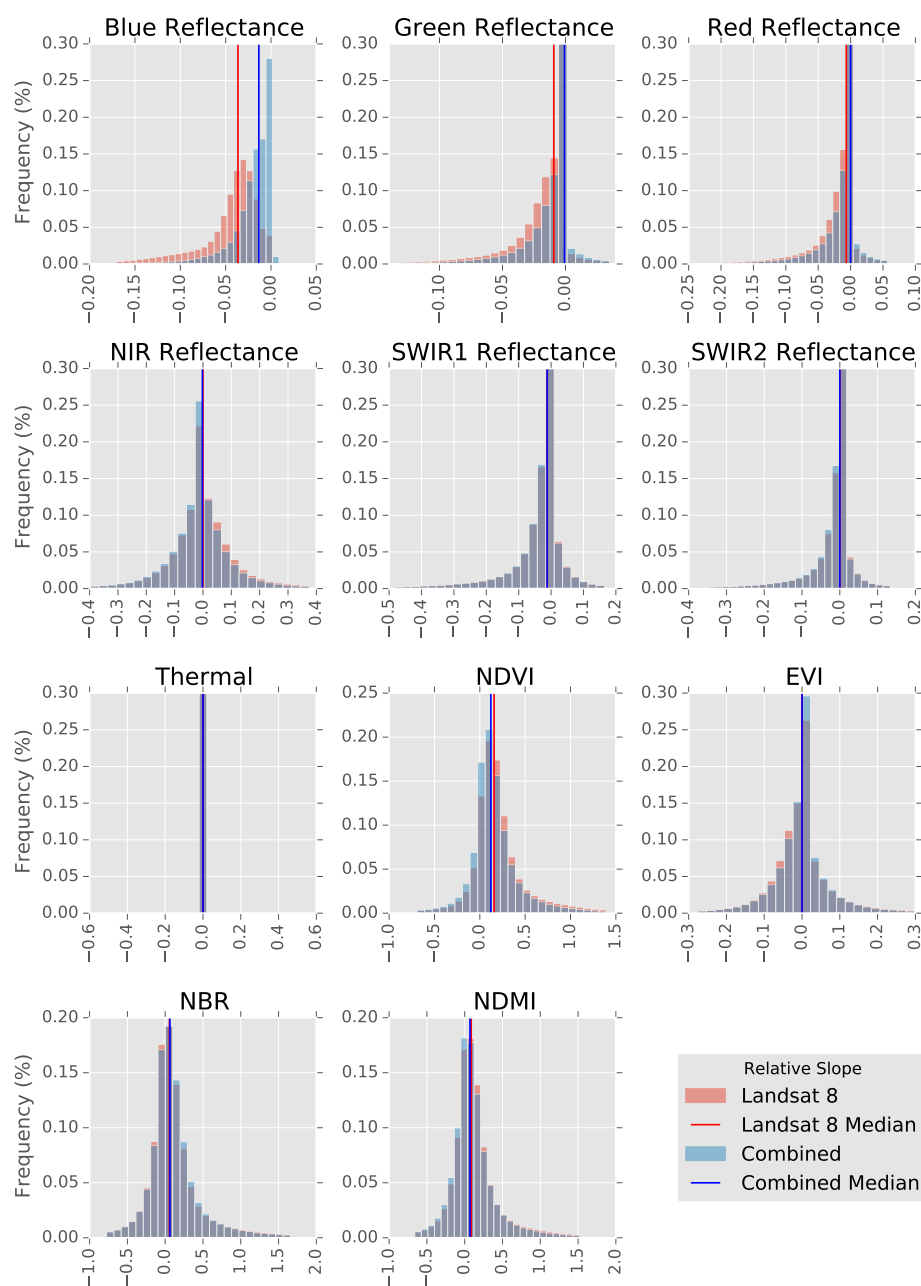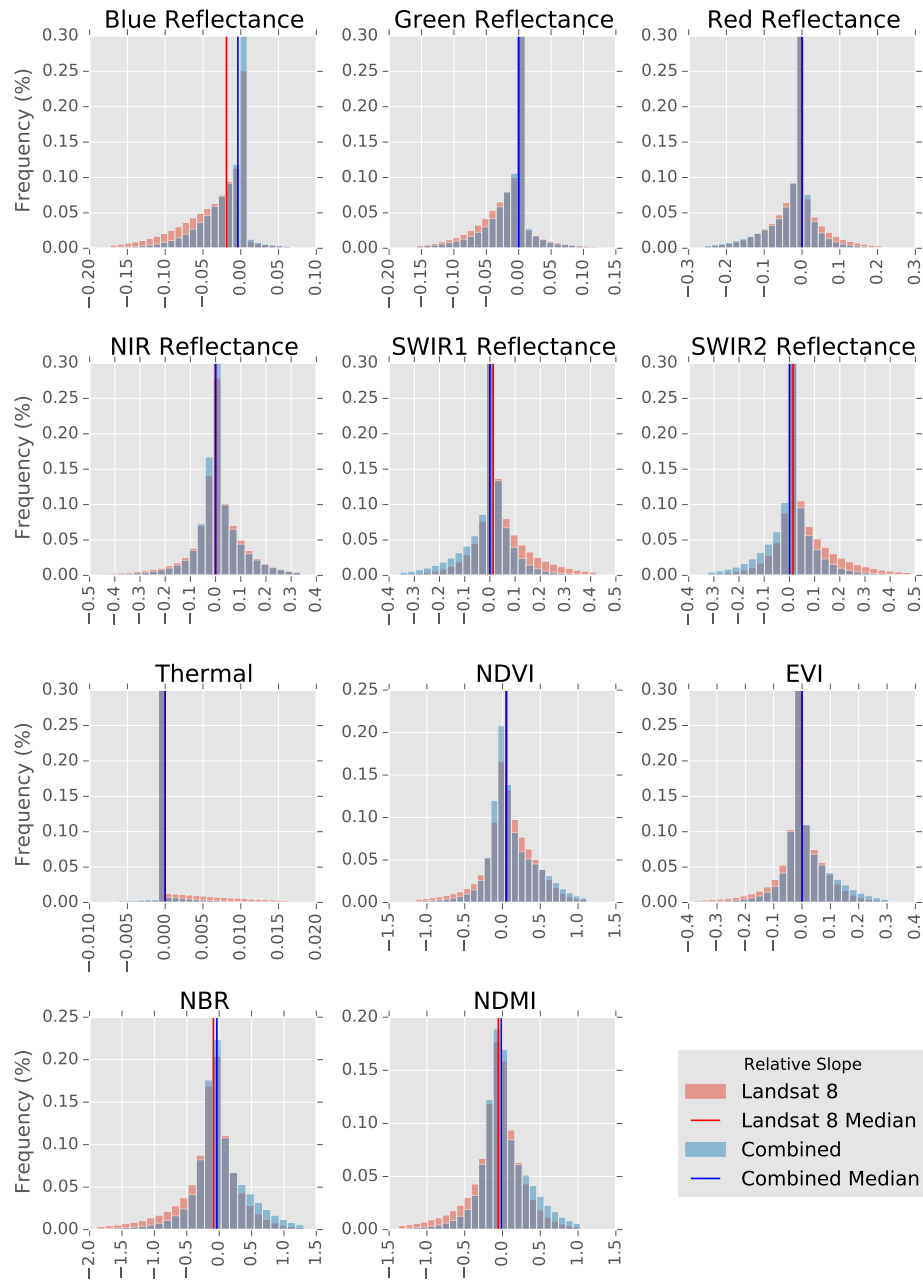
Figure A.3.8: Normalized slope estimates for time series models in California, United States (P043R034) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 slope estimates were normalized by subtracting the slopes from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
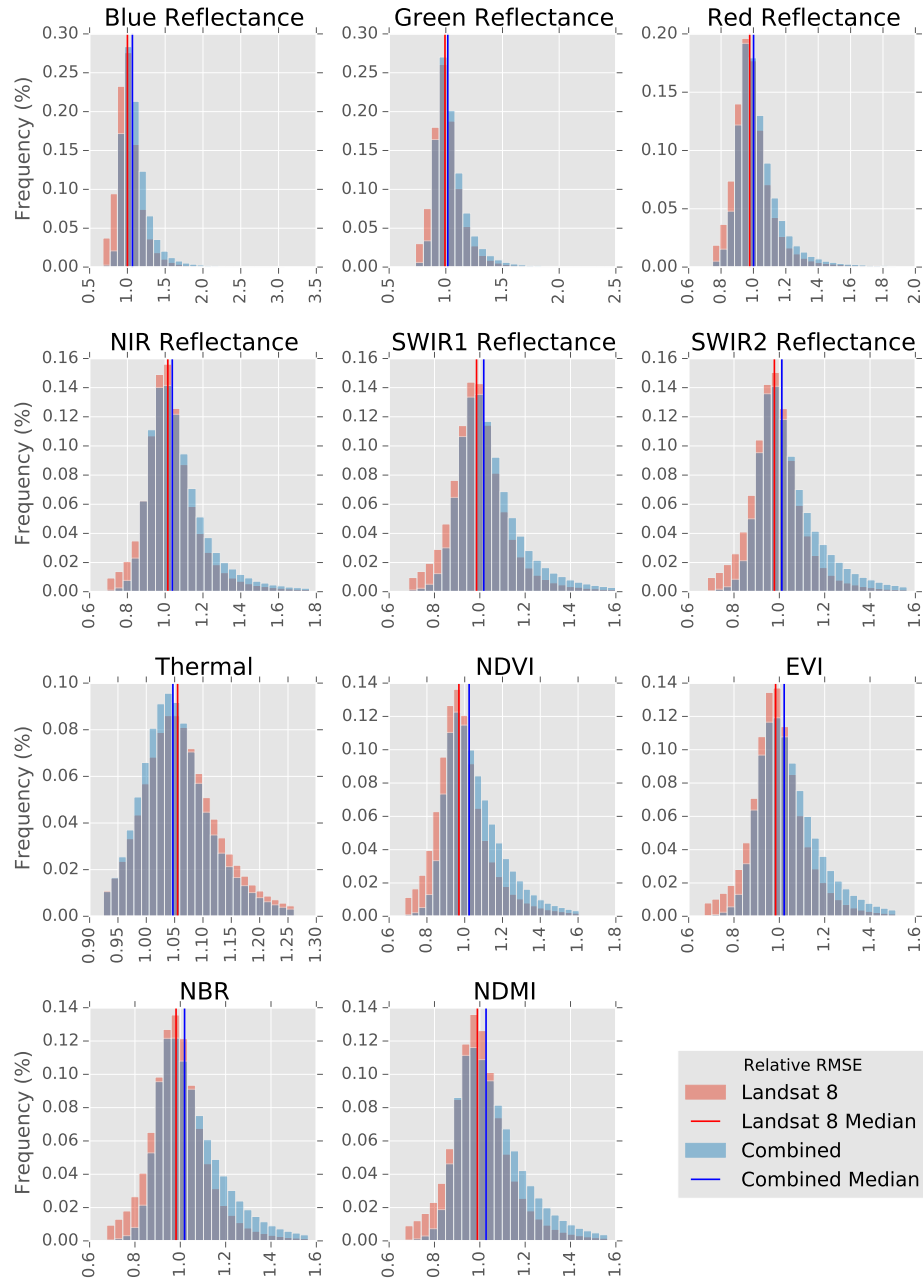
Figure A.3.9: Normalized RMSE estimates for time series models in California, United States (P043R034) used to assess spectral differences caused by including Landsat 8 data. The Landsat 8 only and combined Landsat 8 and Landsat 7 RMSE estimates were normalized by dividing by the RMSE from time series models using only Landsat 7 data. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

A.4 Time Series Cirrus Cloud Results

This section contains supplementary figures for the Landsat time series cirrus band comparison.

Figure A.4.1: Normalized intercept estimates for time series models in Florida, United States (P016R041) used to assess the influence of Landsat 8's cirrus band. Intercept estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by intercepts from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

Figure A.4.2: Normalized slope estimates for time series models in Florida, United States (P016R041) used to assess the influence of Landsat 8's cirrus band. Slope estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by subtracting slopes from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

Figure A.4.3: Normalized RMSE estimates for time series models in Florida, United States (P016R041) used to assess the influence of Landsat 8's cirrus band. RMSE estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by RMSE from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
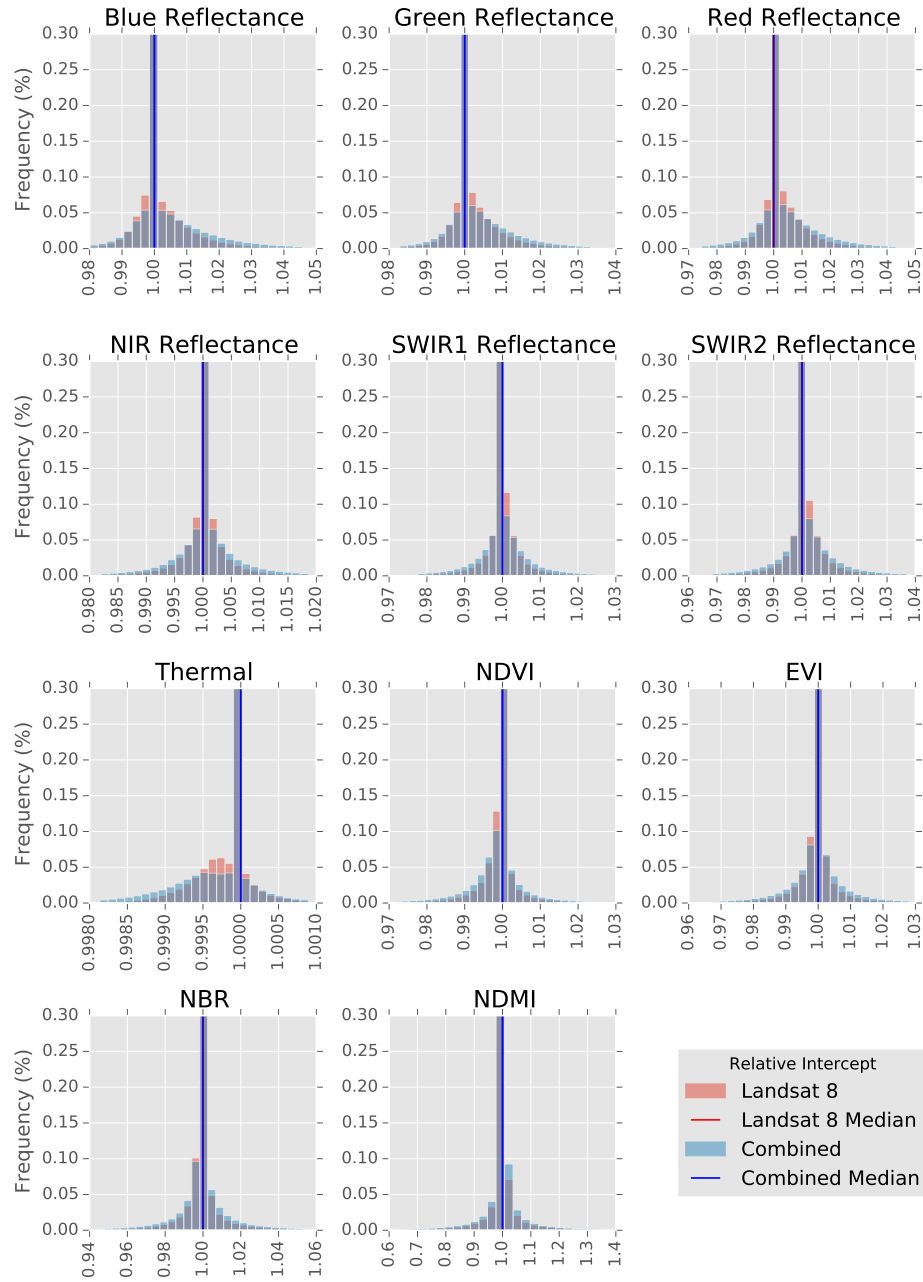
Figure A.4.4: Normalized intercept estimates for time series models in Mississippi, United States (P023R037) used to assess the influence of Landsat 8's cirrus band. Intercept estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by intercepts from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
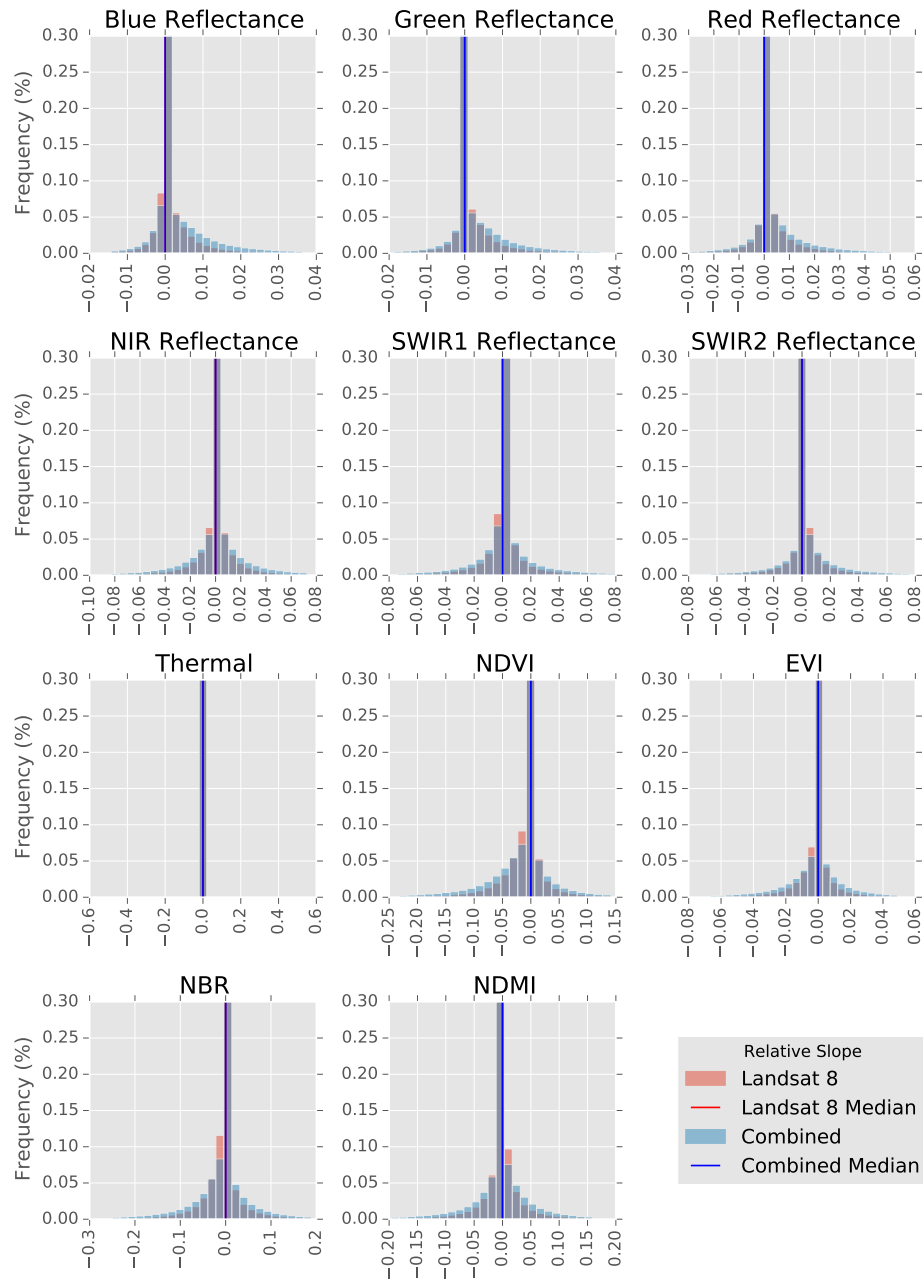
Figure A.4.5: Normalized slope estimates for time series models in Mississippi, United States (P023R037) used to assess the influence of Landsat 8's cirrus band. Slope estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by subtracting slopes from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
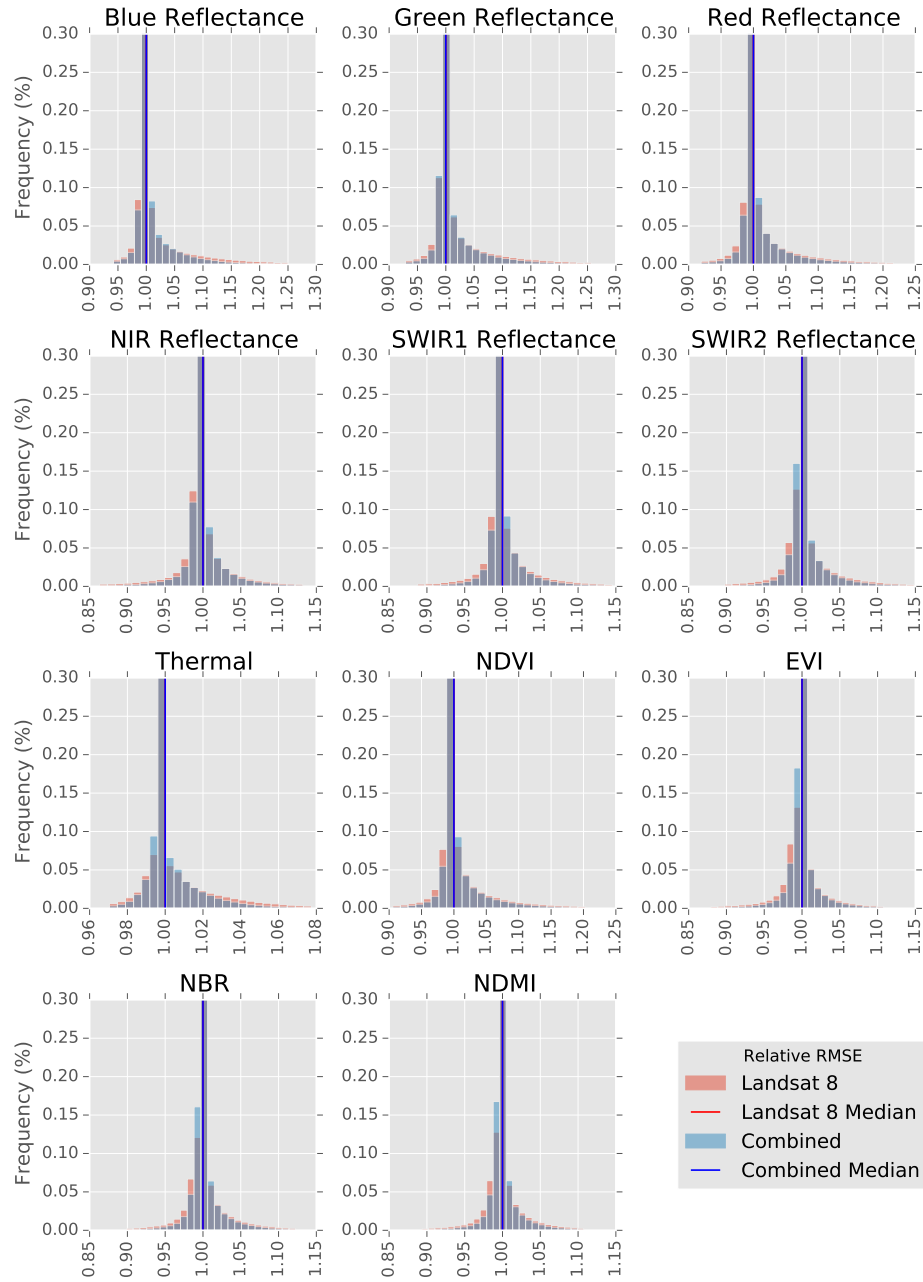
Figure A.4.6: Normalized RMSE estimates for time series models in Mississippi, United States (P023R037) used to assess the influence of Landsat 8's cirrus band. RMSE estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by RMSE from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
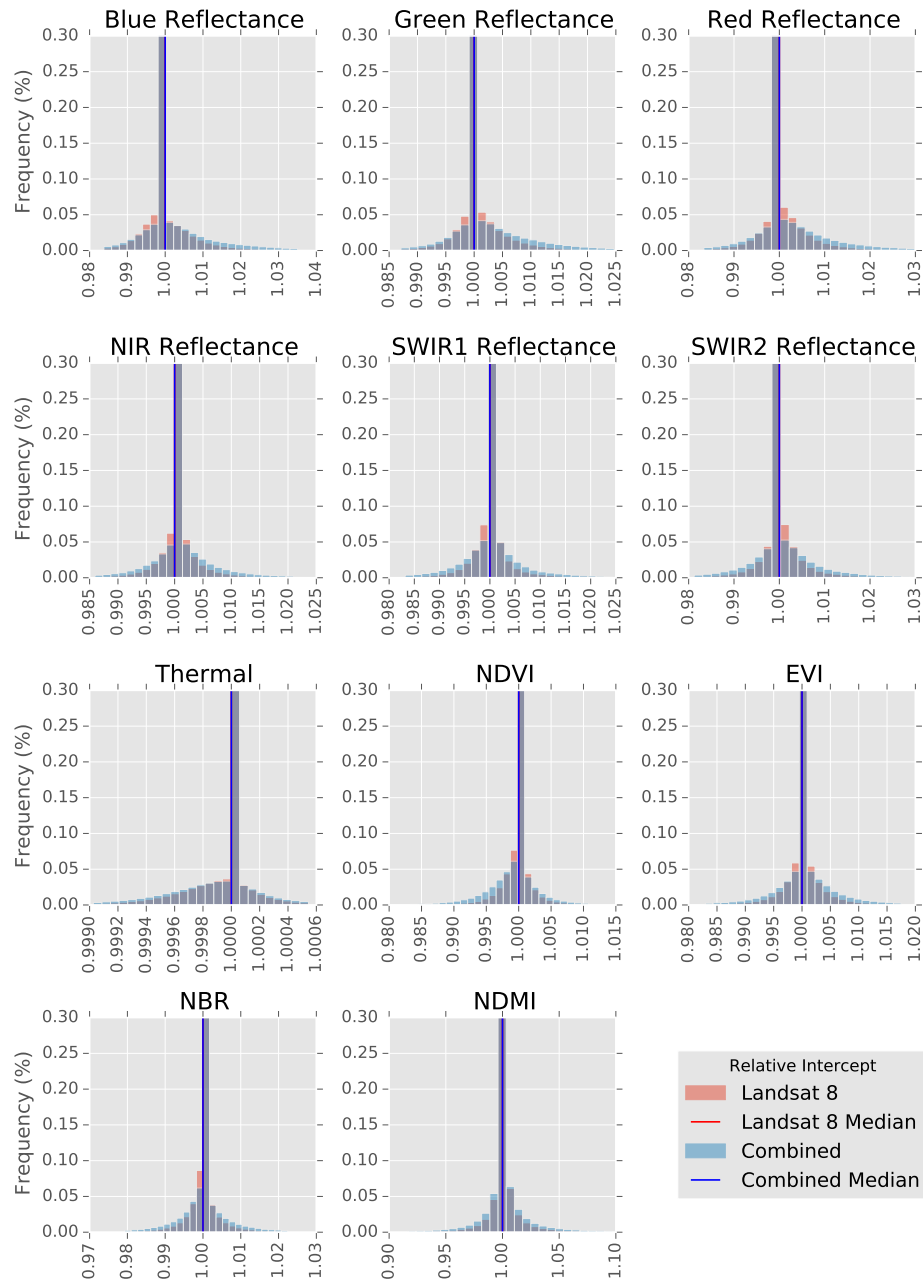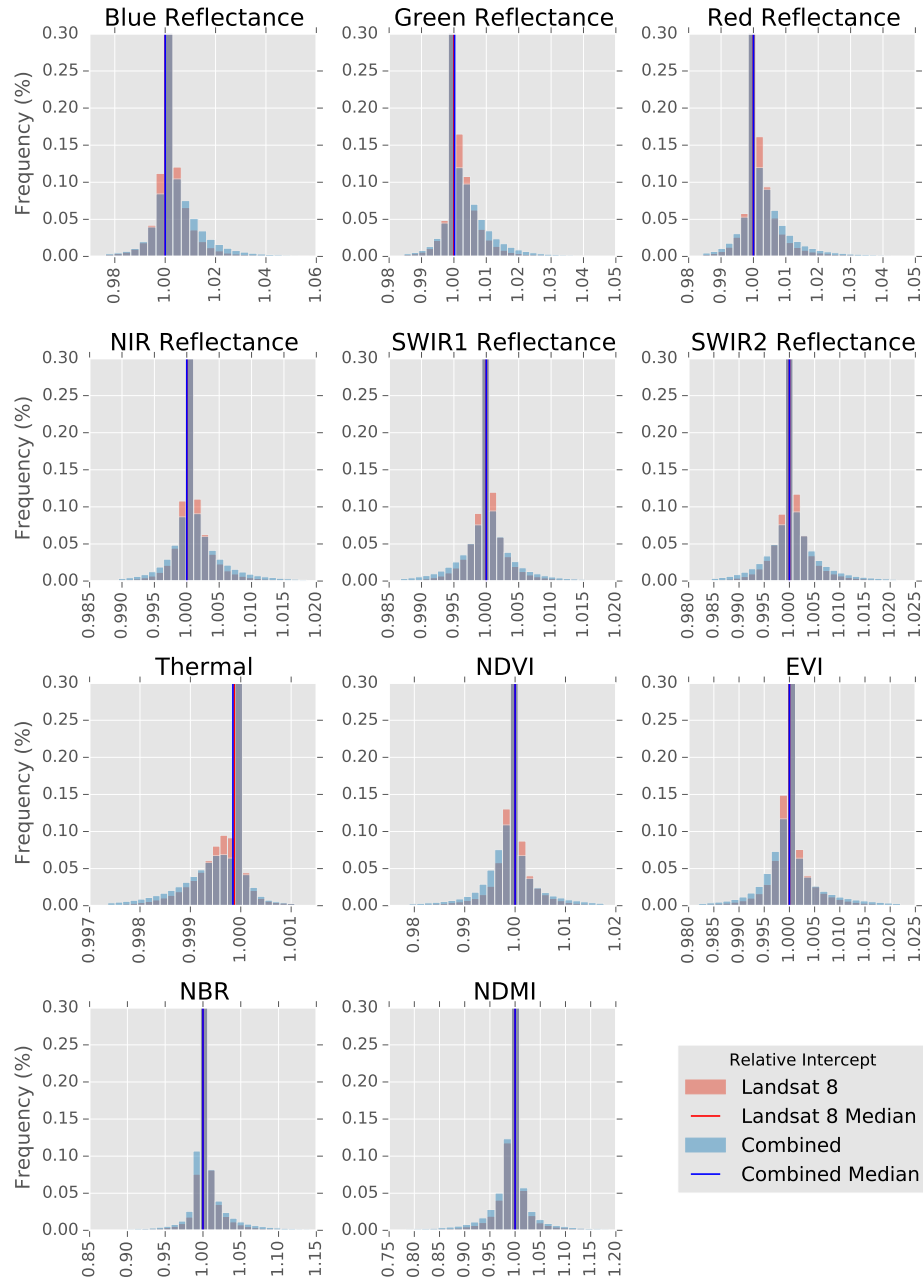
Figure A.4.7: Normalized intercept estimates for time series models in California, United States (P043R034) used to assess the influence of Landsat 8's cirrus band. Intercept estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by intercepts from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
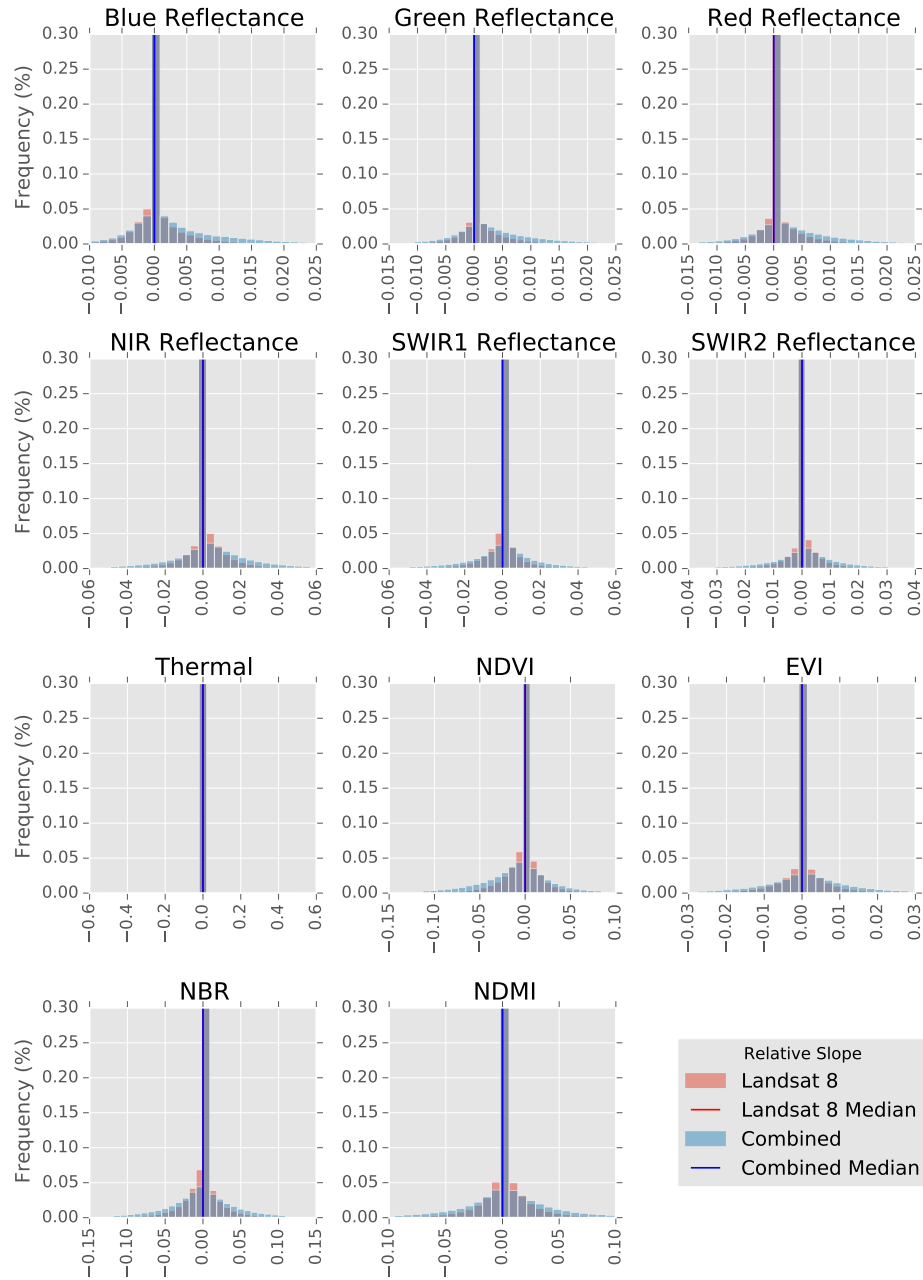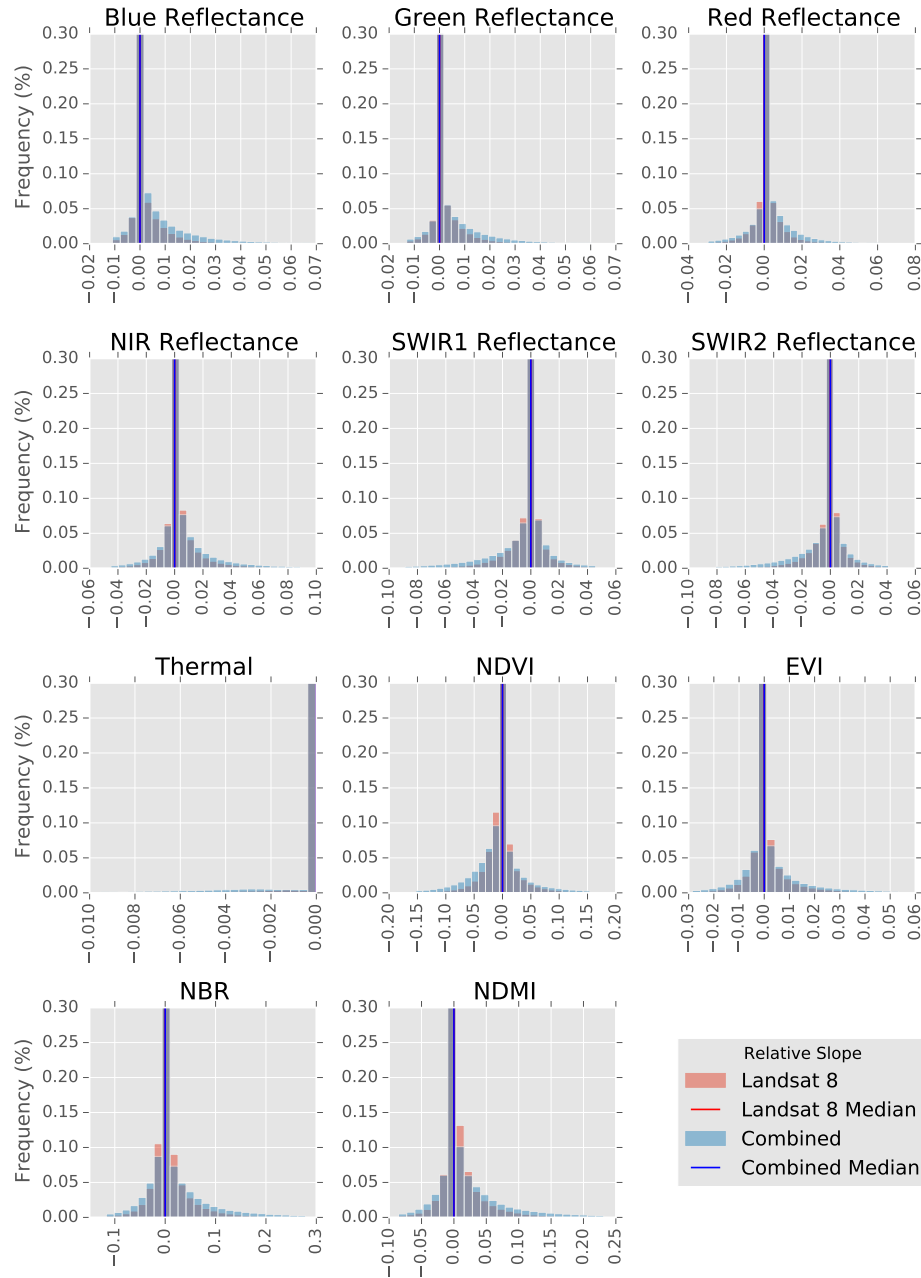
Figure A.4.8: Normalized slope estimates for time series models in California, United States (P043R034) used to assess the influence of Landsat 8's cirrus band. Slope estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by subtracting slopes from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.
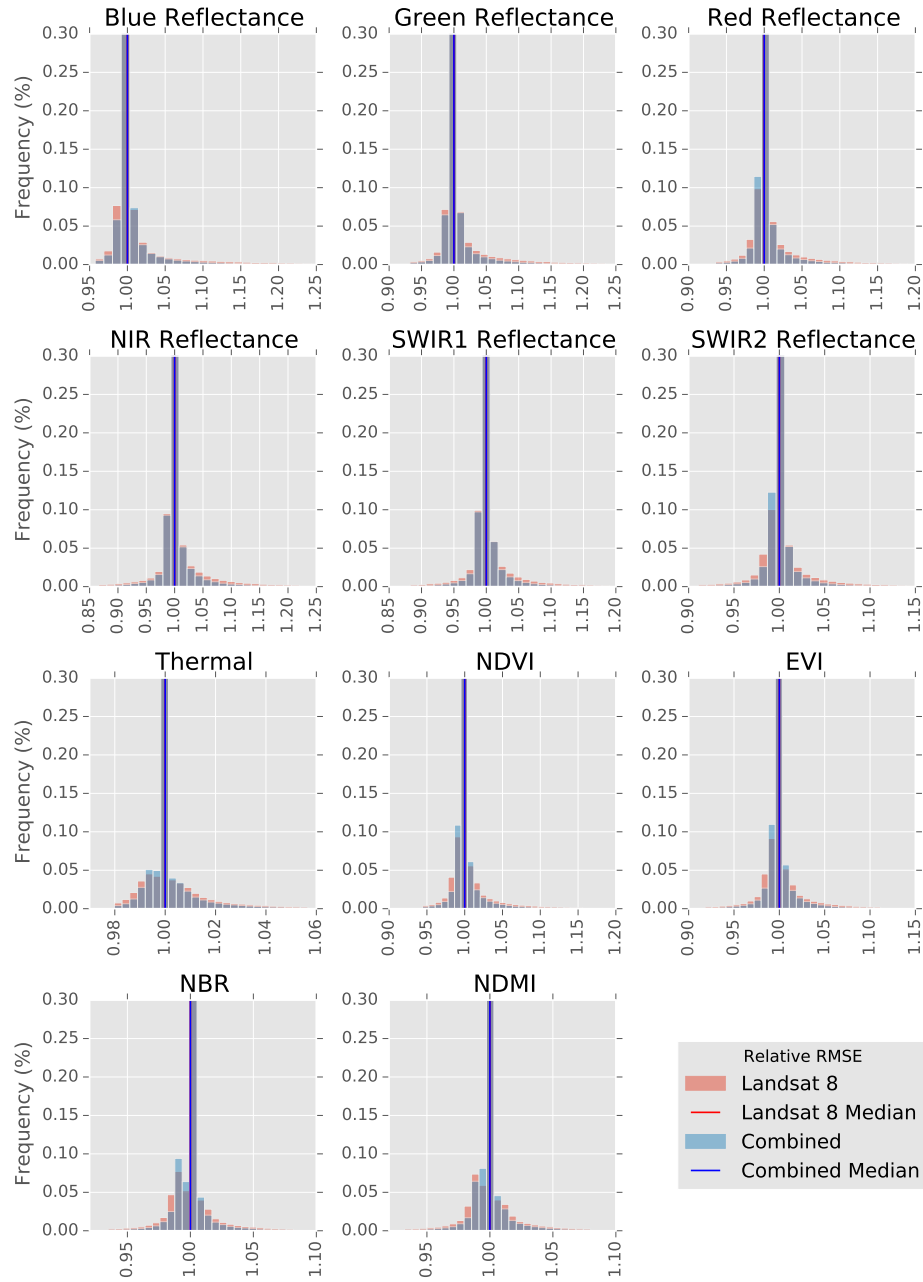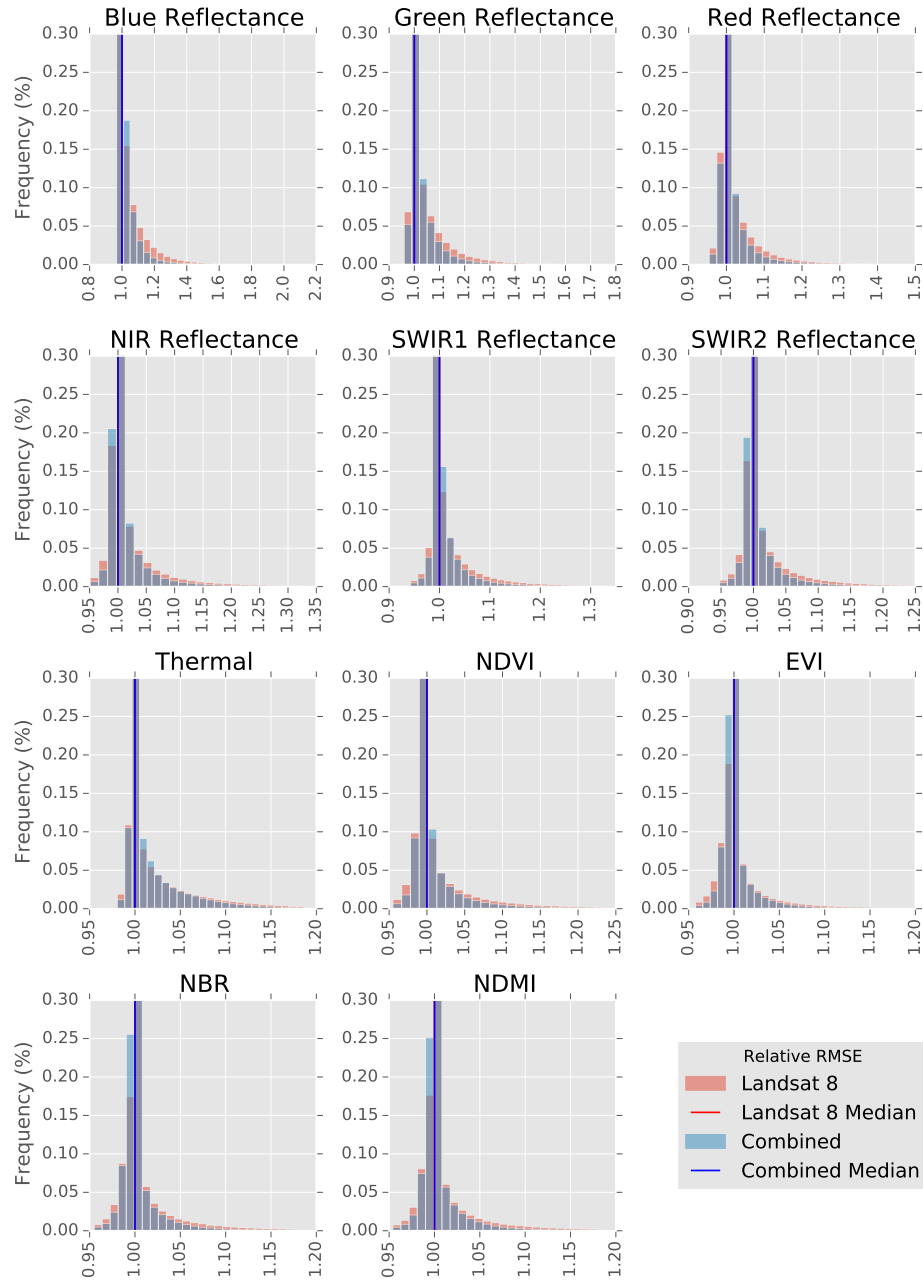
Figure A.4.9: Normalized RMSE estimates for time series models in California, United States (P043R034) used to assess the influence of Landsat 8's cirrus band. RMSE estimates from time series not using Landsat 8's cirrus band in cloud masks were scaled by dividing by RMSE from time series using the cirrus band in Fmask. The Y-axis was truncated at 30% to give better resolution to smaller frequency bins.

Appendix B

Chapter 4: Landsat Time Series Meteorological Data Fusion In Semi-Arid Ecosystems

B.1   Results

B.1.1   Change detection accuracy

Accuracy Assessment

We plotted the distribution of accuracy metrics for all time series models ("Harmonic", "Harmonic + PPT", and "Harmonic + SPEI") for the three land cover classes analyzed in this study using both change detection methods (CCDC and Bai-Perron). These mean values of the groups of data plotted here are summarized in Table 4.2 for CCDC and Table 4.2 for the Bai-Perron algorithm.
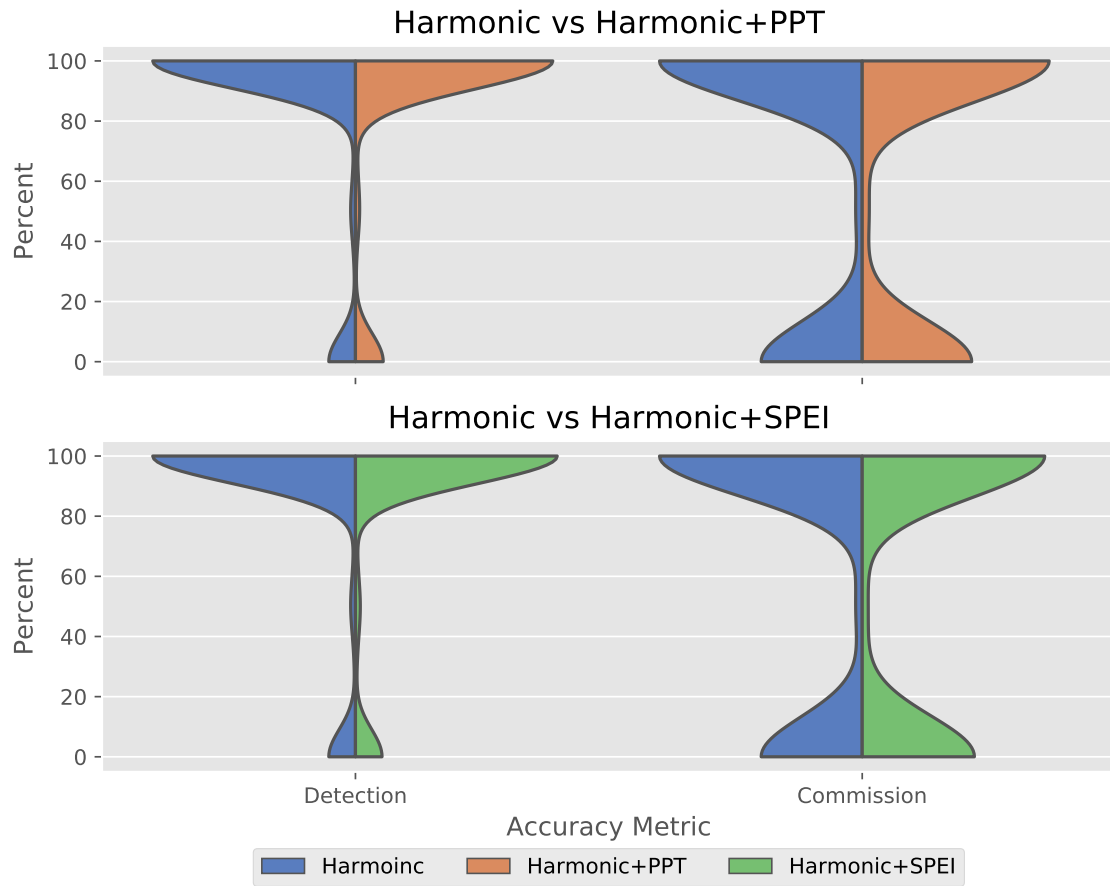
Figure B.1.1: Distribution of accuracy metrics for the CCDC change detection algorithm. These data are summarized as mean values in Table 4.1.
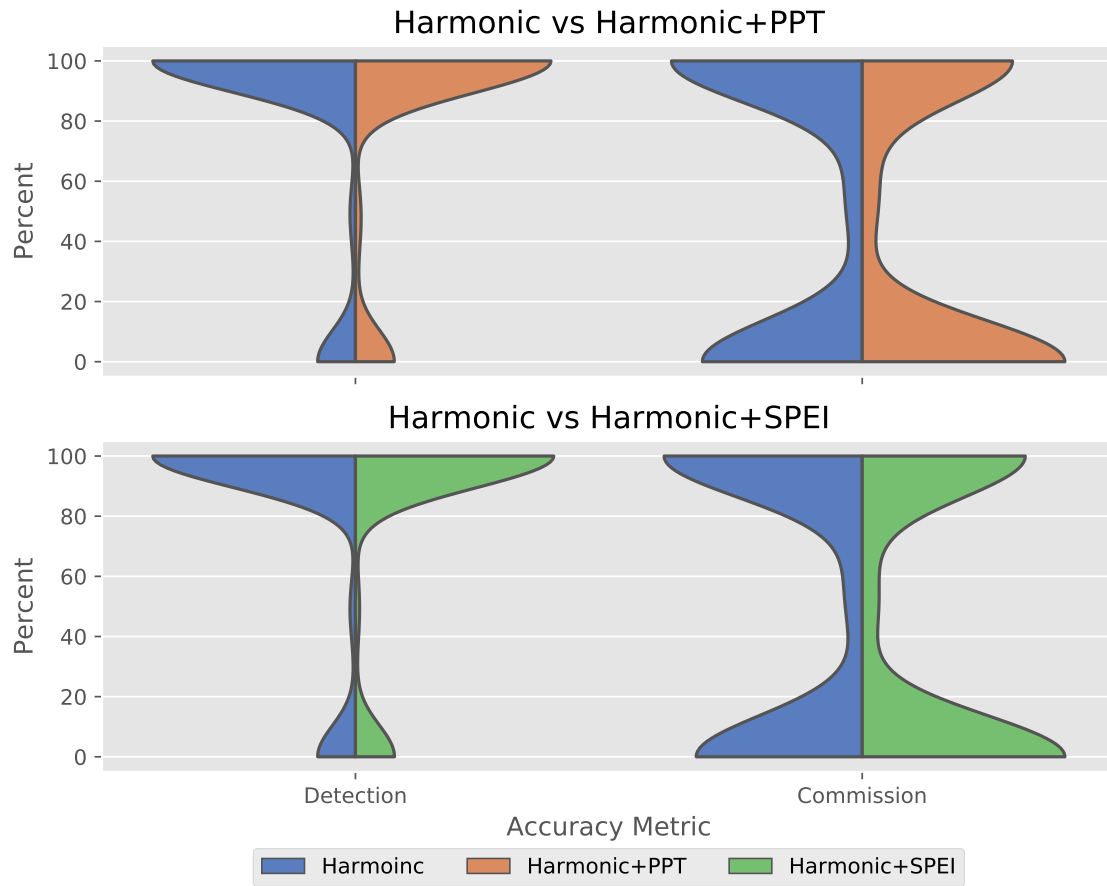
Figure B.1.2: Distribution of accuracy metrics for the Bai-Perron change detection algorithm. These data are summarized as mean values in Table 4.2.

Bibliography

Abatzoglou, J. T., D. J. McEvoy, and K. T. Redmond (2017). "The west wide drought tracker: drought monitoring at fine spatial scales". In: *Bulletin of the American Meteorological Society* 98.9, pp. 1815–1820.

Abercrombie, S. P. and M. A. Friedl (2016). "Improving the consistency of multitemporal land cover maps using a Hidden Markov Model". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.2, pp. 703–713.

Andrews, D. W. (1993). "Tests for parameter instability and structural change with unknown change point". In: *Econometrica: Journal of the Econometric Society*, pp. 821–856.

ASF DAAC (2015). *ALOS-1 PALSAR_Radiometric_Terrain_Corrected_low_res.* Includes Material ©JAXA/METI 2007. Accessed through ASF DAAC 27 August, 2017. DOI: `https://www.asf.alaska.edu/doi/105067/z97hfcnkr6va/`.

Asner, G. P., D. E. Knapp, A. Balaji, and G. Páez-Acosta (2009). "Automated mapping of tropical deforestation and forest degradation: CLASlite". In: *Journal of Applied Remote Sensing* 3.1, p. 033543.

Bai, J. and P. Perron (1998). "Estimating and testing linear models with multiple structural changes". In: *Econometrica*, pp. 47–78.

Birtwistle, A. N., M. Laituri, B. Bledsoe, and J. M. Friedman (2016). "Using NDVI to measure precipitation in semi-arid landscapes". In: *Journal of Arid Environments* 131, pp. 15–24.

Breiman, L. (2001). "Random forests". In: *Machine learning* 45.1, pp. 5–32.

Broich, M., M. G. Tulbure, J. Verbesselt, Q. Xin, and J. Wearne (2018). "Quantifying Australia's dryland vegetation response to flooding and drought at sub-continental scale". In: *Remote Sensing of Environment* 212, pp. 60–78.

Brooks, E. B., R. H. Wynne, V. A. Thomas, C. E. Blinn, and J. W. Coulston (2014). "On-the-fly massively multitemporal change detection using statistical quality

control charts and Landsat data". In: *IEEE Transactions on Geoscience and Remote Sensing* 52.6, pp. 3316–3332.

Bullock, E. L., C. E. Woodcock, and P. Olofsson (2018). "Monitoring Tropical Forest Degradation using Spectral Unmixing and Landsat Time Series Analysis". In: *Remote Sensing of Environment.*

Burga Ríos, M. (2016). "Incremento de la deforestación y sus consecuencias en la pérdida de biomasa en los bosques de la provincia Alto Amazonas del departamento de Loreto, 2000-2014". http://repositorio.ucp.edu.pe/handle/UCP/107. Bachelor's Thesis. Universidad Científica del Perú. Facultad de Ciencias e Ingeniería.

California Department of Forestry and Fire Protection (2018). *Fire Perimeters Version 17$_1$.* `http://frap.fire.ca.gov/data/frapgisdata-sw-fireperimeters_download`.

Cañote Amaya, B. (2018). "Impacto de la instalación de monocultivos de palma aceitera en los bosques húmedos del departamento de Loreto, periodo 1995-2016". http://repositorio.usil.edu.pe/handle/USIL/3206. Bachelor's Thesis.

Canty, M. J., A. A. Nielsen, and M. Schmidt (2004). "Automatic radiometric normalization of multitemporal satellite imagery". In: *Remote Sensing of Environment* 91.3-4, pp. 441–451.

Caruana, R. and A. Niculescu-Mizil (2006). "An empirical comparison of supervised learning algorithms". In: *Proceedings of the 23rd international conference on Machine learning.* ACM, pp. 161–168.

Chen, C., A. Liaw, and L. Breiman (2004). "Using random forest to learn imbalanced data". In: *University of California, Berkeley* 110, pp. 1–12.

Chen, Y., D. Lu, L. Luo, Y. Pokhrel, K. Deb, J. Huang, and Y. Ran (2018). "Detecting irrigation extent, frequency, and timing in a heterogeneous arid agricultural region using MODIS time series, Landsat imagery, and ancillary data". In: *Remote Sensing of Environment* 204, pp. 197–211.

Chow, G. C. (1960). "Tests of equality between sets of coefficients in two linear regressions". In: *Econometrica: Journal of the Econometric Society*, pp. 591–605.

Chu, C.-S. J., K. Hornik, and C.-M. Kaun (1995). "MOSUM tests for parameter constancy". In: *Biometrika* 82.3, pp. 603–617.

Claverie, M., J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, and C. Justice (2018). "The Harmonized Landsat and Sentinel-2 surface reflectance data set". In: *Remote Sensing of Environment* 219, pp. 145–161.

Claverie, M., J. G. Masek, J. Ju, and J. L. Dungan (2017). "Harmonized Landsat-8 Sentinel-2 (HLS) Product User's Guide". In: *National Aeronautics and Space Administration (NASA): Washington, DC, USA*.

Cleland, E. E., I. Chuine, A. Menzel, H. A. Mooney, and M. D. Schwartz (2007). "Shifting plant phenology in response to global change". In: *Trends in ecology & evolution* 22.7, pp. 357–365.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. Terpenning (1990). "STL: A Seasonal-Trend Decomposition". In: *Journal of Official Statistics* 6.1, pp. 3–73.

Cohen, W. B., Z. Yang, and R. Kennedy (2010). "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync—Tools for calibration and validation". In: *Remote Sensing of Environment* 114.12, pp. 2911–2924.

Crist, E. P. (1985). "A TM tasseled cap equivalent transformation for reflectance factor data". In: *Remote Sensing of Environment* 17.3, pp. 301–306.

Cutler, M., D. Boyd, G. Foody, and A. Vetrivel (2012). "Estimating tropical forest biomass with a combination of SAR image texture and Landsat TM data: An assessment of predictions between regions". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 70, pp. 66–77.

Daly, C., G. Taylor, and W. Gibson (1997). "The PRISM approach to mapping precipitation and temperature". In: *Proc., 10th AMS Conf. on Applied Climatology*. Citeseer, pp. 20–23.

Dutrieux, L. P., J. Verbesselt, L. Kooistra, and M. Herold (2015). "Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 107, pp. 112–125.

Dwyer, J., D. Roy, B. Sauer, C. Jenkerson, H. Zhang, and L. Lymburner (2018). "Analysis Ready Data: Enabling Analysis of the Landsat Archive". In: *Remote Sensing* 10.9, p. 1363.

Fernández-Delgado, M., E. Cernadas, S. Barro, and D. Amorim (2014). "Do we need hundreds of classifiers to solve real world classification problems?" In: *The Journal of Machine Learning Research* 15.1, pp. 3133–3181.

Flood, N. (2014). "Continuity of reflectance data between Landsat-7 ETM+ and Landsat-8 OLI, for both top-of-atmosphere and surface reflectance: a study in the Australian landscape". In: *Remote Sensing* 6.9, pp. 7952–7970.

Flood, N., T. Danaher, T. Gill, and S. Gillingham (2013). "An operational scheme for deriving standardised surface reflectance from Landsat TM/ETM+ and SPOT HRG imagery for Eastern Australia". In: *Remote Sensing* 5.1, pp. 83–109.

Formica, A. F., R. J. Burnside, and P. M. Dolman (2017). "Rainfall validates MODIS-derived NDVI as an index of spatio-temporal variation in green biomass across non-montane semi-arid and arid Central Asia". In: *Journal of Arid Environments* 142, pp. 11–21.

Frankenberg, C., C. O'Dell, J. Berry, L. Guanter, J. Joiner, P. Köhler, R. Pollock, and T. E. Taylor (2014). "Prospects for chlorophyll fluorescence remote sensing from the Orbiting Carbon Observatory-2". In: *Remote Sensing of Environment* 147, pp. 1–12.

Frey, O., M. Santoro, C. L. Werner, and U. Wegmuller (2013). "DEM-based SAR pixel-area estimation for enhanced geocoding refinement and radiometric normalization". In: *IEEE Geoscience and Remote Sensing Letters* 10.1, pp. 48–52.

Friedl, M. A., J. M. Gray, E. K. Melaas, A. D. Richardson, K. Hufkens, T. F. Keenan, A. Bailey, and J. O'Keefe (2014). "A tale of two springs: using recent climate anomalies to characterize the sensitivity of temperate forest phenology to climate change". In: *Environmental Research Letters* 9.5, p. 054006.

Friedl, M. A., D. K. McIver, J. C. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, et al. (2002). "Global land cover mapping from MODIS: algorithms and early results". In: *Remote sensing of Environment* 83.1-2, pp. 287–302.

Gao, B.-C., A. F. Goetz, and W. J. Wiscombe (1993). "Cirrus cloud detection from airborne imaging spectrometer data using the 1.38 μm water vapor band". In: *Geophysical Research Letters* 20.4, pp. 301–304.

Gao, B.-C. and Y. J. Kaufman (1995). "Selection of the 1.375-μm MODIS channel for remote sensing of cirrus clouds and stratospheric aerosols from space". In: *Journal of the atmospheric sciences* 52.23, pp. 4231–4237.

Gao, B.-C., Y. J. Kaufman, W. Han, and W. J. Wiscombe (1998). "Corection of thin cirrus path radiances in the 0.4–1.0 μm spectral region using the sensitive 1.375 μm cirrus detecting channel". In: *Journal of Geophysical Research: Atmospheres* 103.D24, pp. 32169–32176.

Gao, B.-C. and R.-R. Li (2000). "Quantitative improvement in the estimates of NDVI values from remotely sensed data by correcting thin cirrus scattering effects". In: *Remote Sensing of Environment* 74.3, pp. 494–502.

Gao, F., J. Masek, M. Schwaller, and F. Hall (2006). "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance". In: *IEEE Transactions on Geoscience and Remote sensing* 44.8, pp. 2207–2218.

Gens, R. (2015). "Algorithm Theoretical Basis Document: ASF radiometric terrain corrected products (Revision 1.2)". In: pp. 1–16.

Goetz, S. J., A. G. Bunn, G. J. Fiske, and R. Houghton (2005). "Satellite-observed photosynthetic trends across boreal North America associated with climate and fire disturbance". In: *Proceedings of the National Academy of Sciences* 102.38, pp. 13521–13525.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". In: *Remote Sensing of Environment* 202, pp. 18–27.

Griffiths, P., T. Kuemmerle, R. E. Kennedy, I. V. Abrudan, J. Knorn, and P. Hostert (2012). "Using annual time-series of Landsat images to assess the effects of forest restitution in post-socialist Romania". In: *Remote Sensing of Environment* 118, pp. 199–214.

Griffiths, P., S. van der Linden, T. Kuemmerle, and P. Hostert (2013). "A pixel-based Landsat compositing algorithm for large area land cover mapping". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.5, pp. 2088–2101.

Guzmán, W. and L. Arévalo (2003). "Servicios ambientales de almacenamiento de carbono activo para el desarrollo en la Amazonia Peruana: Avances y retos". Seminario Permanente de Investigación Agraria. (10, 2003. Pucallpa, Perú).

Hamunyela, E., J. Reiche, J. Verbesselt, and M. Herold (2017). "Using space-time features to improve detection of forest disturbances from landsat time series". In: *Remote Sensing* 9.6, p. 515.

Hamunyela, E., J. Verbesselt, S. De Bruin, and M. Herold (2016). "Monitoring deforestation at sub-annual scales as extreme events in Landsat data cubes". In: *Remote Sensing* 8.8, p. 651.

Hamunyela, E., J. Verbesselt, and M. Herold (2016). "Using spatial context to improve early detection of deforestation from Landsat time series". In: *Remote Sensing of Environment* 172, pp. 126–138.

Hansen, M. C. and T. R. Loveland (2012). "A review of large area monitoring of land cover change using Landsat data". In: *Remote sensing of Environment* 122, pp. 66–74.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. Turubanova, A. Tyukavina, D. Thau, S. Stehman, S. Goetz, T. Loveland, et al. (2013). "High-resolution global maps of 21st-century forest cover change". In: *science* 342.6160, pp. 850–853.

Hansen, M., A. Egorov, P. Potapov, S. Stehman, A. Tyukavina, S. Turubanova, D. P. Roy, S. Goetz, T. Loveland, J. Ju, et al. (2014). "Monitoring conterminous United States (CONUS) land cover change with web-enabled Landsat data (WELD)". In: *Remote sensing of Environment* 140, pp. 466–484.

Hausner, M. B., J. L. Huntington, C. Nash, C. Morton, D. J. McEvoy, D. S. Pilliod, K. C. Hegewisch, B. Daudert, J. T. Abatzoglou, and G. Grant (2018). "Assessing the effectiveness of riparian restoration projects using Landsat and precipitation data from the cloud-computing application ClimateEngine. org". In: *Ecological Engineering* 120, pp. 432–440.

Healey, S. P., W. B. Cohen, Z. Yang, C. K. Brewer, E. B. Brooks, N. Gorelick, A. J. Hernandez, C. Huang, M. J. Hughes, R. E. Kennedy, et al. (2018). "Mapping forest change using stacked generalization: An ensemble approach". In: *Remote Sensing of Environment* 204, pp. 717–728.

Hilker, T., A. I. Lyapustin, C. J. Tucker, F. G. Hall, R. B. Myneni, Y. Wang, J. Bi, Y. M. de Moura, and P. J. Sellers (2014). "Vegetation dynamics and rainfall sensitivity of the Amazon". In: *Proceedings of the National Academy of Sciences* 111.45, pp. 16041–16046.

Holden, C. E. and C. E. Woodcock (2016). "An analysis of Landsat 7 and Landsat 8 underflight data and the implications for time series investigations". In: *Remote Sensing of Environment* 185, pp. 16–36.

Homer, C., J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown (2015). "Completion of the 2011 National Land Cover Database for the conterminous United States–representing a decade of land cover change information". In: *Photogrammetric Engineering & Remote Sensing* 81.5, pp. 345–354.

Houghton, R. A., J. L. Hackler, and K. T. Lawrence (1999). "The US carbon budget: contributions from land-use change". In: *Science* 285.5427, pp. 574–578.

Huang, C., S. N. Goward, J. G. Masek, N. Thomas, Z. Zhu, and J. E. Vogelmann (2010). "An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks". In: *Remote Sensing of Environment* 114.1, pp. 183–198.

Huang, C., B. Wylie, L. Yang, C. Homer, and G. Zylstra (2002). "Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance". In: *International Journal of Remote Sensing* 23.8, pp. 1741–1748.

Huete, A. R., K. Didan, Y. E. Shimabukuro, P. Ratana, S. R. Saleska, L. R. Hutyra, W. Yang, R. R. Nemani, and R. Myneni (2006). "Amazon rainforests green-up with sunlight in dry season". In: *Geophysical research letters* 33.6.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira (2002). "Overview of the radiometric and biophysical performance of the MODIS vegetation indices". In: *Remote sensing of environment* 83.1-2, pp. 195–213.

Hufkens, K., T. F. Keenan, L. B. Flanagan, R. L. Scott, C. J. Bernacchi, E. Joo, N. A. Brunsell, J. Verfaillie, and A. D. Richardson (2016). "Productivity of North American grasslands is increased under future climate scenarios despite rising aridity". In: *Nature Climate Change* 6.7, p. 710.

Irons, J. R., J. L. Dwyer, and J. A. Barsi (2012). "The next Landsat satellite: The Landsat data continuity mission". In: *Remote Sensing of Environment* 122, pp. 11–21.

Joshi, N., M. Baumann, A. Ehammer, R. Fensholt, K. Grogan, P. Hostert, M. R. Jepsen, T. Kuemmerle, P. Meyfroidt, E. T. Mitchard, et al. (2016). "A review

of the application of optical and radar remote sensing data fusion to land use mapping and monitoring". In: *Remote Sensing* 8.1, p. 70.

Ju, J. and J. G. Masek (2016). "The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data". In: *Remote Sensing of Environment* 176, pp. 1–16.

Kankaku, Y., M. Sagisaka, and S. Suzuki (2014). "PALSAR-2 launch and early orbit status". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*. IEEE, pp. 3410–3412.

Kaufman, Y. J., A. E. Wald, L. A. Remer, B.-C. Gao, R.-R. Li, and L. Flynn (1997). "The MODIS 2.1-/spl mu/m channel-correlation with visible reflectance for use in remote sensing of aerosol". In: *IEEE transactions on Geoscience and Remote Sensing* 35.5, pp. 1286–1298.

Kauth, R. J. and G. Thomas (1976). "The tasselled cap–a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat". In: *LARS Symposia*, p. 159.

Kennedy, R. E., W. B. Cohen, and T. A. Schroeder (2007). "Trajectory-based change detection for automated characterization of forest disturbance dynamics". In: *Remote Sensing of Environment* 110.3, pp. 370–386.

Kennedy, R. E., Z. Yang, and W. B. Cohen (2010). "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms". In: *Remote Sensing of Environment* 114.12, pp. 2897–2910.

Kennedy, R. E., Z. Yang, W. B. Cohen, E. Pfaff, J. Braaten, and P. Nelson (2012). "Spatial and temporal patterns of forest disturbance and regrowth within the area of the Northwest Forest Plan". In: *Remote Sensing of Environment* 122, pp. 117–133.

Kennedy, R. E., S. Andréfouët, W. B. Cohen, C. Gómez, P. Griffiths, M. Hais, S. P. Healey, E. H. Helmer, P. Hostert, M. B. Lyons, et al. (2014). "Bringing an ecological view of change to Landsat-based remote sensing". In: *Frontiers in Ecology and the Environment* 12.6, pp. 339–346.

Key, C. and N. Benson (2005). "Landscape assessment: remote sensing of severity, the normalized burn ratio and ground measure of severity, the composite burn index".

In: *FIREMON: Fire effects monitoring and inventory system Ogden, Utah: USDA Forest Service, Rocky Mountain Res. Station.*

Kim, D.-H., J. O. Sexton, P. Noojipady, C. Huang, A. Anand, S. Channan, M. Feng, and J. R. Townshend (2014). "Global, Landsat-based forest-cover change from 1990 to 2000". In: *Remote Sensing of Environment* 155, pp. 178–193.

Knight, E. J. and G. Kvaran (2014). "Landsat-8 operational land imager design, characterization and performance". In: *Remote Sensing* 6.11, pp. 10286–10305.

Kovalskyy, V. and D. P. Roy (2013). "The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation". In: *Remote Sensing of Environment* 130, pp. 280–293.

Kovalskyy, V. and D. P. Roy (2015). "A one year Landsat 8 conterminous United States study of cirrus and non-cirrus clouds". In: *Remote Sensing* 7.1, pp. 564–578.

Kuplich, T., P. J. Curran, and P. M. Atkinson (2005). "Relating SAR image texture to the biomass of regenerating tropical forests". In: *International Journal of Remote Sensing* 26.21, pp. 4829–4854.

LeCun, Y., Y. Bengio, and G. Hinton (2015). "Deep learning". In: *nature* 521.7553, p. 436.

Lesica, P. and P. Kittelson (2010). "Precipitation and temperature are associated with advanced flowering phenology in a semi-arid grassland". In: *Journal of Arid Environments* 74.9, pp. 1013–1017.

Lewis, A., S. Oliver, L. Lymburner, B. Evans, L. Wyborn, N. Mueller, G. Raevksi, J. Hooke, R. Woodcock, J. Sixsmith, et al. (2017). "The Australian geoscience data cube—Foundations and lessons learned". In: *Remote Sensing of Environment* 202, pp. 276–292.

Li, P., L. Jiang, and Z. Feng (2013). "Cross-comparison of vegetation indices derived from Landsat-7 enhanced thematic mapper plus (ETM+) and Landsat-8 operational land imager (OLI) sensors". In: *Remote Sensing* 6.1, pp. 310–329.

Liu, Y., M. J. Hill, X. Zhang, Z. Wang, A. D. Richardson, K. Hufkens, G. Filippa, D. D. Baldocchi, S. Ma, J. Verfaillie, et al. (2017). "Using data from Landsat, MODIS, VIIRS and PhenoCams to monitor the phenology of California oak/grass

savanna and open grassland across spatial scales". In: *Agricultural and Forest Meteorology* 237, pp. 311–325.

Loveland, T. R. and J. L. Dwyer (2012). "Landsat: Building a strong future". In: *Remote Sensing of Environment* 122, pp. 22–29.

Lu, D. (2005). "Aboveground biomass estimation using Landsat TM data in the Brazilian Amazon". In: *International Journal of Remote Sensing* 26.12, pp. 2509–2525.

Lu, D. and Q. Weng (2007). "A survey of image classification methods and techniques for improving classification performance". In: *International journal of Remote sensing* 28.5, pp. 823–870.

Lucas, J. M. and M. S. Saccucci (1990). "Exponentially weighted moving average control schemes: properties and enhancements". In: *Technometrics* 32.1, pp. 1–12.

Lucas, R., J. Armston, R. Fairfax, R. Fensham, A. Accad, J. Carreiras, J. Kelley, P. Bunting, D. Clewley, S. Bray, et al. (2010). "An evaluation of the ALOS PALSAR L-band backscatter—Above ground biomass relationship Queensland, Australia: Impacts of surface moisture condition and vegetation structure". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3.4, pp. 576–593.

Luckman, A., A. Frery, C. Yanasse, and G. Groom (1997). "Texture in airborne SAR imagery of tropical forest and its relationship to forest regeneration stage". In: *International Journal of Remote Sensing* 18.6, pp. 1333–1349.

Malenovskỳ, Z., H. Rott, J. Cihlar, M. E. Schaepman, G. Garćıa-Santos, R. Fernandes, and M. Berger (2012). "Sentinels for science: Potential of Sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land". In: *Remote Sensing of Environment* 120, pp. 91–101.

Masek, J. G., S. N. Goward, R. E. Kennedy, W. B. Cohen, G. G. Moisen, K. Schleeweis, and C. Huang (2013). "United States forest disturbance trends observed using Landsat time series". In: *Ecosystems* 16.6, pp. 1087–1104.

Masek, J. G., C. Huang, R. Wolfe, W. Cohen, F. Hall, J. Kutler, and P. Nelson (2008). "North American forest disturbance mapped from a decadal Landsat record". In: *Remote Sensing of Environment* 112.6, pp. 2914–2926.

Masek, J. G., E. F. Vermote, N. E. Saleous, R. Wolfe, F. G. Hall, K. F. Huemmrich, F. Gao, J. Kutler, and T.-K. Lim (2006). "A Landsat surface reflectance dataset for North America, 1990-2000". In: *IEEE Geoscience and Remote Sensing Letters* 3.1, pp. 68–72.

McIver, D. and M. Friedl (2002). "Using prior probabilities in decision-tree classification of remotely sensed data". In: *Remote sensing of Environment* 81.2-3, pp. 253–261.

Meigs, G. W., R. E. Kennedy, and W. B. Cohen (2011). "A Landsat time series approach to characterize bark beetle and defoliator impacts on tree mortality and surface fuels in conifer forests". In: *Remote Sensing of Environment* 115.12, pp. 3707–3718.

Melaas, E. K., M. A. Friedl, and Z. Zhu (2013). "Detecting interannual variation in deciduous broadleaf forest phenology using Landsat TM/ETM+ data". In: *Remote Sensing of Environment* 132, pp. 176–185.

Melaas, E. K., D. Sulla-Menashe, J. M. Gray, T. A. Black, T. H. Morin, A. D. Richardson, and M. A. Friedl (2016). "Multisite analysis of land surface phenology in North American temperate and boreal deciduous forests from Landsat". In: *Remote Sensing of Environment* 186, pp. 452–464.

Meroni, M., M. Rossini, L. Guanter, L. Alonso, U. Rascher, R. Colombo, and J. Moreno (2009). "Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications". In: *Remote Sensing of Environment* 113.10, pp. 2037–2051.

Mishra, N., M. O. Haque, L. Leigh, D. Aaron, D. Helder, and B. Markham (2014). "Radiometric cross calibration of Landsat 8 operational land imager (OLI) and Landsat 7 enhanced thematic mapper plus (ETM+)". In: *Remote Sensing* 6.12, pp. 12619–12638.

Morfitt, R., J. Barsi, R. Levy, B. Markham, E. Micijevic, L. Ong, P. Scaramuzza, and K. Vanderwerff (2015). "Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit". In: *Remote Sensing* 7.2, pp. 2208–2237.

Motohka, T., Y. Kankaku, S. Suzuki, and M. Shimada (2017). "Status of the advanced land observing satellite-2 (ALOS-2) and its follow-on L-band SAR mission". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2017 IEEE International*. IEEE, pp. 2427–2429.

Moulin, S., L. Kergoat, N. Viovy, and G. Dedieu (1997). "Global-scale assessment of vegetation phenology using NOAA/AVHRR satellite measurements". In: *Journal of Climate* 10.6, pp. 1154–1170.

Myneni, R. B., C. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani (1997). "Increased plant growth in the northern high latitudes from 1981 to 1991". In: *Nature* 386.6626, p. 698.

Niculescu-Mizil, A. and R. Caruana (2005). "Predicting good probabilities with supervised learning". In: *Proceedings of the 22nd international conference on Machine learning.* ACM, pp. 625–632.

Olofsson, P., G. M. Foody, S. V. Stehman, and C. E. Woodcock (2013). "Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation". In: *Remote Sensing of Environment* 129, pp. 122–131.

Page, E. S. (1954). "Continuous inspection schemes". In: *Biometrika* 41.1/2, pp. 100–115.

Pasquarella, V. J., C. E. Holden, L. Kaufman, and C. E. Woodcock (2016). "From imagery to ecology: leveraging time series of all available Landsat observations to map and monitor ecosystem state and dynamics". In: *Remote Sensing in Ecology and Conservation* 2.3, pp. 152–170.

Pasquarella, V. J., C. E. Holden, and C. E. Woodcock (2018). "Improved mapping of forest type using spectral-temporal Landsat features". In: *Remote Sensing of Environment* 210, pp. 193–207.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pflugmacher, D., W. B. Cohen, and R. E. Kennedy (2012). "Using Landsat-derived disturbance history (1972–2010) to predict current forest structure". In: *Remote Sensing of Environment* 122, pp. 146–165.

Platt, J. et al. (1999). "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods". In: *Advances in large margin classifiers* 10.3, pp. 61–74.

Ploberger, W. and W. Krämer (1992). "The CUSUM test with OLS residuals". In: *Econometrica: Journal of the Econometric Society*, pp. 271–285.

Provost, F. and P. Domingos (2003). "Tree induction for probability-based ranking". In: *Machine learning* 52.3, pp. 199–215.

Quegan, S. and J. J. Yu (2001). "Filtering of multichannel SAR images". In: *IEEE Transactions on Geoscience and Remote Sensing* 39.11, pp. 2373–2379.

Rabiner, L. R. (1989). "A tutorial on hidden Markov models and selected applications in speech recognition". In: *Proceedings of the IEEE* 77.2, pp. 257–286.

Reiche, J., S. de Bruin, D. Hoekman, J. Verbesselt, and M. Herold (2015). "A Bayesian approach to combine Landsat and ALOS PALSAR time series for near real-time deforestation detection". In: *Remote Sensing* 7.5, pp. 4973–4996.

Reiche, J., E. Hamunyela, J. Verbesselt, D. Hoekman, and M. Herold (2018). "Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2". In: *Remote Sensing of Environment* 204, pp. 147–161.

Reiche, J., J. Verbesselt, D. Hoekman, and M. Herold (2015). "Fusing Landsat and SAR time series to detect deforestation in the tropics". In: *Remote Sensing of Environment* 156, pp. 276–293.

Reiche, J., R. Lucas, A. L. Mitchell, J. Verbesselt, D. H. Hoekman, J. Haarpaintner, J. M. Kellndorfer, A. Rosenqvist, E. A. Lehmann, C. E. Woodcock, et al. (2016). "Combining satellite data for better tropical forest monitoring". In: *Nature Climate Change* 6.2, p. 120.

Rosen, P. A., Y. Kim, R. Kumar, T. Misra, R. Bhan, and V. R. Sagi (2017). "Global persistent SAR sampling with the NASA-ISRO SAR (NISAR) mission". In: *Radar Conference (RadarConf), 2017 IEEE*. IEEE, pp. 0410–0414.

Rosen, P., S. Hensley, S. Shaffer, W. Edelstein, Y. Kim, R. Kumar, T. Misra, R. Bhan, and R. Sagi (2017). "The NASA-ISRO SAR (NISAR) mission dual-band radar instrument preliminary design". In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, pp. 3832–3835.

Rosenqvist, A., M. Shimada, S. Suzuki, F. Ohgushi, T. Tadono, M. Watanabe, K. Tsuzuku, T. Watanabe, S. Kamijo, and E. Aoki (2014). "Operational performance

of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2". In: *Remote Sensing of Environment* 155, pp. 3–12.

Roy, D. P., V. Kovalskyy, H. Zhang, E. F. Vermote, L. Yan, S. Kumar, and A. Egorov (2016). "Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity". In: *Remote Sensing of Environment* 185, pp. 57–70.

Roy, D. P., H. Zhang, J. Ju, J. L. Gomez-Dans, P. E. Lewis, C. Schaaf, Q. Sun, J. Li, H. Huang, and V. Kovalskyy (2016). "A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance". In: *Remote Sensing of Environment* 176, pp. 255–271.

Roy, D. P., M. Wulder, T. R. Loveland, C. Woodcock, R. Allen, M. Anderson, D. Helder, J. Irons, D. Johnson, R. Kennedy, et al. (2014). "Landsat-8: Science and product vision for terrestrial global change research". In: *Remote sensing of Environment* 145, pp. 154–172.

Salberg, A.-B. and Ø. D. Trier (2011). "Temporal analysis of forest cover using hidden Markov models". In: *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*. IEEE, pp. 2322–2325.

Samanta, A., S. Ganguly, H. Hashimoto, S. Devadiga, E. Vermote, Y. Knyazikhin, R. R. Nemani, and R. B. Myneni (2010). "Amazon forests did not green-up during the 2005 drought". In: *Geophysical research letters* 37.5.

Samanta, A., S. Ganguly, E. Vermote, R. R. Nemani, and R. B. Myneni (2012). "Why is remote sensing of amazon forest greenness so challenging?" In: *Earth Interactions* 16.7, pp. 1–14.

Schneider, A. (2012). "Monitoring land cover change in urban and peri-urban areas using dense time stacks of Landsat satellite data and a data mining approach". In: *Remote Sensing of Environment* 124, pp. 689–704.

Schneider, A., M. A. Friedl, D. K. McIver, and C. E. Woodcock (2003). "Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data". In: *Photogrammetric Engineering & Remote Sensing* 69.12, pp. 1377–1386.

Seddon, A. W., M. Macias-Fauria, P. R. Long, D. Benz, and K. J. Willis (2016). "Sensitivity of global terrestrial ecosystems to climate variability". In: *Nature* 531.7593, p. 229.

Sexton, J. O., X.-P. Song, M. Feng, P. Noojipady, A. Anand, C. Huang, D.-H. Kim, K. M. Collins, S. Channan, C. DiMiceli, et al. (2013). "Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error". In: *International Journal of Digital Earth* 6.5, pp. 427–448.

Sharma, P., J. R. Doubleday, and S. Shaffer (2018). "Instrument commissioning timeline for NASA-ISRO Synthetic Aperture Radar (NISAR)". In: *2018 IEEE Aerospace Conference*. IEEE, pp. 1–13.

Shen, M., Y. Tang, J. Chen, X. Zhu, and Y. Zheng (2011). "Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau". In: *Agricultural and Forest Meteorology* 151.12, pp. 1711–1722.

Shewhart, W. A. (1931). *Economic control of quality of manufactured product*. ASQ Quality Press.

Shimada, M., T. Itoh, T. Motooka, M. Watanabe, T. Shiraishi, R. Thapa, and R. Lucas (2014). "New global forest/non-forest maps from ALOS PALSAR data (2007–2010)". In: *Remote Sensing of Environment* 155, pp. 13–31.

Silverman, B. (1986). "Density Estimation for Statistics and Data Analysis". In:

Simard, M., S. S. Saatchi, and G. De Grandi (2000). "The use of decision tree and multiscale texture for classification of JERS-1 SAR data over tropical forest". In: *IEEE Transactions on Geoscience and Remote Sensing* 38.5, pp. 2310–2321.

Solberg, A. H. S., T. Taxt, and A. K. Jain (1996). "A Markov random field model for classification of multisource satellite imagery". In: *IEEE transactions on geoscience and remote sensing* 34.1, pp. 100–113.

Solberg, R., R. B. Huseby, H. Koren, and E. Malnes (2008). "Time-series fusion of optical and SAR data for snow cover area mapping". In: *Proceedings of the EARSeL Land Ice and Snow Special Interest Group Workshop, Berne, Switzerland*. Vol. 2224. Citeseer.

Song, C., C. E. Woodcock, K. C. Seto, M. P. Lenney, and S. A. Macomber (2001). "Classification and change detection using Landsat TM data: when and how to correct atmospheric effects?" In: *Remote sensing of Environment* 75.2, pp. 230–244.

Sotelo Montes, C. and J. C. Weber (1997). "Priorización de especies arbóreas para sistemas agroforestales en la selva baja del Perú". In: *Agroforestería en las Américas* 4.14. http://repositorio.bibliotecaorton.catie.ac.cr/handle/11554/6701.

Souza Jr, C. M., D. A. Roberts, and M. A. Cochrane (2005). "Combining spectral and spatial information to map canopy damage from selective logging and forest fires". In: *Remote Sensing of Environment* 98.2-3, pp. 329–343.

Storey, J., M. Choate, and K. Lee (2014). "Landsat 8 Operational Land Imager on-orbit geometric calibration and performance". In: *Remote sensing* 6.11, pp. 11127–11152.

Sulla-Menashe, D., M. A. Friedl, and C. E. Woodcock (2016). "Sources of bias and variability in long-term Landsat time series over Canadian boreal forests". In: *Remote sensing of environment* 177, pp. 206–219.

Sulla-Menashe, D., C. E. Woodcock, and M. A. Friedl (2018). "Canadian boreal forest greening and browning trends: an analysis of biogeographic patterns and the relative roles of disturbance versus climate drivers". In: *Environmental Research Letters* 13.1, p. 014007.

Tang, X. (2018). "Near Real-Time Monitoring of Forest Disturbance: A Multi-Sensor Remote Sensing Approach and Assessment Framework". PhD thesis. Boston University.

Thapa, R. B., M. Watanabe, T. Motohka, and M. Shimada (2015). "Potential of high-resolution ALOS–PALSAR mosaic texture for aboveground forest carbon tracking in tropical region". In: *Remote Sensing of Environment* 160, pp. 122–133.

Thomas, N., R. Lucas, T. Itoh, M. Simard, L. Fatoyinbo, P. Bunting, and A. Rosenqvist (2015). "An approach to monitoring mangrove extents through time-series comparison of JERS-1 SAR and ALOS PALSAR data". In: *Wetlands ecology and management* 23.1, pp. 3–17.

Thompson, J. R., J. S. Plisinski, P. Olofsson, C. E. Holden, and M. J. Duveneck (2017). "Forest loss in New England: A projection of recent trends". In: *PloS one* 12.12, e0189636.

Tibshirani, R. (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

Torres, R., P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, et al. (2012). "GMES Sentinel-1 mission". In: *Remote Sensing of Environment* 120, pp. 9–24.

Trier, Ø. D. and A.-B. Salberg (2011). "Time-series analysis of satellite images for forest cover change monitoring in Tanzania". In: *1st EARSeL Workshop on Operational Remote Sensing in Forest Management, Prague, Czech Republic.* Citeseer, pp. 2–3.

Tucker, C. J. (1979). "Red and photographic infrared linear combinations for monitoring vegetation". In: *Remote sensing of Environment* 8.2, pp. 127–150.

Urquiza Muñoz, J. D. and M. Burga Ríos (2016). "Incremento de la deforestación y sus consecuencias en la pérdida de biomasa en los bosques de la región Loreto, 2000-2014". http://repositorio.unapiquitos.edu.pe/handle/UNAP/5002. MA thesis. Universidad Nacional de la Amazonía Peruana. Escuela de Posgrado.

U.S. Geological Survey (2015). "PRODUCT GUIDE: Landsat Surface Reflectance products courtesy of the U.S. Geological Survey". In: pp. 1–27.

Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor (2010). "Detecting trend and seasonal changes in satellite image time series". In: *Remote sensing of Environment* 114.1, pp. 106–115.

Verbesselt, J., R. Hyndman, A. Zeileis, and D. Culvenor (2010). "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series". In: *Remote Sensing of Environment* 114.12, pp. 2970–2980.

Verbesselt, J., A. Zeileis, and M. Herold (2012). "Near real-time disturbance detection using satellite image time series". In: *Remote Sensing of Environment* 123, pp. 98–108.

Vermote, E., C. Justice, M. Claverie, and B. Franch (2016). "Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product". In: *Remote Sensing of Environment* 185, pp. 46–56.

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno (2010). "A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index". In: *Journal of climate* 23.7, pp. 1696–1718.

Viena Vela, H. (2011). "Estimación de la biomasa y almacenamiento de carbono bajo cuatro sistemas de uso de la tierra en la Provincia de Leoncio Prado."

http://repositorio.unas.edu.pe/handle/UNAS/108. Bachelor's Thesis. Universidad Nacional Agraria de la Selva. Facultad de Agronomía.

Vogeler, J. C., J. D. Braaten, R. A. Slesak, and M. J. Falkowski (2018). "Extracting the full value of the Landsat archive: Inter-sensor harmonization for the mapping of Minnesota forest canopy cover (1973–2015)". In: *Remote Sensing of Environment* 209, pp. 363–374.

Vogelmann, J. E., G. Xian, C. Homer, and B. Tolk (2012). "Monitoring gradual ecosystem change using Landsat time series analyses: Case studies in selected forest and rangeland ecosystems". In: *Remote Sensing of Environment* 122, pp. 92–105.

Walker, W. S., C. M. Stickler, J. M. Kellndorfer, K. M. Kirsch, and D. C. Nepstad (2010). "Large-area classification and mapping of forest and land cover in the Brazilian Amazon: A comparative analysis of ALOS/PALSAR and Landsat data sources". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3.4, pp. 594–604.

Welty, J., D. Pilliod, and R. Arkle (2017). *Combined wildfire dataset for the United States and certain territories, 1870–2015: US Geological Survey data release.* DOI: 10.5066/F75H7F5M. URL: https://doi.org/10.5066/F75H7F5M.

White, J., M. Wulder, G. Hobart, J. Luther, T. Hermosilla, P. Griffiths, N. Coops, R. Hall, P. Hostert, A. Dyk, et al. (2014). "Pixel-based image compositing for large-area dense time series applications and science". In: *Canadian Journal of Remote Sensing* 40.3, pp. 192–212.

Wulder, M. A., J. G. Masek, W. B. Cohen, T. R. Loveland, and C. E. Woodcock (2012). "Opening the archive: How free data has enabled the science and monitoring promise of Landsat". In: *Remote Sensing of Environment* 122, pp. 2–10.

Wulder, M. A., J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, W. B. Cohen, E. A. Fosnight, J. Shaw, J. G. Masek, and D. P. Roy (2016). "The global Landsat archive: Status, consolidation, and direction". In: *Remote Sensing of Environment* 185, pp. 271–283.

Xin, Q., M. Broich, P. Zhu, and P. Gong (2015). "Modeling grassland spring onset across the Western United States using climate variables and MODIS-derived phenology metrics". In: *Remote Sensing of Environment* 161, pp. 63–77.

Xin, Q., P. Olofsson, Z. Zhu, B. Tan, and C. E. Woodcock (2013). "Toward near real-time monitoring of forest disturbance by fusion of MODIS and Landsat data". In: *Remote Sensing of Environment* 135, pp. 234–247.

Xu, L., A. Samanta, M. H. Costa, S. Ganguly, R. R. Nemani, and R. B. Myneni (2011). "Widespread decline in greenness of Amazonian vegetation due to the 2010 drought". In: *Geophysical Research Letters* 38.7.

Yan, L. and D. Roy (2015). "Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction". In: *Remote Sensing of Environment* 158, pp. 478–491.

Yang, Y., M. C. Anderson, F. Gao, B. Wardlow, C. R. Hain, J. A. Otkin, J. Alfieri, Y. Yang, L. Sun, and W. Dulaney (2018). "Field-scale mapping of evaporative stress indicators of crop yield: An application over Mead, NE, USA". In: *Remote Sensing of Environment* 210, pp. 387–402.

Yang, Y., M. C. Anderson, F. Gao, C. R. Hain, K. A. Semmens, W. P. Kustas, A. Noormets, R. H. Wynne, V. A. Thomas, and G. Sun (2017). "Daily Landsat-scale evapotranspiration estimation over a forested landscape in North Carolina, USA, using multi-satellite data fusion". In: *Hydrology and Earth System Sciences* 21.2, pp. 1017–1037.

Zadrozny, B. and C. Elkan (2001). "Obtaining calibrated probability estimates from decision trees and naive Bayesian classifiers". In: *Icml*. Vol. 1. Citeseer, pp. 609–616.

Zeileis, A. (2005). "A unified approach to structural change tests based on ML scores, F statistics, and OLS residuals". In: *Econometric Reviews* 24.4, pp. 445–466.

Zeileis, A., C. Kleiber, W. Krämer, and K. Hornik (2003). "Testing and dating of structural changes in practice". In: *Computational Statistics & Data Analysis* 44.1-2, pp. 109–123.

Zhang, L., L. Zhang, and B. Du (2016). "Deep learning for remote sensing data: A technical tutorial on the state of the art". In: *IEEE Geoscience and Remote Sensing Magazine* 4.2, pp. 22–40.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. Hodges, F. Gao, B. C. Reed, and A. Huete (2003). "Monitoring vegetation phenology using MODIS". In: *Remote sensing of environment* 84.3, pp. 471–475.

Zhang, X., M. A. Friedl, C. B. Schaaf, A. H. Strahler, and Z. Liu (2005). "Monitoring the response of vegetation phenology to precipitation in Africa by coupling MODIS and TRMM instruments". In: *Journal of Geophysical Research: Atmospheres* 110.D12.

Zhao, W. and S. Du (2016). "Learning multiscale and deep representations for classifying remotely sensed imagery". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 113, pp. 155–165.

Zhu, Z. (2017). "Change detection using landsat time series: A review of frequencies, preprocessing, algorithms, and applications". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 130, pp. 370–384.

Zhu, Z., Y. Fu, C. E. Woodcock, P. Olofsson, J. E. Vogelmann, C. Holden, M. Wang, S. Dai, and Y. Yu (2016). "Including land cover change in analysis of greenness trends using all available Landsat 5, 7, and 8 images: A case study from Guangzhou, China (2000–2014)". In: *Remote sensing of environment* 185, pp. 243–257.

Zhu, Z., S. Wang, and C. E. Woodcock (2015). "Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images". In: *Remote Sensing of Environment* 159, pp. 269–277.

Zhu, Z. and C. E. Woodcock (2012). "Object-based cloud and cloud shadow detection in Landsat imagery". In: *Remote sensing of environment* 118, pp. 83–94.

Zhu, Z. and C. E. Woodcock (2014). "Continuous change detection and classification of land cover using all available Landsat data". In: *Remote sensing of Environment* 144, pp. 152–171.

Zhu, Z., C. E. Woodcock, C. Holden, and Z. Yang (2015). "Generating synthetic Landsat images based on all available Landsat data: Predicting Landsat surface reflectance at any given time". In: *Remote Sensing of Environment* 162, pp. 67–83.

Zhu, Z., C. E. Woodcock, and P. Olofsson (2012). "Continuous monitoring of forest disturbance using all available Landsat imagery". In: *Remote Sensing of Environment* 122, pp. 75–91.

Zhu, Z., C. E. Woodcock, J. Rogan, and J. Kellndorfer (2012). "Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data". In: *Remote Sensing of Environment* 117, pp. 72–82.