USING LANDSAT AND A BAYESIAN HARD CLASSIFIER TO STUDY FOREST CHANGE IN THE SALMON CREEK WATERSHED AREA FROM 1972–2013

By

David Stone Mullis

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

June 2014

DEDICATION

I dedicate this document to Spencer and Kaison, who must be tired of hearing Daddy talk about homework all the time. You mean the world to me.

ACKNOWLEDGEMENTS

I would like to acknowledge Dr. Travis Longcore for his patience and guidance during this difficult time. Dr. Tarek Rashed and Dr. Su Jin Lee have also provided valuable critiques which are very much appreciated.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CWD- Coarse Woody Debris

DoC- Department of Commerce

ERTS- Earth Resources Technology Satellite

ETM+- Enhanced Thematic Mapper Plus

GIS- Geographic Information Science

KIA- Kappa Index of Agreement

LDCM- Landsat Data Continuity Mission

MSS- Multispectral Scanner

NASA- National Aeronautic and Space Administration

NDVI- Normalized Difference Vegetation Index

NOAA- National Oceanic and Atmospheric Administration

OLI- Operational Land Imager

PCA- Principal Component Analysis

TIRS- Thermal Infrared Sensor

TDRSS- Tracking and Data Relay Satellite System

TM- Thematic Mapper

USGS- United States Geological Survey

UTM- Universal Transverse Mercator

VCT- Vegetation Change Tracker

VI- Vegetation Index

Abstract

The Salmon Creek Watershed in Sonoma County, California, USA, is home to a variety of wildlife, and many of its residents are mindful of their place in its ecology. In the past half century, several of its native and rare species have become threatened, endangered, or extinct, most notably the once common Coho salmon and Chinook salmon. The cause of this decline is believed to be a combination of global climate change, local land use, and land cover change. More specifically, the clearing of forested land to create vineyards, as well as other agricultural and residential uses, has led to a decline in biodiversity and habitat structure. I studied sub-scenes of Landsat data from 1972 to 2013 for the Salmon Creek Watershed area to estimate forest cover over this period. I used a maximum likelihood hard classifier to determine forest area, a Mahalanobis distance soft classifier to show the software's uncertainty in classification, and manually digitized forest cover to test and compare results for the 2013 30 m image. Because the earliest images were lower spatial resolution, I also tested the effects of resolution on these statistics. The images before 1985 are at 60 m spatial resolution while the later images are at 30 m resolution. Each image was processed individually and the training data were based on knowledge of the area and a mosaic of aerial photography. Each sub-scene was classified into five categories: water, forest, pasture, vineyard/orchard, and developed/barren. The research shows a decline in forest area from 1972 to around the mid-1990s, then an increase in forest area from the mid-1990s to present. The forest statistics can be helpful for conservation and restoration purposes, while the study on resolution can be helpful for landscape analysis on many levels.

Chapter 1: Introduction

Ecological processes have been constrained due to anthropogenic causes worldwide, leading to a decline in biodiversity and extinction of many species. Landsat satellite data are freely available, and with data from 1972 to present, they are commonly used to monitor surface conditions including forest change and ecosystem attributes. Several species residing in Sonoma County have been classified as rare, threatened, or endangered either by the Federal Endangered Species Act or by the California Endangered Species Act (California Department of Fish and Wildlife 2014). These species include the Marbled Murrelet, Baker's Larkspur, the California Red-legged Frog, the California Tiger Salamander, Yellow Larkspur, Chinook salmon, Coho salmon, and Steelhead trout, each of which is sensitive to changes in land cover, especially forest. I studied the forest area in the Salmon Creek Watershed and the surrounding areas using Landsat satellite data dating from 1972 to the present to determine forest area and the places of change over the years. I used the maximum likelihood hard classifier to determine forest area, a Mahalanobis distance soft classifier to determine uncertainty in classification, and I manually digitized the forest area for the 2013 image for comparison with the 2013 maximum likelihood forest classification.

The Salmon Creek Watershed covers 56.3 km² of the 323.9 km² study area in southwestern Sonoma County and contains critical habitats for several species, but is often overlooked in research and studies for larger adjacent river systems (i.e., the adjacent Russian River Watershed), or is combined with other watersheds for analysis. For instance, the Department of Commerce (DoC) and the National Oceanic and Atmospheric Administration (NOAA) combine the Salmon Creek, Estero Americano, and adjoining small watersheds into

Bodega Hydraulic Unit 1115 when it designated critical habitats for salmon and steelhead in California (DoC and NOAA 2005).

The Salmon Creek Watershed retains unique characteristics from either of the adjoining watersheds. The Estero Americano Watershed is mostly rolling grassy hills and does not have the elevation differential of Salmon Creek. The Russian River Watershed covers a vastly larger area and supplies much of the region with agricultural and drinking water, as well as having dams and water pumped to it from the Eel River. The Salmon Creek Water Conservation Plan was developed in a collaborative effort primarily to restore historic late-spring and summer surface flows in the watershed (Prunuske Chatham, Inc. 2010). The species highlighted in the study are primarily aquatic such as Coho and Chinook salmon, Steelhead trout, tidewater goby, California red-legged frog, California freshwater shrimp, and the northwestern pond turtle. Contributing to the essential environments of these species is complexity in the stream environment, sufficient riparian cover, and strong groundwater retention leading to stronger summer flows. Forests provide all of these in riparian zones, as well as nesting and rearing habitats (Bragg and Kershner 1999) and in the broadest sense the entirety of the Salmon Creek Watershed can be considered riparian. The watershed and landscape level are the best for local salmon restoration projects (Feist et al. 2003), and little research has been focused solely on the Salmon Creek Watershed, although the extreme decline in salmon counts over the years has shown that the watershed needs immediate attention. This study was designed to focus on the Salmon Creek Watershed and the surrounding areas in order to fill in some gaps for researchers of the area.

Conservation efforts are often focused on areas containing the most biodiversity (Wiens et al. 2009) and landscape connectivity has been shown to strengthen species richness

(Damschen and Brudvig 2012). Of primary concern to conservationists in the Salmon Creek Watershed are habitats for endangered Chinook and Coho salmon and Steelhead trout, who are reared in freshwater inland streams before making their way to sea, finally coming back to their original nesting stream to spawn and die. The habitat for these anadromous fish has been shown to be characterized by cooler temperatures, riparian cover, and coarse woody debris (CWD), and it has been degraded by anthropogenic disturbances such as roads, clear-cutting, and extensive urban and residential development. It is estimated that one-third of all Pacific Salmon populations have become extinct (Gustafson et al. 2007). In turn, riparian forests provide connectivity and many benefits that regulate stream and biotic processes by normalizing light and temperature fluctuations, providing nourishment, providing physical structural complexity, filtering sediment, reducing runoff, providing links between aquatic and terrestrial habitats, regulating stream flow, and benefitting overall water quality (Naiman, Décamps, and Pollock 1993; Naiman and Décamps 1997; Robinson, Tockner, and Ward 2002; Tomer et al. 2008; and Atkinson, Hunter, and English 2010).

Complexity refers to variations in the arrangement and makeup of physical structures allowing for greater biodiversity a more stable ecosystem (Kovalenko, Thomaz, and Warfe 2012). A complex habitat has impediments of sight and movement, variance in type, size, and arrangement of objects, and is characterized by a diversity of flora and fauna. Old-growth forests tend to be physically and biologically complex environments, whereas clear-cutting and mono-crop agriculture are examples of practices creating a simplistic environment not capable of sustaining the needs of multiple diverse creatures. In streams, complexity includes stream morphology and changes in depth, which are increased with adequate amounts of CWD provided by healthy forests.

CWD is the large dead wood which provides shelter and sustenance for a diverse biology and adds complexity to waterway morphology (Bragg and Kershner 1999). CWD is strongly associated with polypore communities (Edman, Kruys, and Jonsson 2004), which in turn are associated with strong ecosystem health (Lonsdale, Pautasso, and Holdenrieder 2008). The volume of CWD is often used as an indicator for forest biodiversity (Rondeux and Sanchez 2010; Lassauce et al. 2011) and is positively correlated with fungal species richness (Edman, Kruys, and Jonsson 2004; Lonsdale, Pautasso, and Holdenrieder 2008). The highest priority sites for reforestation are more likely to be near riparian or large forested areas (Holzmueller, Gaskins, and Mangun 2011). Parcel conservation adjacent to nature reserves can reduce fragmentation and help preserve environmental processes (Rissman et al. 2006). It is generally concluded that sites rich in CWD are of primary conservation interest and that restoration projects would be most effective in sites in the riparian corridor adjacent to CWD-rich areas. CWD-rich areas correspond strongly with aged forests, so much so that modern logging practice involves leaving certain amounts of CWD behind to help maintain forest health (Lonsdale, Pautasso, and Holdenrieder 2008; Junninen and Komonen 2010), and the Normalized Difference Vegetation Index (NDVI) was shown to be the best predictor of CWD after a disturbance (White, McGill, and Lechowicz 2012). The chronological as well as spatial continuity of deadwood is important for ecological processes and is associated with riparian and mature forest areas (Lonsdale, Pautasso, and Holdenrieder 2008).

In a recent study (Pflugmacher, Cohen, and Kennedy 2012), Landsat time stacks were used to determine the degree and frequency of canopy disturbance, the changes associated with the death of trees, to estimate CWD. The time stacks were shown to function almost as well as the spatially and temporally limited LiDAR data sets at estimating forest structure, and were

better at estimating CWD. LiDAR is available for only a small portion of Earth's surface, and it is expensive to obtain new images, making Landsat time series a feasible alternative in areas lacking LiDAR data, such as most of the Salmon Creek Watershed.

Habitat loss is a known danger to biological and genetic diversity worldwide and preserving parcels is the main conservation strategy used worldwide for reduction in biodiversity loss (Mönkkönen et al. 2010). Coastal California counties near major urban centers are considered highly vulnerable areas for many sensitive species due to the rapid encroachment of human habitats in the ecosystem (Newburn, Berck, and Merenlender, 2006). Landscape fragmentation from anthropogenic uses (deforestation, urban growth, agriculture) is a major cause of habitat degradation, as it restricts the movement of many species and reduces habitat complexity associated with species richness and biodiversity (Malanson and Cramer 1999; Kovalenko, Thomaz, and Warfe 2012). Due to extensive clear-cutting, lands have been converted to agricultural and residential uses, leading to declines in biodiversity and habitat structure.

Forests are essential for maintaining ecological balance through providing habitat structure and sustenance for many flora and fauna, essential detritus and debris, and by sequestering carbon from the air and contaminants from the water and soil. Some species depend on spatial and temporal continuity of forests, in particular some specialized wood-decaying species of fungus (Hottola, Ovaskainen, and Hanski, 2009) and coho salmon (Andrew and Wulder 2011). Another local example is the Marbled Murrelet, which is a seabird that was discovered to have its nesting grounds in old-growth redwood forests extending as far as 80 km inland. Trees provide erosion control, as well, which prevents harmful sediment from contaminating streams and adds to stream geomorphology. Demand for timber and conversion

of land for residential and agricultural purposes has led to certain types of disasters worldwide, including landslides and flooding, leading forest conservation to be the main approach to reduce disasters related to the atmosphere, water, land, and biodiversity (Bhagat 2009).

Remote sensing can be an effective and efficient way to observe and analyze forest characteristics, and Landsat data has been used in 42 percent of landscape studies in a recent survey of the journal *Landscape Ecology* (Newburn, Berck, and Merenlender, 2006). Remotely sensed data has been used commonly in landscape characterization for ecological mapping. Remote sensing allows a quick alternative to ground surveys and other conventional methods for landscape analysis, and often proves to be an affordable tool for land managers. Landsat is perhaps the most widely used remotely sensed data for vegetation and land-cover categorization due to its spatial and temporal continuity, and its availability for free download on the Internet. Spatial resolutions of 60 m for the Multispectral Sensor (MSS) of Landsat missions 1-3 and 30 m for the Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) are sufficient for many landscape management purposes.

Chapter 2: Background

2.1 Landsat History

The first satellite was originally named Earth Resources Technology Satellite 1(ERTS-1) and was the first satellite aimed at studying Earth's land (NASA 2013). It was launched July 23, 1972 and stayed in commission until January 1978. The sensors aboard the satellites were designed to capture specific wavelengths of the electromagnetic spectrum (Figure 1), called bandwidths. The sensors passively detect reflected electromagnetic energy within each bandwidth, allowing greater analysis of the land. Its instruments were the Return Beam Vidicon (RBV) and highly experimental Multispectral Scanner (MSS), and while the RBV was anticipated as the primary tool, the MSS proved more useful. The still-experimental Landsat 2 was launched January 22, 1975 and had the same instruments as ERTS-1, now known as Landsat 1. It lasted seven years and operations were terminated on February 25, 1982 (Tables 1&2).

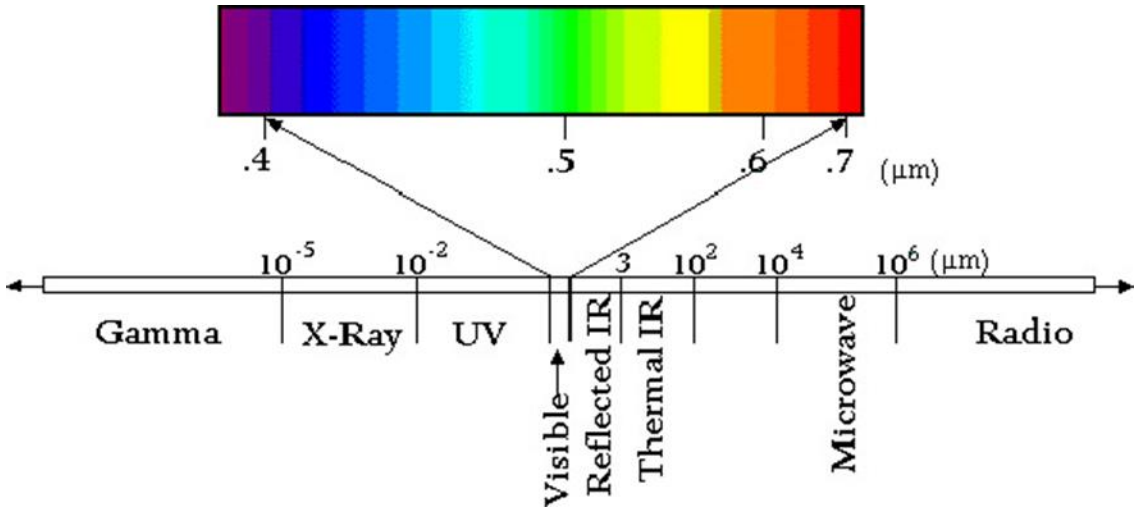

Figure 1. The electromagnetic spectrum with the visible wavelengths highlighted. From http://landsat.usgs.gov/images/Spectral_Bands.jpg**.**

Landsat 3 launched March 5, 1978 and had the RBV and MSS, as well. The RBV was changed so that it had greater spatial resolution and captured a broad spectral range, while the

MSS had an added thermal band, although the thermal infrared band malfunctioned after liftoff. Landsat 3 continued for five years before being decommissioned September 7, 1983.

Landsat 4 launched July 16, 1982, and while still having the MSS instrument, it was significantly different from the previous ones in the addition of the Thematic Mapper (TM), which replaced the RBV. The TM had finer spatial resolution and effectively had three more bands gathering data from a wider range of bandwidths (a thermal infrared band and two mid-infrared bands were added). Landsat 4 suffered technical damages less than a year into its mission and stopped gathering data in 1993, but was kept in orbit for organizational purposes until it was decommissioned in 2001.

The Landsat 5 satellite and instruments were assembled together with Landsat 4 to the same specifications, and it was launched less than two years later on March 1, 1984. Three years after launch, the Tracking and Data Relay Satellite System (TDRSS) Ku-transmitter failed, leaving Landsat 4 to downlink international data until 1993. The MSS was turned off in 1995 until 2012, shortly after the TM malfunctioned; however, the satellite was decommissioned later that year because of a broken gyroscope.

Landsat 6 was launched October 5, 1993, but never reached orbit. It had a new instrument, the Enhanced Thematic Mapper (ETM), which contained a panchromatic band at 15 m resolution for sharpening images. Had it not been for Landsat missions 4–5 outlasting their designed life-span, we would have data gaps that Landsat 6 was designed to fill, and this thesis would perhaps be one comparing data from two data eras.

Landsat 7 launched on April 15, 1999, carrying the ETM+ instrument which had a new panchromatic band at 15 m resolution and a 60 m resolution thermal infrared band, as well as improvements in data recording and radiometric calibration. The satellite experienced a

hardware failure in the scan line corrector (SLC) in May 2003, which gives the data wedge-shaped areas of missing pixels totaling about 22% of the entire image. The satellite continues to function properly otherwise and its highly accurate calibration makes Landsat 7 the go-to satellite for validating data from other sensors. Landsat 7 data were originally offered at $600 per scene, a reduced price from previous missions, until the USGS made all Landsat 7 data free for public download in October 2008. The demand for data downloads led the USGS to offer all Landsat data for free in January of the next year, leading to an explosion in data downloads.

The eighth Landsat satellite was launched February 11, 2013 and called the Landsat Data Continuity Mission (LDCM) until the USGS took over operations on May 30, 2013, when it became known as Landsat 8. Two new instruments onboard Landsat 8 provide added coverage of the electromagnetic spectrum: the Operational Land Imager (OLI) adds a coastal/aerosol detecting band 1 and a cirrus detecting band 9, while the Thermal Infrared Sensor (TIRS) uses two thermal bands (10 & 11) instead of one. The bandwidths are also slightly adjusted from the ETM+ sensor.

Table 1. Landsat satellites and sensor with associated bandwidths and spatial resolutions.

Satellite	Sensor	Bandwidth (µm)	Resolution (m)	Satellite	Sensor	Bandwidth (µm)	Resolution (m)
Landsat 1-2	**RBV**	(1)0.48-0.57	80	***Landsat 6***	***ETM***	*Bands 1-7 the same as TM*	
		(2)0.58-0.68	80			*(8)0.52-0.90 PAN*	*15*
		(3)0.70-0.83	80	**Landsat 7**	**ETM+**	(1)0.45-0.52	30
	MSS	(4)0.50-0.60	79			(2)0.52-0.60	30
		(5)0.60-0.70	79			(3)0.63-0.69	30
		(6)0.70-0.80	79			(4)0.76-0.90	30
		(7)0.80-1.10	79			(5)1.55-1.75	30
Landsat 3	**RBV**	(1)0.505-0.75	40			(6)10.4-12.5	60
	MSS	(4)0.50-0.60	79			(7)2.08-2.35	30
		(5)0.60-0.70	79			PAN 0.52-0.90	15
		(6)0.70-0.80	79	**Landsat 8**	**OLI**	(1)0.434-0.451	30
		(7)0.80-1.10	79			(2)0.452-0.512	30
		(8)10.4-12.6	240			(3)0.533-0.590	30
Landsat 4-5	**MSS**	(4)0.50-0.60	82			(4)0.636-0.673	30
		(5)0.60-0.70	82			(5)0.851-0.879	30
		(6)0.70-0.80	82			(6)1.566-1.651	30
		(7)0.80-1.10	82			(7)2.107-2.294	30
	TM	(1)0.45-0.52	30			(8)0.503-0.676	15
		(2)0.52-0.60	30			(9)1.363-1.384	30
		(3)0.63-0.69	30		**TIRS**	(10)10.60-11.19	100
		(4)0.76-0.90	30			(11)11.50-12.51	100
		(5)1.55-1.75	30				
		(6)10.4-12.5	120				
		(7)2.08-2.35	30				

Table 2. Landsat mission characteristics.

System	*Launch- Date of Decommission*	*Instrument(s)*	*Altitude (km)*	*Return (days)*
Landsat 1	23 July 1972- 06 January 1978	RBV and MSS	917	18
Landsat 2	22 January 1975- 25 February 1982	RBV and MSS	917	18
Landsat 3	05 March 1978- 31 March 1983	RBV and MSS	917	18
Landsat 4	16 July 1982 - 15 June 2001	MSS and TM	705	16
Landsat 5	01 March 1984- 05 June 2013	MSS and TM	705	16
Landsat 6	05 October 1993- 05 October 1993	ETM	705	16
Landsat 7	15 April 1999- present	ETM+	705	16
Landsat 8	11 February 2013- present	OLI and TIRS	705	16

2.2 Landsat in Forest Mapping

Remote sensing is commonly used for vegetation mapping, having been applied to land cover classification, forest analysis, and change analysis, among many other uses. Since the first satellite was launched in 1972, the data from what would become known as the Landsat missions has been widely used for forest and vegetation mapping. With over 40 years of collecting data methodically, Landsat is the longest-lasting of the Earth monitoring satellite missions and has played a significant role in land management and ecology (Cohen and Goward 2004).

Landsat's multiple bands allow for detection of features such as water, mineral, or vegetation with specific reflectance properties. For instance, the near infrared bands are well-suited for vegetation detection because the "green" chlorophyll absorbs red wavelengths and has its highest reflectance in the near infrared wavelengths (Özyavuz 2010). Band combinations are also used for detection of various phenomena, and certain mathematical and algebraic transformations can produce revealing information including a range of vegetation indices (VIs).

The Normalized Difference Vegetation Index (NDVI) is a widely used VI and is expressed in the formula $\frac{NIR-Red}{NIR+Red}$. For example, the NDVI for the TM and ETM+ would be $\frac{Band4-Band\ 3}{Band\ 4+Band\ 3}$, and would reveal the degree of vegetation present in an area. While bands 4 and 3 of the TM are used for NDVI, bands 4 and 2 of the MSS are preferred.

Landsat digital data can be manipulated in many ways to extract valuable data. Histograms of DNs can narrow down certain ranges of data for display or evaluation. Transformations on DNs can be performed to such indexes as wetness, greenness, or brightness, as in the Tasseled Cap transformation (Healy et al. 2005). Image classification is one of the more common applications of Landsat data, and the most commonly used supervised classification method since the 1970's has been the maximum likelihood classifier due to Gaussian probabilities of each class having a similar spectral pattern (Roberts 2013).

2.3 Change Detection

Because of Landsat's orbital pattern, data for the same place on earth can be obtained every 16-18 days, depending on the satellite (Table 2). As the sensors scan the earth in a "push-broom" fashion, they gather data for entire surface of the earth until the satellite comes back to the same point in just more than two weeks. The data are indexed by path and row, with each path/row combination covering approximately the same geographical area and are given in the Universal Transverse Mercator (UTM) projected coordinate system. This allows for a solid comparison of images from multiple dates.

Before automated change detection methods were introduced, change detection consisted of placing a transparent negative and positive plastic sheet of Landsat images (Crapper and Hynson 1983). In an instance of no change, the two sheets would cancel each other out,

revealing no information; in instances of change, such as images separated by time, the areas of variation will be exposed while unchanged areas remain neutral. Automated evaluations of Landsat time-stacks have become common in remote sensing and have led to a greater understanding of the world at large, and they have led to many advances in climate change detection.

Automated vegetation change detection algorithms are introduced fairly regularly in attempts to improve time and processing constraints, as well as change detection accuracy (e.g. Huang et al. 2010). Many change detection techniques have been used including thresholding (areas of change are given a value of 1, while areas of no change are given a background value of 0), univariate image differencing, image regression, image ratioing (Howarth and Wickware 1981), decision trees (Belward and de Hoyos 1987), vegetation index differencing, Principal Component Analysis (PCA), post-classification comparison, direct multi-date classification, change vector analysis, and image subtraction (see Singh 1989). Newer forest change detection techniques which have proven successful include the vegetation change tracker (VCT) (Huang et al. 2010), and temporal segmentation algorithms (Kennedy, Yang, and Cohen 2010).

Due to its spatial resolution, footprint (area covered by a single image), global coverage, and availability, studies involving forest change using Landsat data often cover larger regions than what this study covers. At regional, continental, and global scales, Landsat provides sufficient accuracy for forest change detection (Huang et al. 2008), and many common uses of Landsat data are at the footprint scale or broader (e.g. Kennedy, Yang, and Cohen 2010; Masek et al. 2013; and Tan et al. 2013). Use of Landsat sub-scenes (scenes within a greater Landsat image) often is improved with ancillary data such as field data, GPS, analytic software, and DEMs (e.g. Mihal, Savulescu, and Sandric 2007).

Change detection using Landsat time stacks has improved to detect trends (e.g. Kennedy, Yang, and Cohen 2010), as well as finding causality for certain phenomena. For instance, forest disturbance in the western United States has been linked to western wildfires using Landsat time series analysis (Masek et al. 2013).

Table 3. Studies in forest change detection and the scale and region studied.

Author(s) - Date	Subject/ Technique	Scale	Study Region
Howarth and Wickware (1981)	Band ratioing and supervised post-classification change detection	Single scene	Peace-Athabasca Delta, northern British Columbia, Canada
Belward and de Hoyos (1987)	Comparison of maximum likelihood with supervised decision tree	Single scene	Farmlingham, Suffolk, UK
Singh (1989)	Review of change detection techniques	N/A	N/A
Mihal, Savulescu, and Sandric (2007)	Evolution of vegetation zones	Sub-scene (<500km²)	Iezer Mountains, Southern Carpathians, Romania
Keleş et al. (2008)	Urbanization and forest cover change- maximum likelihood and fragmentation analytic software	Single scene	Regional Directorate of Trabzon Forestry, NE Turkey
Nelson et al. (2009)	Spatial aggregation and resolution effects on forest area estimation	Two Landsat scenes	Minnesota
Huang et al. (2010)	Vegetation Change Tracker (VCT)	Multiple Landsat scenes	Eastern Virginia and Oregon
Kennedy, Yang, and Cohen (2010)	LandTrendr: temporal segmentation algorithms	Multiple Landsat scenes	Pacific Northwest

Chapter 3: Methods

3.1 Study Area

Sonoma County, California, USA is on the north-central coast of California in the San Francisco Bay area. It is home to a rich diversity of terrains, land cover, and species. According to the Sonoma County Administrator's Office (2011) based on the 2010 US Census, Sonoma County is projected to experience over a 5% population growth by 2015, causing environmentalists to be concerned about urbanization. The Salmon Creek Watershed is a small, predominantly rural, coastal watershed emptying just north of Bodega Bay. Extending less than $1/11^{th}$ of a Landsat image, the study area covers 323.9 km² at a scale of 1:144,841, and using the UTM-10N projected coordinate system it has minimum and maximum *x* coordinates of 491265 and 512685, and minimum and maximum *y* values of 4239195 and 4254315, respectively.

Coastal redwood, various pine, Douglas fir, and California bay as well as deciduous oaks and maple are common trees throughout the study area. Alder, willow, and cedar are also present in the riparian areas. Sonoma County has become known for its wine and dairy products, and apple orchards are seen throughout much of the study area based on ground observation. The coastal regions and the south of the study area are dominated by rolling grassy hills, many of which are being used for grazing by cattle or sheep. Coastal scrub brush is also present, especially on the coastal grassy hills. Five small towns are located in the study area: Bodega, Bodega Bay, Camp Meeker, Freestone, and Occidental.

Figure 2. The Salmon Creek Watershed study area, Sonoma County, California. Image is a window, or sub-scene, of Landsat 8's OLI instrument Path 045, Row 033 from bands 2, 3, and 4, day 289. The Salmon Creek Watershed is outlined in red.

3.2 Data

The Landsat images were downloaded from the USGS EarthExplorer website (http://earthexplorer.usgs.gov/) for WRS1 path 048, row 033 and WRS2 path 045, row 033 starting in 1972. All images were gathered for mid- to late-summer clear days in order to enable easier distinction between grasses, cultivated land, and forested land (Table 3). The images from before 1985 were collected at 80 m spatial resolution and processed and distributed at 60 m resolution; the images since 1985 are at 30 m spatial resolution. Six USGS Digital Orthophoto Quadrangles (DOQs) from 11–13 July 1993 were also downloaded from the USGS EarthExplorer website and are in 1 m spatial resolution. All images are publicly available for free download.

Table 4. Image dates, Landsat mission, and instrument used in analysis of the Salmon Creek Watershed study area. * denotes both days were combined into a single image due to the SLC-off anomaly.

Year	Day	Landsat Mission	Instrument
1972	299	1	MSS
1975	211	1	MSS
1977	155	2	MSS
1980	266	2	MSS
1982	192	3	MSS
1985	276	5	TM
1987	186	5	TM
1990	178	5	TM
1992	248	5	TM
1995	176	5	TM
1997	213	5	TM
2000	230	7	ETM+
2002	187	7	ETM+
2005	267	5	TM
2007	241	5	TM
2010	249	5	TM
2012	231/ 263*	7	ETM+ (SLC-off)
2013	289	8	OLI

3.3 Tools

IDRISI is image processing and geographic information systems software with strong raster processing capabilities. The Selva edition of the software provides many tools for analyzing spatial data, including surface data, statistical analysis, change and time-series exploration, and hard and soft classifiers. Mathematical and logical processes can be applied to raster images to determine relationships between data and for pattern analysis. IDRISI also offers convenient import and export formats compatible with the most common formats. IDRISI Selva tools were used for the examination of the Landsat images, for analyzing the resulting rasters, and for assessing the accuracy of the classified images (Table 5).

Table 5. Tools used in IDRISI Taiga.

Tool	Usage
COMPOSITE	Uses a 3-band combination representing red, green, and blue to create a color composite image. In this case, I used MSS bands 4, 5, and 7, and TM and ETM+ bands 2, 3, and 4 to create a false-color composite highlighting vegetative growth.
EXPAND	Increases image resolution by a given factor.
MAHALCLASS	Creates a new set of signature classes by determining Mahalanobis distance.
MAKESIG	Creates signature files of classes based on the spectral reflectance signatures in the training data used for hard classification.
MOSAIC	Creates an image from overlapping images of the same data type, referencing system, and pixel resolution. The tool offers the option to match image grey levels, creating what looks like a more natural single image. I used this tool for six DOQs covering the study area.
RESAMPLE	This tool is a "rubber sheet" transformation that registers the data of one grid system into that of another when the PROJECT tool is not sufficient. I used the RESAMPLE tool to correct the position of the 1975 image.
WINDOW	Extracts a subset of a larger image by coordinates or row/column positions.

The Image Calculator can generate new images using either a mathematical or logical expression. I used the AREA tool to generate a table of the total area in square kilometers for forests in each classified image. I also used it for comparing images or determining overall trends, as with the resulting images of the MAHALCLASS tool.

The MAXLIKE maximum likelihood classification tool is a Bayesian classifier determining classes based on the probability of each pixel belonging to the predefined class signatures. Bayes' Theorem infers the probability of an occurrence given a set of possible occurrences and is expressed in the following equation for the probability of *x* given *y*:

$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$ where $x \neq y$ and p is the probability of an occurrence.

The MAHALCLASS tool is a soft classifier, one which decides the degree of membership in a particular class of an image. The same signature files are used as those which are generated from the MAKESIG tool to classify all pixels in an image, but contrary to other classifiers, the MAHALCLASS tool generates a series of rasters showing the degree to which each pixel is typical of the training data for each class based on Mahalanobis distance. Mahalanobis distance determines the similarity of each pixel to the training data and is equivalent to *z*-scores (standard deviations from the mean) for multiple variables.

3.4 Image Preparation

Some preprocessing must be performed on the Landsat images before they will be useful for analysis. Atmospheric correction was not used because the training data are at the same scale as the image data and are taken from the images themselves (Song et al. 2001). First, each Landsat image was cropped, band by band, to an area containing the Salmon Creek Watershed study area using the WINDOW tool. One image was found to be incorrectly registered, having coordinates off by as much as 1.5 miles. This was corrected by using the RESAMPLE tool and creating ground control points for reference with another image.

Second, the images from before 1985 are not in the same spatial resolution as the later images. The images must all be in the same resolution before they can be compared. Since Landsat 1, 2, and 3 data are distributed by the USGS in 60 m spatial resolution (although they are gathered in 80 m resolution) and the latter Landsat data are in 30 m resolution, and in order to enable further analysis, the resolution of the latter images was enlarged to 60 m resolution using the EXPAND tool, expanding the image by a factor of two in width and length, combining the

smaller pixels to create an aggregated image. This created two different sets of images for Landsat data obtained at 30 m spatial resolution, 30 m and 60 m resolution datasets.

Using the COMPOSITE tool, a false-color composite of the windowed images was created for each year from bands 4, 5, and 7 for MSS images and bands 2, 3, and 4 for TM, ETM+, and OLI images. Each of the DOQs was pieced together using the MOSAIC tool so that they can be used as a reference for classification of the Landsat images.

3.5 Training Data and Hard Classification

The windowed composite images were used to define sites as training data to digitize vector files of five land cover categories: water, forest, pasture/grassland, vineyard/orchard, and developed/barren. The water in the images has obvious visual color differences, since some of the water pixels are open ocean, some are shallow/murky bay, and some are in streams and estuaries. The forested areas, which are the focus of this study, dominate the mid- and northern regions of the study area. Because the forest class was the focus of interest, chaparral areas were considered as part of the pasture class, opting for a simpler classification structure to determine forested land. The pasture land, including scrub brush, dominates the southern and coastal areas, and seems to fill many of the gaps between forests. Most of the vineyard/orchard land is located in the northeastern quarter of the study area. The developed/barren areas were beaches, bare land, roads, and buildings, and are scattered throughout the scene.

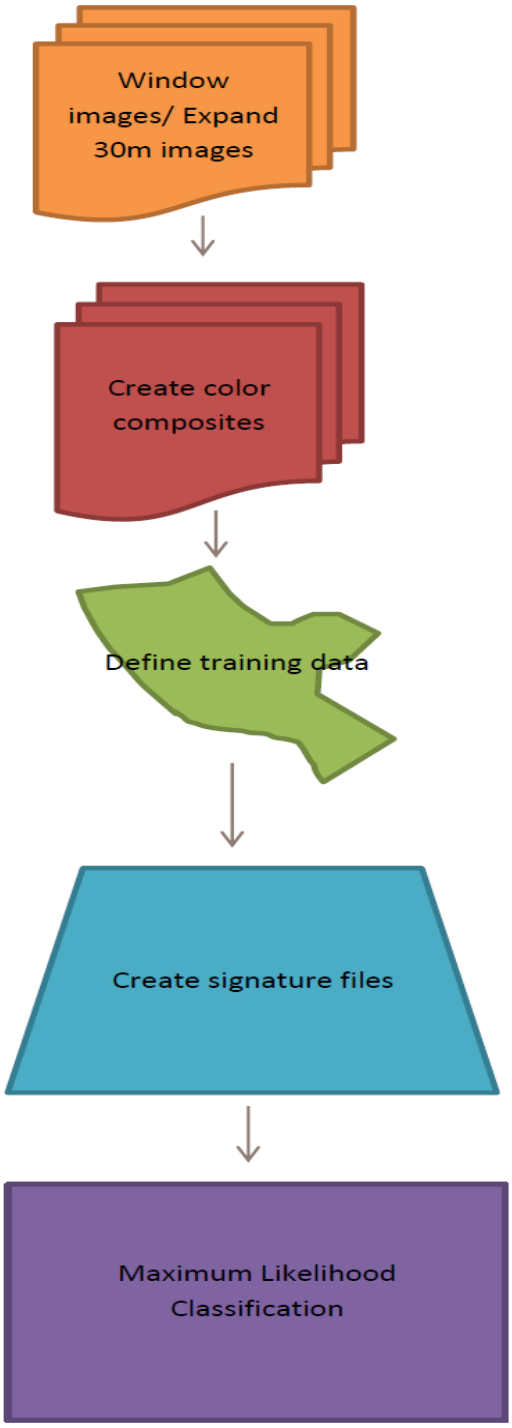

Figure 3. Flowchart for the process of creating an image using the maximum likelihood classifier.

Training data were created by carefully selecting representative samples of the selected land cover for each image. Using new training data for each image is recommended to avoid complications from atmospheric effects (Huang et al. 2008). The DOQ mosaic was used as a reference to find areas with "pure" pixels for each category. Training data were manually digitized using representative areas for each different category. The digital number (DN) associated with each pixel in the image is placed in one of the above categories by the maximum likelihood classifier based on its similarity to the training data DNs. After this comparison, each pixel is placed in the category that it is most likely to fall into. Sometimes there are mixed pixels (pixels containing two or more types of land cover) or otherwise confusing DNs (i.e. cloud cover), which may cause some complications.

Signature files were then created from the vector training data polygons by using the signature development tool MAKESIG, which uses the spectral signatures from the training data applied to the various bands of the Landsat image to extract the spectral values for that land cover type for that particular image. To create a classified image using the maximum likelihood classification method, the signature files were entered as input to the MAXLIKE tool with equal probabilities that a pixel's DN could fall into each class. A color palette file was created and named *LandClassification* using the "Symbol Workshop" tool for consistent visual interpretation.

3.6 Digitization of Forest Area

To compare the automated forest classification with ground truth data, forest area was manually digitized according to the author's best judgment for the 2013 image, and cross-referenced with the DOQ mosaic (Figure 4). This was done by using the Digitize function in IDRISI, creating a vector polygon file. Once the forest area was digitized, the RASTERVECTOR tool was used to transform the vector file to a raster file. The AREA tool was then used to generate a tabular file showing how many square kilometers the digitized forest area covers. While this approach may have the drawback of missing individual pixels, it is a useful tool for showing what the author considers forest area compared to what the software may include. It is one way of checking for and describing the error found in the automated maximum likelihood method.

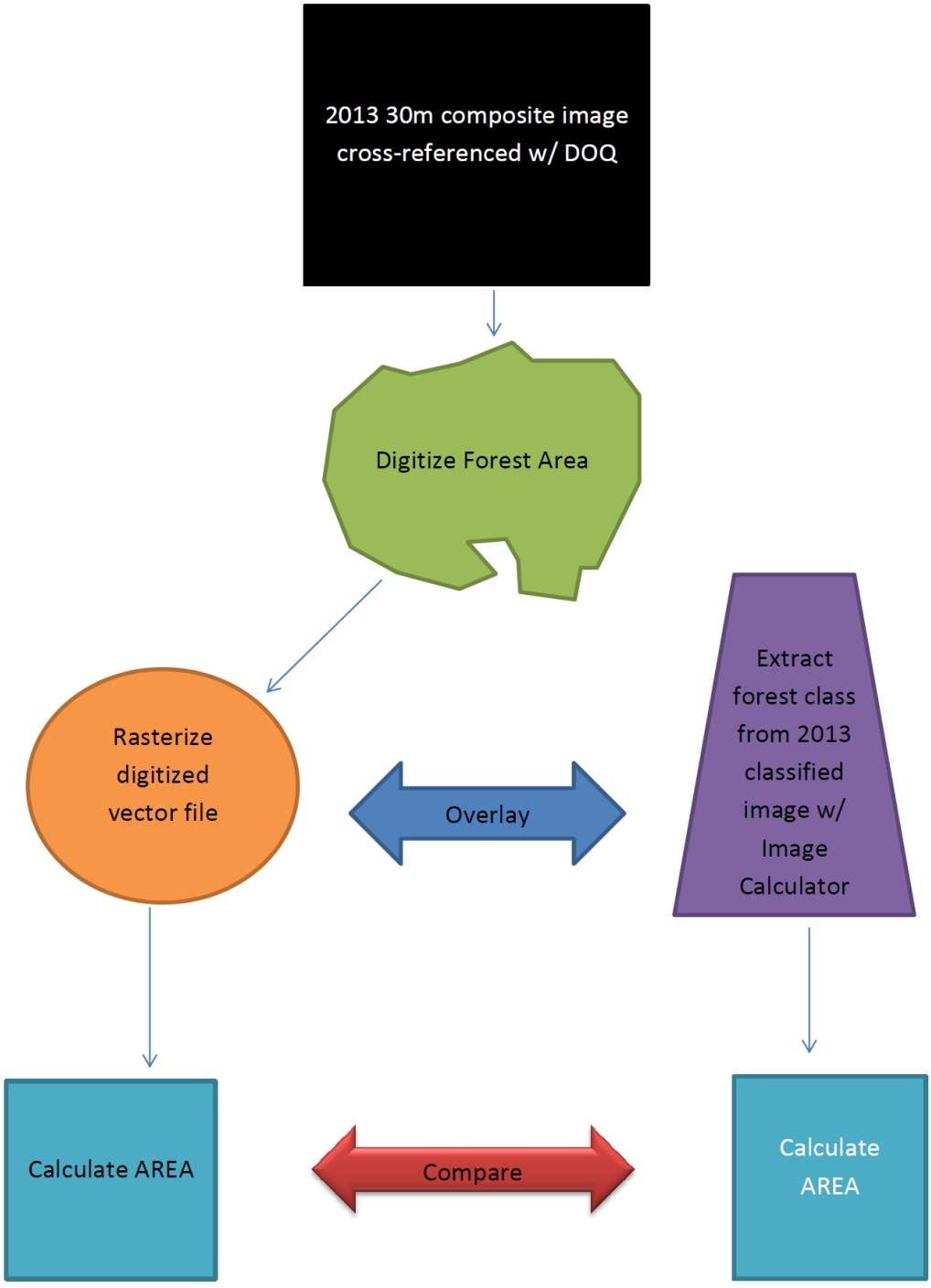

Figure 4. Flowchart for manual digitization of the 2013 30 m resolution composite image and comparison with the 2013 30 m resolution maximum likelihood forest classification.

To compare the manually digitized image with the automated classification, the Image Calculator was used to show only the forest area (pixels in category 2, which is forest) from the 2013 30 m maximum likelihood classified image. The Image Calculator was again used to subtract the automated pixel values from the digitized pixel values, which showed me where there was disagreement between what the author considers forest and what the tool considers forest based on the training data used.

3.7 Soft classification

IDRISI's MAHALCLASS soft classifier was used with the intension of visualizing the uncertainty that the software had in determining classes. The same signature data files were used as for the MAXLIKE tool (Figure 5), and the result is an image showing the degree of uncertainty involved in the classification, both for individual classes and for all classes combined. The MAHALCLASS tool was used for every image and both 30 m and 60 m resolutions. After the images were processed in this manner, the Image Calculator was used to add the combined uncertainty for each year for each resolution. This gave the overall uncertainty involved for each pixel in all categories in each resolution. Because the forest class was the sole focus, the Image Calculator was used to add only the uncertainty from forest classes for each year at both resolutions, which gave the uncertainty through the years involved only in categorizing forest classes at each resolution.

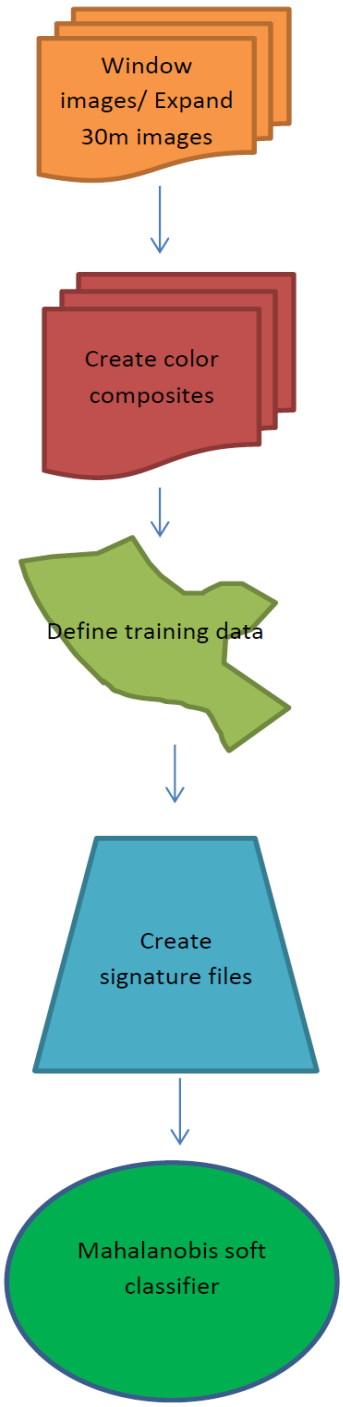

Figure 5. Flowchart for creating Mahalanobis distance soft classification images to visualize uncertainty in classification.

3.8 Error Analysis

Error analysis is an important part of GIS and remote sensing classification. It is a way of statistically determining the extent of error to be expected in the classified images, and therefore, assessing the final classified images' accuracy. Since 300 test points are recommended per scene, and because of time limitations, only six scenes were selected for error assessment. IDRISI has built in to its software an accuracy assessment tool that generates an error matrix, *ERRMAT*, based on manual visual interpretation of a representative sample of points in the study area. The process is laid out in Warner and Campagna, *Ch. 8* (2009), which I summarize here.

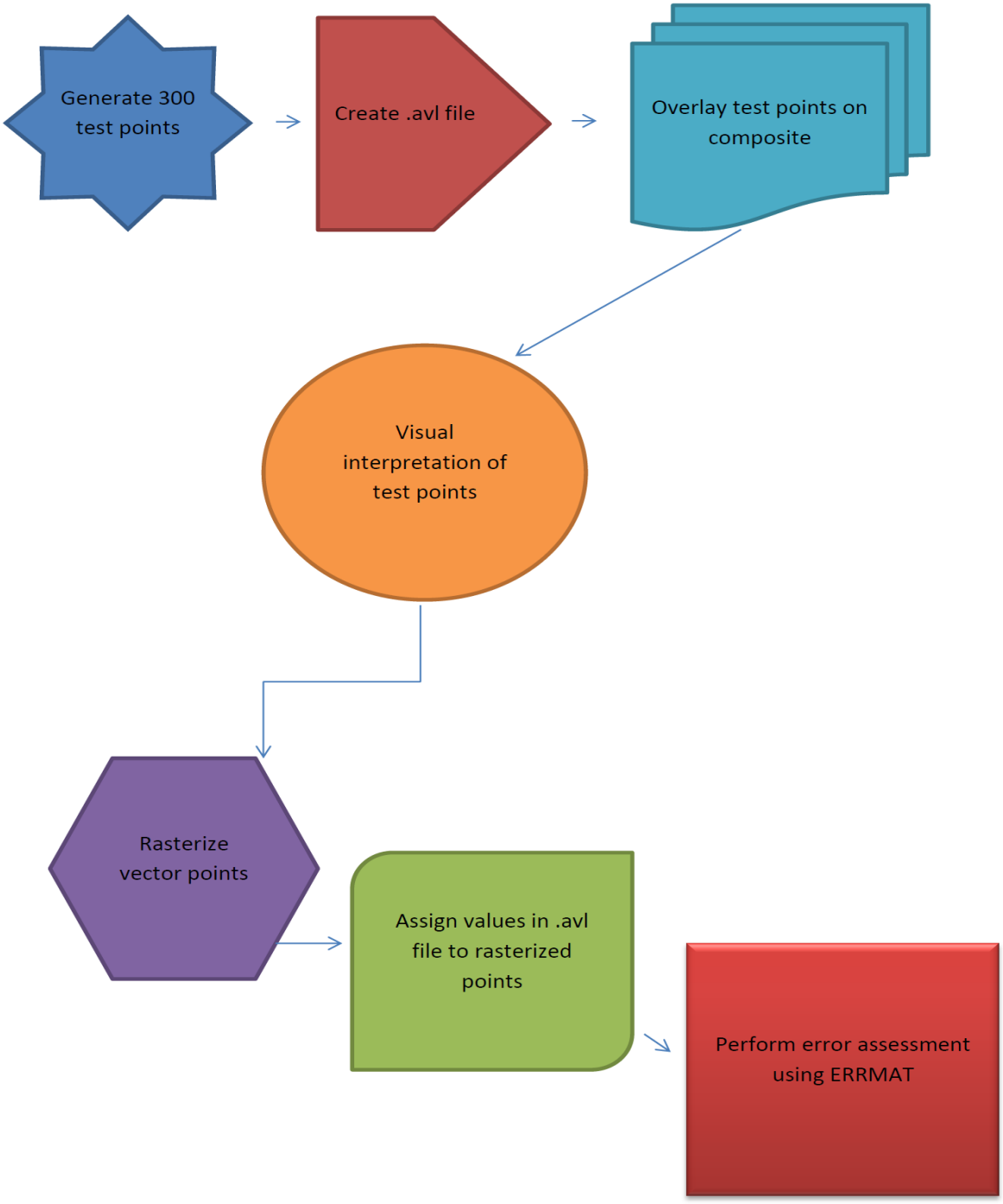

Figure 6. Flowchart for the task of generating an error matrix based on stratified random sampling in IDRISI Selva.

The process is begun by the SAMPLE tool to create a vector file of random test points (Figure 6). The recommended 300 test points per scene were used for stratified random sampling. Stratified random sampling spreads the sample points randomly, while assuring each class is well represented. Then, an *Attribute values file* (.avl) was created using the EDIT tool to match points 1–300 with the class most representative of the point. The random test points are overlaid both on the color composite from which the classified image was derived and on the DOQ of the study area. Each point is systematically visually interpreted and categorized based on a comparison of their place in the DOQ and composite images. This can be a somewhat troublesome task as some pixels are mixed; that is, they represent the reflected radiance of more than one class. For instance, how is a pixel to be assigned when a road lined with small trees on one side and a pasture on the other is present, which can certainly be the case when the resolution is 60 m or 30 m? A common country road surely is only 1/3 of a 60 m pixel at most. In these situations, the *most* representative class was chosen, or a "*0"* was entered to prevent that pixel from being computed in the error matrix. The .avl file was updated with each number, one-300, being assigned the category most representative of the point on the image through visual interpretation. A raster was then created from the vector layer of interpreted sample points using the tool RASTERVECTOR. A "ground-truth" image was created with the updated attribute values file (.avl) using ASSIGN, which assigns photo-interpreted values for the pixels associated with the sample points. Finally, the accuracy of the classified images was calculated with the ERRMAT tool.

The ERRMAT tool creates an error matrix in which the columns represent the ground-truth samples measured against the rows which contain the classifications as generated by the maximum likelihood classification for the corresponding pixels. The error matrix shows errors

of omission and errors of commission for each category in the classified image. At the bottom of each column shown as a proportion are the total errors of omission; that is, the pixels which were assigned to one category in the ground-truth data, and to another category in the classified image. At the end of each row are the errors of commission, or the proportion of pixels assigned incorrectly in the ground-truth data. The diagonal line of numbers from the upper left to lower right represent correctly classified pixels. In the bottom right-hand corner of the matrix is the total classification error, also shown as a proportion. The ERRMAT tool also calculates the overall *Kappa*, which is a measure of difference in agreement between the overall accuracy of a classified image and the amount of agreement one should expect by a random comparison of two maps.

3.9 Study Assumptions

In the process of the study, I made several assumptions that affected its direction. The first assumption was that atmospheric correction is not necessary. The training data are at the same resolution as the images and were chosen from each year's image. Training data from one year and resolution was not used for an image from another year or resolution. This means that an atmospheric condition in one image would not carry over to another image.

Another assumption is that by expanding the 30 m resolution images to match the resolution of the 60 m images, the expanded image would be equivalent to an original 60 m resolution image. The aggregated image inherently involves mixing pixels, and mixed pixels are a common problem in remote sensing. But, by decreasing the resolution of the 30 m images, it was assumed that a basic comparison could be made with the already too coarse earlier images.

Due to high error in preliminary results, it was assumed that a soft classifier would help determine where the problem areas are. Also, the forest area was manually digitized for the 2013 scene with the assumption that the author's knowledge of the area may be more accurate overall, or at the very least, helpful ancillary data.

An important assumption was that the soft classifier would be helpful by showing the areas that the software had the most difficulty in determining classes. The purpose of the soft classifier was not to determine forest cover, but to help visualize uncertainty or confusion. An image with high levels of uncertainty can mean corrupted training data, so it was believed to be a helpful double-check.

The author also assumed that the DOQ and personal knowledge of the area are sufficient information to generate training data for the earlier images. Much of the forested area, based on ground research, is comprised of large redwoods and firs that are at least 50–75 years old. The assumption was that the large trees and forested area visible in the DOQ from 1993 were in fact predominantly of this type from 1972–1992.

Chapter 4: Results

4.1 Training Data

The training data was carefully chosen to include different types of forest, pasture, vineyard, water, and developed or barren land based on a DOQ mosaic and the author's knowledge of the area (Figure 7).

Figure 7. The 2013 Landsat subscene with training data outlined. Blue=water; green=forest; yellow=pasture; magenta=vineyard/orchard; and cyan=developed/barren.

4.2 Maximum Likelihood Hard Classifier

The images resulting from the maximum likelihood classifier showed obvious differences in resolution (i.e. Figures 8 and 9), and the variation from image to image was a concern (compare Figures 9 and 10).

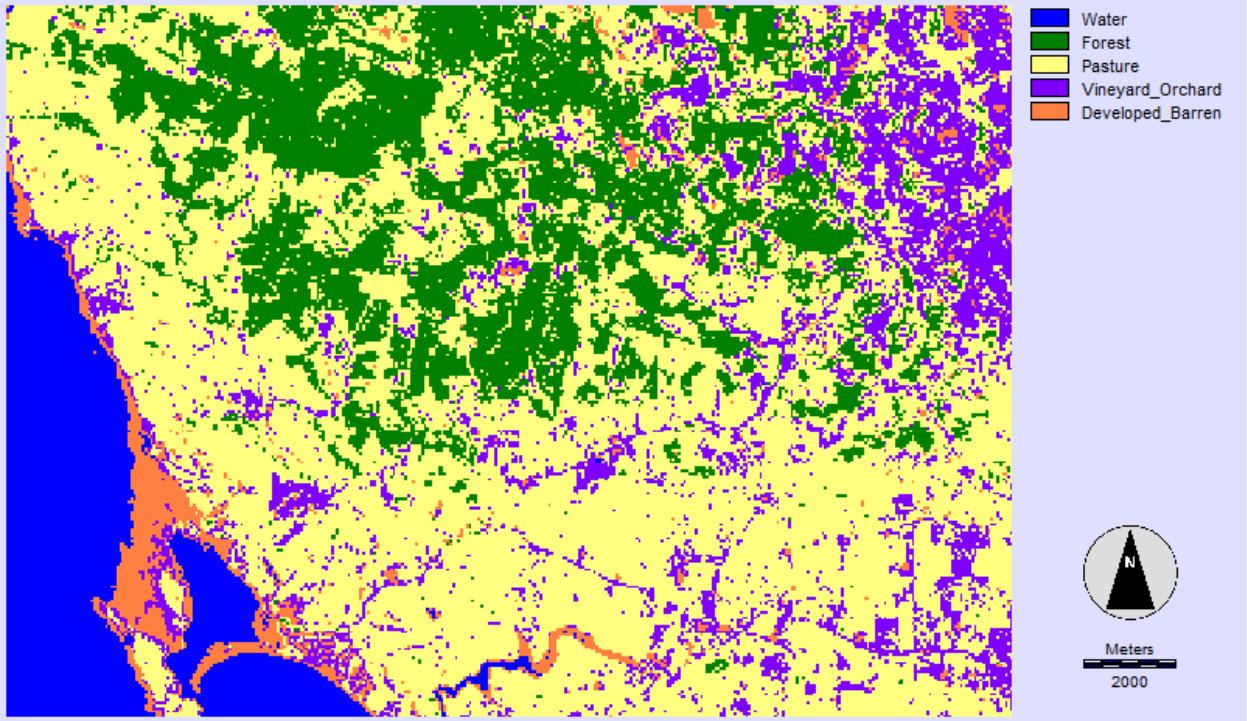

Figure 8. The results of the maximum likelihood classifier for the 60 m resolution 1995 Landsat subscene of WRS Path 45/ Row 33.

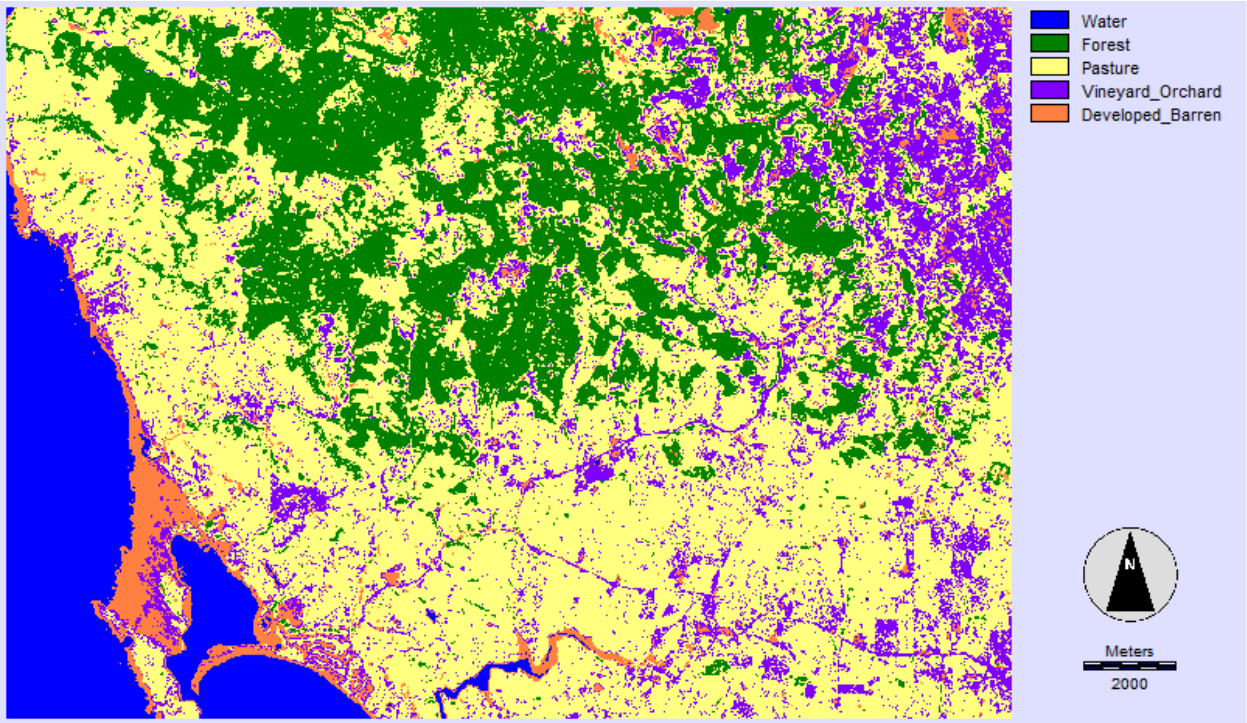

Figure 9. The results of the maximum likelihood classifier for the 30 m resolution 1995 Landsat subscene of WRS Path 45/ Row 33.

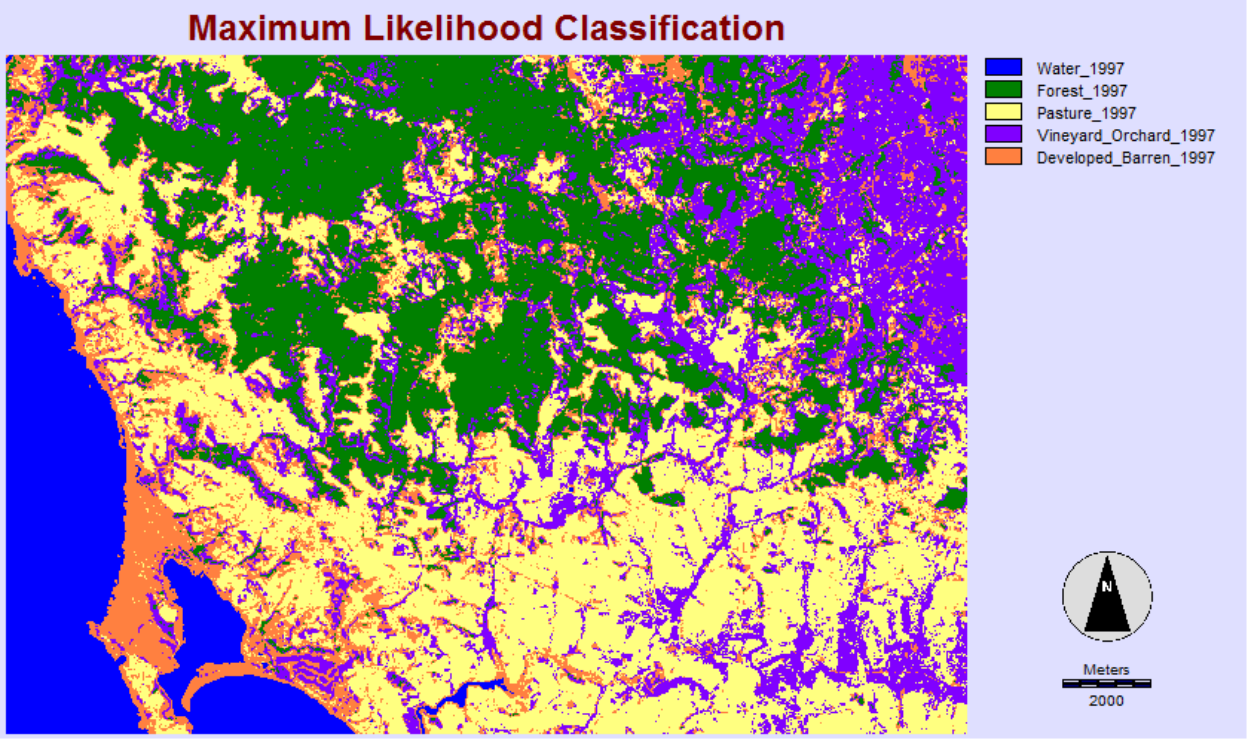

Figure 10. The results of the maximum likelihood classifier for the 30 m resolution 1997 Landsat subscene of WRS Path 45/ Row 33.

A comparison of forest area from year to year shows significant variation, but the core forest area remains essentially the same throughout the different images. A comparison of the first two images, years 1972 and 1975, shows a difference of –9.2592 km² in forest area (Figure 11, see Figure 16). A comparison of change between the 1985 classified forest area (the first year of 30 m resolution images) and the 2013 classified forest area shows a more detailed change image (Figure 12) depicting an overall gain in forest area of 8.2476 km² during that time period. The overall forest change based on post-classification comparison is a net loss of 3.7512 km² of forest area since 1972 based on 60 m resolution analysis (Figure 13). The 30 m resolution images, however, consistently showed greater forest area than the aggregated 60 m resolution images (see Figure 16), and a comparison of the 1972 60 m resolution forest area with the 2013 30 m resolution forest area shows a net gain in forest area of 2.2599 km².

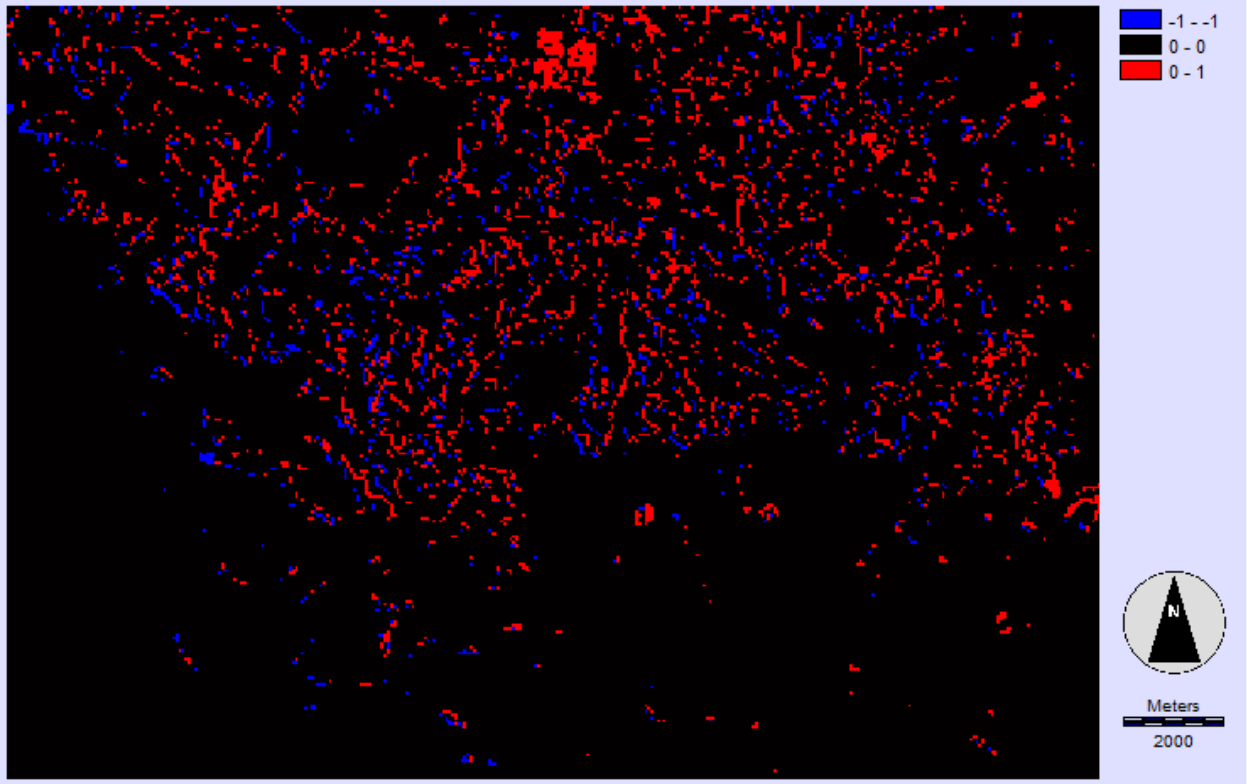

Figure 11. The forest change from post classification change detection. The image represents the forest area from 1972 minus the forest area from 1975 after the maximum likelihood classification in 60 m resolution. Blue areas represent forest loss, black represents no change, and red represents forest gain from 1972–1975. Forest area of of 9.2592 km² was lost.

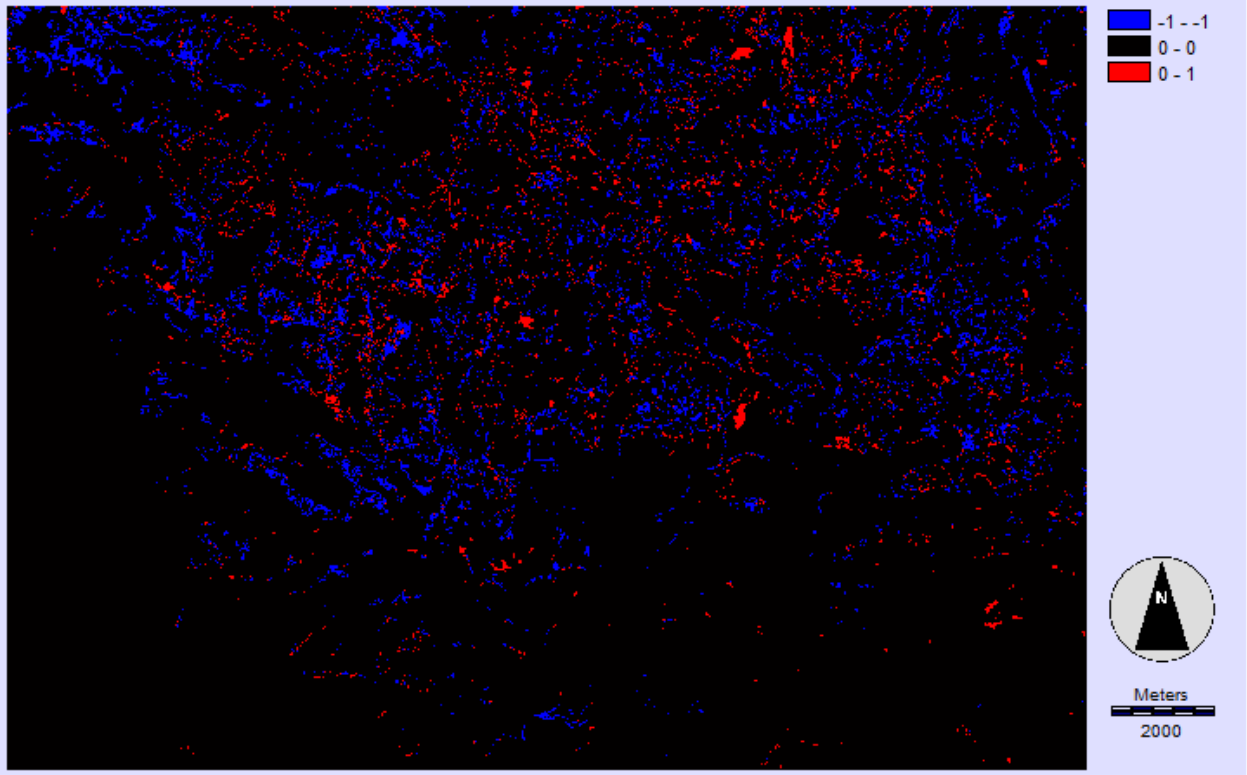

Figure 12. Forest change between the 1985 maximum likelihood classified forest area and the 2013 classified forest area at 30m resolution. Blue represents areas of forest loss, black represents no change, and red represents forest gain. A total of 8.2476km^2 of forest area was gained.

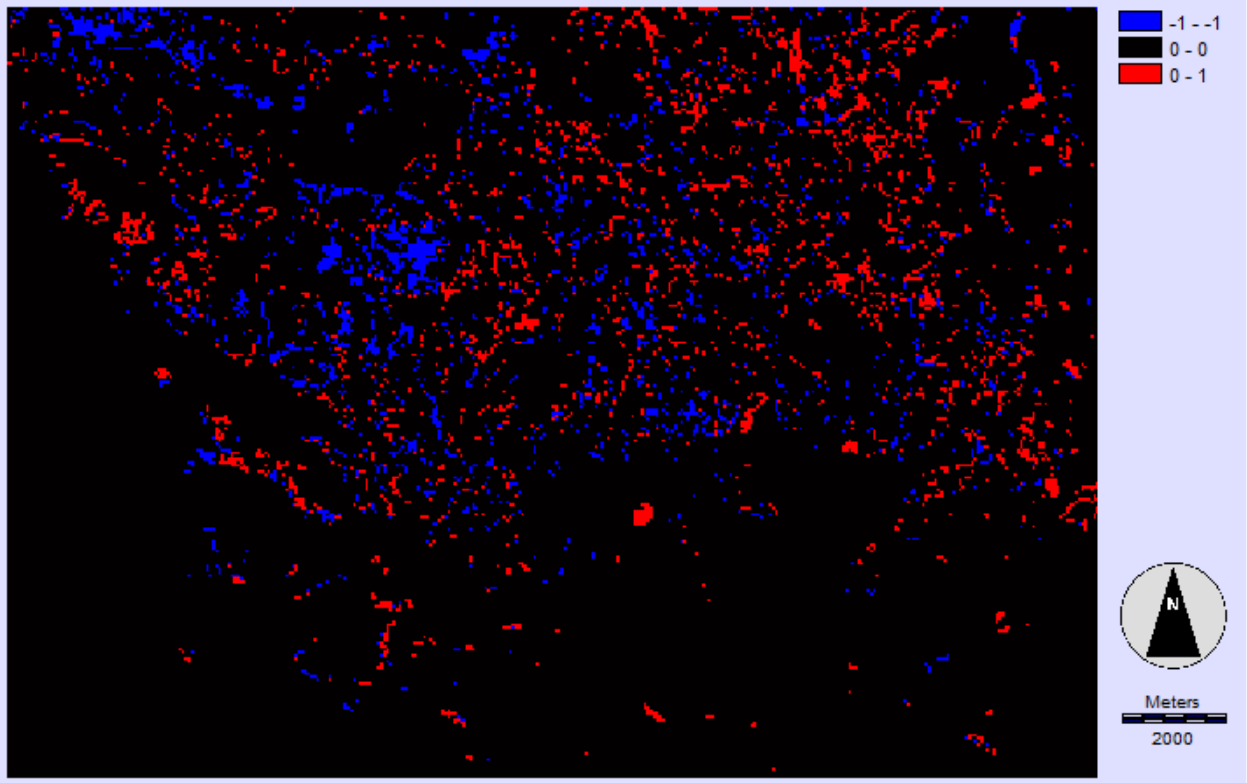

Figure 13. Forest change between 1972 and 2013 for 60 m maximum likelihood forest classification. Blue represents areas of forest loss, black represents no change, and red represents forest gain. At this resolution, a total of 3.7512 km² of forest area was shown to be lost.

Estimates of forest area varied substantially from year to year but appear to trend downward from 1972 to around 1987, then upward from 1987 to the present, with a peak around 2007 (Figure 16). Digitized forest area for 2013 at 30 m resolution is about 100 km², which is within the upper limit of the error associated with the maximum likelihood classified images. By adding the images of forest area for each year, it is possible to see the core forest areas, those areas which have not major canopy disturbance events, and the areas which have experienced some form of change in the time period of 1972 to 2013 (see Figure 14 for change at 30 m resolution since 1985; see Figure 15 for change at 60 m resolution since 1972).

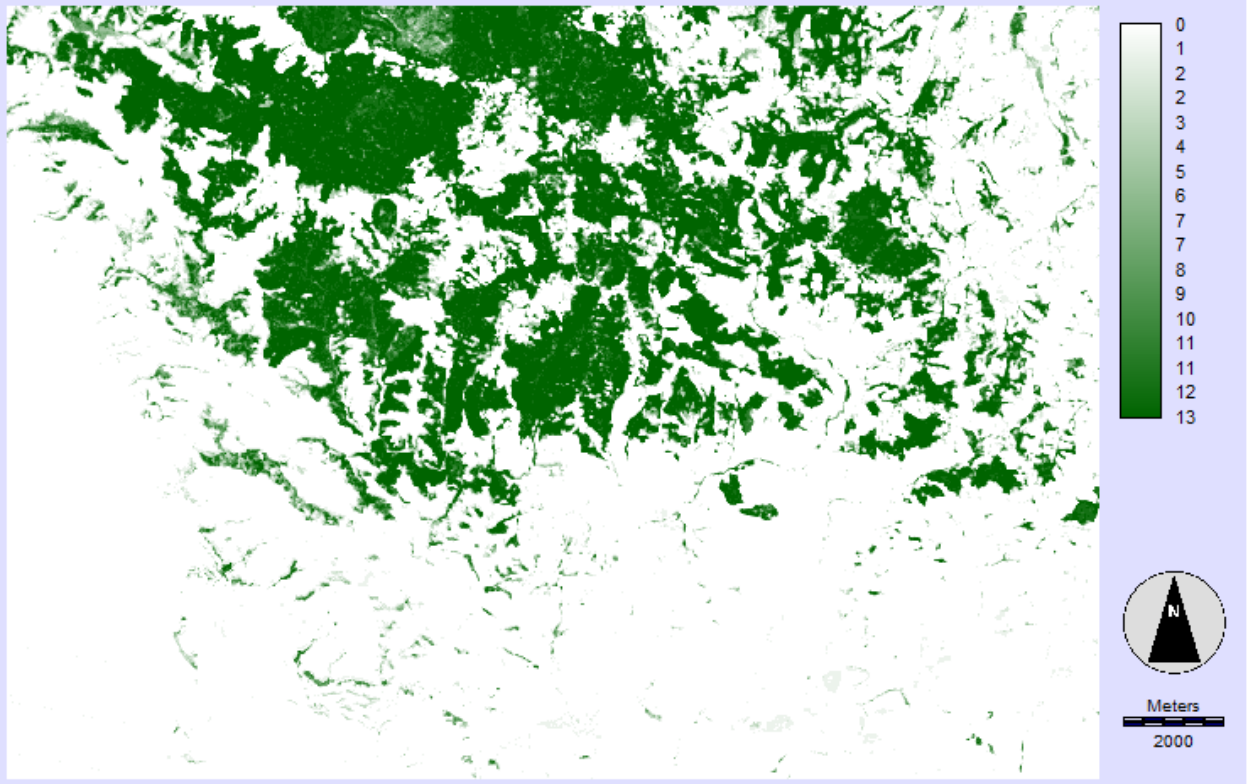

Figure 14. Total forest change from 1985–2013 at 30 m resolution. Darker green areas represent those pixels which were more consistently categorized as forest. Lighter green colors were categorized as forest less frequently, and white areas were not classified as forest during this time period.

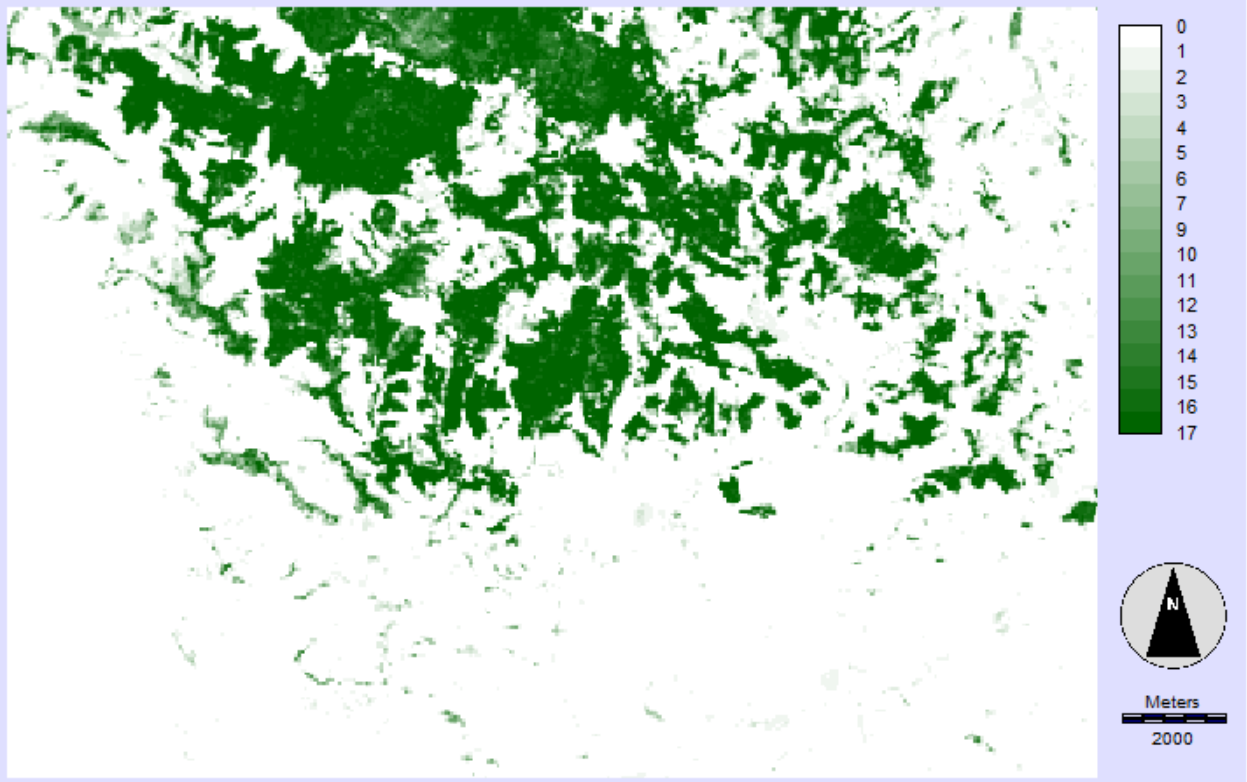

Figure 15. Total forest change from 1972–2013 at 60 m resolution. Darker green areas represent those pixels which were more consistently categorized as forest. Lighter green colors were categorized as forest less frequently, and white areas were not classified as forest during this time period.

When comparing the original 30 m resolution classified images with those made from the 60 m resolution aggregated images, it can be seen that the 60 m aggregated images consistently underestimate forest area. Some variation can be found from image to image, with the 1992 images having the widest gap at 12.9708 km² difference in forest area between the 30 m and 60 m classified images. The smallest gap was in 1985 with only 1.8747 km² difference. The average difference between 30 m resolution forest area and aggregated 60 m resolution forest area was 7.1346 km².

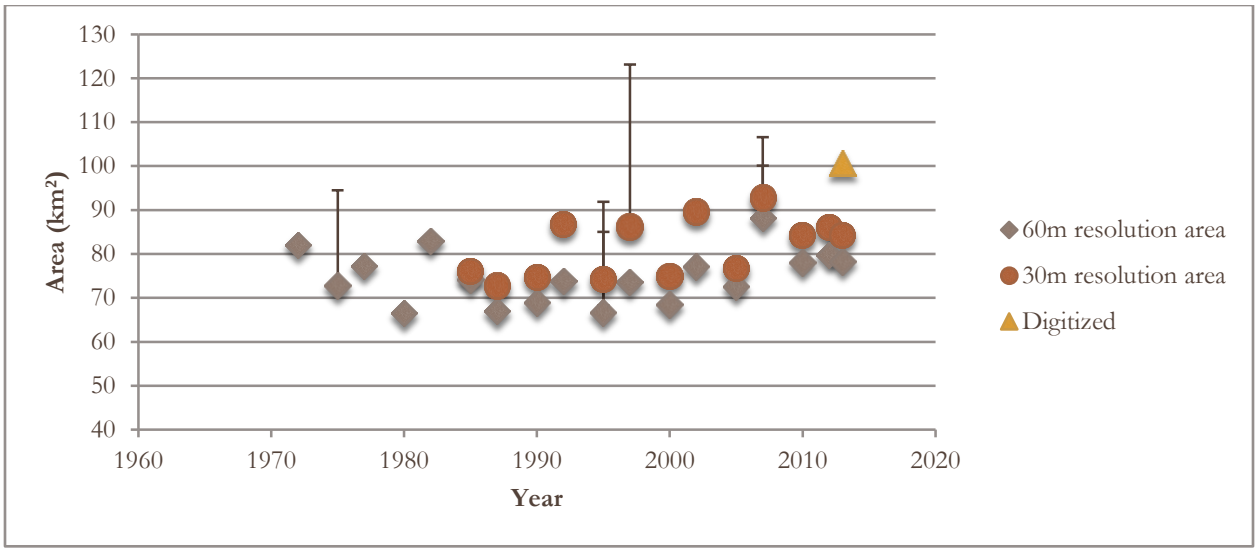

Figure 16. Total forest area for 30m and 60m resolution images after the maximum likelihood classification and error associated with forest classification. The error bars below the points are errors of commission, while the error bars above are errors of omission.

4.4 Error Analysis

The error of omission was a major concern (Tables 6, 7, & 8), showing that the maximum likelihood classification consistently underestimated forest area. This could be due to a limited classification scheme, incomplete or otherwise erroneous training data, spatial resolution that is too coarse for this scale, or a perhaps some types of forest have spectral signatures that are more similar another class in the classification scheme (for instance, riparian willows and alders may be getting confused with orchard or pasture).

Table 6. The error matrix from the 1975 Landsat maximum likelihood classified image (60 m) for the five coded classifications: 1= water, 2= forest, 3= pasture, 4= vineyard/orchard, 5= developed/barren. The error and KIA are expressed as proportions for each category.

Category	Water	**Forest**	Pasture	Vineyard /Orchard	Developed/ Barren	Total	ErrorC	KIA compared with classified image
Water	**26**	0	0	0	0	26	0	1
Forest	0	**68**	0	0	0	68	**0**	**1**
Pasture	0	11	**122**	4	1	138	0.1159	0.7697
Vineyard/Orchard	0	18	19	**16**	1	54	0.7037	0.246
Developed/Barren	1	0	8	0	**5**	14	0.6429	0.3418
Total	27	97	149	20	7	**300**		
ErrorO	0.037	**0.299**	0.1812	0.2	0.2857		**Overall Error 0.21**	
KIA compared with truth sample	0.9594	**0.6134**	0.6644	0.7561	0.7003			**Overall Kappa 0.69**

Because error of omission was unacceptably high, the training data were tested to determine whether they were satisfactory by more than doubling the training data for the 1997 image, which had unusually high error after the first run of ERRMAT. After classification and error analysis, the overall error and forest error of omission were still very similar and unacceptably high, but the places where the error occurred changed (Tables 7 and 8). The original image (Table 7) showed that the MAXLIKE classification tool classified the forest as vineyard/orchard (category 4) 28 times and as developed/barren 23 times out of 299 total points

tested, leading to a 42.86% error of omission, with no error of commission. After more than doubling the training data for forest, the MAXLIKE tool classified forest as pasture 18 times (compared to 0 in the original), still classified forest as vineyard/orchard about the same at 26 times, and as developed/barren only twice (Table 8). The error of omission was only slightly lower at 38.66%, while the error of commission remained at 0%. The overall error improved some from 31.44% to 23.41%, which is still unacceptably high, leading me to believe that it is unlikely that the training data was primary source of error.

Table 7. The error matrix associated with the 1997 image at 30 m resolution with the original training data. Bold data in columns and rows labeled "2" are associated with forest. ErrorC is error of commission; ErrorO is error of omission for 299 pixels tested.

Category	Water	**Forest**	Pasture	Vineyard/Orchard	Developed/Barren	Total	ErrorC
Water	32	**0**	0	0	1	33	0.0303
Forest	**0**	**68**	**0**	**0**	**0**	**68**	**0.0000**
Pasture	0	**0**	81	1	4	86	0.0581
Vineyard/Orchard	0	**28**	21	19	4	72	0.7361
Developed/Barren	1	**23**	11	0	5	40	0.8750
----------	--	**--**	--	--	--	--	--
Total	33	**119**	113	20	14	299	
ErrorO	0.0303	**0.4286**	0.2832	0.0500	0.6429		**0.3144**

Table 8. Error matrix associated with 1997 image at 30 m resolution after doubling the training data. Bold data in columns and rows labeled "2" are associated with forest. ErrorC is error of commission; ErrorO is error of omission for 299 pixels tested.

Category	Water	**Forest**	Pasture	Vineyard/Orchard	Developed/Barren	Total	ErrorC
Water	32	**0**	0	0	1	33	0.0303
Forest	**0**	**73**	**0**	**0**	**0**	**73**	**0.0000**
Pasture	0	**18**	100	1	4	123	0.1870
Vineyard/Orchard	0	**26**	11	19	4	60	0.6833
Developed/Barren	1	**2**	2	0	5	10	0.5000
----------	--	--	--	--	--		
Total	33	**119**	113	20	14	299	
ErrorO	0.0303	**0.3866**	0.1150	0.0500	0.6429		**0.2341**

4.2 Soft Classifier

The results of the MAHALCLASS tool showed that the main areas of uncertainty are the coast where the waves meet the sand, riparian areas (especially those to the south of the main forest areas), some of the pasture area in the southeast corner of the images, and almost the entirety of the forest area (Figure 17). Of the areas with the highest degree of uncertainty, the riparian and forest areas (even some areas that had been used for the training data) showed high degrees of uncertainty for the forest class, with pixels on the edge of the forest and some vineyard/orchard also being a concern (Figure 18).

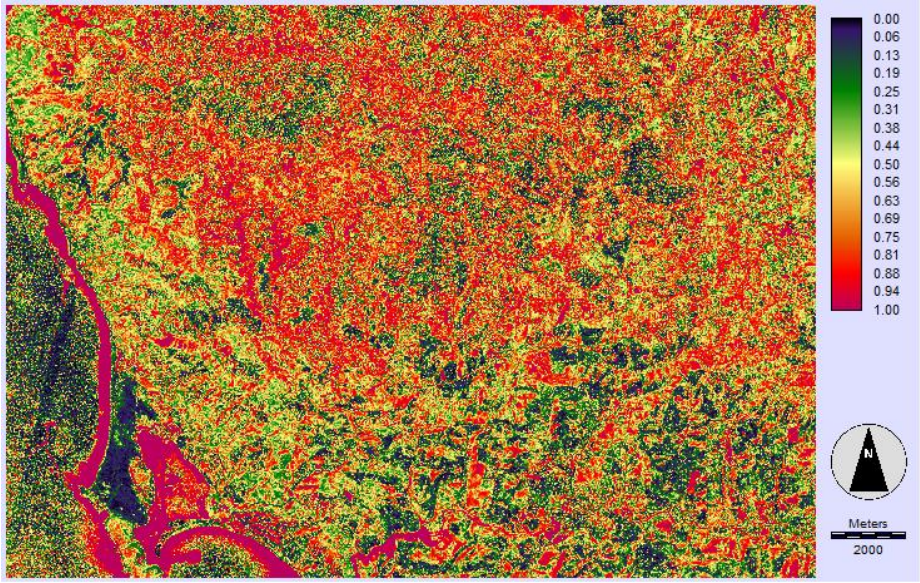

Figure 17. The classification uncertainty for all classes for the 2013 30m resolution image; 0.00= no uncertainty, 1.00= 100% uncertainty.

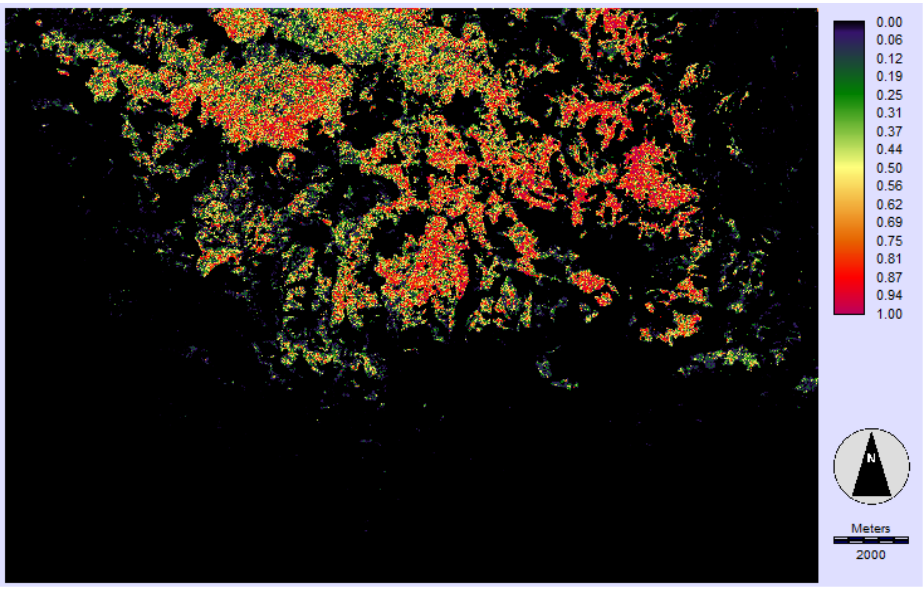

Figure 18. Uncertainty involved in forest classification for 2013 30 m resolution image after application of MAHALCLASS tool. 0.00= no uncertainty, 1.00= 100% uncertainty.

After adding the results of each year using the Image Calculator, a clearer picture of the uncertainty can be seen. The total uncertainty for all classes of all 30m resolution images reveals that the coast, some chaparral areas just inland from the coast, riparian areas, and the same

pasture land in the southeast corner of the image prove to be areas of greatest uncertainty (Figure 19). The total uncertainty for forest classification for each resolution compounded through time shows that the most heavily forested areas are associated with the highest uncertainty (Figures 20 and 21).

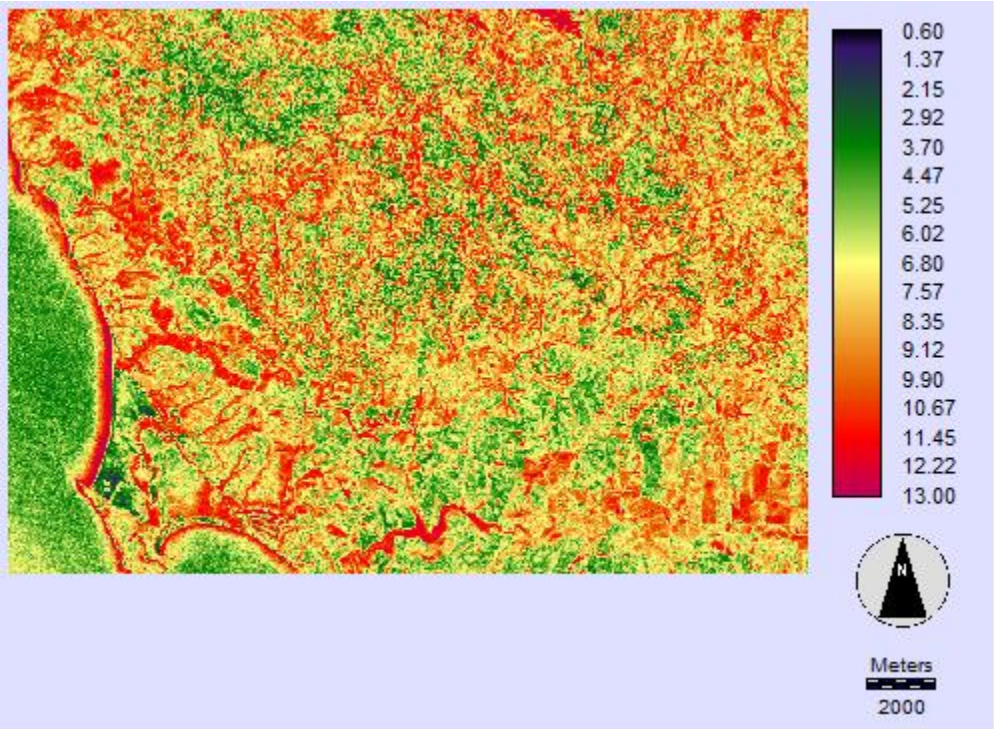

Figure 19. The accumulation of Mahalanobis distance classification uncertainty for all 30 m resolution images. Blue= least uncertainty, Magenta= most uncertainty.

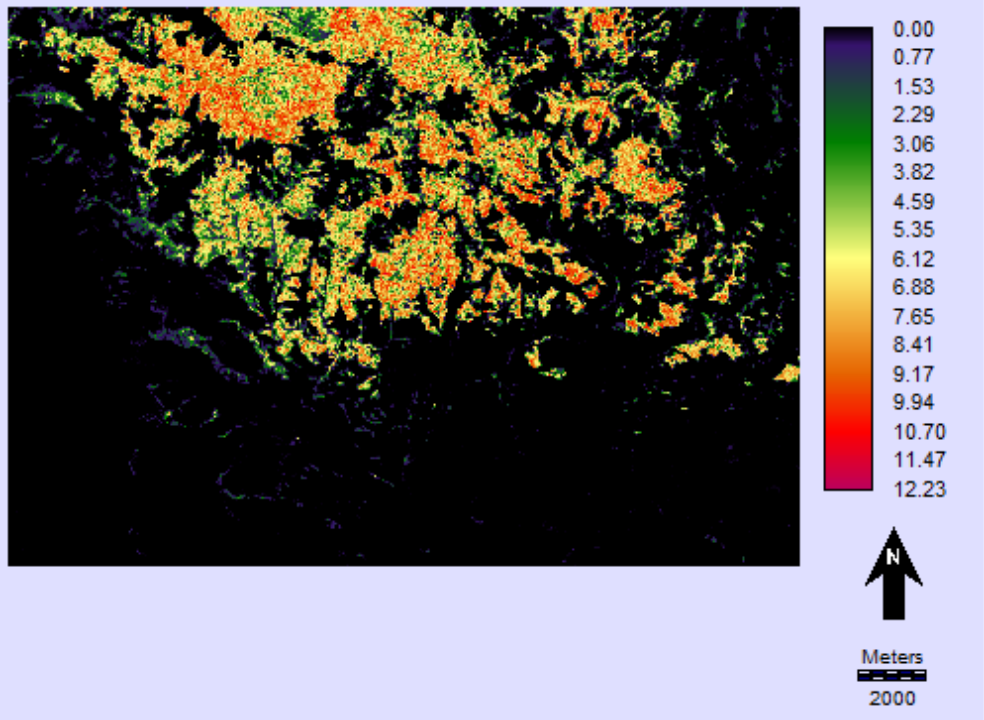

Figure 20. Uncertainty associated with forest classification for all years at 30 m resolution using the MAHALCLASS soft classifier in IDRISI Selva. Black= lowest uncertainty, Magenta= highest uncertainty.

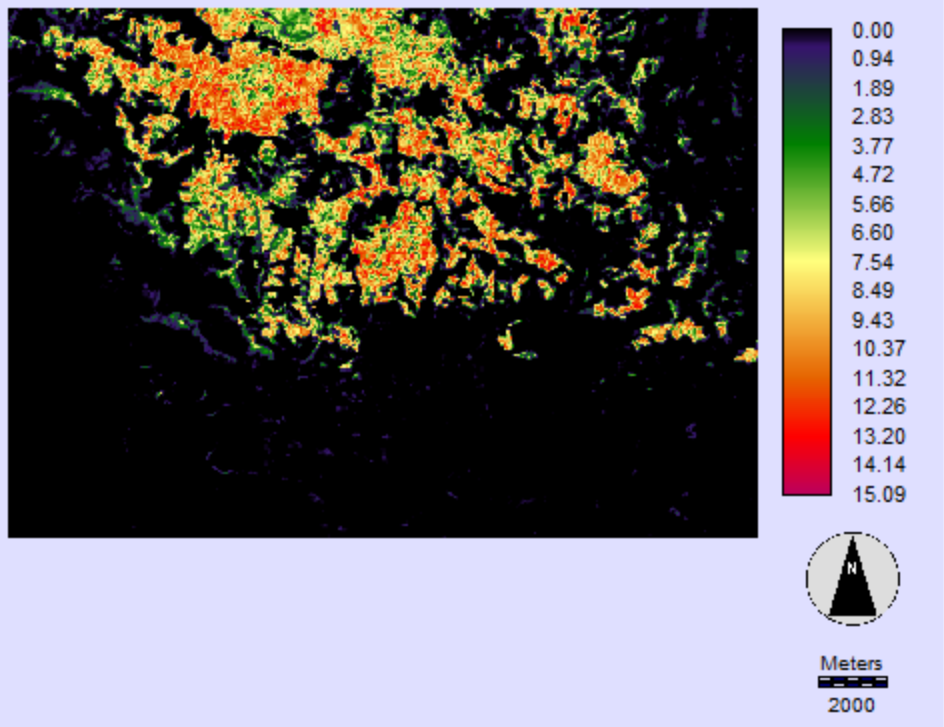

Figure 21. Uncertainty associated with forest classification for all years at 60 m resolution using the MAHALCLASS soft classifier in IDRISI Selva. Black= lowest uncertainty, Magenta= highest uncertainty.

4.3 Digitized Forest Area

When compared to the automatically classified 2013 image, the manually digitized 2013 image had approximately 16 km² more land classified as forest (100.7217 km², compared with 84.2823 km²), which was within the error associated with the automatically classified images. After subtracting the automated forest area (Figure 22) from the manually digitized forest area (Figure 23), two images were produced showing the areas of disagreement (Figures 24 and 25). The main areas of disagreement are riparian areas along the Salmon Creek between the towns of Bodega and Freestone, some pockets of within forested areas containing a few pixels of other types, and some coastal chaparral.

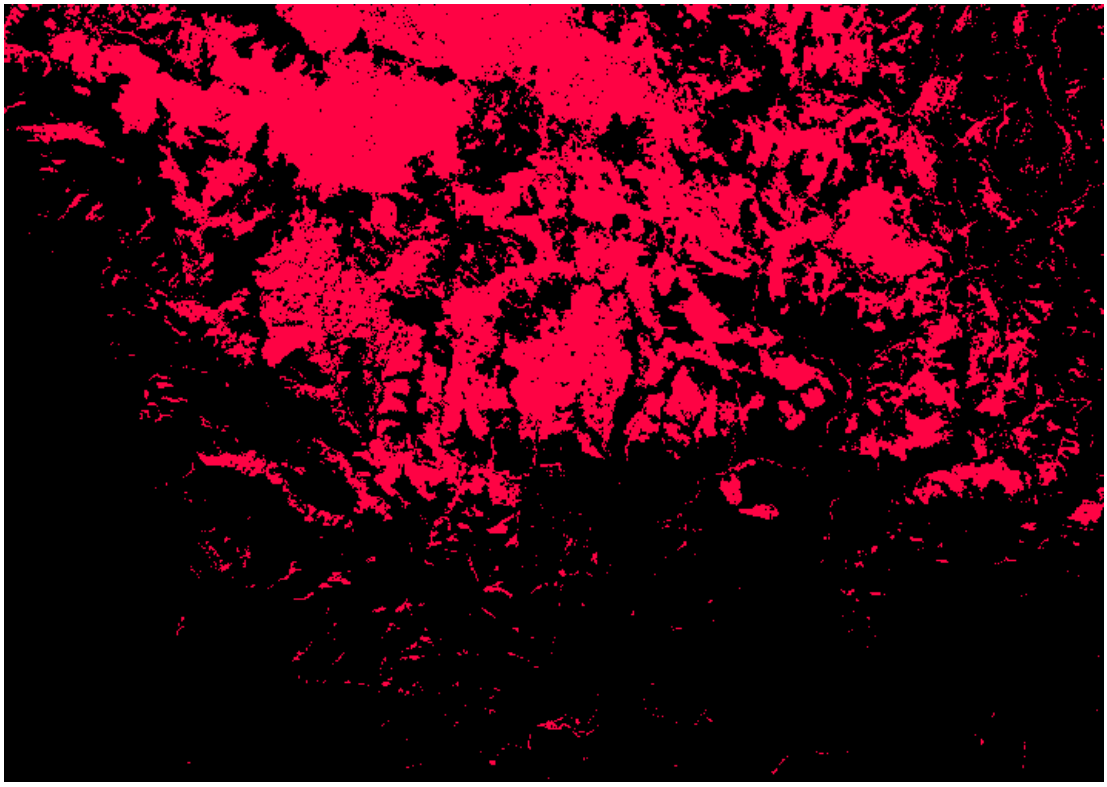

Figure 22. The 2013 forest area in red as defined by the maximum likelihood tool, MAXLIKE, in IDRISI Selva.

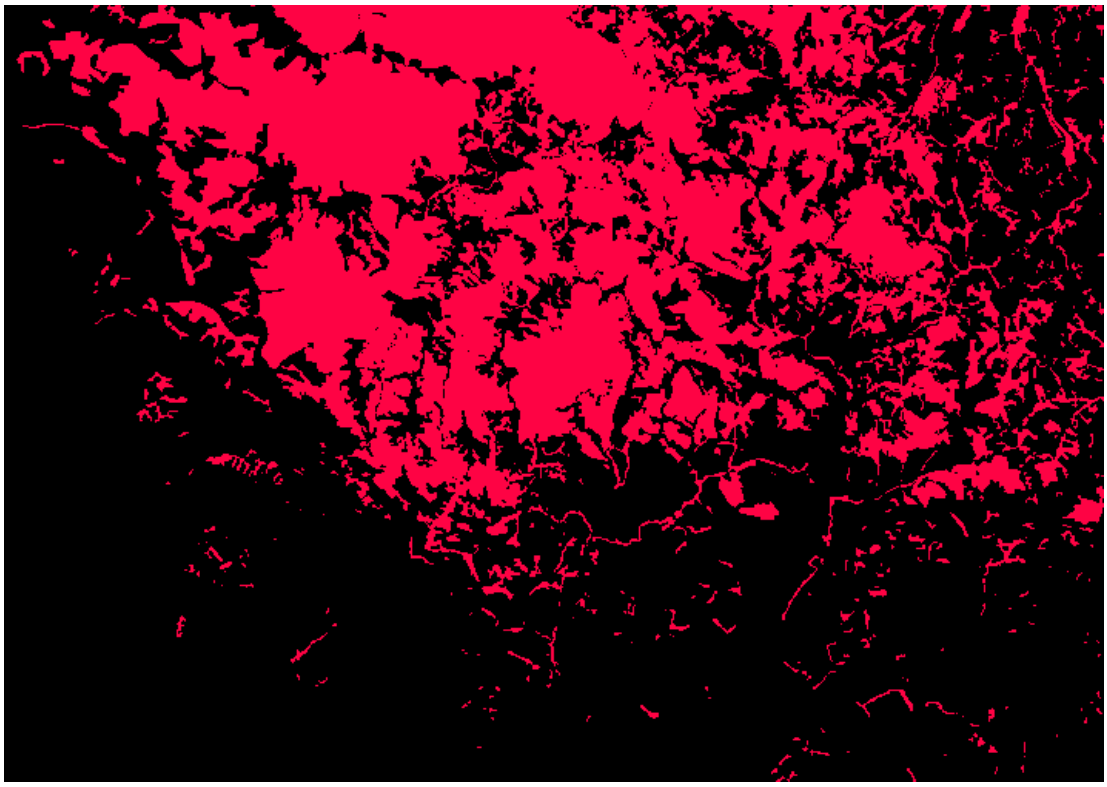

Figure 23. The 2013 forest area in red as defined by the author by manual digitization using the Digitize function in IDRISI Selva.

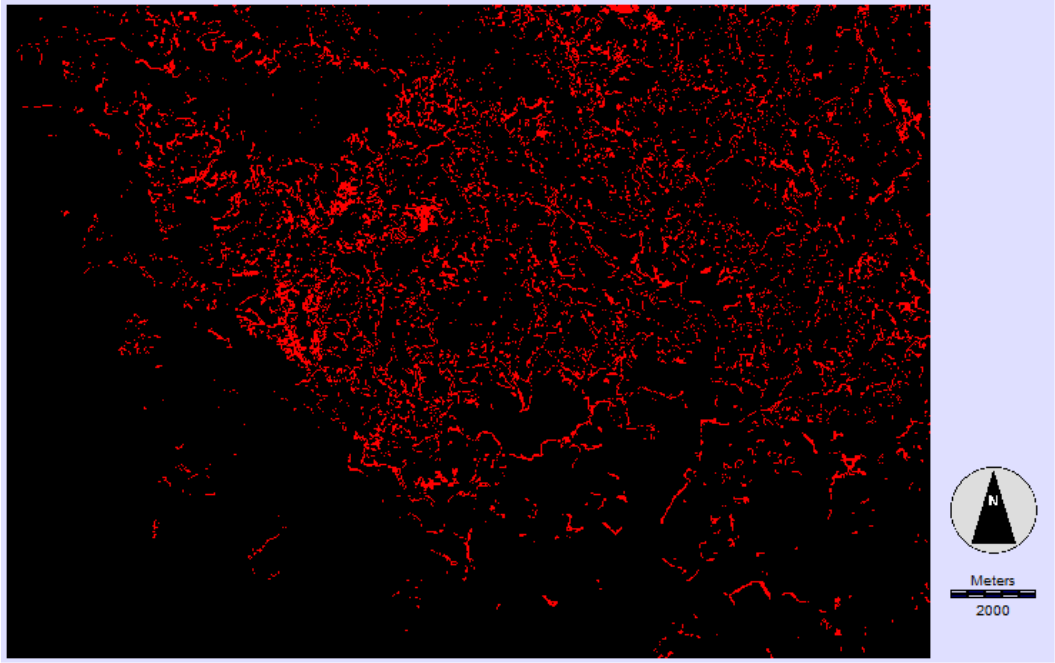

Figure 24. The results of the comparison of the manually digitized forest area with the automatically classified forest area for the 2013 scene. Red areas are those which I classified as forest, but which were classified otherwise by the software.

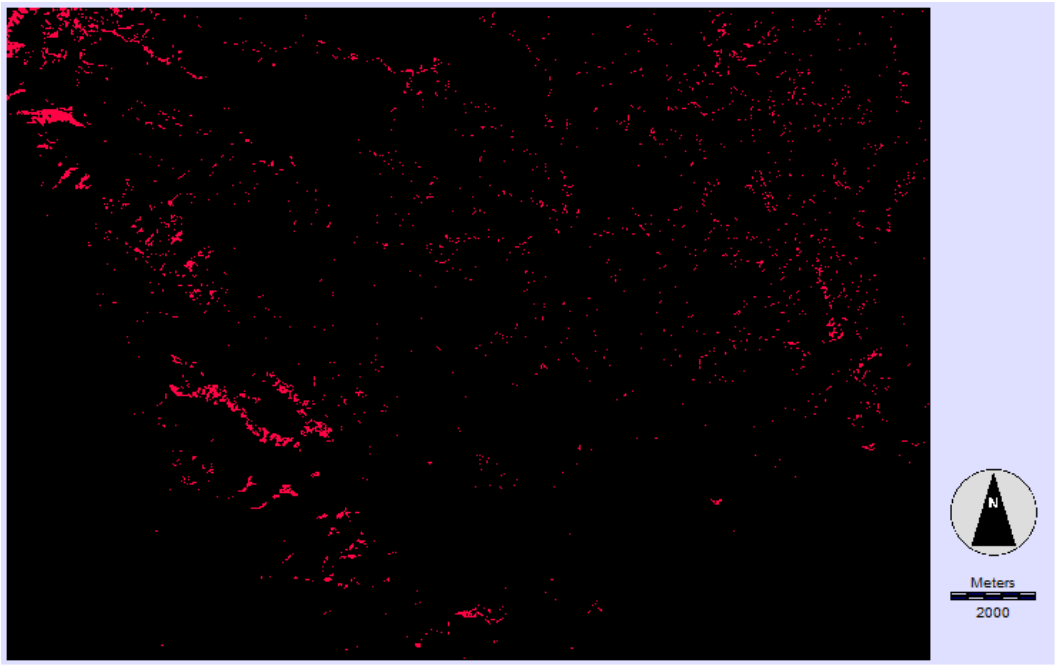

Figure 25. The results of the comparison of the manually digitized forest area with the automatically classified forest area for the 2013 scene. The areas in red correspond to areas classified by the software as forest, but which I did not consider to be forest.

Chapter 5: Discussion

The main task of this thesis was to study the forest cover change for a time stack of subscenes of Landsat data to discover the story of the forests in the Salmon Creek Watershed and vicinity since 1972. The primary interest was to estimate forest cover change in the Salmon Creek Watershed and its immediate surroundings with an aim to present new knowledge to those interested in conservation or restoration in the area. The maximum likelihood classifier was the main tool used for assessment of forest area, while the soft classifier aided in visualizing uncertainty in classification, and a manually digitized 2013 scene revealed the author's interpretation of forest area for that scene. The results that stand out most are the consistent errors of omission in forest cover associated with the maximum likelihood classifications. The maximum likelihood classifier consistently underestimated forest cover, but there seemed to be enough of a pattern to see a "U"-shaped trend in forest area. The trend seems to bottom out in the mid-1990's before beginning to rise, which would seem to coincide with logging regulations in the mid-1990's sparked by the spotted owl controversies (U.S. Government 1995).

Total forest area from 1972 to 2013 appears to have slightly risen by 2.2599 km². However, if the 60 m resolution images from the MSS instrument underestimate forest area in the same way as the aggregated images do, then there may be a net decrease in forest area since 1972. The aggregated 60 m images underestimated forest area compared to the original 30 m images by on average 7.1346 km², meaning that there could be a net loss of about 5 km² forest area.

The chosen classification scheme did not allow for mixed pixels, pixels showing reflectance from two or more land cover types, which was exacerbated in the 30 m resolution images that were degraded to 60m resolution images. The images which were aggregated to 60

m resolution consistently underestimated forest area compared to the original 30 m resolution images, and draws into question whether the MSS 60 m resolution images also underestimates forest area. The error analysis consistently showed very few errors of commission, but significant omission errors, meaning that the data should have been classified as forest much more frequently than they were for all years. The manually classified image showed that what I believe to be the actual forest area is significantly greater than what IDRISI Selva shows it to be with the MAXLIKE tool, the training data I gave it, and the resolution of the Landsat images. The forest area was consistently underestimated at these resolutions.

5.1 Training Data

The training data chosen for the forest class showed to be sufficient, although, in retrospect, another category for chaparral may have helped. After more than doubling the training data on an image with especially high overall error, the error only decreased slightly. The choice to add chaparral to the pasture class seemed to cause confusion not only between forest and pasture, but between pasture and vineyard/orchard, as well. Applying a terrain correction algorithm may have produced better results, since shadows and slope differences can affect the results in mountainous terrain (Tan et al. 2013).

5.2 Hard Classifier

The maximum likelihood classifier alone did not produce sufficient results. This could be due to the classification scheme, an inability to discern mixed pixels, and/or insufficient spatial resolution. The variation from year to year made it difficult to make an accurate assessment of forest area. Areas that showed the greatest variation are mixed pixels, pastures,

and vineyard/orchards. Much of this variation may be due to the amount of moisture in the soil. For instance, some open, grassy fields are on irrigation systems and some are not. While the fields may appear visually similar, they are distinguishable by the different bands available with Landsat.

5.3 Error Analysis

The error analysis was conducted on only a few select years due to the time constraints involved in classifying 300 points per scene. It would have been only slightly less time-consuming than digitizing forest area for all years. It was high error of omission from preliminary error analysis which led to use of a soft classifier and manual digitization.

5.4 Soft Classifier

The results of the MAHALCLASS analysis show that there were areas of consistent confusion; even some of the same places that were used as training data were associated with great uncertainty. The coastal brush seems to have less uncertainty in the forest classification than there is for the more heavily forested areas, adding to my belief that the training data was not the source of confusion. Based on this and the fact that the coastal brush areas show great uncertainty when all five categories are accounted for, I conclude that the coastal brush was more likely confused with the vineyard/orchard and/or pasture classes than with the forest class. The use of the soft classifier was interesting, but proved to be little help as far as actually determining forest area, and it is the author's opinion that this step could have been skipped entirely with little difference in the overall results of the study.

5.5 Digitization

Another factor that may be involved is perceived contiguity as a factor in determining classes. For instance, in my manual digitization of the forest area, I may have considered riparian areas as more likely to be forest, whereas the DN associated with these pixels may more closely represent the vineyard/orchard class due to moisture and canopy density. My decision to classify these riparian areas as forest was perhaps more an issue of function rather than the density of trees. The presence of a few trees in riparian areas may have been more likely to be considered forest, whereas the presence of a few trees in developed areas may still be likely to receive the developed/barren designation. Many residences are also located amongst the redwoods and other tall trees, which may have the visual appearance of pure forest. In these cases, I chose to classify the pixels as forest rather than developed/barren because the forest canopy is more easily observed; although, there are aspects such as impervious surfaces and fences which hinder full forest functionality below the canopy.

Digitizing was a tedious and time-consuming task, and I was aware that I inevitably enclosed some non-forest pixels in the digitized polygons (shown by sparse pixelated areas in Figure 24). This is by no means the only difference between the digitized and automated, as I took great care in keeping as many of these areas outside the digitized polygons as possible.

5.6 Further Directions

Even with the variation from image to image, I believe there is enough evidence to suggest that the downward and then upward trend in forest area is essentially correct, though underestimated. The consistently low error of commission is my main evidence for this

conclusion, supported strongly by the manual digitization of the forest area. The soft classifier showed that there were areas of uncertainty in determining forested pixels, but that the chaparral was not a major source of confusion (see Figure 20).

One on the most common and successful techniques for determining vegetation cover and change is the NDVI, which would have been a much simpler and more direct approach to this project. A comparison of NDVIs would have given much more information than what I was able to determine based on the methods laid out above, and the procedure is fairly simple and straightforward without taking much time to process.

A Landsat time lapse from 1984 to 2012 is available on the internet via Google's Earth Engine at https://earthengine.google.org/#intro/v=38.5779555,-122.98883189999998,7.946972730143203 (Google 2012). Upon inspection, a few areas of development and some conversion of grassland to vineyard/orchard are apparent, but there are no obvious signs of clear-cutting, and only a few small areas seem to have been converted from forest to developed/barren. It is interesting to note that MSS data were not used in the above time-lapse.

Most Landsat time series studies use full or mosaicked images to capture the significance of regional dynamics, while some such as this search for significance at a scale within a sub-scene of a full image. This presents difficulties as the 30 m and 60 m resolution images do not capture the detail desired for assessment of a small area, although finer spatial resolution will not necessarily mean higher map accuracy (see Thompson and Gergel 2008). The availability and history of the Landsat data contribute heavily in our foundational understanding of Earth's surface, especially at medium to broad scales, and has over 40 years of consistent data. With its many benefits, Landsat essentially lacks the ability to capture nuanced details necessary for a

thorough investigation of forest dynamic in smaller areas. For instance, several smaller trees or undergrowth could be "thinned" from a forest background without much of a change in an individual pixel's value at 30 m resolution, yet the ecology has changed. At 60 m resolution, a fairly large redwood could be removed without being detected.

While it has been shown sufficient to aggregate pixels to coarser resolutions to estimate forest land where larger patches (9–25 ha) are concerned (Nelson et al. 2009), this says little about the utility of 30 m and 60 m resolution images for smaller land areas. The greatest benefit to this project of the Landsat data was it temporal history, but for improved change detection at the small watershed scale, data with broader temporal resolution and finer spatial resolution would be required.

Further study should be conducted in whether the apparent downward then upward trend in forest cover is associated with salmon counts in the Salmon Creek Watershed during this time.

References

Andrew, M. E., and M. A. Wulder. 2011. Idiosyncratic responses of Pacific salmon species to land cover, fragmentation, and scale. *Ecography* 34: 780–797. DOI: 10.1111/j.1600-0587.2010.06607.x.

Atkinson, S. F., B. A. Hunter, and A. R. English. 2010. Prioritizing riparian corridors for water quality protection in urbanizing watersheds. *Journal of Water Resource and Protection* 2: 675–682. DOI: 10.4236/jwarp.2010.27078.

Bhagat, V. S. 2009. Use of Landsat ETM+ data for detection of potential areas for afforestation. *International Journal of Remote Sensing* 30 (10): 2607–2617. DOI: 10.1080/01431160802552793.

Bragg, D. C., and J. L. Kershner. 1999. Coarse woody debris in riparian zones: Opportunity for interdisciplinary interaction. *Journal of Forestry* 97 (4): 30–35.

Clark Labs. 2013. IDRISI Selva v. 17.02. Clark University, Worchester, MA, USA.

Cohen, W. B., and S. N. Goward. 2004. Landsat's role in ecological applications of remote sensing. *Bioscience* 54 (6): 535–545.

Crapper, P. F., and K. C. Hynson. 1983. Change detection using Landsat photographic imagery. *Remote Sensing of Environment* 13: 291–300.

Damschen, E. I., and L. A. Brudvig. 2012. Landscape connectivity strengthens local–regional richness relationships in successional plant communities. *Ecology* 93 (4): 704–710.

Department of Commerce and National Oceanic and Atmospheric Administration. 2005. Endangered and threatened species; designation of critical habitat for seven evolutionarily significant units of Pacific salmon and steelhead in California. *Federal Register* 70 (170): 52488–52627.

Edman, M., N. Kruys, and B. G. Jonsson. 2004. Local dispersal sources strongly affect colonization patterns of wood-decaying fungi on spruce logs. *Ecological Applications* 14 (3): 893–901.

Feist, B. E., E. A. Steel, G. R. Pess, and R. E. Bilby. 2003. The influence of scale on salmon habitat restoration priorities. *Animal Conservation* 6: 271–282. DOI: 10.1017/S1367943003003330.

Google. 2012. Earth Engine: Landsat time lapse 1984–2012. https://earthengine.google.org/#intro/v=38.5779555,-122.98883189999998,7.946972730143203. Last accessed 31 March 2014.

Gustafson, R. G., R. S. Waples, J. M. Myers, L. A. Weitkamp, G. J. Bryant, O. W. Johnson, and J. J. Hard. 2007. Pacific salmon extinctions: quantifying lost and remaining diversity. *Conservation Biology* 21 (4): 1009–1020. DOI: 10.1111/j.1523-1739.2007.00693.x.

Healey, S. P., W. B. Cohen, Y. Zhiqiang, and O. N. Krankina. 2005. Comparison of Tasseled Cap-based Landsat data structures for use in forest disturbance detection. *Remote Sensing of Environment* 97: 301–310. DOI: 10.1016/j.rse.2005.05.009.

Holzmueller, E. J., M. D. Gaskins, and J. C. Mangun. 2011. A GIS approach to prioritizing habitat for restoration using Neotropical migrant songbird criteria. *Environmental Management* 48: 150–157. DOI: 10.1007/s00267-011-9660-1.

Hottola, J., O. Ovaskainen, and I. Hanski. 2009. A unified measure of the number, volume and diversity of dead trees and the response of fungal communities. *Journal of Ecology* 97: 1320–1328. DOI: 10.1111/j.1365-2745.2009.01583.x.

Howarth, P. J. and G. M. Wickware. 1981. Procedures for change detection using Landsat digital data. *International Journal for Remote Sensing* 2 (3): 277–291.

Huang, C., K. Song, S. Kim, J. R. G. Townshend, P. Davis, J. G. Masek, and S. N. Goward. 2008. Use of a dark object concept and support vector machines to automate forest cover change analysis. Remote Sensing of Environment 112: 970–985. DOI: 10.1016/j.rse.2007.07.023.

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. *Remote Sensing of Environment* 114: 183–198. DOI: 10.1016/j.rse.2009.08.017.

IDRISI Selva. Clark Labs, Clark University, Worcester, MA, USA.

Junninen, K., and A. Komonen. 2010. Conservation ecology of boreal polypores: A review. *Biological Conservation* 144: 11–20. DOI: 10.1016/j.biocon.2010.07.010.

Keleş, S., F. Sivrikaya, G. Çakir, and S. Köse. 2008. Urbanization and forest cover change in regional directorate of Trabzon forestry from 1975 to 2000 using Landsat data. *Environmental Monitoring Assessment* 140: 1–14. DOI 10.1007/s10661-007-9845-5.

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. *Remote Sensing of Environment* 114: 2897–2910. DOI: 10.1016/j.rse.2010.07.008.

Kovalenko, K. E., S. M. Thomaz, and D. M. Warfe. 2012. Habitat complexity: approaches and future directions. *Hydrobiologia* 685: 1–17. DOI: 10.1007/s10750-011-0974-z.

Lassauce, A., Y. Paillet, H. Jactel, and C. Bouget. 2011. Deadwood as a surrogate for forest biodiversity: Meta-analysis of correlations between deadwood volume and species

richness of saproxylic organisms. *Ecological Indicators* 11: 1027–1039. DOI:10.1016/j.ecolind.2011.02.004.

Lonsdale, D., M. Pautasso, and O. Holdenrieder. 2008. Wood-decaying fungi in the forest: conservation needs and management options. *European Journal of Forest Restoration* 127: 1–22. DOI: 10.1007/s10342-007-0182-6.

Malanson, G. P. and B. E. Cramer. 1999. Landscape heterogeneity, connectivity, and critical landscapes for conservation. *Diversity and Distributions* 5: 27–39.

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. *Ecosystems* 16: 1087–1104. DOI: 10.1007/s10021-013-9669-9.

Mihal, B., I. Savulescu, and I. Sandric. 2007. Change detection analysis (1986–2002) of vegetation cover in Romania: a study of alpine, subalpine, and forest landscapes in the Iezer Mountains, Southern Carpathians. *Mountain Research and Development* 27 (3): 250–258.

Mönkkönen, M., P. Reunanen, J. S. Kotiaho, A. Juutinen, O. -P. Tikkanen, and J. Kouki. 2011. Cost-effective strategies to conserve boreal forest biodiversity and long-term landscape-level maintenance of habitats. *European Journal of Forest Restoration* 130: 717–727. DOI: 10.1007/s10342-010-0461-5.

Naiman, R. J., H. Décamps, and M. Pollock. 1993. The role of riparian corridors in maintaining regional biodiversity. *Ecological Applications* 3 (2): 209–212.

——. 1997. The ecology of interfaces: riparian zones. Annual Review of Ecological Systems 28: 621–658.

National Aeronautics and Space Administration. 2011. Landsat 7 Science Data Users Handbook. Prepared by the Landsat Project Science Office at NASA's Goddard Space Flight Center, Greenbelt, Maryland. 1–186.

——. 2013. Landsat Science webpage: http://landsat.gsfc.nasa.gov/. Accessed 04 Dec 2013.

Natural Resources Conservation Service. 2008. California Watershed Boundary Dataset. Accessed on 17 March 2013 from http://ceres.ca.gov/search#search/watershed+boundary/%255BRecommend(int)%253E0%255D/false/none/false/1/6.

Nelson, M. D., R. E. McRoberts, G. R. Holden, and M. E. Bauer. 2009. Effects of satellite image spatial aggregation and resolution on estimates of forest land area. *International Journal of Remote Sensing* 30 (8): 1913–1940. DOI: 10.1080/01431160802545631.

Newburn, D. A., P. Berck, and A. M. Merenlender. 2006. Habitat and Open Space at Risk of Land-Use Conversion: Targeting Strategies for Land Conservation. *American Journal of Agricultural Economics* 88 (1): 28–42. http://www.jstor.org/stable/3697964.

Özyavuz, M. 2010. Analysis of changes in vegetation using multitemporal satellite imagery, the case of Tekirdağ coastal town. *Journal of Coastal Research* 26 (6): 1038–1046. ISSN 0749-0208.

Pflugmacher, D., W. B. Cohen, and R. E. Kennedy. 2012. Using Landsat-derived disturbance history (1972–2010) to predict current forest structure. *Remote Sensing of Environment* 122: 146–165. DOI: 10.1016/j.rse.2011.09.025.

Prunuske Chatham, Inc. 2010. Salmon Creek Water Conservation Plan. Virginia Porter Consulting, and Occidental Arts and Ecology Center's WATER Institute for the Occidental Arts and Ecology Center's WATER Institute with funding from the California State Coastal Conservancy.

Richards, J. A. 2013. *Remote sensing digital image analysis: an introduction*. Fifth Edition. Springer-Verlag Berlin Heidelberg: 90–92. DOI 10.1007/978-3-642-30062-2.

Rissman, A. R., L. Lozier, T. Comendant, P. Kareiva, J. M. Keisecker, M. R. Shaw, and A. M. Merenlender. 2007. Conservation easements: biodiversity protection and private use. *Conservation Biology* 21 (3): 709–718. DOI: 10.1111/j.1523-1739.2007.00660.x.

Robinson, C. T., K. Tockner, and J. V. Ward. 2002. The fauna of dynamic riverine landscapes. *Freshwater Biology* 47: 661–677.

Rondeux, J., and C. Sanchez. 2010. Review of indicators and field methods for monitoring biodiversity within national forest inventories. Core variable: Deadwood. *Environmental Monitoring Assessment* 164: 617–630. DOI: 10.1007/s10661-009-0917-6.

Singh, A. 1989. Digital change detection techniques using remotely-sensed data. International *Journal of Remote Sensing* 10 (6): 989–1003.

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? *Remote Sensing of Environment* 75: 230–244.

Sonoma County Administrator's Office. 2011. Population. http://www.sonoma-county.org/cao/citizens_guide/sonoma_county_population.htm (last accessed 17 March 2013).

State of California Natural Resources Agency, Department of Fish and Wildlife. 2014. *State and Federally Listed Endangered & Threatened Animals of California*. http://www.dfg.ca.gov/biogeodata/cnddb/pdfs/TEAnimals.pdf. Last accessed 19 May 2014.

Tan, B., J. G. Masek, R. Wolfe, F. Gao, C. Huang, E. F. Vermote, J. O. Sexton, G. Ederer. 2013. Improved forest change detection with terrain illumination corrected Landsat images. *Remote Sensing of Environment* 136: 469–483. DOI: 10.1016/j.rse.2013.05.013.

Thompson, S. D., and S. E. Gergel. 2008. Conservation implications of mapping rare ecosystems using high spatial resolution imagery: recommendations for heterogeneous and fragmented landscapes. *Landscape Ecology* 23: 1023–1037. DOI: 10.1007/s10980-008-9263-2.

Tomer, M. D., M. G. Dosskey, M. R. Burkart, D. E. James, M. J. Helmers, and D. E. Eisenhauer. 2009. Methods to prioritize placement of riparian buffers for improved water quality. *Agroforestry Systems* 75: 17–25. DOI: 10.1007/s10457-008-9134-5.

United States Geological Survey. 2013. Landsat data downloaded from http://earthexplorer.usgs.gov/ on various dates in 2013 and 2014.

United States Government. 1995. California Spotted Owl Recovery Plan. U.S. Government Printing Office. Washington, D.C. ISBN 0-16-047668-2.

Warner, T. A., D. J. and Campagna. 2009. Remote Sensing with IDRISI® Taiga: A Beginner's Guide. Geocarto International Centre 180–231. Hong Kong.

Wiens, J., R. Sutter, M. Anderson, J. Blanchard, A. Barnett, N. Aguilar-Amuchastegui, C. Avery, and S. Laine. 2009. Selecting and conserving lands for biodiversity: The role of remote sensing. *Remote Sensing of Environment* 113: 1370–1381. DOI: 10.1016/j.rse.2008.06.020.

White, P. J. T., B. J. McGill, and M. J. Lechowicz. 2012. Detecting changes in forest floor habitat after canopy disturbance. *Ecological Restoration* 27: 397–406. DOI: 10.1007/s11284-011-0911-7.