University of Tennessee, Knoxville
TRACE: Tennessee Research and Creative Exchange

Masters Theses | Graduate School

8-2016

Developing Historical Productivity Maps for the Blue Oak Woodlands in California using Remote Sensing and Dendrochronology

Kyle Lawrence Landolt
University of Tennessee, Knoxville, landolt.kl@gmail.com

Follow this and additional works at: https://trace.tennessee.edu/utk_gradthes

Recommended Citation

Landolt, Kyle Lawrence, "Developing Historical Productivity Maps for the Blue Oak Woodlands in California using Remote Sensing and Dendrochronology. " Master's Thesis, University of Tennessee, 2016.
https://trace.tennessee.edu/utk_gradthes/4011

This Thesis is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a thesis written by Kyle Lawrence Landolt entitled "Developing Historical Productivity Maps for the Blue Oak Woodlands in California using Remote Sensing and Dendrochronology." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Geography.

Robert A. Washington-Allen, Major Professor

We have read this thesis and recommend its acceptance:

Nicholas Nagle, Henri Grissino-Mayer

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Developing Historical Productivity Maps for the Blue Oak Woodlands in California using Remote Sensing and Dendrochronology

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Kyle Lawrence Landolt

August 2016

© by Kyle Lawrence Landolt, 2016

All Rights Reserved.

This thesis is dedicated to my family. I wouldn't be here today without their support and patience.

With all the love in my heart

- Kyle Landolt

Acknowledgements

First and foremost, I would like to express my full appreciation and gratitude towards my committee members: Dr. Washington-Allen, Dr. Nicholas Nagle, and Dr. Henri Grissino-Mayer. Their support and dedication to my academic career made the time spent during my master's degree a truly valuable experience. Furthermore, I would like to personally thank Dr. Washington-Allen for everything he has done for me. Ever since I was his student at Texas A&M University (*Gig 'em*), he has provided an extraordinary amount of opportunities for research, adventure, and education. I can say with true sincerity that without the guidance of Dr. Washington-Allen, I wouldn't be here writing my master's thesis.

I would like to thank my good friend, Lauren Stachowiak, for helping me with my thesis no matter what her own responsibilities were. The assistance she provided me was solely out of her own brilliance and kindness and she serves as an excellent example of how selfless one can be during a very hard and struggling time. Along with her, I would like to thank all of my other friends, peers, and mentors in the department that have helped me through my thesis.

I would also like to thank my mentor, Alfredo Delgado, for his guidance and support and inspiring me to follow a path of science and research. Whether it be trekking through the plains of west Texas or wading through the wetlands near the gulf, Alfredo has been supportive with every research endeavour I've been a part of.

Lastly, I would like to thank my loving mom, dad, brother, and friends back in Texas for all the support they've given me. They've gone out of their way to help me and I would be lost without them. I will forever be grateful for the love they've given me.

Abstract

Global terrestrial carbon dynamics play a critical role in the Earth system. Increased net primary productivity (NPP) associated with terrestrial ecosystems can influence the amount of atmospheric carbon dioxide by sequestering carbon through photosynthesis. Drylands, in particular, are an important terrestrial component that can act as a source or sink for carbon storage depending on various environmental factors that influence vegetation and soil patterns. Blue oak woodlands represent a large component of the terrestrial vegetation in California and provide an opportunity to study the historical variation in NPP by combining dendrochronological measurements and a satellite-based proxy for NPP: the remotely sensed time-integrated NDVI (iNDVI). The purpose of this research was to investigate the possible correlation of iNDVI to blue oak chronology measurements to reconstruct iNDVI of the blue oak woodlands from 1700 to 2003. We developed linear regressions to explain iNDVI using two types of blue oak chronologies, elevation, slope, aspect, cation exchange capacity, and ecological sections and compared our results to reconstructed Palmer Drought Severity Index (PDSI). A spatial scale study was done

to test for the optimal scale at which iNDVI can be predicted. Our reconstructions were able to predict iNDVI (Adjusted R-squared = 0.715; $p < 0.001$) and were also correlated to PDSI values ($r = 0.8$; $p < 0.001$), finding notably high and low values of iNDVI at 1789 and 1934, respectively. We also found that the residual chronology and a spatial scale of 3136 square kilometers predicted iNDVI most accurately. These findings produce a novel time-series of iNDVI that can be used to determine how historical vegetation productivity has fluctuated and sequestered carbon in the blue oak woodlands and provides a framework for global reconstructions of vegetation productivity.

Table of Contents

List of Figures

List of Attachments

Chapter 1

Introduction

1.1 Background

The carbon dynamics of terrestrial ecosystems play a critical role in the Earth and global climate systems (Heimann and Reichstein, 2008; Rockström et al., 2009; Keeling and Shertz, 1992). The dominant processes that influence these dynamics are the creation or destruction of organic matter through two main critical chemical reactions: photosynthesis and respiration. Vast amounts of carbon are stored in living vegetation and soil organic matter through photosynthesis and carbon is lost in the form of CO_2 primarily through respiration. The rate of accumulation of carbon in the biosphere is referred to as net primary productivity (NPP, g m^{-2} yr^{-1}) (Leith and Whittaker, 1975) and is considered a critical measurement to describe the habitability of Earth (Running, 2012).

Ecologists began to measure global satellite-based estimates of NPP using remote sensing data since the early 1980s (Running et al., 2004). In particular, the development of the normalized difference vegetation index (NDVI) led to many estimates of NPP, phytomass, and vegetation cover (Sellers, 1985; Pettorelli et al., 2005; Campbell and Wynne, 2011). The first integration of NDVI over a time series into iNDVI (integrated NDVI) was developed by Tucker et al. (1981) to remotely sense the accumulation of above ground dry biomass or NPP. iNDVI was also used to produce a continental scale land-cover classification map of Africa (Tucker et al., 1985). However, drylands have not been the specific target for similar studies.

Dryland ecosystems represent approximately 47.2% of terrestrial ecosystems (Hillel and Rosenzweig, 2002; Middleton and Thomas, 1992) and may play a vital role in reducing atmospheric CO_2 through increased NPP (Lal, 2004; Poulter et al., 2014). California and various parts of the southwestern U.S. are composed of an arrangement of dryland ecosystems that are subject to desertification and degradation that may result in increased emissions of carbon. Nonetheless, little is known about the role of dryland carbon dynamics in the Earth system (Schimel, 2010).

The blue oak woodlands of California are an ecologically unique landscape that can be used to investigate the carbon dynamics of drylands. Endemic to California, the blue oak woodlands (*Quercus douglasii* Hook & Arn.) form one of the major oak woodland communities in the state extending along the Sierra Nevada-Cascade-Coast Range foothills of the Sacramento and San Joaquin Valley (Figure 2.2). They are found on a wide variety soils, slopes, and elevations (Barbour et al., 2007). Blue

oak trees are sensitive recorders of past climates and have been used previously in reconstructions of drought, salinity, and precipitation (Stahle et al., 2001; Meko et al., 2011).

The purpose of this thesis is to combine dendrochronological measurements of blue oak and satellite-derived iNDVI measurements to develop historical reconstructions of productivity in the drylands of California. Although carbon sequestration and NPP by terrestrial ecosystems are regarded as highly important topics for the global system (Running, 2012), little research has been done to spatially reconstruct measurements of historical carbon sequestration.

1.2 Thesis Synopsis

This study investigated whether a functional relationship exists between iNDVI and dendrochronological measurements of blue oak woodlands, and given that there is a relationship, to develop reconstructed maps of iNDVI from 1700 to 2003 if such a relationship exists. Due to the lack of clarity from previous research done to correlate NDVI and iNDVI to chronology data, I initiated an analysis to develop technical methods to correlate iNDVI produced from contemporary satellite imagery with blue oak chronology data using two chronology types and four different spatial scales. I used a multivariate linear regression analysis to determine the most effective combination of these two variables to produce the most accurate reconstructions of iNDVI. I performed an accuracy assessment to determine how accurate reconstructed

iNDVI values were from their true values. Furthermore, I used regressional cokriging to interpolate iNDVI values across the state of California. I then compared my reconstructed values of iNDVI with the reconstructed Palmer Drought Severity Index values developed by (Cook et al., 2004) using bilinear regression.

The model I developed accurately reconstructed iNDVI values and had a significant linear relationship between iNDVI and the residual chronology data ($p < 0.001$). The accuracy assessment showed no significant difference between predicted iNDVI values and actual iNDVI values using a paired t-test ($p = 0.8775$). An area of 3136 km^2 produced the highest adjusted R^2 value (Adjusted $R^2 = 0.799$) when developing a model to reconstruct iNDVI, although I determined that the 64 km^2 model produced a good enough fit for this study (Adjusted $R^2 = 0.715$). Lastly, I found a significant linear relationship between my reconstructed iNDVI values and the reconstructed PDSI values by (Cook et al., 2004) (Adjusted $R^2 = 0.6388$; $p < 0.001$).

1.3 Research Questions

The blue oak woodlands provide a unique opportunity to study the historical carbon dynamics of drylands ecosystems. These woodlands possess a large amount of dominant individuals with ages up to 400 years old (Stahle et al., 2001), providing an extended temporal reconstruction period. With the inclusion of modern satellite imagery, I can develop reconstructions of iNDVI to further our knowledge of carbon

dynamics. However, little research exists to guide the use of technical methods required to reconstruct iNDVI. Consequently, I analyzed the most optimal method to reconstruct iNDVI using various chronology and spatial scales. I also correlated my reconstructions of iNDVI to Cook et al. (2004)'s reconstructions of Palmer Drought Severity Index. The following research questions were addressed:

1. Which chronology type shows the highest degree of association with iNDVI?

2. Which spatial scale provides the highest degree of association between tree-ring indices and iNDVI?

3. Will reconstructed historical maps of iNDVI based on tree-ring data correspond to similar reconstructions of drought in California?

4. Why reconstruct historical iNDVI?

Chapter 2

Literature Review

2.1 Study area and species: The blue oak savannah and woodlands of California

Oak woodlands and savannas occupy approximately 4 million hectares in California (Griffin, 1977; Bolsinger, 1988; FRAP, 2003; Barbour et al., 2007). Generally, these areas are composed of an overstory tree canopy and an annual grassland understory, with shrubs and perennial grasses in some areas (Figure 2.1) (Griffin, 1973; Bartolome, 1987; Holmes, 1990). Compared to other vegetation types in the California, they possess the greatest species richness with over 300 vertebrate, 5,000 invertebrate, and 2,000 plant species (Barrett, 1979; Verner, 1979; Garrison, 1996). Oak savannas and woodlands are found at 60 – 700 meters in elevation and have a Mediterranean climate with precipitation mainly occurring between October and May (Barbour et al., 2007).

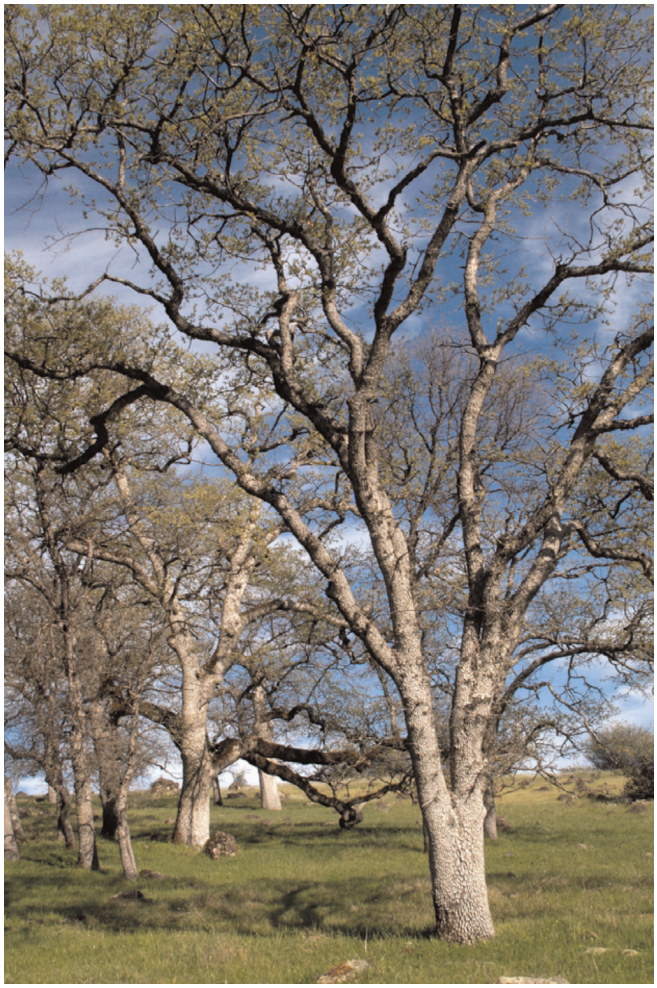

Figure 2.1: Blue oak woodland in Bear Creek Canyon near the northern limit of the species distribution. (Photo credit: Stahle et al., 2013)

Approximately 73% of oak savannas and woodlands are privately owned (FRAP, 2003) with primary economic products ranging from livestock production, firewood, wildlife, water, to recreation (McClaran and Bartolome, 1985; Standiford et al., 1996). Barbour et al. (2007) describe five communities of oak woodlands in California: the valley oak woodland, coastal oak woodland, montane hardwood forest, blue oak-foothill pine woodland, and the blue oak woodland. My thesis is focused on the blue oak woodland and blue oak-foothill pine woodland.

Figure 2.2: The blue oak distribution in California as developed by Griffin and Critchfield (1972).

Blue oak (*Quercus douglasii* Hook. & Arn.) is a deciduous tree species endemic to the state of California and forms one of the major oak woodland communities that extends along the Sierra Nevada-Cascade-Coast Range foothill of the Sacramento and San Joaquin Valleys (Figure 2.2) (Barbour et al., 2007). This area includes one of the most productive agricultural regions in the world.

Commonly associated tree species are valley oak (*Q. lobata* Nee), foothill pine (*Pinus sabiniana* Dougl.), and the California black oak (*Q. kelloggii* Newb.) (Allen, 1989; Ritter, 1988; Neal, 1980). On gentle slopes, blue oak woodlands are generally found in large blocks but are found in small patches with varying understory on

steeper slopes. Blue oak woodlands occur on a wide range of soils but these are typically well drained, infertile, and shallow (Barbour et al., 2007). Canopy cover of blue oak woodland varies from sparse to nearly closed while basal area can range from 5 m^2/ha to 11 m^2/ha (Barbour et al., 2007). The understory composition and productivity are significantly affected by blue oak trees, depending on tree density and annual precipitation. Areas of precipitation greater than 50 cm annually have suppressed understory biomass during the growing season (Barbour et al., 2007) while drier sites can have productivity under blue oak canopies twice that of open grassland (Holland, 1980).

Blue oak woodlands provide a spatially extensive and unique opportunity for long-term reconstruction of past environmental factors. Although blue oak woodlands are increasingly being used for agricultural and urban development (Stahle et al., 2013), dominant individual blue oak tree are widespread in California with ages ranging from 150 to 400 years old (Stahle et al., 2001). Disturbances of the fire regime (Swetnam et al., 2009), the invasion of nonnative grasses (Seabloom et al., 2003), and livestock grazing (Mensing, 1992) have not influenced a major change in the mean growth or inter-annual variability of blue oak chronologies (Stahle et al., 2013). Stahle et al. (2013) found that the genetic variability of this species led to the spatial distribution of blue oaks found in arid conditions to be related to the amount of precipitation received while blue oak sites found in high elevation mesic conditions have weaker associations with precipitation. Furthermore, blue oak trees are sensitive recorders

of past climates and have been used in numerous reconstructions of precipitation, salinity, and drought (Stahle et al., 2001; Meko et al., 2011; Gervais, 2006).

2.2 Remote sensing of vegetation indices for carbon studies

Attempts to measure the contemporary carbon dynamics of terrestrial ecosystems have been made since the mid-1970s (Running et al., 2004). Leith and Whittaker (1975) produced the earliest estimates of global NPP using meteorological data gathered from weather stations. Various improvements to methods used for estimating NPP were made with the recent inclusion of increased atmospheric flask sampling, basic ecological land surface process modeling, and the use of satellite derived vegetation indices that include the normalized difference vegetation index (NVDI) (Rouse Jr et al., 1974; Tucker, 1979; Kerr and Ostrovsky, 2003; Pettorelli et al., 2005) which was derived from readily available meteorological data (Running et al., 2004).

NDVI has become one of the most extensively used vegetation indices to study NPP, phytomass, and vegetation cover (Sellers, 1985; Pettorelli et al., 2005; Campbell and Wynne, 2011). NDVI was developed in the early 1970s by Rouse Jr et al. (1974) to measure the difference in growth of vegetation between two pastures that had different stocking rates of livestock. NDVI uses the red (RED) and near-infrared (NIR) reflectance and is calculated as:

$$NDVI = (NIR - RED)/(NIR + RED) \tag{2.1}$$

NDVI takes advantage of the difference in absorption and reflectance of red and near-infra red radiation by plant canopies where leaf pigments in the palisade mesophyll cells of vegetation absorbs more red radiation compared to the near-infrared radiation that is highly reflected by the spongy mesophyll tissues. NDVI values range from -1 and $+1$ with values of 0.05 corresponding to sparse vegetation and 0.7 to dense vegetation cover and negative values generally corresponding to barren rock and snow (Tucker and Sellers, 1986). Since these early studies, NDVI has been applied globally as a measure of the fraction of photosynthetically active radiation (fPAR) (Goward et al., 1985; Tucker and Sellers, 1986; Prince, 1991). Integration or summation of an annual time series of NDVI measurements (365 scenes) or summation of the growing season NDVI produces integrated NDVI (iNDVI) (Tucker et al., 1981). iNDVI has served many uses for the large-scale study of terrestrial vegetation and carbon dynamics, being first developed by Tucker et al. (1981) to remotely sense the accumulation of above-ground dry biomass or NPP. Tucker et al. (1985) continued to use iNDVI derived from time series of the NOAA Advanced Very High-Resolution Radiometer (AVHRR) on a continental scale to produce a land-cover classification map of Africa. Further work by Goward et al. (1985) showed that growing season iNDVI values are highly correlated with both gross primary productivity (GPP) and NPP of major biomes (r = 0.94).

2.3 Vegetation indices and dendrochronological studies

A number of studies have been performed to relate NDVI and dendrochronological measurements of tree-ring width. Beginning in the late 1980s, researchers began to associate measurements of tree-ring width to the rate of atmospheric CO_2 sequestration and remotely sensed photosynthetic activity (D'Arrigo et al., 1987). Other dendrochronological measurements, such as maximum latewood density, have been shown to have positive correlations with NPP measurements derived from NDVI (D'Arrigo et al., 2000). Relationships have also been found with time-series analyses of tree-ring width and NDVI to explain growth patterns of forests (Hunt et al., 1991; Malmstrm et al., 1997). Further research has been conducted to include iNDVI, linkages between heterogeneous environments, and scaling concerns.

Tree-ring indices serve as proxies for local environmental NDVI characteristics of terrestrial environments. Forested environments show associations between NDVI and tree rings in northern latitudes (D'Arrigo et al., 2000). Kaufmann et al. (2004) determined that NDVI measurements of northern forests approximate the physiological status of the tree expressed through tree vigor and tree rings. Other terrestrial ecosystems, including grasslands, have also shown NDVI patterns that suggest a relationship with local tree-ring chronologies. He and Shao (2006) used principal components analysis to derive relationships of maximum value composite NDVI images to chronologies of Qilian juniper (*Juniperus przewalskii* Kom.) in

Delingha, China. In the semi-arid grasslands of north China, iNDVI is highly correlated with the tree-ring width index of Meyer spruce (*Picea meyeri* Rehd. et Wils.) and can serve as a proxy for above-ground biomass (Liang et al., 2005). Though correlations and highly significant regression models have been developed between tree-ring data and NDVI, an explanatory causal mechanism that is rooted in tree physiology has yet to be established (Hunt et al., 1991; D'Arrigo et al., 2000; Kaufmann et al., 2004).

Methods for correlating iNDVI values and tree-ring indices vary between studies. For example, some studies vary the chronology type used, the spatial resolution or grain of NDVI and iNVDI, and how the iNDVI values were produced. Wang et al. (2004) conducted a study in Kansas that examined the correlation of NDVI values at different levels of resolution varying from 1.2 km^2 to 146.4 km^2 to values of a standardized tree-ring index and found that intermediate scales ($\sim$ 50 km^2) provided higher correlations than the nominal scale of local NDVI values (1.2 km^2). However, the potential to continue exploring correlations of iNDVI to tree-ring indices at finer levels of resolutions still remain.

Chapter 3

Developing Historical Productivity Maps for the Blue Oak Woodlands in California

This chapter is intended for publication in the journal *Global Change Biology*. I developed the research topic with my committee members and following authors, Dr. Robert A. Washington-Allen, Dr. Nicholas Nagle, and Dr. Henri Grissino-Mayer. The use of we within the text refers to me and my committee members, who assisted with project development, methodology, data processing, and text editing. My contributions to this chapter include data processing, interpretation, developing graphic displays of results, and writing the manuscript.

3.1 Abstract

Global terrestrial carbon dynamics play a critical role in Earth's climate patterns. Increased net primary productivity (NPP) associated with terrestrial ecosystems can influence the amount of atmospheric carbon dioxide by sequestering carbon through photosynthesis. Drylands, in particular, are an important terrestrial component that can act as a source or sink for carbon storage depending on various environmental factors that influence vegetation and soil patterns. Blue oak woodlands represent a large component of the terrestrial vegetation in California and provide an opportunity to study the historical variation in NPP by combining dendrochronological measurements and a satellite-based proxy for NPP: the remotely sensed time-integrated NDVI (iNDVI). The purpose of this research was to investigate the possible correlation of iNDVI to blue oak chronology measurements to reconstruct iNDVI of the blue oak woodlands from 1700 to 2003. We developed linear regressions to explain iNDVI using two types of blue oak chronologies, elevation, slope, aspect, cation exchange capacity, and ecological sections and compared our results to reconstructed Palmer Drought Severity Index (PDSI). A spatial scale study was done to test for the optimal scale at which iNDVI can be predicted. Our reconstructions were able to predict iNDVI (Adjusted R-squared = 0.715; $p < 0.001$) and were also correlated to PDSI values ($r = 0.8$; $p < 0.001$), finding notably high and low values of iNDVI at 1789 and 1934, respectively. We also found that the residual chronology and a spatial scale of 3136 square kilometers predicted iNDVI most accurately. These

findings produce a novel time-series of iNDVI that can be used to determine how historical vegetation productivity has fluctuated and sequestered carbon in the blue oak woodlands and provides a framework for global reconstructions of vegetation productivity.

3.2 Introduction

The carbon dynamics of terrestrial ecosystems play a critical role in the Earth and global climate systems (Heimann and Reichstein, 2008; Rockström et al., 2009; Keeling and Shertz, 1992). The dominant processes that influence these dynamics are the creation or destruction of organic matter through two main critical chemical reactions: photosynthesis and respiration. Vast amounts of carbon are stored in living vegetation and soil organic matter through photosynthesis and carbon is lost in the form of CO_2, primarily through respiration. The rate of accumulation of carbon in the biosphere is referred to as net primary productivity (NPP, g m^{-2} yr^{-1}) (Leith and Whittaker, 1975) and is considered a critical measurement that describes the habitability of Earth (Running, 2012). NPP is subject to various environmental controls that influence the amount of carbon that is either stored or released into the atmosphere, changing over spatial and temporal scales. Goward et al. (1985) found that NPP is linearly correlated to iNDVI ($r = 0.97$), a temporal integration of NDVI, showing that spectral vegetation index measurements from satellite imagery provide a consistent method to conduct global vegetation studies.

Drylands represent approximately 47.2% of terrestrial ecosystems (Hillel and Rosenzweig, 2002; Middleton and Thomas, 1992) and may play a vital role in reducing atmospheric CO_2 through increased NPP (Lal, 2004; Poulter et al., 2014). However, little is known about the role of dryland carbon dynamics in the Earth system relative to other ecosystems (Schimel, 2010). Even so, drylands constitute a major component in reducing the rate of atmospheric CO_2 through soil organic and inorganic carbon (Lal, 2002). Desertification and degradation reduce the ability for drylands to store carbon and may even lead to CO_2 emission. California and various parts of the southwestern U.S. are composed of an arrangement of dryland ecosystems that are subject to desertification and degradation that may result in increased emissions of carbon.

Satellite imagery and spatially explicit environmental data of terrestrial vegetation productivity are becoming increasingly available for use in geographical studies. Before the 1980s, studies of NPP were limited to small scale field plots (Running et al., 2004), but the development of these remotely sensed data has allowed for regional based quantitative studies of vegetation productivity and carbon sequestration (Chen et al., 2012; D'Arrigo et al., 2000; Fuentes et al., 2006). Recent improvements of the Advanced Very High Resolution Radiometer (AVHRR) dataset by Pinzon and Tucker (2014) can now provide accurate NDVI products from 1982 to 2012. This data product can be further integrated to iNDVI for use in addressing numerous questions regarding the carbon dynamics of drylands, an understudied landscape in carbon research and an important topic for global change. Furthermore, the development of historical

carbon dynamics of drylands can provide a robust understanding of terrestrial carbon research. To address this, we use a combination of dendrochronological measurements of tree-ring widths and modern satellite imagery.

Tree-ring indices can serve as proxies for local environmental NDVI characteristics of terrestrial environments and the rate of CO_2 sequestration (D'Arrigo et al., 1987). NDVI approximates the physiological status of trees expressed through tree vigor and tree growth (Kaufmann et al., 2004). D'Arrigo et al. (2000) correlated maximum latewood density measurements from tree rings with NPP measurements derived from NDVI in the boreal forest, suggesting that maximum latewood density may be an approximate index for production. iNDVI also was found to be highly correlated with the tree-ring widths of Meyer spruce (*Picea meyeri* Rehd. et Wils.) and can serve as a proxy for above-ground biomass (Liang et al., 2005). Another study by He and Shao (2006) found relationships between maximum value composite NDVI images to chronology values in Delingha, China, showing that NDVI has a good relationship with the above ground biomass of grassland regions. The challenge is to determine which spatial scale NDVI provides the highest degree of association with tree-ring data, although one study by Wang et al. (2004) shows NDVI of grassland landscapes in Kansas were correlated best at approximate 50 km^2. The type of chronology used in these correlations is also unclear as many studies do not describe the type of tree-ring data used in detail.

The blue oak woodlands ecosystem covers a large extent of the drylands in California and provides a spatially extensive and unique opportunity for carbon

studies. Blue oak (*Quercus douglasii* Hook & Arn.) is a deciduous tree species endemic to California and forms one of the major oak woodland community types in the state (Figure 2.2) (Barbour et al., 2007). Although blue oak woodlands are increasingly being used for agricultural and urban development (Stahle et al., 2013), dominant blue oak trees remain wide-spread in California with ages ranging from 150 to 400 years old (Stahle et al., 2013). Disturbances of the fire regime (Swetnam et al., 2009), the invasion of nonnative grasses (Seabloom et al., 2003), and livestock grazing (Mensing, 1992) have also not had a major influence on the mean growth or inter-annual variability of blue oak trees (Stahle et al., 2013). Lastly, blue oak trees are sensitive recorders of the environment and have been used in numerous reconstructions of precipitation, salinity, and drought (Gervais, 2006; Meko et al., 2011; Stahle et al., 2013).

Our study aims to investigate whether a functional relationship exists between contemporary satellite-based measurements of iNDVI and dendrochronological measurements of blue oak growth. By including covariates of elevation, slope, aspect, cation exchange capacity, and ecological section data, we wish to reconstruct historical iNDVI maps from 1700 to 2003 using dendrchronological measurements of blue oak trees. We have three research goals: 1) Which chronology type shows the highest degree of association with iNDVI, 2) which spatial scale provides the highest degree of association between tree-ring indices and iNDVI, 3) will reconstructed historical maps of iNDVI correspond to similar reconstructions of drought in California, and 4) how does iNDVI fluctuate over time and what are potential implications of these

changes? The third question concerns the historical spatial reconstruction of the Palmer Drought Severity Index (PDSI) maps that were derived from tree-ring data (Cook et al., 2004). I hypothesize spatial and temporal connections exist between historically low iNDVI values with reconstructed Palmer Drought Severity Indices developed by Cook et al. (1999). This study uses approximately 22 years of NDVI data and 304 years of blue oak chronology data, thus concerning itself with the temporal variation of iNDVI, but including covariates that exist in geographical space will allow for some spatial analysis.

This research will provide novel maps of reconstructed iNDVI from 1700 to 2003. These maps will give insight to how the carbon dynamics of drylands varied over space and time. Moreover, we will establish a technical precedent for other reconstructions of iNDVI as our work will determine the most effective chronology type and spatial scale to use for reconstructions of iNDVI. Lastly, we aim to correspond our work to another well established reconstruction of PDSI by Cook et al. (2004). This correspondence will allow our work to be validated against a well known environmental reconstruction. Although our work reconstructs iNDVI, we assume a similar relationship exists with NPP.

3.3 Data and Methods

3.3.1 Data

Spatially explicit data for iNDVI, blue oak chronologies, topography, cation exchange capacity, and ecological section data were assembled for the state of California. Selection of these factors was based on a conceptual theory that explains how iNDVI is associated with tree growth values using the other data layers as covariates in a multivariate regression. We also assume that these variables stay relatively constant throughout time, meaning that our constructed model will remain consistent when reconstructing iNDVI. All variables were from direct data products (Table 3.1) excluding the slope and aspect layer.

3.3.2 Blue oak Chronologies

Blue oak chronologies were acquired from the International Tree-Ring Database (ITRDB). Only locations with a complete residual and ARSTAN chronology files were used. Both chronology types are composed of standardized, unitless values, with the residual chronology values processed using autoregressive modeling and the ARSTAN reincorporating the pooled autoregression (Cook, 1985). We chose these chronology types because they are commonly used as temperature and precipitation proxies. Chronologies were further filtered to those that had a minimum sample

Table 3.1: The spatial data set used in this study to reconstruct historical integrated normalized difference vegetation index (iNDVI) or net primary productivity from 1700 to 2003.

Name	**Units or Type**	**Original Resolution**	**Source**	**Notes**
Chronologies	Vector shapefile	Points	International Tree-ring Data Bank (ITRDB)	Blue oak chronologies that have a sample depth of at least 10 between years 1700 to 2003
NDVI	Spectral index	64 km^2	NDVI3g	Developed by Pinzon and Tucker (2014); accounts for calibration loss, orbital drift, volcanic eruptions, etc.
DEM - Slope - Aspect	meters degrees N,S,E,W	1 km^2	Global 30 Arc-Second Elevation (GTOPO30)	Derived from a variety of raster and vector sources of topographic information; used to develop elevation, slope, and aspect layers
CEC	cmol/kg	1 km^2	ISRIC	Global soils data at 1 km^2 resolution; generated at ISRIC World Soil Information
Ecological Sections	Vector shapefile	Polygons	USFS ECOMAP	Developed to encourage ecosystem-based approaches to forest land conservation and management
PDSI	Drought Index	2.5 x 2.5 degrees	(Cook et al., 2004)	Reconstructed Palmer Drought Severity Index values by Cook et al. (2004) to observe periods of drought

depth of 10 series from years 1700 to 2003. Thirty-three chronologies matched these conditions and were used in the study (Appendix A: Table A1).

3.3.3 iNDVI

The NDVI dataset used in this study was the third generation version of the Global Inventory Modeling and Mapping Studies program (GIMMS, (Pinzon and Tucker, 2014)). This dataset, termed NDVI3g, is an atmospherically-corrected 15 day maximum value composite of NDVI at 64 km^2 pixel resolution that was developed from NOAA's Advanced Very High Resolution Radiometer (AVHRR) for the period 1982 to 2011. We used the bimonthly data from years 1982 to 2003, limited by the time-range of our chronology values.

We filtered both datasets from April to September to ensure NDVI values produced during the growing season were captured. NDVI values were integrated per growing season to create a total of 22 iNDVI raster layers. To integrate NDVI values over time, we found the area of the plotted NDVI curve under the temporally composited images (Figure 3.1).

3.3.4 Topography, cation exchange capacity, and ecological sections

The Global 30 Arc-Second Elevation (GTOPO30) digital elevation model was used to develop three topographic layers of California: elevation, slope, and aspect.

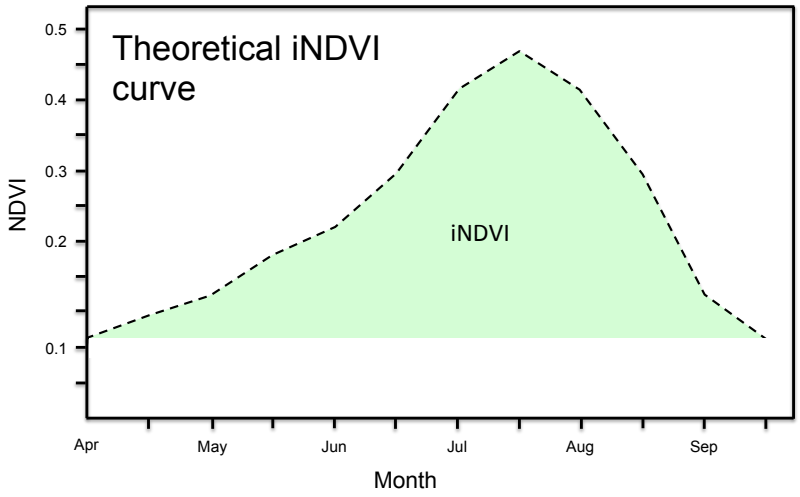

Figure 3.1: A theoretical example of how to calculate iNDVI over a bimonthly time series of NDVI from April to September. The area under the curve is represented in green and is used to calculate iNDVI.

GTOPO30 is a global data set with a horizontal grid spacing of 30-arc seconds (0.0083 degrees; approximately 1 km) and ocean areas are assigned as "no data" and given the value of –9999. To produce the elevation layer, we masked the DEM by the California state boundary and removed values that were less than zero. Slope and aspect were produced by using the *terrain* function found in the *raster* package in R Software (http:// www.r-project.org/) according to algorithms described by Horn (1981). Slope and aspect were calculated in degrees. Aspect was converted to categorical data of north, east, south, or west depending on if the degree value was between 315° to 45°, 45° to 135°, 135° to 225°, or 225° to 315°, respectively.

The cation-exchange capacity (CEC) data was developed using the SoilGrids1km product developed by Hengl et al. (2014) ISRIC/WDC-Soils. The SoilGrids1km is a global 3D soil information system at 1 km^2 spatial resolution containing spatial

predictions for various soil properties. For each soil property, predictions are made for 6 depths (0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm, 100–200 cm) and with a mean and upper and lower limits of a 90% confidence interval. For our study, we used the mean CEC data product estimated for the top 5 cm of soil.

Lastly, we used ecological section data developed by ECOMAP, the USDA Forest Service initiative to map ecological units. Hierarchical regional land classifications exist for resource planning and management, although we used the data as a categorical metric for delineating the unique environments in which the chronologies are located. Ecological sections are broad areas of similar sub-regional climate, geographic processes, stratigraphy, geologic origin, topography, and drainage networks (Cleland et al., 1997).

3.3.5 Data Preprocessing

Extensive effort was taken to ensure data were assembled, created, and projected in a way to preserve data quality and accuracy. All data were projected into an equivalent coordinate system of latitude and longitude values by Geodetic Reference System 1980 (GRS80) using the *rgdal* package in R developed by Bivand et al. (2015). Raster data (i.e. iNDVI, CEC, elevation, slope, and aspect) were reprojected using the nearest neighbor algorithm at the data's initial scale and then resampled to 64 km^2 to reduce error. Vector data (i.e. ecological sections) were converted to raster data by taking the maximum value per 64 km^2 pixel using the *raster* package developed by Hijmans

(2015). All data were clipped by a common California shapefile boundary provided by the U.S. Census Bureau.

3.3.6 Methods

Our analysis was divided into three main components: to determine which chronology type has a higher correlation to iNDVI, using that chronology type to determine which spatial scale has the highest association to iNDVI, and then using the respective chronology type and spatial scale size to reconstruct historical iNDVI maps of California. Conventional ordinary least squares (OLS) multivariate regression was used in this study to model iNDVI as a factor of topography, CEC, and ecological sections as independent variables. This method allowed us to determine how correlated iNDVI is to each chronology type and which spatial scale provides the highest degree of correlation.

A calibration period was established from years 1982 to 1998 to develop our regression models. For each year during the period, the respective chronology values were selected and mapped using the coordinates provided by the ITRDB. Then the overlapping pixel value of the respective iNDVI, topography, CEC, and ecological section layer were extracted at each chronology location. Once pixel values were extracted by the chronology locations, a multivariate linear model was constructed to explain iNDVI using topography, CEC, and ecological sections as independent variables.

We developed a total of five models to explain iNDVI, two to compare chronology types and three to compare spatial resolutions. The first two models were constructed using the initial spatial resolution, 64 km^2, the only difference being that one used ARSTAN chronology values and the second used residual chronology values. An Akaike information criterion (AIC) was used to determine the model that had highest degree of association with iNDVI. We used the chronology type that provided the highest degree of association for the spatial scale study.

To test for spatial scale relationships, three models were developed using similar steps found above but at an incrementally higher spatial resolutions of 576 km^2, 1600 km^2, and 3136 km^2. Coarser resolutions were obtained by using a 3x3, 5x5, and 7x7 window, respectively, and finding the mean values of all raster layers with continuous data. The ecological section layer was coarsened by taking the maximum value found within each window. Once each model was produced, a qualitative comparison was made between the spatial resolution and the Adjusted R^2 statistic to determine the most desired and effective model to use in further studies.

3.3.7 Accuracy Assessment

Various methods exist to test the accuracy of remotely sensed data (Congalton and Green, 2008) and historically reconstructed spatial data (Li et al., 2010). Using the multivariate regression model with the lowest AIC value, we predicted the iNDVI value at each chronology location from 1999 to 2003. These data weren't included in producing the models. We then used a paired t-test that compared the predicted

iNDVI values to the actual satellite-derived iNDVI values from 1999 to 2003 at each chronology location. Our null hypothesis stated that the difference in the means is equal to zero (predicted = actual) while our alternative hypothesis stated that the difference in the means is not equal to zero.

After our reconstructions were developed, we compared our iNDVI reconstructions to actual iNDVI measurements from 1982 to 2003. To do this, we developed a mean relative error map to visualize where out reconstructions differed from actual iNDVI values during this time period.

3.3.8 Reconstruction of historical iNDVI

Using the model of the appropriate chronology type and spatial resolution, we reconstructed historical iNDVI of California from 1700 to 2003. An empty raster grid of California was developed at the given spatial resolution. iNDVI values were predicted at each chronology location from 1700 to 2003 and values were applied to the overlapping grid cell. To interpolate iNDVI values between chronology locations, regressional cokriging was used to take advantage of intermediate data between chronology locations. We chose kriging over other methods of interpolation (such as splining) because kriging allows for higher accuracy (Dubrule, 1984). For each year, we ran regressional cokriging between grid cells with iNDVI values. Ecological section and aspect data were not included in the kriging procedure because they were not continuous integers.

To better visualize fluctuations in the reconstructed iNDVI, anomaly maps were made using the reconstructed iNDVI layers. We found the mean iNDVI value of each pixel from 1700 to 2003, subtracting the pixel's given iNDVI value from its mean value for each year. If the pixel value of an anomaly map was close to zero, then the iNDVI value of that given year is close to average. If the pixel value is less than zero or greater than zero, then the iNDVI value was lower or higher than average, respectively. Anomaly maps are at the same spatial and temporal resolution as our standard reconstructed iNDVI maps.

3.3.9 Validation to reconstructed PDSI

The Palmer Drought Severity Index (PDSI) is used to estimate relative dryness using temperature and precipitation data and has been used to signify long-term drought as an anthropologically related global-warming phenomena (Dai et al., 2004). We compared reconstructed historical iNDVI to the PDSI developed by Cook et al. (2004) using linear regression to determine their relationship.

To do this, we downloaded the yearly reconstructed PDSI maps from 1700 to 2003 and reprojected them to match the iNDVI layer. Because the PDSI data were developed in a 2.5° x 2.5° resolution grid, we aggregated the iNDVI layer by taking a 19 x 19 cell window and applying the mean value of those cells to new the larger cell. We then resampled each iNDVI layer to match the PDSI layer using the nearest neighborhood algorithm. We then found the mean value of iNDVI and PDSI for each year and ran a linear regression model using these values.

3.4 Results

The 33 blue oak chronologies were mapped across California (Figure 3.2). The mapped chronologies showed a similar spatial distribution as that found by Little and Viereck (1971), stretching from the Sierra Nevada-Cascade-Cast Range foothill of the San Joaquin Valleys, although the Bear Creek Canyon (BCC) chronology is distant from the general geographic region.

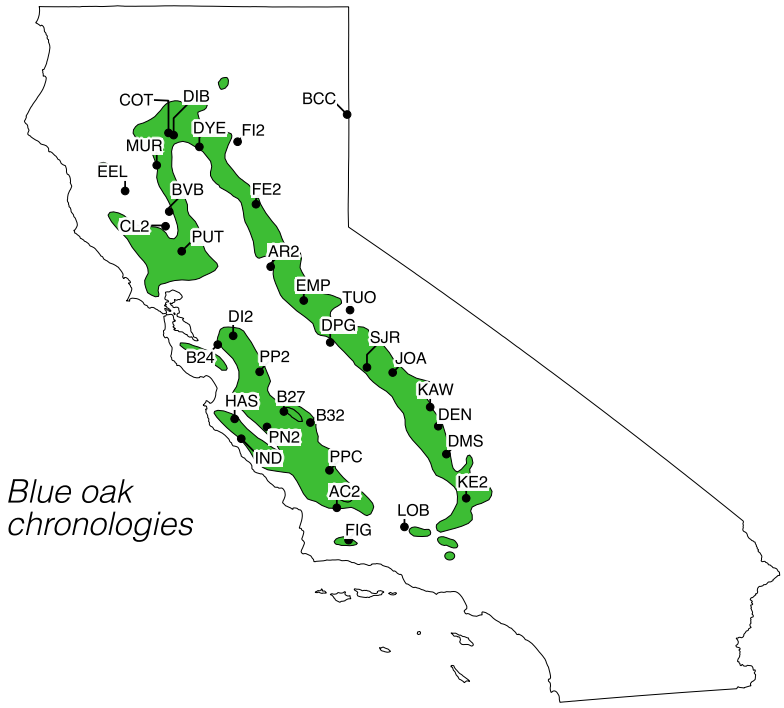

Figure 3.2: The blue oak chronologies used to reconstruct iNDVI from 1700 to 2003. Chronology locations were from the International Tree-ring Data Bank and the green polygon distribution map is from Little and Viereck (1971).

Elevation, aspect, slope, CEC, and ecological section data were formatted into maps, each at 64 km^2 spatial resolution (Figure 3.3). Aspect was mapped in degrees

but was converted to north, south, east, or west values before being used in analyses. Black points on the maps are chronology locations.

Chronologies are located across a wide variety of topography in California. The mean elevation value that the chronologies were located in was 731 meters with a range from 69 to 2132 meters. The mean slope value that the chronologies were located in is 5 degrees but range from 0 degrees to 11 degrees. The mean chronology CEC value was 22.08 cmol/kg and ranged from 12.51 to 34.37 cmol/kg. Seven chronologies were located on east facing slopes, 21 were on south facing slopes, and five were on west facing slopes. Two chronologies are located in ecoregion 261A (Central California Coast), one chronology in 261B (Southern California Coast), one chronology in 262A (Great Valley), one chronology in 342B (Northwestern Basin and Range), one chronology in M261B (Northern California Coast Ranges), six chronologies in M261C (Northern California Interior Coast Ranges), one chronology in M261D (Southern Cascades), five chronologies in M261E (Sierra Nevada), seven chronologies in M261F (Sierra Nevada Foothills), and eight chronologies in M262A (Central California Coast Ranges).

3.4.1 Chronology type association with iNDVI

When constructing the multivariate linear regression models, we modified the variables to determine the possibility of increasing our Adjusted R^2 value. We found that removing aspect while using the remaining variables to explain iNDVI increased

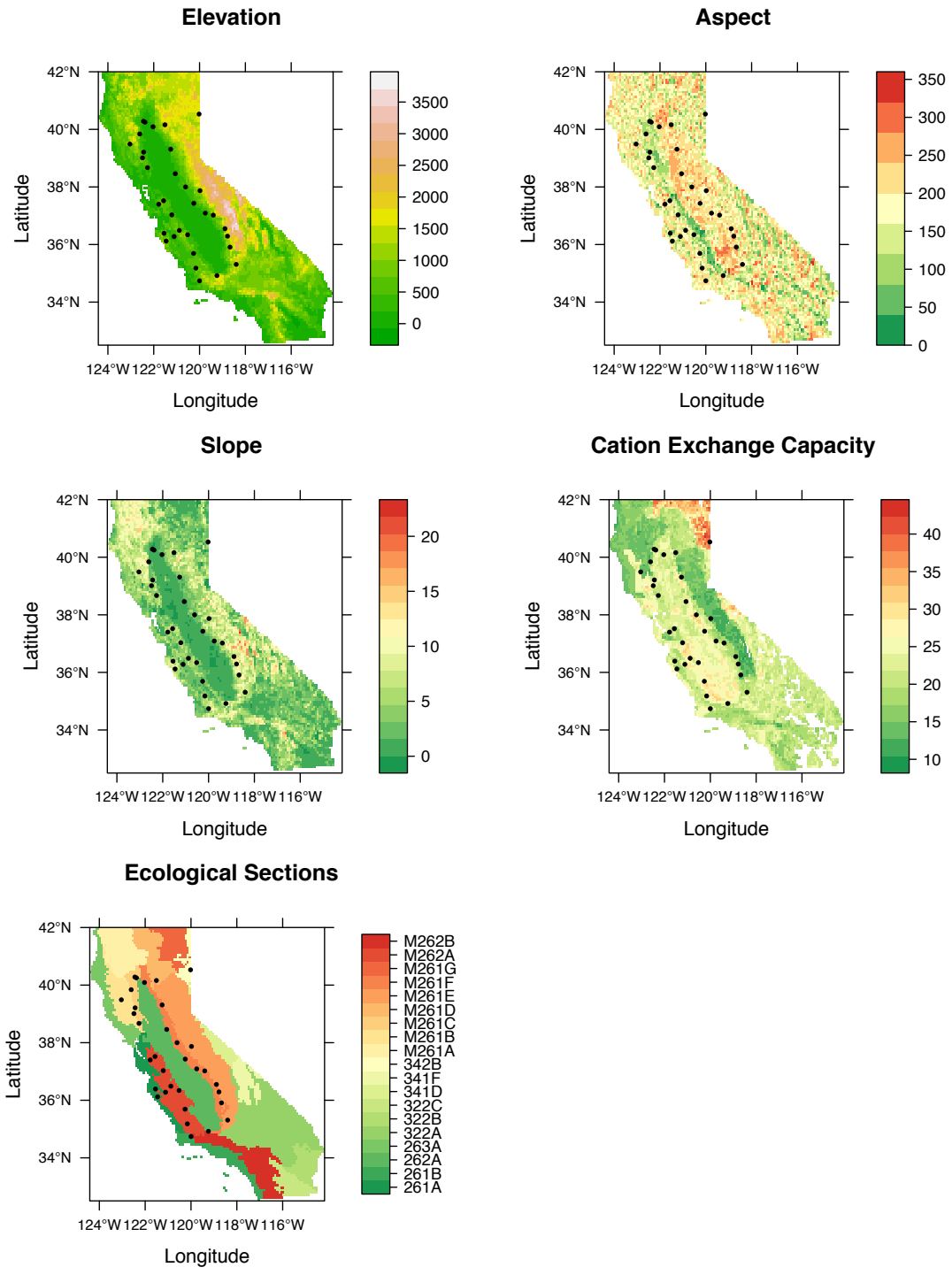

Figure 3.3: The data used to explain iNDVI: elevation (meters), aspect (degrees), slope (degrees), cation exchange capacity (cmol/kg), and ecological sections. The black points are chronology locations.

the R^2 value of both models. Therefore, we only used slope, elevation, CEC, ecological section, and chronology values to explain iNDVI in our regression analysis.

Both models had similar statistical output and coefficient significance. For each model, increases in elevation ($p < 0.001$) and CEC ($p < 0.001$) had a significant negative relationship with iNDVI while increases in slope had a positive relation ($p < 0.001$) to iNDVI. The model using residual chronology values (Adjusted $R^2 = 0.715$) explained the same amount of variance as did the model using the ARSTAN chronology value (Adjusted $R^2 = 0.714$) (Appendix: Table A2). Furthermore, all ecological sections had a significantly negative relationship with iNDVI except for section M261B where the Eel River chronology was located. The model with the residual chronology value had a slightly lower AIC value (1316.863) compared to ARSTAN chronology model (1318.327). Residual chronology values were used to produce all further results including the spatial scale analysis, reconstructing iNDVI, and comparison of the reconstructed iNDVI values to the reconstructed PDSI values.

3.4.2 Accuracy Assessment

We found no significant difference between predicted iNDVI values ($\overline{X} = 6.543$, SD = 1.183) and actual iNDVI values ($\overline{X} = 6.553$, SD = 1.45) from 1999 to 2003 at chronology locations. The paired t-test results had a t-value of 0.15443 and a p-value of 0.8775. Given this, we fail to reject the null hypothesis that the means of our predicted values differ from the actual values. We can say with certain confidence that we can accurately predict iNDVI values at chronology locations.

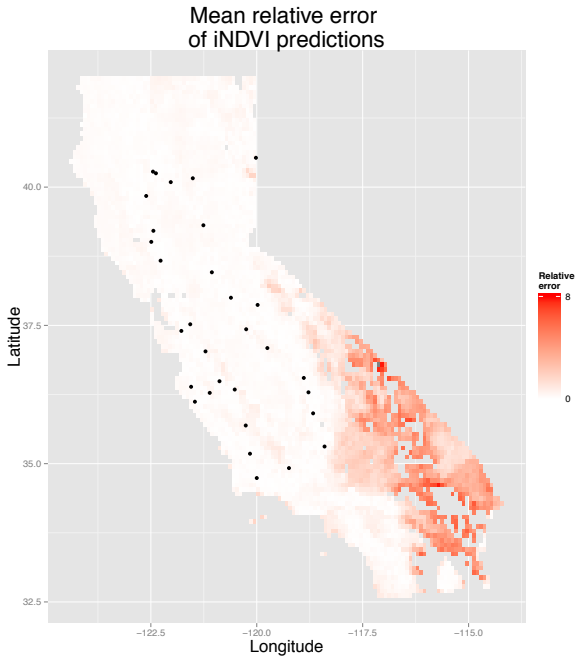

Figure 3.4: The mean relative error of iNDVI reconstructions from 1982 to 2003. Higher relative error is expressed as red and the black points are chronology locations.

Lastly, we found high predictive accuracy in the regions of California near our chronology locations with low relative mean error (Figure 3.4). We found low predictive accuracy in the arid regions of south western California with high mean relative error.

3.4.3 Spatial Scale Analysis

The spatial scale analysis produced three models to explain iNDVI at increasingly larger spatial resolutions of 576 km^2, 1600 km^2, and 3136 km^2 and compared them to the base model of 64 km^2 (Appendix: Table A3). The Adjusted R^2 values of the 576 km^2, 1600 km^2, and 3136 km^2 model were 0.746, 0.700, and 0.799, respectively, while the base 64 km^2 model had an Adjusted R^2 value of 0.715. The ecological sections

used within each model varied depending on spatial scale. The 576 km^2 model used sections 261B, 342B, M261B, M261C, M261E, M261F, M262A, and M262B. The 1600 km^2 model used sections M261B, M261C, M261E, M261F, M261G, M262A, and M262B. The 3136 km^2 model used M261C, M261F, M261G, M262A, and M262B. The covariates of our models expressed varying significance values. All covariates in the 576 km^2 model were significant except for Section 261B, Section M261C, and CEC. All covariates in the 1600 km^2 model were significant except for Section M261B. All covariates in the 3136 km^2 model were significant except for Section M261G. We chose to use the 64 km^2 model to reconstruct iNDVI because we want to develop reconstructions at a finer spatial scale.

3.4.4 Reconstruction of historical iNDVI

We produced 304 maps of reconstructed growing season iNDVI from years 1700 to 2003*. To display these data in a condensed format, we developed a 9 year simple moving average smoothed time series of iNDVI using the mean iNDVI value per year (Figure 3.5). iNDVI fluctuated around a mean of 6.677 $\pm$ 0.094 with the largest peak of iNDVI occurring in 1789 and the largest dip occurring in 1934. We selected these years to display individually as reconstructed iNDVI maps and anomaly maps (Figures 3.6 and 3.7).

*A complete set of the reconstructed iNDVI maps are located in the online thesis attachments (File 1)

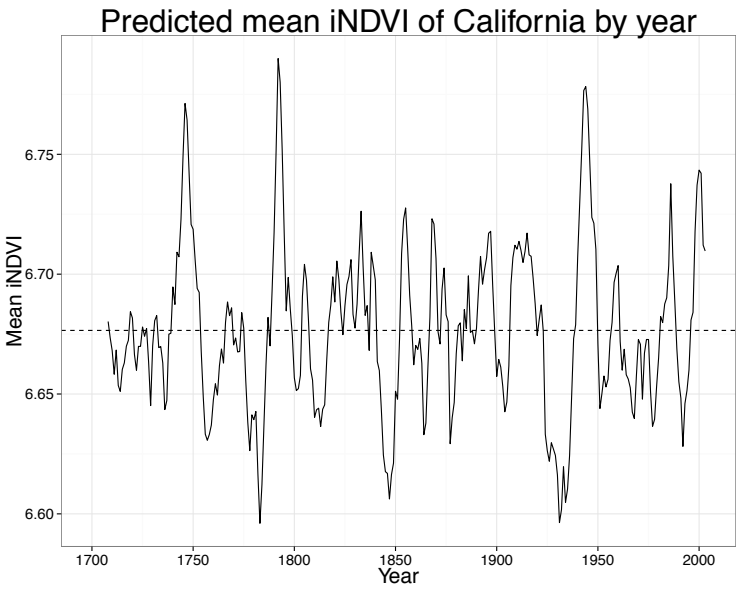

Figure 3.5: The 9 year simple moving average smoothed time series of predicted mean iNDVI from 1700 to 2003. The dashed horizontal line represents the mean value of iNDVI throughout the entire time series.

The 1789 reconstructed iNDVI and anomaly maps (Figure 3.6) show California displaying higher than average values of iNDVI throughout all regions except for portions of the Sierra Nevada mountain range. The 1934 reconstructed iNDVI and anomaly maps (Figure 3.7) shows California displaying lower than average values of iNDVI throughout all regions except for the western edge of the central valley.

3.4.5 Comparison to reconstructed PDSI

A significant linear relationship exists between mean reconstructed PDSI and mean reconstructed iNDVI (Figure 3.8). The output of our regression analysis showed that for every one unit increase in PDSI, there was an increase of iNDVI by 0.037 (Adjusted $R^2 = 0.6388$; $p < 0.001$).

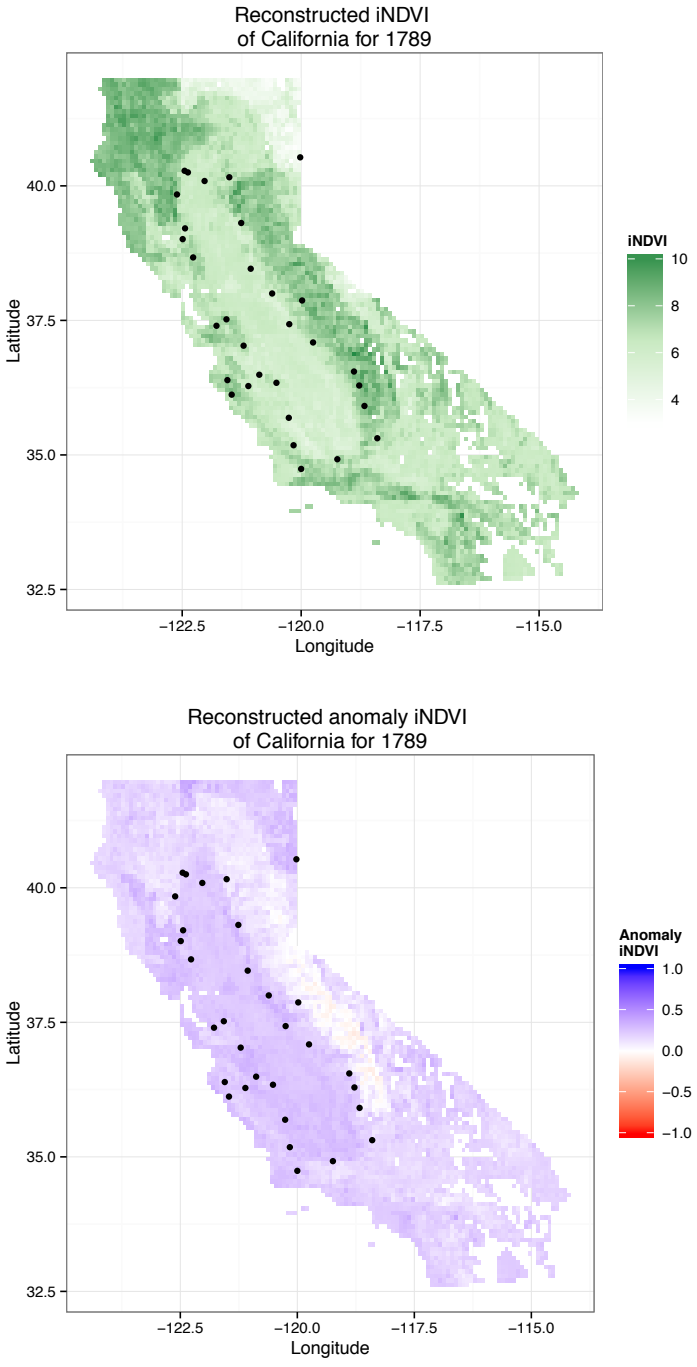

Figure 3.6: The reconstructed iNDVI and the anomaly iNDVI of California for 1789 that shows relatively high iNDVI values. Black points are chronology locations.

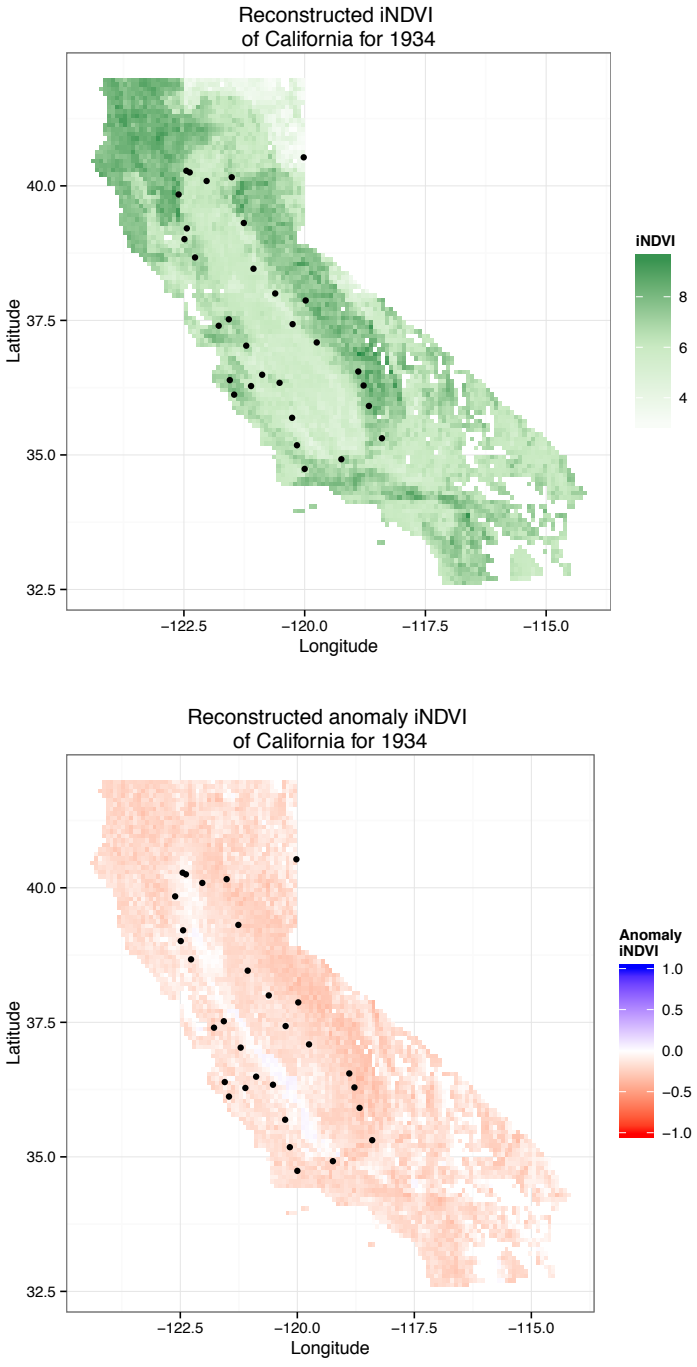

Figure 3.7: The reconstructed iNDVI and the anomaly iNDVI of California for 1934 that shows relatively low iNDVI values. Black points are chronology locations.

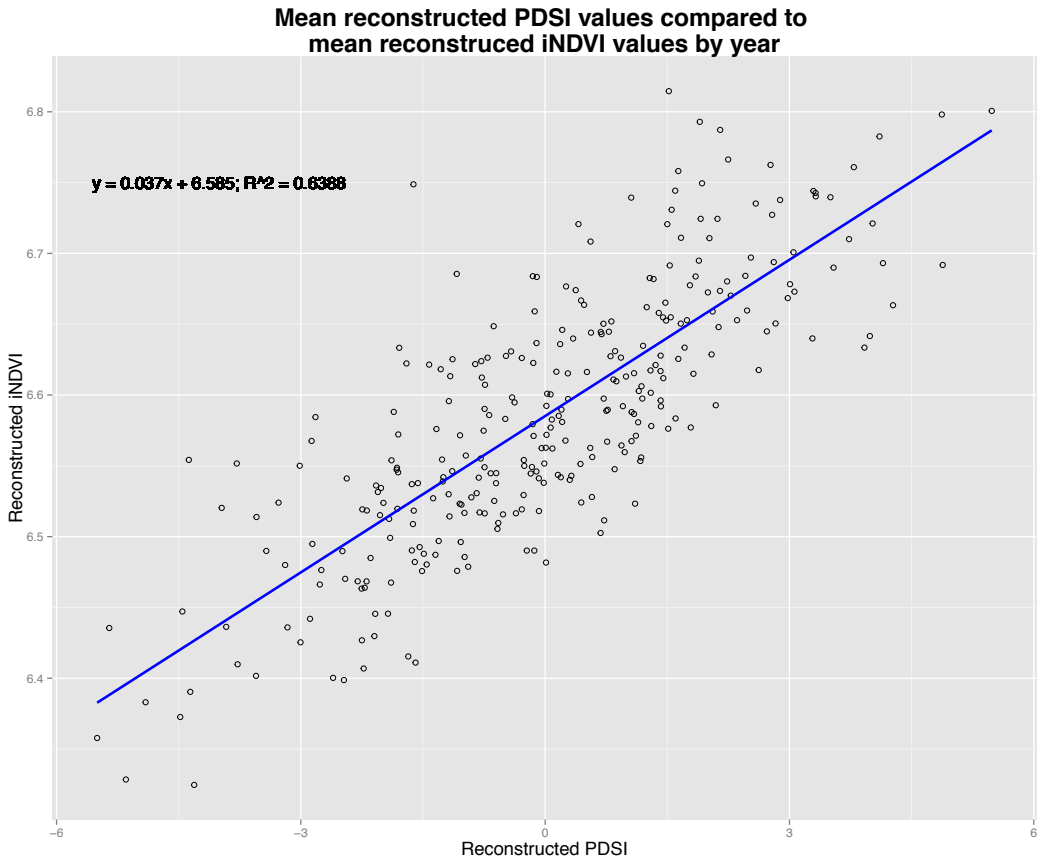

Figure 3.8: The linear regression between reconstructed PDSI values produced by Cook et al. (2004) to reconstructed iNDVI values.

Because of the results shown in Figure 3.5, we developed Figures 3.9 and 3.10 to compare iNDVI and PDSI for years 1789 and 1934, respectively. Visually, each map contains 7 grid cells of values that overlay the state of California. The PDSI map of 1789 displays high values among northern grid cells while the iNDVI map displays similar results but is more varied in its values (Figure 3.9). The PDSI map for 1934 displays low values throughout the state while similar results are found in the iNDVI map (Figure 3.10).

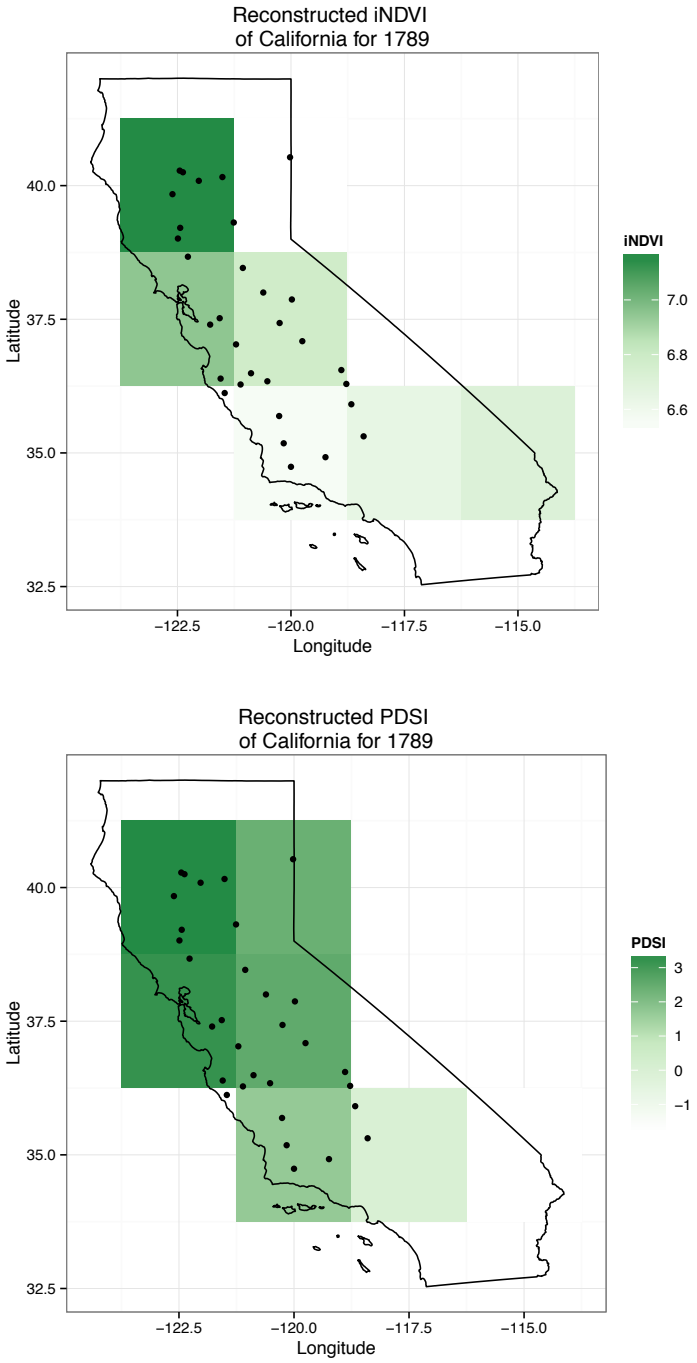

Figure 3.9: A comparison of reconstructed iNDVI to reconstructed PDSI values of California for 1789.

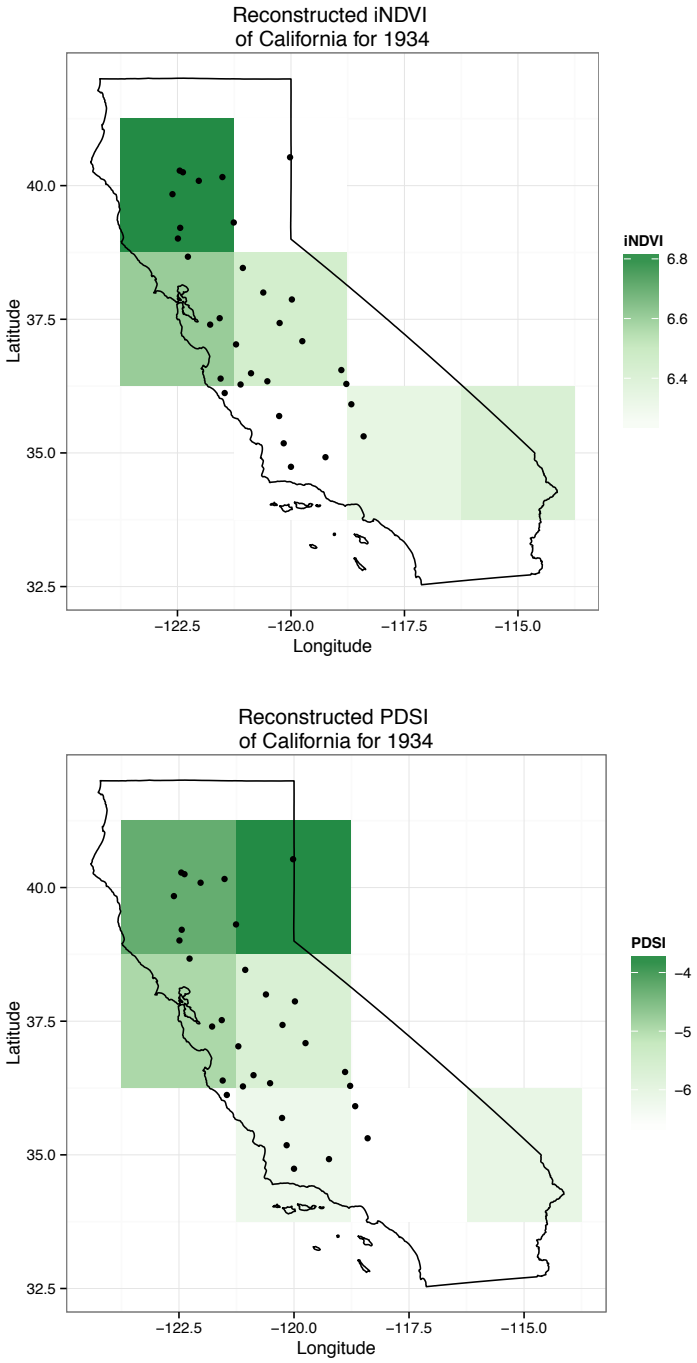

Figure 3.10: A comparison of reconstructed iNDVI to reconstructed PDSI values of California for 1934.

3.5 Discussion and Conclusions

3.5.1 Chronology type association with iNDVI

Our analysis suggests that residual chronology values have a higher association with iNDVI than ARSTAN chronology values by using slope, elevation, CEC, and ecological section data as covariates in a linear regression model. Aspect was not included in the regression model because it reduced the Adjusted R^2 value, potentially due to errors of coarsening degree values. However, the difference between the final model's Adjusted R^2 value is small and may not be critical to develop historical iNDVI. Our AIC values were also very similar because we didn't change the complexity of the models, as only the chronology types used changed. Furthermore, Wang et al. (2004) simply used the standardized chronology to correlate with NDVI with successful results.

Both chronologies are related to iNDVI which is highly associated with the physical growth of the tree. Alternatively, residual and ARSTAN chronologies have been used to reconstruct varying climate and environmental factors that may not be as highly associated with tree-ring width. However, this study only used blue oak trees in a limited geographical area. Extending this study to other regions with different species and environmental variables may alter which chronology is more associated to iNDVI.

Other chronology metrics can be also be used related to iNDVI which may provide more significant results. Our study strictly used tree-ring width, but similar work by

D'Arrigo et al. (2000) used maximum latewood density of annual tree rings. These metrics may also be used to produce estimates of carbon storage.

3.5.2 Accuracy Assessment

iNDVI was predicted with significant accuracy at chronology locations from years 1999 to 2003. However, we did not distinguish between chronologies, so we're unable to determine if certain chronology locations have a higher accuracy than others. We assume that this level of accuracy exists throughout our entire reconstruction time-frame between years 1700 to 2003. Fortunately, our covariates of slope, elevation, CEC, and ecological section values theoretically remain stable during our reconstruction period which allows for some confidence that our current model will be temporally consistent. Lastly, our mean relative error map showed high predictive accuracy for most of the state and low predictive accuracy in the arid portion of south western California. We believe there is higher mean relative error here because our model doesn't include data in from region. Furthermore, as distance from our chronology locations increase, the error of our reconstructions increase.

3.5.3 Spatial Scale Analysis

Our spatial scale analysis is necessary when comparing spatially explicit iNDVI values to chronology values that express large-scale relationships with their environment. The spatial extent at which our individual chronologies represent is unknown and

the coordinate provided by the ITRDB is only in a generalized area. To remediate this, we ran a similar technique as did Wang et al. (2004) to adjust the window size. We found that the spatial scales at which the model produced the highest to lowest adjusted R^2 value are at 3136 km^2, 576 km^2, 64 km^2, and 1600 km^2, respectively. Although the most coarse spatial scale used in this study produced the highest R^2 value, there is no clear gradient or pattern.

When assessing the spatial scale results, we chose to reconstruct iNDVI at 64 km^2 resolution. Considering the trade-off between fine resolution iNDVI data and the overall effectiveness of the model, we determined that the 64 km^2 model was suitable. However, if other studies can allow for more coarse resolution, then choosing to reconstruct iNDVI at 3136 km^2 will be more accurate at chronology locations. Other tree species and geographical areas may produce different results.

3.5.4 Reconstruction of historical iNDVI

Over the course of our reconstruction period, iNDVI is shown to have been dynamic throughout the blue oak woodlands and the state of California (Figure 3.5). When observing the reconstructed iNDVI maps for 1789 and 1934 (Figures 3.9 and 3.10), we see the spatial patterns of how iNDVI fluctuates throughout the state. We found periods of high iNDVI values that could be associated with an increase in vegetation productivity. This, in turn, could be associated with an increase in terrestrial carbon storage. Alternatively, periods of low productivity were found. A large period of low iNDVI occurs around 1934 where written records of a severe drought and lack

of vegetation growth known as the Dust Bowl occurred in the region. The anomaly maps clearly show this deviance from the mean.

However, our methods for interpolating iNDVI between chronology locations does contain uncertainty. Regressional cokriging is not capable of accurately estimating iNDVI as distance increases from a chronology location. Furthermore, we were not able to include categorical variables in the regressional cokriging process, leaving out ecological section data as a covariate in the interpolation. Future studies can include more chronologies of other tree species to reduce the interpolation distance and alternative methods for interpolation like indicator cokriging can be used to include categorical variables.

Our reconstructions of iNDVI provide an insight to the historical carbon dynamics of drylands. Understanding how biological productivity of terrestrial vegetation is a concern for the development of human activity (Running et al., 2000; Running, 2012). Melillo et al. (1993) describes the importance of terrestrial vegetation as fundamental to humans as plant life supports the majority of food, fuel, and resources. Historical reconstructions of iNDVI provide a more robust data set that can be used for understanding the relationships of climate on vegetation productivity and what should be expected in the future given the current phenomena of global warming and drought. Furthermore, historical iNDVI maps give a novel view on how human activity may be altering the historical trajectory of carbon dynamics and will allow for a more refined approach to remediate potentially negative impacts.

3.5.5 Comparison to reconstructed PDSI

The significant correlation ($r = 0.8$; $p < 0.001$) between reconstructed yearly mean iNDVI and PDSI (Figure 3.8) shows that drought is linked to a decrease in iNDVI. This can signal a relationship between dry conditions and decreasing rates of terrestrial carbon storage while wet conditions may signal increasing rates of vegetation growth and carbon storage.

Some issues arise when comparing reconstructed PDSI and reconstructed iNDVI. We observed an edge effect when re-sampling the 64 km^2 values of iNDVI which altered the output values near the eastern side of California. Also, both PDSI and iNDVI were developed using tree-ring chronologies, so we expect comparable results when developing the reconstructed metrics. However, we were able to develop a similar reconstruction of iNDVI using a single tree species compared to PDSI in which Cook et al. (2004) used multiple tree species.

3.5.6 Future research

We successfully reconstructed iNDVI from 1700 to 2003 to further understand the carbon dynamics of blue oak drylands in California. This reconstructed data set can be used as a proxy for NPP and converted into estimated weights of carbon allocation. However, future work will be required to adapt our current results.

Our model to reconstruct iNDVI used data from 1982 to 1998, excluding any variation that may exist outside of that time-line, such as invasive species and urban

development. Invasive terrestrial plant species have been prolific in California since European settlement; 9.2 million hectares of land have been invaded by exotic annual grasses and forbs (Swetnam et al., 2009). Urban development has occurred rapidly throughout the state, removing vegetation from the landscape. Both of these factors alter actual iNDVI values but will not be present in reconstructed iNDVI maps. Including the spread of invasive species and urban development as covariates in the reconstructed iNDVI maps will allow for more accurate estimates. Lastly, ground-based measurements of atmospheric conditions by the flux-tower network can be implemented to refine our relationships of NDVI, tree-ring width, and carbon storage. The flux-tower network is an expanding collection of sensors that offer field-based measurements of carbon sequestration. Integrating these measurements with satellite-based imagery can serve as calibration points for estimating carbon storage.

Chapter 4

Conclusions and Future Research

4.1 Major Conclusions

This study provides a reconstruction of yearly growing season iNDVI from 1700 to 2003 in California using the blue oak woodlands as a case study. My objectives were to refine technical methodologies to reconstruct iNDVI and to provide a correlation to a well-known reconstructed drought metric, PDSI, developed by Cook et al. (2004). Overall, my thesis is the first of its kind to develop historical iNDVI by combining dendrochronological measurements, satellite-derived NDVI products, and an assortment of environmental factors including topography, CEC, and ecological sections. Furthermore, my thesis expands on the knowledge of carbon dynamics of drylands, an understudied topic of terrestrial carbon dynamics.

Relative to other measurements of the environment such as temperature and precipitation, little research has been done in reconstructing historical iNDVI using

dendrochronology and remote sensing. Although Wang et al. (2004) observed variations of spatial scale, they did not test different chronology types and their degree of correlation to iNDVI. D'Arrigo et al. (2000) also ran their analysis based on 1° x 1° NDVI, but didn't test for variations of larger spatial scales. Because of this lack of uniformity between studies, two of my research questions determine the optimal chronology type and spatial scale to use for reconstructing iNDVI. Then to integrate my research into the broad spectrum of environmental reconstructions using dendrochlogical measuremnts, I correlated my reconstructions to PDSI values. This correlation will also provide a brief example of the effect of drought on vegetation productivity which can be associated with a decline of carbon sequestration by terrestrial vegetation.

My first research question was to determine which chronology type shows the highest degree of association with iNDVI. Despite having close values, I found that the residual chronology type has a higher adjusted R^2 value to iNDVI compared to the ARSTAN chronology. This finding sets a precedent on which chronology type should be used in further research to reconstruct iNDVI. Also interesting is the fact that both chronology types produced a relatively high adjusted R^2 value. The similarities of these values may be a reflection that iNDVI is a measurement of vegetation productivity, which tree-ring width is a fairly direct measurement. However, the local geographical nature of my thesis is limiting in how it could apply to other studies. Varying geographic regions and other tree species may influence which chronology

type is more associated with iNDVI due to differences in topography, climate, and physiology.

The second research question I asked was to determine the spatial scale that provides the highest degree of association between tree-ring indices and iNDVI. Finding the optimal spatial scale at which iNDVI can be predicted was necessary to develop accurate reconstructions. Furthermore, I found it necessary to determine the spatial extent at which chronology reflect their local iNDVI values. I found that the largest spatial scale used in this study, 3136 km^2, can most accurately predict iNDVI. However, increasing spatial scale does not consistently increase our degree of correlation to iNDVI. The models also can not be directly associated with each other as they are each developed using different spatial scales. I made the decision to use 64 km^2 to reconstruct iNDVI even though the model had a slightly lower adjusted R^2 value. This decision was based on having the ability to reconstruct iNDVI at fine scales which can provide more information at local levels than coarse scale information.

My third research question was to compare reconstructed iNDVI and reconstructed PDSI. When comparing yearly iNDVI and PDSI over the entire state of California I found a significant correlation between reconstructed PDSI and iNDVI, linking drought to a decrease in iNDVI. This can signal a relationship with drought and a decrease in terrestrial carbon sequestration while wet conditions may allow for an increase in vegetation growth and carbon storage. It's interesting to note that I found a significant relationship between iNDVI and PDSI even though the iNDVI

reconstructions used less tree-ring chronologies and were only from blue oak species. This shows some evidence that local studies using fewer data can provide the same results without the need for large-scale data-sets used in Cook et al. (2004). An edge effect occurred when coarsening iNDVI to match the PDSI resolution which may diminish the linear relationship. However, both data-sets are constructed using tree-ring width chronologies which may artificially inflate the true relationship between these two variables.

My fourth research question was to ask why historical iNDVI should be reconstructed for drylands. It is well understood that the productivity of terrestrial vegetation plays a critical role for humankind (Melillo et al., 1993; Running, 2012). Running et al. (2000) states that biological productivity is fundamental to the habitability of Earth. Given the necessity of vegetation productivity on human development, reconstructions of historical iNDVI provide an insight to the drivers vegetation productivity and how terrestrial ecosystems respond to certain conditions. Negative impacts of these conditions on agricultural, range, and forest products can be detrimental for human development on a global scale.

In conclusion, my work expands on the knowledge of carbon dynamics in dryland ecosystems by providing reconstructed iNDVI maps from 1700 to modern day, giving insight to how terrestrial vegetation productivity has led to the sequestration of carbon throughout this time period. This is a novel approach to determining historical carbon dynamics of drylands by taking advantage dendrochronological measurements, satellite-derived NDVI products, and an assortment of environmental

factors including topography, CEC, and ecological sections. Notable cases of high iNDVI values occur in 1789 and low iNDVI values occur in 1934, insinuating a period of high carbon sequestration and low carbon sequestration, respectively. By interpolating my data using regressional cokriging, it's also possible to interpret carbon dynamics outside of dryland ecosystems allowing for the possibility of extending this study to a larger regional scale. A potential weakness in this study is how I omitted other chronologies from different tree species that could be used in reconstructing iNDVI. iNDVI measurements were also never converted to more direct measurements of carbon storage such as NPP, which would be necessary when quantifying carbon storage. Lastly, I selected factors on which to model iNDVI on based on theoretical rationale without testing for correlated or confounding variables.

4.2 Future Research

The next stage for future research will be to improve estimates of reconstructed carbon dynamics and to expand the geographical area of study. By refining modeling estimates, altering statistical techniques to more appropriate methods, and including a larger data-set, I will be able to reconstruct global measurements of iNDVI and convert those values into NPP (g m^{-2} yr^{-1}). Distinguishing how dryland carbon dynamics differ from other ecosystems can also provide some insight as to how drylands differ in periods of drought or productivity. Lastly, I wish to incorporate data into my models that provide other facets of reconstruction. FLUXNET data

can be used as ground-truth measurements of CO_2 sequestration. The flux-tower network provides local atmospheric conditions which can be used to calibrate satellite imagery and allow for more refined models. Using NDVI products from the Moderate Resolution Imaging Spectroradiometer (MODIS) can develop carbon dynamic maps at a much finer resolution than current AVHRR NDVI3g data.

Bibliography

Allen, B. H. (1989). *Rangeland Cover Type Descriptions for California Hardwood Rangelands: Review Draft*. California Department of Forestry and Fire Protection. 8

Barbour, M., Keeler-Wolf, T., and Schoenherr, A. (2007). *Terrestrial vegetation of California*. University of California Press, Berkeley, Los Angeles, London, third edition. 2, 6, 7, 8, 9, 19

Barrett, R. H. (1979). Mammals of california oak habitats: management implications. In *Proceedings of the Symposium on the Ecology, Management, and Utilization of California oaks*, pages 275–291. 6

Bartolome, J. W. (1987). California annual grassland and oak savannah. *Rangelands*, pages 122–125. 6

Bivand, R., Keitt, T., and Rowlingson, B. (2015). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 1.0-4. 25

Bolsinger, C. L. (1988). The hardwoods of california's timberlands, woodlands, and savannas. Resource Bulletin PNW-RB-148, U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. 6

Campbell, J. and Wynne, R. (2011). *Introduction to Remote Sensing, Fifth Edition*. Guilford Publications, New York. 2, 10

Chen, Z. J., Li, J. B., Fang, K. Y., Davi, N. K., He, X. Y., Cui, M. X., Zhang, X. L., and Peng, J. J. (2012). Seasonal dynamics of vegetation over the past 100 years inferred from tree rings and climate in Hulunbei'er steppe, northern China. *Journal of Arid Environments*, 83:86–93. 17

Cleland, D. T., Avers, P. E., McNab, W. H., Jensen, M. E., Bailey, R. G., King, T., and Russell, W. E. (1997). National hierarchical framework of ecological units. In Boyce, M. and Haney, A., editors, *Ecosystem management applications for sustainable forest and wildlife resources*, pages 181–200. Yale University Press: New Haven, CT, USA. 25

Congalton, R. G. and Green, K. (2008). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. CRC Press, Boca Raton. 27

Cook, E. R. (1985). *A time series analysis approach to tree-ring standardization*. PhD thesis, Tuscon, Arizona. 21

Cook, E. R., Meko, D. M., Stahle, D. W., and Cleaveland, M. K. (1999). Drought reconstructions for the continental united states. *Journal of Climate*, 12(4):1145–1162. 20

Cook, E. R., Woodhouse, C. A., Eakin, C. M., and Meko, D. M. (2004). Long-Term Aridity Changes in the Western United States. *Science*, 306:1015–1018. x, 4, 5, 20, 22, 29, 39, 46, 48, 51

Dai, A., Trenberth, K. E., and Qian, T. (2004). A global dataset of palmer drought severity index for 1870-2002: Relationship with soil moisture and effects of surface warming. *Journal of Hydrometeorology*, 5(6):1117–1130. 29

D'Arrigo, R., Jacoby, G. C., and Fung, I. Y. (1987). Boreal forests and atmosphere-biosphere exchange of carbon dioxide. *Nature*, 329(6137):321–323. 12, 18

D'Arrigo, R. D., Malmstrom, C. M., Jacoby, G. C., Los, S. O., and Bunker, D. E. (2000). Correlation between maximum latewood density of annual tree rings and NDVI based estimates of forest productivity. *International Journal of Remote Sensing*, 21(11):2329–2336. 12, 13, 17, 18, 43, 49

Dubrule, O. (1984). Comparing splines and kriging. *Computers & Geosciences*, 10(2):327–338. 28

FRAP (2003). Forest and range assessment. the changing California. Assessment Summary. October 2003. 6, 7

Fuentes, D. A., Gamon, J. A., Cheng, Y., Claudio, H. C., Qiu, H.-l., Mao, Z., Sims, D. A., Rahman, A. F., Oechel, W., and Luo, H. (2006). Mapping carbon and water vapor fluxes in a chaparral ecosystem using vegetation indices derived from aviris. *Remote Sensing of Environment*, 103(3):312–323. 17

Garrison, B. (1996). Vertebrate wildlife species and habitat associations. In Standiford, R. B. and Tinnin, P., editors, *Guidelines for Managing California's Hardwood Rangelands*. University of California, Division of Agriculture and Natural Resource, Oakland. 6

Gervais, B. R. (2006). A three-century record of precipitation and blue oak recruitment from the Tehachapi Mountains, Southern California, USA. *Dendrochronologia*, 24(1):29–37. 10, 19

Goward, S. N., Tucker, C. J., and Dye, D. G. (1985). North American vegetation patterns observed with the NOAA-7 advanced very high resolution radiometer. *Vegetatio*, 64(1):3–14. 11, 16

Griffin, J. R. (1973). Valley oaks-the end of an era. *Fremontia*, 1(1):5–9. 6

Griffin, J. R. (1977). Oak woodland. In Barbour, M. and Major, J., editors, *Terrestrial Vegetation of California*, pages 383–416, New York. John Wiley & Sons. 6

Griffin, J. R. and Critchfield, W. B. (1972). The distribution of forest trees in california. *USDA Forest Service Research Paper, Pacific Southwest Forest and Range Experiment Station*, (PSW-82). x, 8

He, J. and Shao, X. (2006). Relationships between tree-ring width index and NDVI of grassland in Delingha. *Chinese Science Bulletin*, 51(9):1106–1114. 12, 18

Heimann, M. and Reichstein, M. (2008). Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature*, 451(7176):289–292. 1, 16

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G., Walsh, M. G., et al. (2014). SoilGrids1km-global soil information based on automated mapping. *PLoS One*, 9(8):e105992. 24

Hijmans, R. J. (2015). *raster: Geographic Data Analysis and Modeling*. R package version 2.5-2. 25

Hillel, D. and Rosenzweig, C. (2002). Desertification in relation to climate variability and change. *Advances in agronomy*, 77:1–38. 2, 17

Holland, V. (1980). Effect of blue oak on rangeland forage production in central california. General Technical Report PSW-44, U.S. Dept. of Agriculture, Berkley, CA. 9

Holmes, T. H. (1990). Botanical trends in northern california oak woodland. *Rangelands*, 12(1):3–7. 6

Horn, B. K. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, 69(1):14–47. 24

Hunt, E. R., Martin, F. C., and Running, S. W. (1991). Simulating the effects of climatic variation on stem carbon accumulation of a ponderosa pine stand: comparison with annual growth increment data. *Tree Physiology*, 9(1-2):161–171. 12, 13

Kaufmann, R. K., D'Arrigo, R. D., Laskowski, C., Myneni, R. B., Zhou, L., and Davi, N. K. (2004). The effect of growing season and summer greenness on northern forests. *Geophysical Research Letters*, 31(9):L09205. 12, 13, 18

Keeling, R. F. and Shertz, S. R. (1992). Seasonal and interannual variations in atmospheric oxygen and implications for the global carbon cycle. *Nature*, 358(6389):723–727. 1, 16

Kerr, J. T. and Ostrovsky, M. (2003). From space to species: ecological applications for remote sensing. *Trends in Ecology & Evolution*, 18(6):299–305. 10

Lal, R. (2002). Carbon sequestration in dryland ecosystems of west Asia and north Africa. *Land Degradation & Development*, 13(1):45–59. 17

Lal, R. (2004). Carbon sequestration in dryland ecosystems. *Environmental Management*, 33(4):528–544. 2, 17

Leith, H. and Whittaker, R. (1975). *Primary Productivity of the Biosphere*. Springer, New York. 1, 10, 16

Li, B., Fang, X., Ye, Y., and Zhang, X. (2010). Accuracy assessment of global historical cropland datasets based on regional reconstructed historical dataa case study in northeast china. *Science China Earth Sciences*, 53(11):1689–1699. 27

Liang, E. Y., Shao, X. M., and He, J. C. (2005). Relationships between tree growth and NDVI of grassland in the semiarid grassland of north China. *International Journal of Remote Sensing*, 26(13):2901–2908. 13, 18

Little, E. L. and Viereck, L. A. (1971). *Atlas of United States trees*, volume 5. US Dept. of Agriculture, Forest Service. x, 30

Malmstrm, C. M., Thompson, M. V., Juday, G. P., Los, S. O., Randerson, J. T., and Field, C. B. (1997). Interannual variation in global-scale net primary production: Testing model estimates. *Global Biogeochemical Cycles*, 11(3):367–392. 12

McClaran, M. P. and Bartolome, J. W. (1985). The importance of oak to ranchers in the california foothill woodland. *Rangelands*, pages 158–161. 7

Meko, D. M., Stahle, D. W., Griffin, D., and Knight, T. A. (2011). Inferring precipitation-anomaly gradients from tree rings. *Quaternary International*, 235(12):89–100. 3, 10, 19

Melillo, J. M., McGuire, A. D., Kicklighter, D. W., Moore, B., Vorosmarty, C. J., and Schloss, A. L. (1993). Global climate change and terrestrial net primary production. *Nature*, 363(6426):234–240. 45, 51

Mensing, S. A. (1992). The impact of European settlement on blue oak (*Quercus douglasii*) regeneration and recruitment in the Tehachapi Mountains, California. *Madroño*, pages 36–46. 9, 19

Middleton, N. and Thomas, D. (1992). *World Atlas of Desertification: United Nations Environmental Programme*. Arnold. 2, 17

Neal, D. L. (1980). Blue oak-digger pine. *Forest cover types of the United States and Canada*, pages 126–127. 8

Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., and Stenseth, N. C. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology & Evolution*, 20(9):503–510. 2, 10

Pinzon, J. E. and Tucker, C. J. (2014). A non-stationary 1981–2012 AVHRR NDVI3g time series. *Remote Sensing*, 6(8):6929–6960. 17, 22, 23

Poulter, B., Frank, D., Ciais, P., Myneni, R. B., Andela, N., Bi, J., Broquet, G., Canadell, J. G., Chevallier, F., Liu, Y. Y., et al. (2014). Contribution of semi-arid ecosystems to interannual variability of the global carbon cycle. *Nature*, 509(7502):600–603. 2, 17

Prince, S. (1991). Satellite remote sensing of primary production: comparison of results for sahelian grasslands 1981-1988. *International Journal of Remote Sensing*, 12(6):1301–1311. 11

Ritter, L. (1988). Blue oak woodland vegetation. In Mayer, K. and Laudenslayer Jr., W., editors, *A Guide to Wildlife Habitats of California*, Sacramento. State of California, Resources Agency, Department of Fish and Game. 8

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin, F. S., Lambin, E. F., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al. (2009). A safe operating space for humanity. *Nature*, 461(7263):472–475. 1, 16

Rouse Jr, J., Haas, R., Schell, J., and Deering, D. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In *Third ERTS Symposium, NASA*, volume 351, pages 309–317, Washington, D.C. NASA, U.S. Govt. Printing Office. 10

Running, S. W. (2012). A measurable planetary boundary for the biosphere. *Science*, 337(6101):1458–1459. 1, 3, 16, 45, 51

Running, S. W., Nemani, R. R., Heinsch, F. A., Zhao, M., Reeves, M., and Hashimoto, H. (2004). A continuous satellite-derived measure of global terrestrial primary production. *BioScience*, 54(6):547–560. 2, 10, 17

Running, S. W., Thornton, P. E., Nemani, R., and Glassy, J. M. (2000). Global terrestrial gross and net primary productivity from the earth observing system. In Sala, O., Jackson, R., and Mooney, H., editors, *Methods in Ecosystem Science*, pages 44–57. Springer–Verlag. 45, 51

Schimel, D. S. (2010). Drylands in the earth system. *Science*, 327(5964):418–419. 2, 17

Seabloom, E. W., Harpole, W. S., Reichman, O., and Tilman, D. (2003). Invasion, competitive dominance, and resource use by exotic and native California grassland species. *Proceedings of the National Academy of Sciences*, 100(23):13384–13389. 9, 19

Sellers, P. J. (1985). Canopy reflectance, photosynthesis and transpiration. *International Journal of Remote Sensing*, 6(8):1335–1372. 2, 10

Stahle, D. W., Griffin, R. D., Meko, D. M., Therrell, M. D., Edmondson, J. R., Cleaveland, M. K., Stahle, L. N., Burnette, D. J., Abatzoglou, J. T., Redmond,

K. T., Dettinger, M. D., and Cayan, D. R. (2013). The ancient blue oak woodlands of California: longevity and hydroclimatic history. *Earth Interactions*, 17(12):1–23. 9, 19

Stahle, D. W., Therrell, M. D., Cleaveland, M. K., Cayan, D. R., Dettinger, M. D., and Knowles, N. (2001). Ancient blue oaks reveal human impact on San Francisco Bay salinity. *Eos, Transactions American Geophysical Union*, 82(12):141–145. 3, 4, 9, 10

Standiford, R. B., Tinnin, P., and Adams, T. (1996). Guidelines for managing California's hardwood rangelands. page 180. University of California, Division of Agriculture and Natural Resources. 7

Swetnam, T. W., Baisan, C. H., Caprio, A. C., Brown, P. M., Touchan, R., Anderson, R. S., Hallett, D. J., et al. (2009). Multi-millennial fire history of the giant forest, sequoia national park, california, usa. *Fire Ecology*, 5(3):120–150. 9, 19, 47

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment*, 8(2):127–150. 10

Tucker, C. J., Holben, B. N., Elgin, Jr., J. H., and McMurtrey III, J. E. (1981). Remote sensing of total dry-matter accumulation in winter wheat. *Remote Sensing of Environment*, 11:171–189. 2, 11

Tucker, C. J. and Sellers, P. J. (1986). Satellite remote sensing of primary production. *International Journal of Remote Sensing*, 7(11):1395–1416. 11

Tucker, C. J., Townshend, J. R. G., and Goff, T. E. (1985). African land-cover classification using satellite data. *Science*, 227(4685):369–375. 2, 11

Verner, J. (1979). Birds of california oak habitats–management implications. In *Proceedings of the Symposium on the Ecology, Management, and Utilization of California Oaks*, volume PSW-44, pages 246–264. U.S. Forest Service. 6

Wang, J., Rich, P. M., Price, K. P., and Kettle, W. D. (2004). Relations between NDVI and tree productivity in the central Great Plains. *International Journal of Remote Sensing*, 25(16):3127–3138. 13, 18, 42, 44, 49

Appendix

Appendix A: Tables

Table A1: The total blue oak chronologies available from the ITRDB.

Site ID	Title	ITRDB Code	Author(s)	Study ID
AC2	American Canyon	CA614	Stahle, D.W.; Therrell, M.D.	4809
CL2	Clear Lake State Park	CA615	Stahle, D.W.; Therrell, M.D.	4832
DPG	Don Pedro Reservoir and Green Valley	CA616	Stahle, D.W.	4837
EEL	Eel River	CA617	Stahle, D.W.; Therrell, M.D.	4844
FE2	Feather River Lake Oroville	CA618	Stahle, D.W.; Therrell, M.D.	4853
FI2	Finley Lake	CA619	Stahle, D.W.; Therrell, M.D.	4854
AR2	Folsom Lake State Rec. Area American River	CA620	Stahle, D.W.; Therrell, M.D.	4855
KE2	Kern River	CA621	Stahle, D.W.	4873
DI2	Mt. Diablo State Park	CA623	Stahle, D.W.; Therrell, M.D.; Gabor, D.; Cook, E.R.	4904
PP2	Pacheco Pass State Park	CA625	Stahle, D.W.; Therrell, M.D.	4916
PN2	Pinnacles National Monument	CA626	Stahle, D.W.; Therrell, M.D.	4922

Continued on next page

Table A1 – *Continued from previous page*

Site ID	**Title**	**ITRDB Code**	**Author(s)**	**Study ID**
JOA	San Joaquin River Millerton Lake	CA627	Stahle, D.W.; Therrell, M.D.	4940
B24	Marys Ranch	CA645	Stahle, D.W.;Griffin, R.D.	8559
B27	Rock Springs Ranch	CA646	Stahle, D.W.;Griffin, R.D.	8564
B32	Wright Mountain	CA647	Stahle, D.W.;Griffin, R.D.; Edmondson, J.R.	8571
BCC	Bear Creek Canyon	CA648	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.;Stahle, L.N.	8543
BVB	Bear Valley Buttes	CA649	Stahle, D.W.; Griffin, R.D.; Therrell, M.D.	8544
COT	South Fork Cottonwood Creek	CA650	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8567
DEN	Dennison Peak	CA651	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8549
DIB	Dibble Creek	CA652	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8550
DMS	Dead Mule Saddle	CA653	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8548
DYE	Dye Creek Preserve	CA654	Stahle, D.W.; Griffin, R.D.; Therrell, M.D.	8551
EMP	Empire Creek	CA655	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8552
FIG	Figueroa Mountain	CA656	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8553
HAS	Hastings Reservation	CA657	Stahle, D.W.; Meko, D.; Griffin, R.D.; Edmondson, J.R.	8556
IND	The Indians	CA658	Stahle, D.W.;Griffin, R.D.; Edmondson, J.R.	8568
KAW	North Fork Kaweah River	CA659	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8561
LOB	Los Lobos Creek	CA660	Stahle, D.W.; Meko, D.;Griffin, R.D.;Edmondson, J.R.	8558
MUR	Murphy Ranch	CA661	Stahle, D.W.; Griffin, R.D.; Therrell, M.D.	8560

Continued on next page

Table A1 – *Continued from previous page*

Site ID	**Title**	**ITRDB Code**	**Author(s)**	**Study ID**
PPC	Palo Prieto Canyon	CA662	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8562
PUT	Putah Creek, Lake Berryessa	CA663	Stahle, D.W.; Griffin, R.D.; Therrell, M.D.	8563
SJR	San Joaquin Experimental Range	CA664	Stahle, D.W.; Griffin, R.D.; Edmondson, J.R.	8565
TUO	Tuolumne River	CA665	Stahle, D.W.; Griffin, R.D.; Therrell, M.D.	8569

Table A2: The models used to determine which chronology value was more highly associated with iNDVI

	Dependent variable:	
	indvi	
	ARSTAN chronology	Residual Chronology
	(1)	(2)
CAQ	0.391*** (0.101)	
CRQ		0.419*** (0.103)
Elevation	−0.001*** (0.0002)	−0.001*** (0.0002)
slope	0.233*** (0.029)	0.232*** (0.029)
Section 261B	−1.066*** (0.275)	−1.076*** (0.274)
Section 262A	−1.435*** (0.261)	−1.445*** (0.261)
Section 342B	−3.848*** (0.357)	−3.856*** (0.356)
Section M261B	0.293 (0.246)	0.300 (0.246)
Section M261C	−0.916*** (0.183)	−0.925*** (0.183)
Section M261D	1.612*** (0.322)	1.601*** (0.321)
Section M261E	−1.243*** (0.165)	−1.252*** (0.164)
Section M261F	−1.745*** (0.171)	−1.752*** (0.171)
Section M262A	−2.220*** (0.162)	−2.228*** (0.162)
CEC	−0.041*** (0.011)	−0.041*** (0.011)
Constant	7.954*** (0.328)	7.936*** (0.328)
Observations	557	557
R^2	0.721	0.721
Adjusted R^2	0.714	0.715
Residual Std. Error (df = 543)	0.779	0.778
F Statistic (df = 13; 543)	107.686***	108.080***
Note:		*p<0.05; **p<0.01; ***p<0.001

Table A3: The models used to determine which spatial scale was more highly associated with iNDVI, 64km^2, 576km^2, 1600km^2, or 3136km^2.

	Dependent variable (iNDVI):			
	64 km^2	576 km^2	1600 km^2	3136 km^2
CRQ	0.419*** (0.103)	0.376*** (0.089)	0.312*** (0.076)	0.337*** (0.079)
Elevation	−0.001*** (0.0002)	−0.002*** (0.0002)	−0.0004* (0.0002)	−0.003*** (0.0002)
slope	0.232*** (0.029)	0.466*** (0.028)	0.115*** (0.029)	0.428*** (0.041)
Section 261B	−1.076*** (0.274)	−0.244 (0.238)		
Section 262A	−1.445*** (0.261)			
Section 342B	−3.856*** (0.356)	−1.327*** (0.350)		
Section M261B	0.300 (0.246)	0.481* (0.240)	0.112 (0.206)	
Section M261C	−0.925*** (0.183)	−0.185 (0.194)	−1.365*** (0.167)	−1.164*** (0.174)
Section M261D	1.601*** (0.321)			
Section M261E	−1.252*** (0.164)	1.388*** (0.266)	−1.496*** (0.245)	
Section M261F	−1.752*** (0.171)	−0.730*** (0.188)	−1.277*** (0.166)	−1.127*** (0.172)
Section M261G			−3.003*** (0.369)	0.580 (0.336)
Section M262A	−2.228*** (0.162)	−1.182*** (0.181)	−1.836*** (0.161)	−1.785*** (0.156)
Section M262B		−1.074*** (0.267)	−1.486*** (0.198)	−1.467*** (0.169)
CEC	−0.041*** (0.011)	−0.014 (0.012)	−0.094*** (0.011)	−0.161*** (0.020)
Constant	7.936*** (0.328)	6.216*** (0.390)	9.685*** (0.350)	11.029*** (0.563)
Observations	557	557	557	423
R^2	0.721	0.751	0.706	0.803
Adjusted R^2	0.715	0.746	0.700	0.799
F Statistic	108.080*** (df = 13; 543)	136.949*** (df = 12; 544)	119.061*** (df = 11; 545)	187.421*** (df = 9; 413)

Note: *p<0.05; **p<0.01; ***p<0.001

Appendix B: R-Code

Processing iNDVI

```
library(Hmisc)
library(caTools)
library(raster)
library(rgdal)
library(dplyr)
library(flux)
library(reshape2)

#creates list of .n[sat][-[VI][version]g files
ndvi3glist <- list.files("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ ndvi/ndvi3g/")

#california extent before reprojection
caliextent <- extent(-126.2854, -112.9501, 27.8707, 43.87317)

#NDVI3gToRast Summary:
# 1. load install.packagesd libraries
# 2. create shortened character without extension
# 3. create transposed image of file x
# 4. give transposedimage an extent, crs, and res
# 5. create two rasters, one with flagged values and actual
    ↪ NDVI values
# 6. write both rasters

NDVI3gToRast <- function(x){
  noext <- substring(x, 1, 11)
  transposedimage <- t(read.ENVI(paste("/Users/ETLab/Dropbox/
      ↪ Thesis_Data/ndvi/ndvi3g/", x, sep=""), "/Users/ETlab/
      ↪ Dropbox/Thesis_Analysis/ReproMod/NDVI3gHead.hdr"))
      ↪ %>% raster()
  extent(transposedimage) <- c(-180, 180, -90, 90)
```

```
  crs(transposedimage) <- "+proj=longlat"
  res(transposedimage) <- 1/12
  ndvi3gValues <- floor(transposedimage/10)/1000
  ndvi3gValues <- crop(ndvi3gValues, caliextent)
  ndvi3gValues <- projectRaster(ndvi3gValues, crs=("+proj=
    ↪ longlat +datum=NAD83 +no_defs +ellps=GRS80 +towgs84
    ↪ =0,0,0"), method="ngb")
  ndvi3gFlags <- transposedimage-floor(transposedimage/10)*10
    ↪  + 1
  ndvi3gFlags <- crop(ndvi3gFlags, caliextent)
  ndvi3gFlags <- projectRaster(ndvi3gFlags, crs=("+proj=
    ↪ longlat +datum=NAD83 +no_defs +ellps=GRS80 +towgs84
    ↪ =0,0,0"), method="ngb")
  writeRaster(ndvi3gValues, paste("/Users/ETLab/Dropbox/
    ↪ Thesis_Data/ndvi/ndvi3gProcessed/", noext, sep=""),
    ↪ format="GTiff")
  writeRaster(ndvi3gFlags, paste("/Users/ETLab/Dropbox/Thesis
    ↪ _Data/ndvi/ndvi3gFlags/", noext, "flags", sep=""),
    ↪ format="GTiff")
}

#loop through ndvi3glist with the NDVI3gToRast function
for (i in ndvi3glist){
  NDVI3gToRast(i)
}

#make shapefile of more correcty california polygon from
  ↪ census
calishp <- readOGR(dsn = "/Users/ETLab/Dropbox/Thesis_Data/
  ↪ carto_bound/california", "cali2014")

#create list of rasters
rasters <- list.files("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
  ↪ ndvi3gProcessed/")

#example of grepl from rasters (run through all years)
rasters82 <- rasters[grepl('*geo82.*.tif$', rasters)]

#filter out growing monthes from rasters# to matches#
growingmonths <- c("apr", "may", "jun", "jul", "aug", "sep")

#example of filtering out just growing months (run through
  ↪ all years)
```

```
matches82 <- unique(grep(paste(growingmonths,collapse="|"),
    ↪ rasters82, value=TRUE))

#list of geo[year]
yearrange <- c(paste("geo", (82:99), sep=""))
yearrange1 <- c(paste("geo", 0, (00:03), sep=""))
yearrange <- c(yearrange, yearrange1)

#beginning to reproject
CrMaRepro <- function(x){
  y <- paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
     ↪ ndvi3gProcessed/", x, sep="")
  reprojected <- projectRaster(raster(y), crs= ("+proj=
     ↪ longlat +datum=NAD83 +no_defs +ellps=GRS80 +towgs84
     ↪ =0,0,0"), method="ngb")
  cropped <- crop(reprojected, calishp)
  masked <- mask(cropped, calishp)
  writeRaster(masked, paste("/Users/ETLab/Dropbox/Thesis_Data
     ↪ /ndvi/ndvi3gcali/", x, sep=""), format="GTiff")
}

#run code through loop (continue for all years)
for (i in matches82){
  CrMaRepro(i)
}

#list of cali ndvi
calirast <- list.files("/Users/ETLab/Dropbox/Thesis_Data/ndvi
    ↪ /ndvi3gcali", full.names=TRUE)

#example of extraction (run for all years)
geo82 <- calirast[grepl('*geo82.*.tif$', calirast)]

#create list of data frames
my.list <- list(geo82, geo83, geo84, geo85, geo86, geo87,
    ↪ geo88, geo89, geo90, geo91, geo92, geo93, geo94, geo95,
    ↪  geo96, geo97, geo98, geo99, geo00, geo01, geo02, geo03
    ↪ )

#use assign in loop above
years <- 82:99
years1 <- c(years,paste(0,(00:03), sep=""))
years <- years1
```

```
x <- paste("matches", years, sep="")

#loop through "my.list" to make each list into data frames
my.list <- lapply(my.list, as.data.frame)

#add new column to list of data frames
monthorder <- c(1, 2, 9, 10, 7, 8, 5, 6, 3, 4, 11, 12)

#add month order column to data frame
for( i in seq_along(my.list)){
  my.list[[i]]$month <- cbind(month = monthorder)
}

#sort data frames by month order
sorted.list <- lapply(my.list, function(my.list){
  my.list[order(my.list$month),]
})

#drop month column
drop <- c("month")
sorted.list <- lapply(sorted.list, function(sorted.list){
  sorted.list[,!(names(sorted.list) %in% drop)]})

#change sort.list factors into character lists matching "
   ↪ matches82" name
for (i in 1:22){
  x <- sorted.list[i]
  assign(paste("geo", years[i], sep=""), as.character(x[[1]])
     ↪ )
}

#stack rasters

for (i in 1:22){
 x <- paste("geo")
}

#example of stacking rasters into raster stack (run for all
   ↪ years)
s82 <- stack(geo82)

#example of the AUC function for year of 1982 (run for all
   ↪ years)
```

```
auc82 <- overlay(s82, fun = function(y)((y[[1]]+y[[2]])/2) +
    ↪ ((y[[2]]+y[[3]])/2) + ((y[[3]]+y[[4]])/2) + ((y[[4]]+y
    ↪ [[5]])/2) + ((y[[5]]+y[[6]])/2) + ((y[[6]]+y[[7]])/2) +
    ↪  ((y[[7]]+y[[8]])/2) + ((y[[8]]+y[[9]])/2) + ((y[[9]]+y
    ↪ [[10]])/2) + ((y[[10]]+y[[11]])/2) + ((y[[11]]+y[[12]])
    ↪ /2))

#an example of writing the rasters to the harddrive (run for
    ↪ all years)
writeRaster(auc82, "/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/ndvi3g_indvi/auc1982", format="GTiff")

#stack raster means
stackauc <- stack(auc82, auc83, auc84, auc85, auc86, auc87,
    ↪ auc88, auc89, auc90, auc91, auc92, auc93, auc94, auc95,
    ↪  auc96, auc97, auc98, auc99, auc00, auc01, auc02, auc03
    ↪ )

auc <- list.files("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/ndvi3g_indvi", full.names = TRUE)
stackauc <- sapply(auc, raster)
```

Processing Blue Oak Chronologies

```
#download necessary packages
library("dplR")
library('dplyr')
library('fields')
library('ggplot2')

#if there aren't files listed under "~/CAqudg/data/itrdb",
    ↪ then create directory, download NOAA file, add
    ↪ destination file, and unzip
if(length(list.files("~/Dropbox/Thesis_Analysis/CAqudg/data/
    ↪ itrdb")) == 0){
  dir.create("~/Dropbox/Thesis_Analysis/CAqudg/data/itrdb")
  download.file(url="http://www1.ncdc.noaa.gov/pub/data/paleo
      ↪ /treering/chronologies/itrdb-v705-usa-crn.zip",
                destfile="~/Dropbox/Thesis_Analysis/CAqudg/
                    ↪ data/itrdb-usa.zip")
  unzip("~/Dropbox/Thesis_Analysis/CAqudg/data/itrdb-usa.zip"
      ↪ ,
        exdir="~/Dropbox/Thesis_Analysis/CAqudg/data/itrdb")
```

```
}

#rip through headers to extract values
read.crn.head <- function(fname){
  header <- readLines(fname, n=4)
  crn <- try(dplR::read.crn(fname))
  if(class(crn)[1]!='try-error')
    return(data_frame(site_id=substr(header[[1]],1,6),
                      site_name=substr(header[[1]], 10, 61),
                      species_code=substr(header[[1]], 62,
                        ↪ 65),
                      optional_id=substr(header[[1]], 66, 80)
                        ↪ ,
                      state=substr(header[[2]], 10, 20),
                      species=substr(header[[2]], 20, 39),
                      elevation=substr(header[[2]], 40, 45),
                      lat_lon=substr(header[[2]], 48,57),
                      years=substr(header[[2]], 68, 76),
                      first=row.names(crn)[1],
                      last=tail(row.names(crn),1),
                      fname=fname))
  else return(NULL)
}

#make list of files
files <- list.files("~/Dropbox/Thesis_Analysis/CAqudg/data/
  ↪ itrdb", full.names=TRUE)

#extract arstan files
caqfiles <- files[grepl('*ca.*(a).crn$', files)]
temp <- lapply(caqfiles,read.crn.head)
temp <- do.call(rbind, temp)
df <- temp %>% filter(grepl(lat_lon, pattern='[[:digit
  ↪ :]]{4,4}-[[:digit:]]{5,5}')) %>% filter(last<2015)
df <- df %>% mutate(lat=as.numeric(substr(lat_lon, 1, 2))+as.
  ↪ numeric(substr(lat_lon, 3, 4))/100,
                    lon=as.numeric(substr(lat_lon, 6, 8))+as.
                      ↪ numeric(substr(lat_lon, 9, 10))/
                      ↪ 100)
caqdf <- as.data.frame(df)

#extract residual files
crqfiles <- files[grepl('*ca.*(r).crn$', files)]
```

```
temp <- lapply(crqfiles,read.crn.head)
temp <- do.call(rbind, temp)
df <- temp %>% filter(grepl(lat_lon, pattern='[[:digit
    ↪ :]]{4,4}-[[:digit:]]{5,5}')) %>% filter(last<2015)
df <- df %>% mutate(lat=as.numeric(substr(lat_lon, 1, 2))+as.
    ↪ numeric(substr(lat_lon, 3, 4))/100,
                        lon=as.numeric(substr(lat_lon, 6, 8))+as.
                            ↪ numeric(substr(lat_lon, 9, 10))/
                            ↪ 100)
crqdf <- as.data.frame(df)

#convert lon to negative value
caqdf$lon <- -abs(caqdf$lon)
crqdf$lon <- -abs(crqdf$lon)

#filter by blue oak sites
caqdf.clean <- caqdf[caqdf$species_code=="QUDG",]
row.names(caqdf.clean) <- 1:nrow(caqdf.clean)
crqdf.clean <- crqdf[crqdf$species_code=="QUDG",]
row.names(crqdf.clean) <- 1:nrow(crqdf.clean)

#remove unnecessary columns
drop <- c("optional_id", "state", "species", "lat_lon", "
    ↪ years")
caqdf <- caqdf.clean[,!names(caqdf.clean) %in% drop]
crqdf <- crqdf.clean[,!names(crqdf.clean) %in% drop]

#write csv for mac
write.table(caqdf, file = "~/Dropbox/Thesis_Analysis/CAqudg/
    ↪ data/caq.csv", sep=",")
write.table(crqdf, file = "~/Dropbox/Thesis_Analysis/CAqudg/
    ↪ data/crq.csv", sep=",")
```

Processing elevation, slope, aspect, cation exchange capacity, and ecological sections

```
library(raster)
library(rgdal)
library(sp)
library(dplyr)
```

```
#cartographic bound of california (polygon shapefile)
calishp <- readOGR(dsn = "/Users/ETLab/Dropbox/Thesis_Data/
    ↪ carto_bound/california", "cali2014")

#DEM of western North America
totdem <- raster("/Users/ETLab/Dropbox/Thesis_Data/DEM/comDEM
    ↪ .tif")

#cropped by extent surrounding california
extent <- extent(-127.1159, -113.0545, 31.1542, 44.21073)
cropped <- crop(totdem, extent)

#reproject DEM by 'calishp' coordinate system
repro <- projectRaster(cropped, crs = ("+proj=longlat +datum=
    ↪ NAD83 +no_defs +ellps=GRS80 +towgs84=0,0,0"), method="
    ↪ ngb")

#crop and mask by calishp
calidem <- crop(repro, calishp)
calidem <- mask(calidem, calishp)

#change null values (-9999; generally regarded as coastal
    ↪ values in the GTOP0 30 data set which aren't
# in reality noData ) to values of 0
ocean <- c(-9999)
values(calidem) <- ifelse(values(calidem) %in% ocean, 0,
    ↪ values(calidem))

#focal statistics on 7x7 grid weighted 1
focdem <- focal(calidem, w=matrix(1, 7, 7), fun=mean)

#resample focal statistics 1 km (30 arc second) DEM by AUC
    ↪ iNDVI layer (8 km)
cali8dem <- resample(focdem, auc82, method="ngb")

#california slope
#calislope <- terrain(cali8dem, opt="slope", unit='degrees',
    ↪ neighbors=8)
#calislope1 <- terrain(cali8dem, opt="slope", unit='degrees',
    ↪  neighbors=4)
calislope <- terrain(calidem, opt="slope", unit='degrees',
    ↪ neighbors=8)
focslope <- focal(calislope, w=matrix(1,7,7), fun=mean)
```

```
cali8slope <- resample(focslope, auc82, method="ngb")
writeRaster()

#california aspect
#caliaspect <- terrain(cali8dem, opt="aspect", unit="degrees
    ↪ ", neighbors=8)
caliaspect <- terrain(calidem, opt="aspect", unit="degrees",
    ↪ neighbors=8)

#interruption for new test
caliaspect1 <- terrain(calidem, opt="aspect", unit="radians",
    ↪  neighbors=8)
testsin <- sin(caliaspect1)
testcos <- cos(caliaspect1)
focsin <- focal(testsin, w=matrix(1,7,7), fun=mean)
foccos <- focal(testcos, w=matrix(1,7,7), fun=mean)
focsign1 <- resample(focsin, auc82, method="ngb")
focsoc1 <- resample(foccos, auc82, method="ngb")
newasp <- atan2(focsign1, focsoc1)

#######
focaspect <- focal(caliaspect, w=matrix(1,7,7), fun=mean)
cali8aspect <- resample(focaspect, auc82, method="ngb")

test <- ratify(cali8aspect)

#NA = 5
#north = 1
#east = 2
#south = 3
#west = 4
for (i in 1:length(test@data@values)){
  if (is.na(test@data@values[i])){
    (is.na(test@data@values[i]))
  }
  else if ((test@data@values[i] > 315) & (test@data@values[i]
    ↪  <= 360)) {
    (test@data@values[i]= 1)
  } else if ((test@data@values[i] >= 0) & (test@data@values[i
    ↪ ] < 45)) {
    (test@data@values[i] = 1)
  } else if ((test@data@values[i] >= 45) & (test@data@values[
    ↪ i] < 135)) {
```

```
    (test@data@values[i] = 2)
  } else if ((test@data@values[i] >= 135) & (test@data@values
    ↪ [i] < 225)) {
    (test@data@values[i] = 3)
  } else {(test@data@values[i] = 4)}
}

my.df <- as.data.frame(test, xy=TRUE)

my.color.aspect <- colorRampPalette(brewer.pal(9, "RdYlGn"))
  ↪ (1000)
ggplot(na.omit(my.df)) + geom_raster(aes(x=x,y=y,fill=factor(
  ↪ cardir_cardir))) + theme_bw()
#eco <- aggregate(caliecoreg, fact=7, fun=max)

##################################################CEC
#load packages; if you're just downloading layers from ftp
  ↪ server, I believe you'll only need dplyr, R.utils, and
  ↪ RCurl
library(raster)
library(rgdal)
library(sp)
library(dplyr)
library(RColorBrewer)
library(snow)
library(R.utils)
library(RCurl)

#if there isn't a CEC directory, populate it with CEC layers
  ↪ downloaded from soilgrids ftp server
if(length(list.files("~/Dropbox/Thesis_Data/soils/CEC")) ==
  ↪ 0){
  dir.create("~/Dropbox/Thesis_Data/soils/CEC/")
  url <- "ftp://soilgrids:soilgrids@ftp.soilgrids.org/data/
    ↪ recent/"
  filenames <- getURL(url, userpwd="soilgrids:soilgrids",
    ↪ dirlistonly=TRUE)
  fullfilenames <- paste(url, strsplit(filenames, "\r*\n")
    ↪ [[1]], sep="")
  shortfilenames <- (strsplit(filenames, "\r*\n")[[1]])
  CECfullfiles <- fullfilenames[grepl('*CEC.*.tif.gz$',
    ↪ fullfilenames)]
```

```
  CECshortfiles <- shortfilenames[grepl('*CEC.*.tif.gz$',
    ↪ shortfilenames)]
  for (i in 1:length(CECfullfiles)){
    download.file(url=CECfullfiles[i], destfile=paste("~/
      ↪ Dropbox/Thesis_Data/soils/CEC/", CECshortfiles[i],
      ↪ sep=""))}
  downfile <- list.files("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ soils/CEC", full.names = TRUE)
  names <- substr(downfile, 44, 68)
  for (i in 1:length(downfile)){
    gunzip(filename= (downfile[i]), destname=(paste("/Users/
      ↪ ETLab/Dropbox/Thesis_Data/soils/CEC/", names[i], ".
      ↪ tif", sep="")), ext="gz", fun=gzfile)
  }
}

#focal statistics on 0 - 2.5 cm CEC of california
CEC0_5 <- raster("~/Dropbox/Thesis_Data/soils/CEC/CEC_sd1_M_
  ↪ 02.tif")

#create extent around California
extent <- extent(-127.0128, -111.5567, 29.91211, 45.36822)

#crop by extent around California
ceccrop <- crop(CEC0_5, extent)

#project raster into common projection
cecrepro <- projectRaster(ceccrop, crs = ("+proj=longlat +
  ↪ datum=NAD83 +no_defs +ellps=GRS80 +towgs84=0,0,0"),
  ↪ method="ngb")

#crop by a California shapefile
ceccrop <- crop(cecrepro, calishp)

#mask by a California shapefile to remove background values
CEC0_5 <- mask(ceccrop, calishp)

#take focal means by a 7 x 7 window (7x7 kilometers)
focCEC <- focal(CEC0_5, w=matrix(1, 7, 7), fun=mean)

#reample by the focal means layer
cec0_58 <- resample(focCEC, auc82, method="ngb")
```

```
##############################
### ecoregions
library(raster)
library(rgdal)
library(dplyr)
library(rasterVis)

calishp <- readOGR(dsn = "/Users/ETLab/Dropbox/Thesis_Data/
    ↪ carto_bound/california", "cali2014")
calieco <- readOGR(dsn = "/Users/ETLab/Dropbox/Thesis_Data/
    ↪ veg", "cali_ecoreg")
calieco1 <- calieco[,-(1:11)]
caliecoreg <- rasterize(calieco1, cali8dem, fun="first")
levelplot(caliecoreg, par.settings=RdBuTheme())
```

Combining iNDVI and Chronologies

```
 #required packages
library("gstat")
library("RColorBrewer")
library("raster")
library("dplR")
library("dplyr")
library("plyr")
library("stringr")
library("magrittr")
library("tidyr")
library("animation")
#sort unique headers by reading .csv (WARNING: NEED TO
    ↪ DOWNLOAD DATA FIRST *SEE "DOWNLOAD_ITRDB2.R")
#mac code
caq <- read.csv("~/Dropbox/Thesis_Analysis/CAqudg/data/caq.
    ↪ csv", stringsAsFactors=FALSE, strip.white=TRUE)
crq <- read.csv("~/Dropbox/Thesis_Analysis/CAqudg/data/crq.
    ↪ csv", stringsAsFactors=FALSE, strip.white=TRUE)

#clean up whitespace
caq$site_id <- str_trim(caq$site_id)
caq$species_code <- str_trim(caq$species_code)

crq$site_id <- str_trim(crq$site_id)
crq$species_code <- str_trim(crq$species_code)
```

```
#make list of chronologies
caq_chrono <- as.list(caq$fname)
caq.list <- lapply(caq_chrono, read.crn)
caq.list <- Map(cbind, caq.list, lat = caq$lat)
caq.list <- Map(cbind, caq.list, lon = caq$lon)
caq.list <- Map(cbind, caq.list, ID = caq$site_id)

crq_chrono <- as.list(crq$fname)
crq.list <- lapply(crq_chrono, read.crn)
crq.list <- Map(cbind, crq.list, lat = crq$lat)
crq.list <- Map(cbind, crq.list, lon = crq$lon)
crq.list <- Map(cbind, crq.list, ID = crq$site_id)

#add rownames as column years
for (i in 1:length(caq.list)){
  caq.list[[i]] <- cbind(year = row.names(caq.list[[i]]), caq
    ↪ .list[[i]])
}

for (i in 1:length(crq.list)){
  crq.list[[i]] <- cbind(year = row.names(crq.list[[i]]), crq
    ↪ .list[[i]])
}

#change column names of ARSTAN values to "CAQ"
for (i in 1:length(caq.list)){
  colnames(caq.list[[i]])[2] <- "CAQ"
}

for (i in 1:length(crq.list)){
  colnames(crq.list[[i]])[2] <- "CRQ"
}

#combined list of data frames
MAcaqlist <- do.call(rbind, caq.list)
MAcaqlist$year <- as.numeric(levels(MAcaqlist$year))[
  ↪ MAcaqlist$year]

MAcrqlist <- do.call(rbind, crq.list)
MAcrqlist$year <- as.numeric(levels(MAcrqlist$year))[
  ↪ MAcrqlist$year]
```

```
#filter out all chronology values with samp.depth great than
   ↪ 10
MAcaqlist10 <- filter(MAcaqlist, samp.depth >= 10)

MAcrqlist10 <- filter(MAcrqlist, samp.depth >= 10)


#chrono by year
#chrono482 <- filter(MAlist10, year==1982)

#make list of years to loop through
years <- as.vector(1982:2003)

#loop through MAlist10, find years 1982:2003, extract rows
   ↪ and save into table
for (i in 1:length(years)){
  listchrono <- filter(MAcaqlist10, year == years[i])
  write.table(listchrono, file = paste("/Users/ETLab/Dropbox/
     ↪ Thesis_Data/NOAA/chronobyyear/", "caq", years[i], sep
     ↪ =""), sep = ",", col.names=TRUE)
}

for (i in 1:length(years)){
  listchrono <- filter(MAcrqlist10, year == years[i])
  write.table(listchrono, file = paste("/Users/ETLab/Dropbox/
     ↪ Thesis_Data/NOAA/chronobyyear/", "crq", years[i], sep
     ↪ =""), sep = ",", col.names=TRUE)
}

#loop through MAlist10, assign object in environment by years
   ↪  1982:2003
for (i in 1:length(years)){
  listchrono <- filter(MAcaqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
     ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  assign(paste("caq", years[i], sep=""), x)
}

for (i in 1:length(years)){
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
     ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  assign(paste("crq", years[i], sep=""), x)
```

```
}

#FOR AVHRR iNDVI, extract auc1982:2003 values by chrono1982
    ↪ :2003, make into final.data list
#ARSTAN values
my.list <- list()
years <- as.vector(1982:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
      ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  listchrono <- filter(MAcaqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
      ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58,
      ↪ caliecoreg, auc)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
      ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
      ↪ layer.5"="ecoreg"))
  names(z)[6] <- "indvi"
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/test
      ↪ /aucchrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

caq.final.data <- do.call(rbind, my.list)

#AVHRR RESIDUAL VALUES
my.list <- list()
years <- as.vector(1982:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
      ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
      ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58,
      ↪ caliecoreg, auc)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
```

```
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
    ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
    ↪ layer.5" = "ecoreg"))
  names(z)[6] <- "indvi"
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/test
    ↪ /aucchrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

crq.final.data <- do.call(rbind, my.list)

#MODIS CAQ
my.list <- list()
years <- as.vector(2000:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/mod_indvi/indvi_mod_", years[i], ".tif", sep=""
    ↪ ))
  listchrono <- filter(MAcaqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
    ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  z <- raster::extract(auc, x)
  z <- cbind(z, as.data.frame(x))
  names(z)[1] <- "indvi"
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/test
    ↪ /mod_indvichrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

caq.mod.data <- do.call(rbind, my.list)

#MODIS CRQ
my.list <- list()
years <- as.vector(2000:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/mod_indvi/indvi_mod_", years[i], ".tif", sep=""
    ↪ ))
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
    ↪ listchrono$lat), data = listchrono, match.ID = "ID")
```

```
  z <- raster::extract(auc, x)
  z <- cbind(z, as.data.frame(x))
  names(z)[1] <- "indvi"
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/test
    ↪ /mod_indvichrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

crq.mod.data <- do.call(rbind, my.list)

#aspect to degrees
crq.final.data$aspect.dir <- NULL
for (i in 1:length(crq.final.data$aspect)){
if (crq.final.data$aspect[i] > 315 & crq.final.data$aspect[i]
   ↪ <= 360) {
    (crq.final.data$aspect.dir[i]="N")
  } else if (crq.final.data$aspect[i] > 0 & crq.final.data$
     ↪ aspect[i] < 45) {
    (crq.final.data$aspect.dir[i] ="N")
  } else if (crq.final.data$aspect[i] >= 45 & crq.final.data$
     ↪ aspect[i] < 135) {
    (crq.final.data$aspect.dir[i] = "E")
  } else if (crq.final.data$aspect[i] >= 135 & crq.final.data
     ↪ $aspect[i] < 225) {
    (crq.final.data$aspect.dir[i] = "S")
  } else {
    (crq.final.data$aspect.dir[i] = "W")
}
}
crq.final.data$aspect.dir <- as.factor(crq.final.data$aspect.
   ↪ dir)

caq.final.data$aspect.dir <- NULL
for (i in 1:length(caq.final.data$aspect)){
  if (caq.final.data$aspect[i] > 315 & caq.final.data$aspect[
     ↪ i] <= 360) {
    (caq.final.data$aspect.dir[i]="N")
  } else if (caq.final.data$aspect[i] > 0 & caq.final.data$
     ↪ aspect[i] < 45) {
    (caq.final.data$aspect.dir[i] ="N")
  } else if (caq.final.data$aspect[i] >= 45 & caq.final.data$
     ↪ aspect[i] < 135) {
    (caq.final.data$aspect.dir[i] = "E")
```

```
  } else if (caq.final.data$aspect[i] >= 135 & caq.final.data
    ↪ $aspect[i] < 225) {
    (caq.final.data$aspect.dir[i] = "S")
  } else {
    (caq.final.data$aspect.dir[i] = "W")
  }
}
caq.final.data$aspect.dir <- as.factor(caq.final.data$aspect.
  ↪ dir)

#sub last five years out
crq.data.sub <- subset(crq.final.data, year <= 1998)
caq.data.sub <- subset(caq.final.data, year <= 1998)


crq.lm2 <- lm(indvi~CRQ + dem + aspect.dir + slope + ecoreg +
  ↪  cec, data=crq.data.sub)
caq.lm2 <- lm(indvi~CAQ + dem + aspect.dir + slope + ecoreg +
  ↪  cec, data=caq.data.sub)

crq.lm3 <- lm(indvi~CRQ + dem + slope + ecoreg + cec, data=
  ↪ crq.data.sub)
caq.lm3 <- lm(indvi~CAQ + dem + slope + ecoreg + cec, data=
  ↪ caq.data.sub)


crq.data.sub %>% select(dem, aspect.dir, slope, cec, ecoreg,
  ↪ indvi, CRQ) %>% ggpairs()

values <- subset(crq.final.data, year==1982)

calieco.att <- caliecoreg@data@attributes
```

Regressional Cokriging

```
#data frame for predictions by years 1700 - 2003
predictyears <- as.vector(1700:2003)
my.historiclist <- list()
for (i in 1:length(predictyears)){
  my.predictyears <- as.character(predictyears[i])
  oldcrn <- filter(MAcrqlist10, year == predictyears[i])
  x <- SpatialPointsDataFrame(coords = cbind(oldcrn$lon,
    ↪ oldcrn$lat), data = oldcrn, match.ID = "ID")
```

```
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58,
     ↪ caliecoreg)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
     ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
     ↪ layer.5" = "ecoreg"))
  my.historiclist[[my.predictyears]] <- z
}

historic.data <- do.call(rbind, my.historiclist)
historic.data$ecoreg <- as.factor(historic.data$ecoreg)


#predict indvi for years 1700:2003 and do paired t-test
   ↪ between values
pindvi <- predict(crq.lm.3g, historic.data)
pindvi <- as.data.frame(pindvi)
predict.data <- cbind(pindvi, historic.data)

#ttest
testsub <- filter(predict.data, year >= 1982)
comdata <- cbind(final.data$indvi, testsub)
comdata <- rename(comdata, c("final.data$indvi" = "indvi"))
comdata1 <- comdata[, c("indvi", "pindvi")]
t.test(comdata1$indvi, comdata1$pindvi ,paired=TRUE)

#chrono1700 <- filter(MAlist10, year == predictyears[1])
#coordinates(chrono1700) <- ~lon+lat

predict.1700 <- filter(predict.data, year == predictyears[1])
predict.1700 <- predict.1700[complete.cases(predict.1700),]
coordinates(predict.1700) <- ~lon+lat

envstack <- stack(cali8dem, cali8aspect, cali8slope, cec0_58,
   ↪  caliecoreg)
env.df <- as.data.frame(rasterToPoints(envstack))
coordinates(env.df) <- ~x+y
gridded(env.df) = TRUE
#env.df <- rename(env.df, c("layer.1"="dem", "layer.2"="cec0_
   ↪ 58"))
```

```
env.df <- rename(env.df, replace = c("layer.1" = "dem", "
    ↪ layer.2" = "aspect", "layer.3" = "slope", "layer.4" = "
    ↪ cec", "layer.5"="ecoreg"))
crs(env.df) <- crs(predict.1700)

#removed ecoreg from vario1 // possibly incorrect
vario1 <- variogram(pindvi ~ dem + aspect + slope + cec,
    ↪ predict.1700)
vario1.fit <- fit.variogram(vario1, model=vgm(1, "Sph", 900,
    ↪ 1))

krg.1700 <- krige(pindvi~dem+aspect+slope+cec, predict.1700,
    ↪ env.df, model=vario1.fit)

#colors
my.color <- colorRampPalette(brewer.pal(9, "RdYlGn"))(1000)

saveGIF({for (i in 1:length(predictyears)){
  my.predict.year <- as.character(predictyears[i])
  predict.year <- filter(predict.data, year == predictyears[i
      ↪ ])
  predict.year <- predict.year[complete.cases(predict.year),]
  coordinates(predict.year) <- ~lon+lat
  crs(predict.year) <- crs(env.df)
  vario1 <- variogram(pindvi~dem + aspect + slope + cec,
      ↪ predict.year)
  vario1.fit <- fit.variogram(vario1, model=vgm(1, "Sph",
      ↪ 900,1))
  krg.year <- krige(pindvi~dem + aspect + slope + cec,
      ↪ predict.year, env.df, model=vario1.fit)
  krg.year.rast <- raster(krg.year)
  writeRaster(krg.year.rast, paste("/Users/ETLab/Dropbox/
      ↪ Thesis_Data/ndvi/indvi/predict_indvi/predict", as.
      ↪ character(predictyears[i]), sep=""), format="GTiff",
      ↪ overwrite=TRUE)
  my.plot <- spplot(krg.year.rast, col.regions=my.color, at=
      ↪ seq(0,12,by=.5), scales=list(draw=TRUE), main= paste(
      ↪ "Predicted iNDVI for", as.character(predictyears[i]),
      ↪  sep= " ")), ylab="Latitude", xlab="Longitude")
  print(my.plot)
}}, interval = 0.25, movie.name = "/Users/ETLab/Desktop/krg_
    ↪ all_test6.gif")
```

```
means <- cellStats(tot.stack, 'mean')
means.df <- as.data.frame(means)
means.df$year <- predictyears

#sum of chronologies by year from MAlist10
chrononum<- as.data.frame(table(MAlist10$year))
chrononum$Year <- as.numeric(as.character(chrononum$Var1))
plot(chrononum$Year, chrononum$Freq, xlab="Year", ylab="
    ↪ Number of Chronologies", type = "s")
abline(v=1700)
abline(v=2003)

#setting up time series
tot.rast <- list.files("/Users/ETLab/Dropbox/Thesis_Data/ndvi
    ↪ /indvi/predict_indvi", full.names=TRUE)
tot.rast1 <- lapply(tot.rast, raster)
tot.stack <- stack(tot.rast1)

predictyears <- as.vector(1700:2003)

my.mean <- calc(tot.stack, mean)

for (i in 1:length(tot.rast1)){
  my.diff <- tot.stack[[i]] - my.mean
   writeRaster(my.diff, filename= paste("/Users/ETLab/Dropbox/
      ↪ Thesis_Data/ndvi/indvi/indvi_anomaly/testanomaly", as
      ↪ .character(predictyears[i]), sep=""), format="GTiff",
      ↪  overwrite=TRUE)
}

#anomoly gif
saveGIF({for (i in 1:length(predictyears)){
  my.diff <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/
      ↪ ndvi/indvi/indvi_anomaly/anomaly", as.character(
      ↪ predictyears[i]), ".tif", sep=""))
  my.plot <- spplot(my.diff, col.regions=my.color, at=seq
      ↪ (-1,1,by=.2), scales=list(draw=TRUE), main= paste("
      ↪ Predicted anomaly of iNDVI for", as.character(
      ↪ predictyears[i]), sep= " "), ylab="Latitude", xlab="
      ↪ Longitude")
  print(my.plot)
}}, interval=0.25, movie.name="/Users/ETLab/Dropbox/GEOG_512/
    ↪ Project/finalpres/indvi_anomaly.gif")
```

```
means <- cellStats(tot.stack, 'mean')
means.df <- as.data.frame(means)
means.df$year <- predictyears
means.df1 <- means.df[complete.cases(means.df),]
myts <- ts(means.df1$means, start=c(1700,1), end=c(2003,1),
    ↪ frequency=1)
smooth.myts <- SMA(myts, n=9)
#save this time series for results of iNDVI
plot(smooth.myts, ylab=(list("Mean iNDVI")), xlab=(list("Year
    ↪ ")), main=(list("Predicted mean iNDVI of California by
    ↪ year")))

#plotting for trans class
plot(envstack, y=3, main = c("Elevation", "Aspect", "Slope",
    ↪ "Cation Exchange Capacity"), col = terrain.colors(20),
    ↪ axes=T) + text(x=chrono1982, labels=(chrono1982$ID),
    ↪ cex=.7, halo=TRUE, hw=0.1, vfont=c("sans serif", "bold"
    ↪ ), pos=3, offset=.2) + points(chrono1982, pch = 20)

plot(auc82, main = c("1982 iNDVI"), col=brewer.pal(9, "Greens
    ↪ "), axes=T, zlim=c(0)) + text(x=chrono1982, labels=(
    ↪ chrono1982$ID), cex=.7, halo=TRUE, hw=0.1, vfont=c("
    ↪ sans serif", "bold"), pos=3, offset=.2) + points(
    ↪ chrono1982, pch = 20)

par(mfrow=c(3,3))
par(mar=c(2,2,2,.8))
for (i in 1:length(years)){
  listchrono <- filter(MAlist10, year == years[i])
  listchrono <- SpatialPointsDataFrame(coords = cbind(
      ↪ listchrono$lon, listchrono$lat), data = listchrono,
      ↪ match.ID = "ID")
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
      ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  z = (plot(auc, y=c(9), legend=FALSE,main = as.character(
      ↪ years[i]), maxnl=9, zlim=0, col=brewer.pal(9, "Greens
      ↪ ")) + points(listchrono, pch=16, cex=listchrono$CAQ +
      ↪  .2))
}
par(mfrow=c(1,1))
par(mar=c(5,5,5,5))
x <- residuals.lm(lm.5)
```

```
hist(x, breaks=20, main="Histogram of lm.5 Residuals")

saveGIF({for (i in 1:length(years)){
  listchrono <- filter(MAlist10, year == years[i])
  listchrono <- SpatialPointsDataFrame(coords = cbind(
    ↪ listchrono$lon, listchrono$lat), data = listchrono,
    ↪ match.ID = "ID")
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  z = spplot(auc, col.regions=my.color, at=seq(0,12,by=.5),
    ↪ scales=list(draw=TRUE), main= paste("iNDVI for", as.
    ↪ character(years[i]), sep= " "), ylab="Latitude", xlab
    ↪ ="Longitude", sp.layout=list('sp.points', listchrono,
    ↪  pch=16, cex=listchrono$CAQ + .2, col="black"))
  print(z)
}}, movie.name = "/Users/ETLab/Dropbox/GEOG_512/Project/
  ↪ finalpres/mygif1.gif")

saveGIF({for (i in 1:length(predictyears)){
  listchrono <- filter(MAlist10, year == predictyears[i])
  listchrono <- SpatialPointsDataFrame(coords = cbind(
    ↪ listchrono$lon, listchrono$lat), data = listchrono,
    ↪ match.ID = "ID")
  plot(calishp) + points(listchrono, pch = 16, cex =
    ↪ listchrono$CAQ)
}}, movie.name = "/Users/ETLab/Desktop/fullhist.gif")

writeOGR(listchrono, "/Users/ETLab/Downloads/", "chrono_rgdal
  ↪ ", driver="ESRI Shapefile")
```

PDSI

```
library(rgdal)
library(raster)
library(dplyr)

locs <- read.table("/Users/ETLab/Dropbox/Thesis_Data/pdsi/dai
  ↪ -namerica-grid.txt")
colnames(locs) <- c("lon", "lat")
locs$grid <- (1:286)

pdsi <- read.csv("~/Dropbox/Thesis_Data/pdsi/namerica-pdsi-
  ↪ recs.txt", sep="", na.strings="-99.999")
pdsi.sub <- subset(pdsi, year >= 1700)
```

```
pdsi.rast <- raster()
res(pdsi.rast) <- 2.5
ext <- c(-140, 64, -55, 45)
pdsi.rast <- crop(pdsi.rast, ext)
proj4string(pdsi.rast) <- CRS("+proj=longlat +ellps=GRS80 +
    ↪ towgs84=0,0,0,0,0,0,0 +no_defs")

years <- 1700:2003
for (i in 1:length(years)){
  by.year <- subset(pdsi.sub, year == years[i])
  by.year <- t(by.year)
  by.year <- by.year[-1,] %>% as.data.frame()
  colnames(by.year) <- paste("pdsi", years[i], sep="")
  by.year$grid <- (1:286)
  joined <- merge(by.year, locs, by="grid")
  joined$grid <- NULL
  coordinates(joined)=~lon+lat
  proj4string(joined) <- CRS("+proj=longlat +ellps=GRS80 +
      ↪ towgs84=0,0,0,0,0,0,0 +no_defs")
  gridded(joined) = TRUE
  r <- raster(joined)
  projection(r) <- CRS("+proj=longlat +ellps=GRS80 +towgs84
      ↪ =0,0,0,0,0,0,0 +no_defs")
  writeRaster(r, filename=paste("/Users/ETLab/Dropbox/
      ↪ Thesis_Data/pdsi/cook/pdsi", years[i], sep=""),
      ↪ format="GTiff")
}

#first aggregate predicted iNDVI values to resolution similar
    ↪  to PDSI
my.pindvi <- list.files("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ ndvi/indvi/predict_indvi/", full.names = TRUE)

cook.compare <- list()
for (i in 1:length(predictyears)){
  x <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
      ↪ indvi/predict_indvi/predict", predictyears[i], ".tif
      ↪ ", sep=""))
  x <- aggregate(x, fact=19, fun=mean)
  pdsi.temp <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data
      ↪ /pdsi/cook_original/pdsi", predictyears[i], ".tif",
      ↪ sep=""))
```

```
  x <- resample(x, pdsi.temp)
  x <- crop(x, calishp)
  x <- mask(x, calishp)
  pdsi.temp <- crop(pdsi.temp, calishp)
  pdsi.temp <- mask(pdsi.temp, calishp)
  x.df <- as.data.frame(x)
  pdsi.df <- as.data.frame(pdsi.temp)
  tot.df <- cbind(x.df, pdsi.df)
  tot.df$pixel <- as.factor(1:12)
  tot.df$year <- as.factor(predictyears[i])
  cook.compare[[i]] <- tot.df

}
col <- c("pNDVI", "pPDSI", "pixel", "year")
cook.compare1 <- lapply(cook.compare, setNames, col)
cook.compare.data <- do.call(rbind, cook.compare1)

cook.compare.final <- cook.compare.data[complete.cases(cook.
    ↪ compare.data),]
mean.by.year <- aggregate(.~year, data=cook.compare.final,
    ↪ mean)
mean.year.lm <- lm(pNDVI~pPDSI, mean.by.year)

by.year.p <- ggplot(data=mean.by.year, aes(x=pPDSI, y = pNDVI
    ↪ )) + geom_point(shape=1) + geom_smooth(method="lm", se=
    ↪ FALSE, color="blue", size=1)
by.year.p <- by.year.p + ggtitle("Mean reconstructed PDSI
    ↪ values compared to \nmean reconstruced iNDVI values by
    ↪ year") + theme(plot.title = element_text(size = 20,
    ↪ face="bold"), axis.title=element_text(size=15))
by.year.p + geom_text(x = -4, y = 6.75, label= "y = 0.037x +
    ↪ 6.585; R^2 = 0.6388") + labs(x="Reconstructed PDSI", y
    ↪ = "Reconstructed iNDVI")

compare.lm <- lm(pNDVI ~ pPDSI, cook.compare.data)

lm_eqn <- function(df){
  m <- lm(pNDVI ~ pPDSI, df);
  eq <- substitute(italic(y) == a + b %.% italic(x)*","~~
      ↪ italic(r)^2~"="~r2,
                   list(a = format(coef(m)[1], digits = 2),
                        b = format(coef(m)[2], digits = 2),
```

```
                                r2 = format(summary(m)$r.squared,
                                    ↪ digits = 3)))
  as.character(as.expression(eq));
}

p <- ggplot(data=cook.compare.final, aes(x=pPDSI, y = pNDVI))
    ↪  + geom_point(shape=1) + geom_smooth(method="lm", se=
    ↪ FALSE, color="blue", size=1)
p <- p + ggtitle("Reconstructed PDSI values compared to \
    ↪ nreconstruced iNDVI values") + theme(plot.title =
    ↪ element_text(size = 13, face="bold"))
p + geom_text(x = -4, y = 7.13, label= "y = 0.032x + 6.585; R
    ↪ ^2 = 0.092") + labs(x="Reconstructed PDSI", y = "
    ↪ Reconstructed iNDVI")
```

Aggregation

```
testag7x7 <- aggregate(test, fact=7, fun=mean)
testag5x5 <- aggregate(test, fact=5, fun=mean)
testag3x3 <- aggregate(test, fact=3, fun=mean)

#make three by three aggregation
#there are no multiple points per pixel! :)
my.list <- list()
years <- as.vector(1982:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
      ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
      ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58, auc)
  y <- aggregate(y, fact=3, fun=mean)
  eco <- aggregate(caliecoreg, fact=3, fun=max)
  y <- stack(y, eco)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
      ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
      ↪ layer" = "ecoreg"))
  names(z)[5] <- "indvi"
```

```
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ aggre/by3aucchrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

crq.by3.data <- do.call(rbind, my.list)


#make five by five aggregation
##there are no multiple points per pixel! :)

my.list <- list()
years <- as.vector(1982:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
    ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58, auc)
  y <- aggregate(y, fact=5, fun=mean)
  eco <- aggregate(caliecoreg, fact=5, fun=max)
  y <- stack(y, eco)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
    ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
    ↪ layer" = "ecoreg"))
  names(z)[5] <- "indvi"
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ aggre/by3aucchrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

crq.by5.data <- do.call(rbind, my.list)

#make seven by seven aggregation
my.list <- list()
years <- as.vector(1982:2003)
for (i in 1:length(years)){
  my.listyears <- as.character(years[i])
```

```
  auc <- raster(paste("/Users/ETLab/Dropbox/Thesis_Data/ndvi/
    ↪ indvi/ndvi3g_indvi/auc", years[i], ".tif", sep=""))
  listchrono <- filter(MAcrqlist10, year == years[i])
  x <- SpatialPointsDataFrame(coords = cbind(listchrono$lon,
    ↪ listchrono$lat), data = listchrono, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58, auc)
  y <- aggregate(y, fact=7, fun=mean)
  eco <- aggregate(caliecoreg, fact=7, fun=max)
  y <- stack(y, eco)
  z <- raster::extract(y, x, cellnumbers=TRUE)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
    ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
    ↪ layer" = "ecoreg"))
  names(z)[6] <- "indvi"
  z <- ddply(z, "cells", numcolwise(mean))
  write.table(z, paste("/Users/ETLab/Dropbox/Thesis_Data/
    ↪ aggre/by3aucchrono", years[i], sep=""))
  my.list[[my.listyears]] <- z
}

crq.by7.data <- do.call(rbind, my.list)

crq.by3.data$aspect.dir <- NULL
for (i in 1:length(crq.by3.data$aspect)){
  if (crq.by3.data$aspect[i] > 315 & crq.by3.data$aspect[i]
    ↪ <= 360) {
    (crq.by3.data$aspect.dir[i]="N")
  } else if (crq.by3.data$aspect[i] > 0 & crq.by3.data$aspect
    ↪ [i] < 45) {
    (crq.by3.data$aspect.dir[i] ="N")
  } else if (crq.by3.data$aspect[i] >= 45 & crq.by3.
    ↪ data$aspect[i] < 135) {
    (crq.by3.data$aspect.dir[i] = "E")
  } else if (crq.by3.data$aspect[i] >= 135 & crq.by3.
    ↪ data$aspect[i] < 225) {
    (crq.by3.data$aspect.dir[i] = "S")
  } else {
    (crq.by3.data$aspect.dir[i] = "W")
  }
}
crq.by3.data$aspect.dir <- as.factor(crq.by3.data$aspect.dir)
```

```
crq.by5.data$aspect.dir <- NULL
for (i in 1:length(crq.by5.data$aspect)){
  if (crq.by5.data$aspect[i] > 315 & crq.by5.data$aspect[i]
    ↪ <= 360) {
    (crq.by5.data$aspect.dir[i]="N")
  } else if (crq.by5.data$aspect[i] > 0 & crq.by5.data$aspect
    ↪ [i] < 45) {
    (crq.by5.data$aspect.dir[i] ="N")
  } else if (crq.by5.data$aspect[i] >= 45 & crq.by5.
    ↪ data$aspect[i] < 135) {
    (crq.by5.data$aspect.dir[i] = "E")
  } else if (crq.by5.data$aspect[i] >= 135 & crq.by5.
    ↪ data$aspect[i] < 225) {
    (crq.by5.data$aspect.dir[i] = "S")
  } else {
    (crq.by5.data$aspect.dir[i] = "W")
  }
}
crq.by5.data$aspect.dir <- as.factor(crq.by5.data$aspect.dir)

crq.by7.data$aspect.dir <- NULL
for (i in 1:length(crq.by7.data$aspect)){
  if (crq.by7.data$aspect[i] > 315 & crq.by7.data$aspect[i]
    ↪ <= 360) {
    (crq.by7.data$aspect.dir[i]="N")
  } else if (crq.by7.data$aspect[i] > 0 & crq.by7.data$aspect
    ↪ [i] < 45) {
    (crq.by7.data$aspect.dir[i] ="N")
  } else if (crq.by7.data$aspect[i] >= 45 & crq.by7.
    ↪ data$aspect[i] < 135) {
    (crq.by7.data$aspect.dir[i] = "E")
  } else if (crq.by7.data$aspect[i] >= 135 & crq.by7.
    ↪ data$aspect[i] < 225) {
    (crq.by7.data$aspect.dir[i] = "S")
  } else {
    (crq.by7.data$aspect.dir[i] = "W")
  }
}
crq.by7.data$aspect.dir <- as.factor(crq.by7.data$aspect.dir)
crq.by3.sub <- subset(crq.by3.data, year <= 1998)
crq.by3.sub$ecoreg <- as.factor(crq.by3.sub$ecoreg)
crq.by5.sub <- subset(crq.by5.data, year <= 1998)
crq.by5.sub$ecoreg <- as.factor(crq.by5.sub$ecoreg)
```

```
crq.by7.sub <- subset(crq.by7.data, year <= 1998)
crq.by7.sub$ecoreg <- as.factor(crq.by7.sub$ecoreg)

crq.lm.by3 <- lm(indvi~CRQ + dem + slope + ecoreg + cec, data
   ↪ =crq.by3.sub)
crq.lm.by5 <- lm(indvi~CRQ + dem + slope + ecoreg + cec, data
   ↪ =crq.by5.sub)
crq.lm.by7 <- lm(indvi~CRQ + dem + slope + ecoreg + cec, data
   ↪ =crq.by7.sub)

stargazer(crq.lm3, crq.lm.by3, crq.lm.by5, crq.lm.by7, column
   ↪ .labels=c("64 km2", "576 km2", "1600 km2", "3136 km2"),
   ↪  single.row=TRUE, star.cutoffs = c(0.05, 0.01, 0.001))
```

Accuracy Assessment

```
#data frame for predictions by years 1700 - 1998
predictyears <- as.vector(1700:2003)
my.historiclist <- list()
for (i in 1:length(predictyears)){
  my.predictyears <- as.character(predictyears[i])
  oldcrn <- filter(MAcrqlist10, year == predictyears[i])
  x <- SpatialPointsDataFrame(coords = cbind(oldcrn$lon,
     ↪ oldcrn$lat), data = oldcrn, match.ID = "ID")
  y <- stack(cali8dem, cali8aspect, cali8slope, cec0_58,
     ↪ caliecoreg)
  z <- raster::extract(y, x)
  z <- cbind(z, as.data.frame(x))
  z <- rename(z, replace = c("layer.1" = "dem", "layer.2" = "
     ↪ aspect", "layer.3" = "slope", "layer.4" = "cec", "
     ↪ layer.5" = "ecoreg"))
  my.historiclist[[my.predictyears]] <- z
}

historic.data <- do.call(rbind, my.historiclist)
historic.data$ecoreg <- as.factor(historic.data$ecoreg)

#predict indvi for years 1700:2003 and do paired t-test
   ↪ between values
pindvi <- predict(crq.lm3, historic.data)
pindvi <- as.data.frame(pindvi)
predict.data <- cbind(pindvi, historic.data)
```

```
#ttest
#be sure to run after mergeCA
testsub <- filter(predict.data, year >= 1999)
testsub1 <- filter(crq.final.data, year >= 1999)
comdata <- cbind(testsub1$indvi, testsub)
comdata <- rename(comdata, c("testsub1$indvi" = "indvi"))
comdata1 <- comdata[, c("indvi", "pindvi")]
t.test(comdata1$indvi, comdata1$pindvi ,paired=TRUE)
```

Vita

Kyle Landolt is from Arlington, Texas, and has a loving older brother, mother, and father. He graduated from James Martin High School in 2010 where he was inspired by his AP environmental studies teacher, Anna Baggett, to pursue a career in earth sciences. Kyle attended Texas A&M University where he met Dr. Washington-Allen during his sophomore year and quickly became involved undergraduate research with his lab. After receiving his Bachelor's of Science in ecological restoration in 2014 (Gig 'em), Kyle continued his work with Dr. Washington-Allen by working towards a Master of Science in geography at the University of Tennessee, Knoxville, with a concentration in remote sensing, physical geography, and biogeography.

At UT, Kyle worked as a Graduate Teaching Assistant from the Fall 2014 to Spring 2016. During that time, Kyle helped teach labs in physical geography, natural hazards, GIS, and remote sensing. During the Summer 2015, Kyle worked as an intern in the Geographic Information Science and Technology group at Oak Ridge National Lab. Kyle wishes to continue his work in geography and earn a Ph.D. in the future.