Title

Historical legacies shape contemporary forests and woodlands: a study of California landscapes integrating historical and modern ecological data

Author

Easterday, Kelly J

Publication Date

2018

Historical legacies shape contemporary forests and woodlands: a study of California landscapes integrating historical and modern ecological data

By

Kelly J Easterday

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Maggi Kelly, Chair
Professor David Ackerly
Professor Scott Stephens

Spring 2018

Historical legacies shape contemporary forests and woodlands: a study of California landscapes integrating historical and modern ecological data

Abstract

Historical legacies shape contemporary forests and woodlands: a study of California landscapes integrating historical and modern ecological data

by

Kelly J Easterday

Doctor of Philosophy in Environmental Science, Policy and Management

University of California, Berkeley

Professor Maggi Kelly, Chair

Global changes in climate and increased anthropogenic activity are fundamentally reshaping the structure and function of ecosystems across scales. The velocity, scale, and intensity of human impact are irrefutable, yet quantifying the effects of anthropogenic activities in relation to natural ecosystem dynamics can be difficult. Understanding the interaction between human activities and landscape change is of paramount importance, especially as anthropogenic driven land cover conversion and disturbances threaten ecosystem biodiversity and natural resources. Our approaches to natural resource management are challenged when environmental outcomes are embedded in complex socio-ecological systems characterized by profound uncertainties and interactive. Interdisciplinary approaches are thus needed to adequately address contemporary environmental problems and evaluate interactions between biophysical and socio-ecological drivers of change.

The research presented in this dissertation has two broad research foci. First, I explore the linkages between human activity and landscape change in the context of California forests and woodlands. I draw on historical and contemporary forest inventory data and investigate more than a century of landscape transformations in California to understand the drivers of change that influence today’s landscape. By studying the changes in forest and woodland distribution and structure across California, I review the often under-evaluated broad scale influence of socio-ecological factors such as land ownership and land management in contributing to forest densification and landscape change. Second, my work contributes to the discussion of technical issues of data availability and data aggregation when historical data are used in modern ecological analysis and combined with contemporary data. My research links historical and contemporary empirical data through data science approaches in data digitization, data aggregation, data sharing, spatial modeling, and species distribution modeling in order to increase the scope and potential of historical data to answer complex environmental problems. At the core of this work is one valuable and recently digitized historical ecological data collection: the California Vegetation Type Mapping (VTM) Project.

My second chapter is motivated by several recent studies that report climate (i.e. climate water deficit (CWD)) as the primary mechanism of large tree decline and change in forest structure in California in the 20th century. Reflecting on these studies and other conflicting opinions of primary drivers of change, I found very few studies that quantify the impact of land ownership and land management on the quantities of large trees and other characteristics of forest structure. Land ownership has been used to understand the long-term effects of and variation in land management practices; especially when spatially explicit data on management practices are unavailable or incomplete. Thus, in Chapter Two, I explicitly investigate 20th century changes in forest structure across six ownership classes in California. In comparing historical and contemporary forest structural data I found that declines in large trees and increases in small tree density were consistent across the state, irrespective of ownership boundaries. However, there were important differences in the magnitude of this change. In particular, this pattern is most pronounced on private timberlands which experience up to 400% regional increases in small tree density since 1930. Nearly all land ownership classes experience declines in large trees however, private timberland, and National Park/Wilderness areas experience a significant reduction of 83% and 73% respectively.

In Chapter Three, I investigate the effects of urban development on changes to the distribution of oak species in California. First, by modeling historical patterns of richness for eight oak species using historical map and plot data from the California Vegetation Type Mapping (VTM) collection I examine spatial intersections between hot spots of historical oak richness and modern urban and conservation land. I found that impacts from development and conservation vary by both species and richness. At the state level, the impact of urban development on oaks has been small within the areas of the highest oak richness but areas of high oak richness are also poorly conserved.

In the first two chapters, I discuss the relationship between social and biophysical drivers of landscape change. This kind of understanding of long-term patterns of change requires data availability and the ability to re-use data. Following from these, Chapter 4 discusses preserving history's place in the growing data landscape, I review three approaches to sharing historical data from field stations using principles from data science. To encourage greater use of historical data across scientific disciplines it is vital to make data findable, accessible, interoperable, and reusable (e.g. the FAIR principles). This summary of three important data collections emerging from the University of California showcase the potential for their use in research and encourages similar ventures that use common archival, geospatial, and data science practices to shepherd historical data out of file drawers and into the contemporary digital data landscape.

Table of Contents

List of Figures

Chapter 2 Figures

Chapter 3 Figures

Chapter 4 Figures

List of Tables

Chapter 2 Tables

Chapter 3 Tables

Chapter 4 Tables

Acknowledgements

First and foremost, I am forever grateful to my adviser Maggi Kelly. Thank you for taking a chance on me and for showing up for me time and time again. You are a true force of nature and I have learned so much from you through your constant professional and personal mentorship. I feel so privileged for all the opportunities you have offered to me, for those you have let me find on my own, and for the ones that are still to come.

I am also deeply grateful to the members of my qualifying and dissertation committee for the many hours they dedicated toward my education and training: Adina Merenlender, Katie Suding, Greg Biging, David Ackerly and Scott Stephens. David, thank you for sticking with me since the beginning and for pushing me to ask and answer questions differently and reach further in my work. Scott, thank you for always lending an ear, your contagious enthusiasm for your work and urgency to communicate its impacts has become a model I hope to emulate.

I am also eternally grateful to have benefited from the tremendous community behind the initial data collection effort and the subsequent preservation effort of the Vegetation Type Mapping dataset. The work I present in this dissertation was made possible due to the foresight and dedication those that participated in the 30 years of detailed field surveys undertaken on foot and horseback and in nearly 30 more years of painstaking digitization to unite the VTM collection. Because I was so fortunate to come into this graduate program at a critical moment in which the whole collection came online, I knew that I would have to find my own way to give homage to these early explorers, those that helped develop this collection, and the future generations that can and should benefit from being able to draw on collections like the VTM. I found this chance with the discovery of the three critical natural resource surveys, the Soil-Vegetation Surveys, the Wieslander Forest Condition surveys, and the John Leiberg- George Sudworth map series in the UC Berkeley Libraries. These collections are invaluable vestiges of what California looked like at the turn of the 19th century. Although these maps never made it into a state where they could become part of this body of work, they were nonetheless a big part of my time at Berkeley. Therefore, I would like to thank Susan Powell and Rebecca Miller for mirroring my affinity for old maps and jumping through hoops to preserve a collection that is likely to be invaluable for the future; as well as the UC Berkeley SPUR program that brought the immense and invaluable participation of 15 undergraduate students. This project is an initial effort to connect three natural resource inventory collections spanning nearly a century (1880-1970) and carry the legacy and dedication of the original mappers into the 21st century.

Furthermore, I am tremendously grateful to all my colleagues and friends. I have benefited from the brilliance of every single one of you, thank you for your enthusiasm and willingness to engage in my early unpolished thoughts and every stage throughout. I am proud to have been a part of the Department of Environmental Science, Policy, and Management and believe I am a better scientist and a better person as a result of the department’s commitment to representing interdisciplinary dialogue. I was also extremely fortunate to spend much of my time working with UC ANR’s Informatics and

GIS crew, Andy Lyons, Shane Feirer, Robert Johnson, and Sean Hogan it has been an absolute blast to have worked alongside you, thank you for your constant encouragement, your knowledge of all things geospatial, and your enthusiasm for creating applied powerful solutions for so many.

There have also been so many powerful women I have the absolute privilege of calling my mentors and friends. Peggy Fiedler found me when I had lost hope, picked me up, dusted me off and encouraged me forward. The past and present Kelly Labers, especially, Alice Kelly and Jenny Palomino (honorary Kelly) whose balance and brilliance created the space I needed to recognize my own voice. Maggi Kelly has done an amazing job of creating a lab culture that empowers and enriches and Erin Beller, Christine and Chippie have been key in carrying that forward. My fearless lady cohort-mates who I shared so many laughs, ideas, and struggles with early on, Kelly I, Carrie, Laura, Chandra, and Natalie. My partly logical and partly biological family, Monica, Jana, and Lauren having you by my side has made all the difference. And, of course, the ladies of girl’s night: Katy Seto, Kate Wilkin, Sara Knox, Lisa Kelley, Ashton Wesner, Stella Cousins, and Kevi Mace I know one day we will combine our forces to take over the world. Katy Seto said it once, and I will say it again this whole journey takes a village, and you all were my village.

I also send my gratitude to the Pitcher and McClure/Easterday families who have provided me endless encouragement along the way. Hank whose artwork peppered my office and encouraged me to look up every so often and Susan who always gave me the biggest hugs. I am so lucky to have a family that has supported me unconditionally throughout my education even if they can never remember what ESPM stands for. My sisters, who have reminded me of life outside the confines of Mulford Hall and provided me with at least 5 reasons to keep on this path. Libby and Phil, who started this all, your unwavering encouragement is truly astounding, and I have you both to thank for so much. Pops, all the years of waking up early to help me study before morning practice have finally paid off. I have never seen a better example of drive and dedication, then I have from you. Mom, your power, and your generosity have pushed me through this in times when no one and nothing else could. You have been the constant voice inside my head vowing to never quit, thank you.

Above all, thank you to my partner Lincoln. We took this leap together, and though I may have beat you to the finish line, I never could have done this without you. You make me happier, stronger, smarter, and more patient. You have been the first one to witness the spark and trepidation of new ideas and the last one to see them through the door. I am so so grateful to you.

Chapter 1 Introduction

Global changes in climate (Parmesan and Yohe 2003) and increased anthropogenic activity (Foley et al. 2005) are fundamentally reshaping the structure and function of ecosystems across scales. The velocity, scale, and intensity of human impact are irrefutable, yet quantifying the effects of anthropogenic activities in relation to natural ecosystem dynamics can be difficult. Understanding the interaction between human activities and landscape change is of paramount importance, especially as anthropogenic driven land cover conversion and disturbance threaten ecosystem biodiversity and natural resources (Foley et al. 2005). Our approaches to natural resource management are therefore challenged when environmental outcomes are embedded in complex socio-ecological systems characterized by profound uncertainties, interactive effects, and positive feedback loops (Liu et al. 2007; J. Dearing et al. 2010; Millar, Stephenson, and Stephens 2007). Interdisciplinary approaches are thus needed to adequately address contemporary environmental problems and evaluate interactions between biophysical and socio-ecological drivers of change (Ostrom 2009; Liu et al. 2007). Understanding and anticipating the outcomes of these dynamics is made difficult by the large number of processes operating over various spatial and temporal scales (Liu et al. 2007), as well as the spatiotemporal mismatch between ecosystem dynamics and data availability (Cadenasso, Pickett, and Grove 2006). Isolating the effects of anthropogenic disturbance from natural ecological mechanisms is not trivial, and a considerable amount of literature has been dedicated to this challenge (e.g. J. A. Dearing et al. 2006; Nelson et al. 2006; Fukami and Wardle 2005; Seidl, Schelhaas, and Lexer 2011; Jackson et al. 2009).

Despite such recent advances, understanding the spatio-temporal interactions between socio-ecological and biophysical drivers is crucial for predicting future landscape dynamics, informing conservation planning, and understanding change in ecosystem function at timescales relevant to society (Svenning and Sandel 2013). In this context, historical data have become increasingly important for investigating long-term processes, documenting rare or extreme events, analyzing ecosystem response to contrasting cultural regimes, and providing key insights into current ecosystem structure, function and response times (Foster 2000). Historical data and a long-term perspective therefore provide three key functions: 1) they elongate the temporal scale of potential scientific inquiry to include otherwise irretrievable past environments; 2) they provide a contextual foundation from which to assess change; and 3) they situate the study of the environment in a wider disciplinary context (Swetnam, Allen, and Betancourt 1999). Legacies of historical land-use have been studied broadly in both aquatic (Allan 2004; Harding et al. 1998) and terrestrial (Petit, Hu, and Dick 2008; Hermy and Verheyen 2007) ecosystems. The mechanisms of anthropogenic impact are varied, and range from urbanization, agricultural expansion, logging, to the modification of disturbance regimes (e.g. fire) (Perring et al. 2016; Foster et al. 2003), among other drivers. These historical legacies create measurable impacts on the landscape long after their direct impact, requiring an integrated approach that leverages historical frameworks as well as contemporary tools and techniques to fully understand current landscapes.

In a time of unprecedented climate change, predicting the resilience of ecosystems in the face of future environmental change can be aided by understanding the range of environmental variation that systems have experienced in the past (Landres, Morgan, and Swanson 1999) and determining the environmental conditions under which those ecosystems arose (Jackson and Overpeck 2000; Jackson 2006). The combination of historical and contemporary data has played a major role in the development of our understanding of ecosystems and ecosystem process in California. In ecosystems with dynamics that play out on long time scales, such as forests, understanding process and interacting drivers of change necessitates a long-term perspective.

In California forests, fragmentation, widespread increases in tree morbidity, shifts in species distribution, composition, and the densification of stand structure have been attributed to several mechanisms including climate change-related temperature increases and declining water availability (Das et al. 2013; McIntyre et al. 2015; J. Lutz, van Wagtendonk, and Franklin 2010); as well as to historical management practices such as fire suppression and logging (Knapp et al. 2013; Laudenslayer and Darr 1990; McKelvey and Johnston 1992; Beesley 1996), and more contemporary patterns of land cover change through rapid urbanization (Radeloff et al. 2005). As forests account for more than half of the world's terrestrial productivity, linking the mechanism of change with a specific impact is integral to enhanced understanding of biogeochemical cycling, regional climate, disturbance regimes, species diversity, ecosystem management, and the response of trees and forests to future changes (Caspersen et al. 2000; Lindenmayer, Laurance, and Franklin 2012; Stephens et al. 2018). It has also long been argued that enhanced understanding of historical and anticipated changes in forested systems is central to designing and implementing resource management policies and ensuring ecosystem resilience (Whitlock et al. 2017; Millar, Stephenson, and Stephens 2007).

The research presented in this dissertation has two broad research foci. First, I explore the linkages between human activity and landscape change in the context of California forests and woodlands. I draw on historical and contemporary forest inventory data and investigate more than a century of landscape transformations in California to understand the drivers of change that influence today’s landscape. By studying the changes in forest and woodland distribution and structure across California, I review the often under evaluated broad scale influence of socio-ecological factors such as land ownership and land management in contributing to forest densification and landscape change. Second, my work contributes to the discussion of technical issues of data availability and data aggregation when historical data are used in modern ecological analysis and combined with contemporary data. My research links historical and contemporary empirical data through data science approaches in data digitization, data aggregation, data sharing, spatial modeling, and species distribution modeling in order to increase the scope and potential of historic data to answer complex environmental problems. At the core of this work is one valuable and recently digitized historical ecological data collection: the California Vegetation Type Mapping (VTM) Project.

1 The Vegetation Type Mapping (VTM) Project and the use of historical data

My dissertation relies on a comprehensive statewide vegetation survey from the 1930s known as the Wieslander or California Vegetation Type Mapping project (VTM). This survey, led by Albert Wieslander of the U.S Forest Service was a nearly 30-year effort to map forested and vegetated areas of the state. The endevour resulted in a collection of maps depicting dominant vegetation types; plot scale information on floristic characteristics including frequency counts of species diameter, height, and site characteristics; landscape photographs; and herbarium specimens (A. E. Wieslander 1935; Kelly et al. 2016; Thorne and Le 2016; Kelly et al. 2017). Each chapter within this dissertation relies on some element of this collection; Chapter 2 focuses on the use of over 9,000 VTM plots to evaluate a century of change in forest structure by ownership; Chapter 3 uses over 85,000 species occurrence records from both vegetation plots and vegetation maps to examine the interaction of urbanization and woodlands with a focus on oak species diversity; and Chapter 4 uses the VTM as a case study to explore best practices for digitizing historical data and making it available for scientific use.

The VTM data and project have a storied legacy that have resulted in contributions to the fields of forest ecology and management (e.g. J. A. Lutz, van Wagtendonk, and Franklin 2009), fire ecology (e.g. Lippitt et al. 2013), vegetation mapping and classification (e.g. Allen-Diaz et al. 1989), ecological modeling (e.g. Dobrowski et al. 2011), conservation biology (e.g. Rubidge et al. 2011), remote sensing (e.g. Wieslander and Wilson 1942), and landscape and urban planning (e.g. Santos et al. 2014). Several comparisons of 1930's VTM inventories and maps have concentrated on the Sierra Nevada, documenting dramatic changes in the distribution of vegetation types over the last 70-80 years (Bouldin 1999; Thorne and Le 2016; Dolanc, Thorne, and Safford 2013; Dolanc et al. 2013). These reports have shown increased dominance of shade-tolerant conifers (especially Fir species) and increase in hardwood dominated forests, loss of blue oak woodland and yellow pine forest and the expansion of subalpine trees into previously unvegetated snowpack areas. Other comparisons have revealed consistent evidence of an increase in young small-diameter trees across the state and decreases in large trees (Minnich et al. 1995; Fellows and Goulden 2008; Goforth and Minnich 2008; Dolanc, Thorne, and Safford 2013; McIntyre et al. 2015; J. A. Lutz, van Wagtendonk, and Franklin 2009). Linking mechanisms to these changes has been complicated even with the reliance on a common historical dataset, with the majority of the studies (~20/39) attributing these changes, regionally, to the alteration of fire regimes and changes in climate.

Working with historical data requires the acknowledgment and examination of challenges such as plot geolocation error and potential bias. In the VTM dataset for example, plot location is derived from original markings on historical topographic maps and positional error is estimated to be ~200 m (Kelly, Allen-Diaz, and Kobzina 2005) per plot which can affect direct plot comparisons or plot re-surveys, especially in highly heterogeneous regions (Keeley 2004). VTM sampling protocols also have raised questions about bias towards sampling undisturbed forests. There is no evidence of bias suggested in the original VTM manual (A. Wieslander 1935) or in the protocol of selecting plot locations to be representative of vegetation types being mapped (A. E. Wieslander 1935), or in the plot distributions (Dolanc et al. 2013). However, there are competing patterns when comparing contemporary forest structure data (as provided in the USFS Forest Inventory and Analysis (FIA) dataset) and VTM. Some studies using alternative comparison data (e.g. 1911 Leiberg data) show similar declines in large trees (e.g. Lydersen et al. 2013; van Mantgem et al. 2009), whereas others have shown declines in large trees on some ownership classes but not on others, and contrasting increases in basal area (Collins et al. 2017; Lydersen et al. 2013). These disparities are difficult to verify as the datasets in question are not directly comparable but do require a cautionary approach to interpreting changes in large trees and biomass and suggest that these patterns and their driving mechanisms maybe scale and region dependent.

2 Land ownership and urbanization as mechanisms of change

Our understanding of the long-term dynamics of landscape change is complicated by the direct and indirect effects between both natural and anthropogenic drivers (Cadenasso, Pickett, and Grove 2006), the variation in the spatiotemporal influence of each driver and their interactive effects on the ecosystem. The fields of remote sensing and spatial data science have contributed to our understanding of the spatial patterning, scale, and intensity of landscape change (Green, Kempka, and Lackey 1994; W. Turner et al. 2003; M. G. Turner 2005) over the last half century, but our understanding of multiple interacting drivers and their consequences increasingly relies on the spatiotemporal reach of historical, socio-political, socio-ecological, and biophysical datasets. More than just dense time series imagery, an evaluation and attribution of the mechanisms of landscape change requires a socio-ecological context.

Generally the academic community attributes patterns of denser forests with more smaller and fewer larger trees and a shift in species composition favoring shade tolerance to three main drivers: 1) climate (McIntyre et al. 2015; Dolanc, Thorne, and Safford 2013; Taylor 2000), 2) fire or lack thereof (Parsons and DeBenedetti 1979; Collins, Everett, and Stephens 2011), and 3) land management including resource extraction by logging (Kaufmann, Regan, and Brown 2000; Naficy et al. 2010; Knapp et al. 2013). However, land ownership institutions play an important role in how land-use decisions translate to land cover (M. G. Turner, Wear, and Flamm 1996; Steen-Adams et al. 2015). Understanding differing patterns of change across a range of landowners (private, state, regional, and federal) is useful scale for the development of actionable management and regulatory policies. In Chapter 2 I assess patterns of forest densification and large tree decline across six land ownership types in California. By comparing Vegetation Type Mapping (VTM) project data from the 1930s with contemporary data from 2010 Forest Inventory and Analysis plots I build upon a sizeable set of literature using this dataset to describe change over time in California forests, and contribute to the discussion by adding an alternative perspective to what might explain contemporary patterns in forest structure.

The analysis provides a depiction of how the numbers and size distribution of trees have changed over the 20th century across six land ownership types and management regimes. I assessed the difference between historical and current forest structure, finding that patterns of forest densification and declines in large trees and basal area are generally consistent across ownership. However, the magnitude of change varies with the greatest densification occurring on Private Timberlands and the greatest loss in large trees on Private Timberlands and unexpectedly in National Parks and Wilderness areas, suggesting that at least at a statewide scale, patterns of land ownership do not explain large tree declines.

Another important driver of landscape change in California is urbanization (Sleeter et al. 2013). Rapid urbanization has heightened concern globally over potential losses of biodiversity and ecosystem services. However, sustainable planning initiatives in conjunction with ecological knowledge can help sustain biodiversity and reduce landscape fragmentation in urban environments. Despite this potential there has been limited adoption and integration of ecological data and methods into the urban planning process. In Chapter 3 I argue that the adoption of historical data in conjunction with modern data science workflows such as species distribution modelling techniques are invaluable in locating hot spots of species richness, understanding where critical habitats and species have been lost, and in providing evidence and incentive to recover what was lost and preserve what still exists. In this chapter, I assessed historical hotspots of oak richness at a statewide scale. Using the plot and map species occurrence records from the 1930s VTM dataset I was able to generate a robust species distribution model to assess historical hotspots of oak richness. Using these probability distributions, I investigated the overlap of predicted areas of historical oak richness with contemporary urban and protected areas to determine regions where land use patterns (urban, protected) either preserved or removed oak woodlands. I found that about a fifth of the area previously containing an oak species in the past is now urban, nearly 20% of the modeled historical range of both Coast Live oak and Engelmann oak is now under the modern urban footprint.

3 History's place in the data landscape

Our ability to understand long term ecosystem trajectories and dynamics and distinguish between drivers of change relies on data availability. Through burgeoning technological developments in sensor networks and mobile data collection platforms, we are able to capture change dynamics in both ecological processes and in our social networks at both fine and coarse scales (Hampton et al. 2013). The data streams these platforms create are immense and only a fraction of what is captured is actually analyzed and turned into information (Hampton et al. 2013). Despite the potential and capabilities of these new platforms, they only include a limited time frame of information - primarily since the 1970s - and do not capture the scope of our anthropogenic legacy and impact. Disciplines generally concentrate on specific time periods. For example, ecology focuses on the immediate past (i.e. 200 years), paleoecology takes a longer temporal viewpoint (across geologic timescales), and ecological informatics and data science concentrate on the most recent decades of data, because that is what is most common and available in a machine-readable format. The disconnect between the temporal foci of these disciplines creates a need for collaboration in order to quantify and understand long term ecosystem dynamics and evaluate the drivers and impacts of land cover change.

The lack of spatially coincident and temporally continuous data available to researchers makes inquiry into landscape patterns and change challenging. Therefore, preserving historical data is essential to preserving our ability to ask complex ecological questions. However, simply preserving data does ensure its use, data preservation must follow simple principles in order to be used in synthetic or transdisciplinary research: data must also be findable, accessible, interoperable, and re-usable. In Chapter 4, I use three case studies that illuminate different but equally valuable approaches to moving data from analog to digital or from the field to the cloud and discuss the steps needed to allow future scientist access to the past. In this discussion I highlight what is currently a gap in data analysis and interdisciplinary discourse, in that historical dark data is underrepresented in the current digital ecological data landscape and is therefore underutilized and can serve a transformative role if stewarded into the 21st century.

In conclusion, Chapter 5 summarizes the key findings of my dissertation as they relate to the two core research foci: 1) the linkages between human activity and landscape change in the context of land ownership, urbanization, and changes to the distribution and structure of California's forests and woodlands, and 2) the technical issues of data availability and data aggregation when historical data are used in modern ecological analysis and combined with contemporary data.

4 References

Allan, J.D., 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. Annu. Rev. Ecol. Evol. Syst. 35, 257–284.

Allen-Diaz, B.H., Evett, R.R., Holzman, B.A., Martin, A.J., 1989. Rangland Cover Type Descriptions for California Hardwood Rangelands. California Department of Forestry and Fire Protection Report 1–319.

Beesley, D., 1996. Past Sierra Nevada Landscapes, in: Sierra Nevada Ecosystem Project: Final Report to Congress, Vol. II, Assessments and Scientific Basis for Management Options. pp. 3–24.

Bouldin, J., 1999. Twentieth-Century Changes in Forests of the Sierra Nevada, California (Doctor of Philosophy). University of California, Davis.

Cadenasso, M.L., Pickett, S.T. a., Grove, J.M., 2006. Dimensions of ecosystem complexity: Heterogeneity, connectivity, and history. Ecol. Complex. 3, 1–12.

Caspersen, J., Pacala, S., Jenkins, J., Hurtt, G., Moorcroft, P., Birdsey, R., 2000. Contributions of Land-Use History to Carbon Accumulation in U.S. Forests. Science 290, 1148–1151.

Collins, B.M., Everett, R.G., Stephens, S.L., 2011. Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests. Ecosphere

Collins, B.M., Fry, D.L., Lydersen, J.M., Everett, R., Stephens, S.L., 2017. Impacts of different land management histories on forest change. Ecol. Appl. 27, 2475–2486.

Das, A.J., Stephenson, N.L., Flint, A., Das, T., van Mantgem, P.J., 2013. Climatic Correlates of Tree Mortality in Water- and Energy-Limited Forests. PLoS One 8. https://doi.org/10.1371/journal.pone.0069917

Dearing, J.A., Battarbee, R.W., Dikau, R., Larocque, I., Oldfield, F., 2006. Human–environment interactions: towards synthesis and simulation. Regional Environ. Change 6, 115–123.

Dearing, J., Braimoh, A., Reenberg, A., Turner, B., van der Leeuw, S., 2010. Complex Land Systems: The Need for Long Time Perspectives to Assess their Future. Ecol. Soc. 15. https://doi.org/10.5751/ES-03645-150421

Dobrowski, S.Z., Thorne, J.H., Greenberg, J.A., Safford, H.D., Mynsberge, A.R., Crimmins, S.M., Swanson, A.K., 2011. Modeling plant ranges over 75 years of climate change in California, USA temporal transferability and species traits. Ecol. Monogr. 81, 241–257.

Dolanc, C.R., Safford, H.D., Dobrowski, S.Z., Thorne, J.H., 2013a. Twentieth century shifts in abundance and composition of vegetation types of the Sierra Nevada, CA, US. Appl. Veg. Sci. 1–14.

Dolanc, C.R., Thorne, J.H., Safford, H.D., 2013b. Widespread shifts in the demographic structure of subalpine forests in the Sierra Nevada, California, 1934 to 2007. Glob. Ecol. Biogeogr. 22, 264–276.

Fellows, A.W., Goulden, M.L., 2008. Has fire suppression increased the amount of carbon stored in western U.S. forests? Geophys. Res. Lett. 35, 1–4.

Foley, J.., Defries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily G.C., Gibbs, H.K., Helkowski, J.H., Holloway, T., Howard, E., Kucharik, C.J., Monfreda, C., Patz, J., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global consequences of land use. Science 309, 570–574.

Foster, D., 2000. Conservation lessons and challenges from ecological history. Forest History Today.

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., Knapp, A., 2003. The Importance of Land-Use Legacies to Ecology and Conservation. Bioscience 53, 77.

Fukami, T., Wardle, D.A., 2005. Long-term ecological dynamics: reciprocal insights from natural and anthropogenic gradients. Proc. Biol. Sci. 272, 2105–2115.

Goforth, B.R., Minnich, R. a., 2008. Densification, stand-replacement wildfire, and extirpation of mixed conifer forest in Cuyamaca Rancho State Park, southern California. For. Ecol. Manage. 256, 36–45.

Green, K., Kempka, D., Lackey, L., 1994. Using remote sensing to detect and monitor land-cover and land-use change. Photogramm. Eng. Remote Sens. 60, 331–337.

Hampton, S.E., Strasser, C. a., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162.

Harding, J.S., Benfield, E.F., Bolstad, P.V., Helfman, G.S., Jones, E.B., 3rd, 1998. Stream biodiversity: the ghost of land use past. Proc. Natl. Acad. Sci. U. S. A. 95, 14843–14847.

Hermy, M., Verheyen, K., 2007. Legacies of the past in the present-day forest biodiversity: a review of past land-use effects on forest plant species composition and diversity. Ecol. Res. 22, 361–371.

Jackson, S.T., 2006. Vegetation, environment, and time: The origination and termination of ecosystems. J. Veg. Sci. 17, 549–557.

Jackson, S.T., Betancourt, J.L., Booth, R.K., Gray, S.T., 2009. Ecology and the ratchet of events: climate variability, niche dimensions, and species distributions. Proc. Natl. Acad. Sci. U. S. A. 106 Suppl 19685–19692.

Jackson, S.T., Overpeck, J.T., 2000. Responses of plant populations and communities to environmental changes of the late Quaternary. Paleobiology 26, 194–220.

Kaufmann, M.R., Regan, C.M., Brown, P.M., 2000. Heterogeneity in ponderosa pine/Douglas-fir forests: age and size structure in unlogged and logged landscapes of central Colorado. Can. J. For. Res. 30, 698–711.

Keeley, J.E., 2004. VTM Plots as Evidence of Historical Change: Goldmine or Landmine? Madrono 51, 372–378.

Kelly, M., Allen-Diaz, B., Kobzina, N., 2005. Digitization of a historic dataset: the Wieslander California vegetation type mapping project. Madroño 52, 191–201.

Kelly, M., Easterday, K., Koo, M., Thorne, J.H., Mukythar, S., Galey, B., 2017. Geospatial Informatics Key to Recovering and Sharing Historical Ecological Data for Modern Use. Photogrammetric Engineering & Remote Sensing 83, 779–786.

Kelly, M., Easterday, K., Rapicullio, G., Koo, M., McIntyre, P., Thorne, J., 2016. Rescuing and Sharing Historical Vegetation Data for Ecological Analysis: The California Vegetation Type Mapping Project. Biodivers. Inf. 11, 40–62.

Knapp, E.E., Skinner, C.N., North, M.P., Estes, B.L., 2013. Long-term overstory and understory change following logging and fire exclusion in a Sierra Nevada mixed-conifer forest. For. Ecol. Manage. 310, 903–914.

Landres, P., Morgan, P., Swanson, F., 1999. Overview of the Natural Variability Concepts in Managing Ecological Systems. Ecol. Appl. 9, 1179–1188.

Laudenslayer, W.F., Darr, H.H., 1990. Historical effects of logging on the forests of the Cascade and Sierra Nevada ranges of California. Transactions of the Western Section of the Wildlife Society 26, 12–23.

Lindenmayer, D.B., Laurance, W.F., Franklin, J.F., 2012. Global Decline in Large Old Trees. Science 338, 1305–1306.

Lippitt, C., Stow, D., O'Leary, J., Franklin, J., 2013. Influence of short-interval fire occurrence on post-fire recovery of fire-prone shrublands in California, USA. Int. J. Wildland Fire 184–193.

Liu, J., Dietz, T., Carpenter, S.R., Alberti, M., Folke, C., Moran, E., Pell, A.N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C.L., Schneider, S.H., Taylor, W.W., 2007. Complexity of coupled human and natural systems. Science 317, 1513–1516.

Lutz, J. a., van Wagtendonk, J.W., Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. J. Biogeogr. 37, 936–950.

Lutz, J.A., van Wagtendonk, J.W., Franklin, J.F., 2009. Twentieth-century decline of large-diameter trees in Yosemite National Park, California, USA. For. Ecol. Manage. 257, 2296–2307.

Lydersen, J.M., North, M.P., Knapp, E.E., Collins, B.M., 2013. Quantifying spatial patterns of tree groups and gaps in mixed-conifer forests: Reference conditions and long-term changes following fire suppression and logging. For. Ecol. Manage. 304, 370–382.

McIntyre, P.J., Thorne, J.H., Dolanc, C.R., Flint, A.L., Flint, L.E., Kelly, M., Ackerly, D.D., 2015. Twentieth-century shifts in forest structure in California: denser forests, smaller trees, and increased dominance of oaks. Proceedings of the National Academy of Sciences 201410186.

McKelvey, K.S., Johnston, J.D., 1992. Historical Perspectives on Forests of the Sierra Nevada and the Transverse Ranges of Southern California: Forest Conditions at the Turn of the Century (No. PSW-GTR-133). USDA Forest Service Gen. Tech. Rep.

Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17, 2145–2151.

Minnich, R.A., Barbour, M.G., Burk, J.H., Fernau, R.F., 1995. Sixty Years of Change in Californian Conifer Forests of the San Bernardino Mountians. Conserv. Biol. 9, 902–914.

Naficy, C., Sala, A., Keeling, E.G., Graham, J., DeLuca, T.H., 2010. Interactive effects of historical logging and fire exclusion on ponderosa pine forest structure in the northern Rockies. Ecol. Appl. 20, 1851–1864.
Nelson, G.C., Dobermann, A., Nakicenovic, N., O'Neill, B.C., 2006. Anthropogenic drivers of ecosystem change: An overview. Ecol. Soc. 11, 29.

Ostrom, E., 2009. A general framework for analyzing sustainability of social-ecological systems. Science 325, 419–422.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37.

Parsons, D.J., DeBenedetti, S.H., 1979. Impact of fire suppression on a mixed-conifer forest. For. Ecol. Manage. 2, 21–33.

Perring, M.P., De Frenne, P., Baeten, L., Maes, S.L., Depauw, L., Blondeel, H., Carón, M.M., Verheyen, K., 2016. Global environmental change effects on ecosystems: the importance of land-use legacies. Glob. Chang. Biol. 22, 1361–1371.

Petit, R.J., Hu, F.S., Dick, C.W., 2008. Forests of the Past: A Window to Future Changes. Science 320.

Radeloff, V.C., Hammer, R.B., Stewart, S.I., Fried, J.S., Holcomb, S.S., McKeefry, J.F., 2005. The wildland-urban interface in the United States. Ecol. Appl. 15, 799–805.

Rubidge, E.M., Monahan, W.B., Parra, J.L., Cameron, S.E., Brashares, J.S., 2011. The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. Glob. Chang. Biol. 17, 696–708.

Santos, M.J., Thorne, J.H., Christensen, J., Frank, Z., 2014. An historical land conservation analysis in the San Francisco Bay Area, USA: 1850–2010. Landsc. Urban Plan. 127, 114–123.

Seidl, R., Schelhaas, M.-J., Lexer, M.J., 2011. Unraveling the drivers of intensifying forest disturbance regimes in Europe. Glob. Chang. Biol. 17, 2842–2852.

Sleeter, B.M., Sohl, T.L., Loveland, T.R., Auch, R.F., Acevedo, W., Drummond, M.A., Sayler, K.L., Stehman, S.V., 2013. Land-cover change in the conterminous United States from 1973 to 2000. Glob. Environ. Change 23, 733–748.

Steen-Adams, M.M., Langston, N., Adams, M.D.O., Mladenoff, D.J., 2015. Historical framework to explain long-term coupled human and natural system feedbacks: Application to a multiple-ownership forest landscape in the Northern Great Lakes region, USA. Ecol. Soc. 20. https://doi.org/10.5751/ES-06930-200128

Stephens, S.L., Collins, B.M., Fettig, C.J., Finney, M.A., Hoffman, C.M., Knapp, E.E., North, M.P., Safford, H., Wayman, R.B., 2018. Drought, Tree Mortality, and Wildfire in Forests Adapted to Frequent Fire. Bioscience 68, 77–88.

Svenning, J.-C., Sandel, B., 2013. Disequilibrium vegetation dynamics under future climate change. Am. J. Bot. 100, 1266–1286.

Swetnam, T., Allen, C.D., Betancourt, J.L., 1999. Applied Historical Ecology: Using the Past to Manage for the Future. Ecol. Appl. 9, 1189–1206.

Taylor, A., 2000. Fire regimes and forest changes in mid and upper montane forests of the southern Cascades, Lassen Volcanic National Park, California, U.S.A. J. Biogeogr. 1905, 87–104.

Thorne, J.H., Le, T.N., 2016. California’s Historic Legacy for Landscape Change, the Wieslander Vegetation Type Maps. Madroño 63, 293–328.

Turner, M.G., 2005. Landscape Ecology: What Is the State of the Science? Annu. Rev. Ecol. Evol. Syst. 36, 319–344.

Turner, M.G., Wear, D.N., Flamm, R.O., 1996. Land Ownership and Land-Cover Change in the Southern Appalachian Highlands and the Olympic Peninsula. Ecol. Appl. 6, 1150–1172.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steininger, M., 2003. Remote sensing for biodiversity science and conservation. Trends Ecol. Evol. 18, 306–314.

van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fule, P.Z., Harmon, M.E., Larson, a. J., Smith, J.M., Taylor, a. H., Veblen, T.T., 2009. Widespread Increase of Tree Mortality Rates in the Western United States. Science 323, 521–524.

Whitlock, C., Colombaroli, D., Conedera, M., Tinner, W., 2017. Land-use history as a guide for forest conservation and management. Conservation Biology 32, 84–97.

Wieslander and Wilson, R. C. Classifying Forest and Other Vegetation from Air Photographs, Photogrammetric Engineering Vol. 8, No. 3, July-August-September 1942, pp. 203-215.

Chapter 2: Land ownership and 20th century changes to forest structure in California

This article has previously been published and is reproduced here with kind permission from co-authors and the Graduate Division

Easterday, Kelly, Patrick McIntyre, and Maggi Kelly. 2018. “Land Ownership and 20th Century Changes to Forest Structure in California.” *Forest Ecology and Management* 422 (August): 137–46.

Abstract

Forests in California have changed dramatically during the 20th century. Shifts in forest structure including densification, declines in large trees and tree basal area have altered the function, productivity, and resilience of modern day forests. Attributing these changes to specific drivers is increasingly important for effective management of healthy and productive forests. Previous studies focus on climatic (temperature, precipitation, climatic water deficit), disturbance (fire), geomorphological (topography, soil types), and anthropogenic (logging, fire suppression) drivers, but few studies evaluate large scale change in forest structure across land ownership type. In this paper, we investigate 20th century changes to forest structure across six land ownership classes in California. We compare historical and contemporary forest structural data and find that declines in large trees and increases in forest density are consistent across the state. This pattern is most pronounced on private timberlands, which experience up to 400% regional increases in small tree (< 10.2 cm) density since 1930. All land ownership classes experience declines in large trees, while private timberlands, national parks, and wilderness areas experience the most extreme change with an average loss of over 83% and 71% respectively. We conclude that understanding patterns of change across land ownership is essential for targeting federal, state, and locally specific policies that foster healthy and resilient forests for the future.

Keywords

Forest structure, forest change, land management, land ownership, California, Vegetation Type Mapping (VTM)

1 Introduction

Present-day forests in California are markedly different from their early 20th century counterparts. Numerous studies show changes in the structure and composition of California's forests by documenting shifts towards more small and fewer large trees (Dolanc et al., 2013a; Lutz et al., 2009; McIntyre et al., 2015); more structurally homogenous stands (Maxwell et al., 2014); and changes in species composition (McIntyre et al., 2015; Minnich et al., 1995; Taylor, 2000; van Mantgem et al., 2013). These changes vary over space and time due to the interaction of biophysical and socio-ecological drivers. In California, large tree decline has been attributed to increased temperatures, variable precipitation, and water deficit (Das et al., 2013; McIntyre et al., 2015; van Mantgem et al., 2013, 2009), as well as historical and contemporary legacies of logging (Knapp et al., 2013; Laudenslayer and Darr, 1990; McKelvey and Johnston, 1992; Beesley, 1996). Large scale forest densification, in part, is the result of nearly a century of widespread fire suppression efforts (Dolanc et al., 2014; 2013b; Lutz et al., 2010; Minnich et al., 1995), with previously logged lands showing greater densities than surrounding landscapes (Naficy et al., 2010). The lack of natural fire and increasing forest density positively correlate with a shift in species composition favoring shade-tolerant species (Miller et al., 2012; Taylor and Skinner, 2003). Such legacies of logging and forest fire suppression have profound impacts that can persist for decades after cessation, altering both the state of contemporary landscapes and influencing future trajectories of change (Perring et al., 2016). These legacies are often specific to the land management practices of a given land owner at a specific time. Given the difficulty in disentangling regional biophysical and socio-ecological drivers, an understanding of forest structure change across land ownership is needed. Additionally, determining how long-term patterns of change vary across ownerships is necessary to help target federal, state, and locally specific management policies that foster healthier more resilient forests for the future.

Land ownership has been used to understand the long-term effects of and variation in land management practices; especially when spatially explicit data on management practices are unavailable or incomplete. In agricultural landscapes for example, Lunt and Spooner (2005) showed that land ownership is predictive of disturbance and therefore can be used to better understand past, current, and future patterns of biodiversity in fragmented areas. In forested landscapes, Turner et al. (1996) showed that property boundaries create quantifiable patterns of land use change and that the similarities in these changes across ownerships are reflective of specific management goals. They showed that while forests on private lands were more fragmented than those on public lands, when areas had a common management goal (e.g. active timber harvesting) forests displayed similar spatial patterns.

Studies documenting changes in forest structure across land ownership at a large scale in California are rare. In this paper, we compared historical 1930s Vegetation Type Mapping project (VTM) forest survey plots with modern 2000s Forest Inventory and Analysis program (FIA) forest inventory plots and examined changes in measures of forest structure: stems per ha per size class (small, medium, large, and total) and total basal area (m^2 /ha).

We assessed change over time in these variables across six California land ownership classes: (1) Private Timberland (PT), (2) NonWilderness National Forest

(NWNF), (3) Non-Wilderness Bureau of Land Management and Tribal Lands (NWBT), (4) Private Protected lands (PP), (5) State and Regional Parks (SR), and (6) National Parks and Wilderness areas (NPW). We distinguished changes in stand density and size class distributions across land ownership and investigated differences in these measures between the six land ownership classes. We address the following questions: (1) have the numbers and sizes of trees changed significantly over time across all six land ownership classes; (2) how do changes in the number of trees per size class and forest densification vary by land ownership; and (3) do these patterns suggest differing land use legacies across ownership classes

2 Methods

2.1 Study Area

Our study area includes the forests of the California floristic province including the Northwestern, Sierra Nevada, Central, and South regions. This area has a Mediterranean climate of dry summers and wet winters. Regional differences in climate and soil characteristics are captured by geomorphic regions that are largely determined by the mountain ranges that divide them. The six land ownership classes investigated cover a range of ecoregions and vegetation types and are also representative of regional characteristics that correspond with spatial patterns of ownership.

California is a complex mosaic of land ownership, with federal, state, tribal, and local entities protecting and managing land. Nearly 150,000 km^2 of forest are managed by distinct ownerships with varying degrees of protection, production, and conversion of forests. 48% of California's forested lands (63,130 km^2) are managed by the U.S. Forest Service as National Forests, 51,395 km^2 (39%) are managed as private timberland encompassing both industrial and non-industrial private forest land. Approximately 8095 km^2 (6%) is set aside as forest reserves and managed through Wilderness designation or as a National Park, while various other private and public entities manage the remaining ~8900 km^2 (7%) (McIver et al., 2015).

PT includes both industrial and non-industrial forest lands, however the majority of plots investigated in this study were on lands managed for industrial timber. Generally, PTs are located on mixed conifer forests in the Northern Sierras, Klamath, and Cascade Ranges, and in the Douglas Fir and Redwood forests of the North and Central Coasts (Stewart et al., 2016), and tend to occur at lower elevations. NWNF areas are also located extensively in mixed conifer forests, interspersed with pockets of Red Fir, Eastside Pine, and Ponderosa Pine and extending into the hardwood forests and woodlands of the Central and South Coast.

NWBT lands are distributed in the low elevations of the North Coast, Mojave, Sierra, Central, and South Coast regions, and in our study area, consist of primarily conifer forests and woodlands concentrated in the Eastern Sierra Nevada, Klamath Ranges, and South Coast Ranges. Very few of our study plots occur on tribal lands, therefore NWBT is primarily illustrative of BLM lands.

PP land is scattered in the matrix of federal, state, and private ownerships, generally representing hardwood woodlands and hardwood forests. SR lands in our study are primarily hardwood woodlands within the greater San Francisco Bay Area. NPW areas are representative of National Parks and all federal agency owned wilderness. These lands are primarily located in the Sierra Nevada region, as well as the Southern Sierra and Transverse Ranges. Forest types in NPW are primarily mixed conifer but also higher elevation Red Fir, Lodgepole Pine, Jeffery Pine, and hardwood forest types. The spatial distribution of ownership types expresses regional concentrations owing to California's complex land settlement history, therefore the patterns represented in this study are reflective of differences across ownership that are particular to the regions where the plots are located.

2.2 Data

2.2.1 Historical and contemporary forest inventory data

The Vegetation Type Mapping (VTM) project is a series of landscape surveys conducted by the US Forest Service that covered ~40% of California between 1928 and 1940 that resulted in a large collection of 350 vegetation maps, 18,000 vegetation plots, over 3000 photographs and ~20,000 herbarium specimens (Wieslander, 1935). These data are digitized and georeferenced (see Kelly et al., 2005; Kelly et al., 2016; Kelly et al., 2017) and available for download via an open API and for download (vtm.berkeley.edu). In this study, we use the vegetation plot data, including geolocated information on numbers, diameter, and species of trees as well as other ancillary environmental information (e.g. elevation) associated with the marked plot location (Fig. 1a). The VTM crews conducted complete inventories of all trees over 10.2 cm diameter at breast height (DBH) within 20 m by 40 m (800m^2) rectangular plots. The trees were tallied by species into four individual size classes: 10.2–30.4 cm, (4–12 in), 30.5–60.9 cm (12–24 in), 61.0–91.3 cm (24–36 in), and > 91.4 cm (> 36 in) (Kelly et al., 2005).

The VTM survey began in 1928, just after the beginning of large scale forest fire suppression across the state and in most areas before the 1940s and 1950s peak in forest harvesting. Today, the VTM collection serves as one of the only comprehensive datasets describing the California landscape in the early 20th century. Working with historical datasets requires the acknowledgment and examination of challenges such as plot geolocation error and potential bias. In the VTM dataset for example, plot location is derived from original markings on historical topographic maps and positional error is estimated to be ~200 m (Kelly et al., 2005) per plot which can affect direct plot comparisons or plot resurveys, especially in highly heterogeneous regions (Keeley, 2004).

The protocols behind VTM methods have raised questions about biased sampling favoring undisturbed forests. However, there is no evidence of bias suggested in the original VTM manual (Wieslander et al., 1933), or in the sample plot distributions, yet there are competing patterns of change when comparing FIA and VTM estimates to other historical comparisons. Some studies using alternative comparison datasets or plot resurvey have shown similar patterns of declines in large trees (e.g. Lydersen et al., 2013; van Mantgem et al., 2009) as the VTM dataset, while other studies have shown declines in large trees on some ownership classes but not on others and increases in basal area across types (Collins et al., 2017; Lydersen et al., 2013). These disparities are difficult to verify as the datasets in question are not directly comparable but do require a cautionary approach to interpreting changes in large trees and biomass. There is no record of intentional bias in the selection of VTM plot locations the locations were chosen as representative samples of vegetation types being mapped (Wieslander et al., 1933), and have been shown to have similar sampling densities across elevation and latitude as FIA plots which are determined randomly using a grid system (Dolanc et al. 2013a). Despite these potential shortcomings, recent work finds utility in the VTM dataset (see references in Kelly et al., 2016; Thorne and Le, 2016) and in comparing forest structure and compositions between VTM and FIA datasets (Dolanc et al., 2013a; McIntyre et al., 2015).

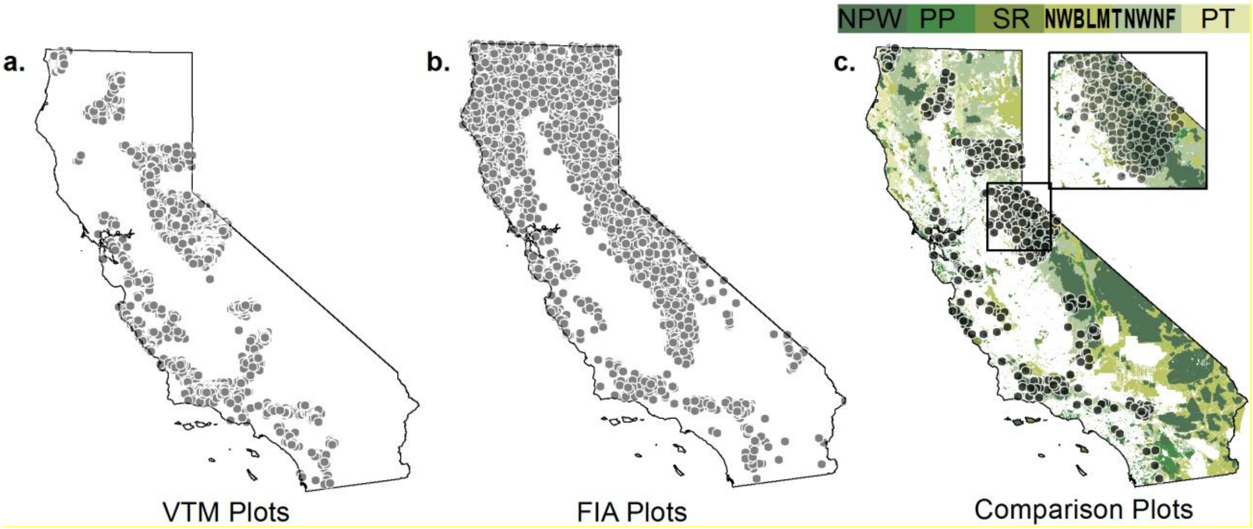


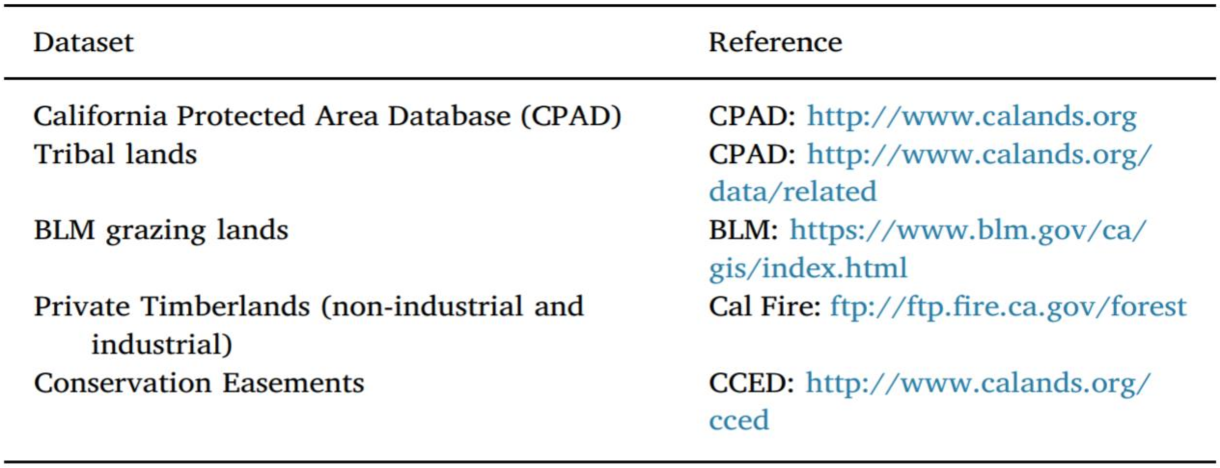
Figure 1 Locations of historical and contemporary forest structure plot data for: (a) VTM (1930s–1940s); (b) FIA (2000–2010); and (c) comparison plots used in analysis and spatial distribution of ownership classes. Map 1(c) shows ownership classes with an inset map of the Sierra Nevada region.

The contemporary FIA dataset is a national inventory program implemented by the U.S. Forest Service, which contains systematically collected detailed monitoring data on forests in some regions since its inception. Surveys included detailed data on species, size, tree condition, and other site factors (Smith, 2002). The 2001–2010 FIA protocol uses circular subplots of 7.3 m (24 ft) radius where every tree greater than > 12.7 cm (5 in) is measured, and within which a microplot (2.1 m) is nested and information on stems > 2.5 cm (> 1.0 in) is collected (Bechtold and Patterson, 2005). This information is collected across four subplots giving a total area of 672.45 m^2. Locations are distributed throughout California's forested areas (Fig. 1b). Although the FIA data collection protocols differ from those used in the earlier VTM surveys, explicit individual tree level information extracted from the FIA dataset allows us to match the size classes determined in the VTM protocol. Data on small trees (> 10.2 cm) from the FIA microplots were added to match the minimum size class of the classes found in the VTM dataset. Due to differences in the VTM and FIA plot sizes, we report accounts of tree numbers per hectare and basal area (m^2/ha). Other studies (e.g. McIntyre et al., 2015; Dolanc et al., 2013a) provide further information for protocols associated with these data transformations.

2.2.2 Spatial depiction of California land ownership

We assembled several freely available datasets to develop a statewide spatial depiction of land ownership classes in California (Fig. 1c, Table 1). We generated categorical descriptions of land management based on ownership using the agency level distinction from the California Protected Area Database (CPAD) to aggregate parcels into six land ownership classes that simplify the diversity of land ownership in California (Table 2). Therefore, the names of the ownership classes may represent multiple owners or distinctions (e.g. BLM includes Tribal lands).

This depiction of California's land ownership suggests a contemporary and static picture; parcel establishment dates and ownership permanence can and have shifted over time within our study period of 1930–2010. However, many of the federal lands were established during the turn of the 20th century. Designation of National Parks and forest reserves between 1880 and 1920 largely protected forest and mountainous landscapes (Santos et al., 2014). These newly protected landscapes were also amongst the first surveyed by the VTM crews and correspond with the highest densities of VTM plots locations. Other areas experienced shifts in land use and tenure. For example, in the North Coast commercial logging in redwood forests dominated the late 19th and early the 20th centuries yet subsequently many areas became federally and locally protected. After the establishment of federally protected forest and National Park lands many of California's conservation lands were acquired through State Park designations in the 1930s (Santos et al., 2014). Therefore, the majority of the plots used in the analysis are located on lands with relatively stable land tenure and the changes to the forest within those boundaries are likely representative of the management legacies of the land owner.



Dataset	Reference
California Protected Area Database (CPAD)	CPAD: http://www.calands.org
Tribal lands	CPAD: http://www.calands.org/data/related
BLM grazing lands	BLM: https://www.blm.gov/ca/gis/index.html
Private Timberlands (non-industrial and industrial)	Cal Fire: ftp://ftp.fire.ca.gov/forest
Conservation Easements	CCED: http://www.calands.org/cced

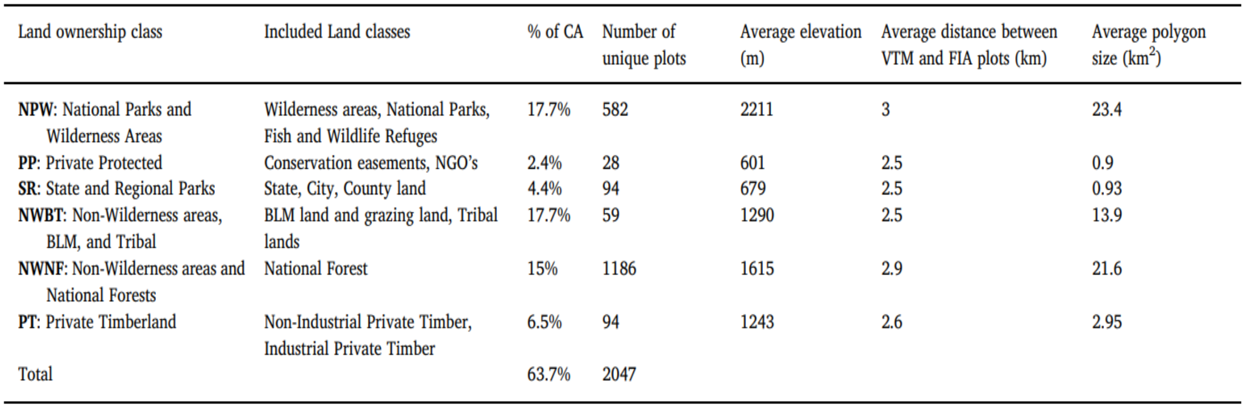
Table 1 The datasets used to create the California land management layer. The California Protected Area Database (CPAD) database was used as the base dataset, the other datasets listed were added to fill in areas where CPAD does not collect information.

2.3 Analysis

2.3.1 Plot comparison

We employ methodology from McIntyre et al. (2015) to calculate measures per species by plot for each dataset (e.g. VTM and FIA): total numbers of trees per ha, numbers of trees per ha within size class 1 (i.e. small trees: 10.2–30.4 cm DBH), numbers of trees in size class 2 (i.e. medium trees: 30.5–60.9 cm DBH), numbers of trees in sizes class 3 (61.0–91.3 cm DBH), numbers of trees per ha within size class 4 (> 91.4 cm DBH), and total basal area (m^2 /ha). Trees within size class 3 and 4 are combined to create a large tree category (> 61 cm DBH).

Per species measures were summed to represent totals by plot and size class for each period and only plots with more than one tree were used in the analysis. Each VTM and FIA comparison plot was assigned an ownership class based on spatial overlap (Fig. 1c). Error in the VTM and FIA plot locations (Waddell, 2013) can contribute to potential erroneous assignment of ownership classes, however the potential positional error of 0.2–1.6 km was much smaller than the average area (5.25 km^2) of individual polygons within ownership classes (Table 2), and further compensated for by the scale of our analysis (Waddell, 2013). Plots where a land ownership description was unavailable were removed from the analysis. We used elevational values derived from a digital elevation model to group plots into 500 m elevation classes. Using these classes, we refined our plot comparison dataset to capture plots that were assigned the same ownership class, elevational class, and were within a 5 km distance threshold. To ensure that the plots adhered to the independence assumptions of our statistical tests, multiple plot matches were evaluated. If the stratification returned more than one FIA plot matching the above criteria, the average of all corresponding plots was used. If a VTM plot matched with a single FIA plot, the plot with the shortest distance was compared. Plot matches were on average no more than 2.9 km apart. From an original dataset of 9388 VTM and 5198 FIA plots, our final comparison dataset was 2047 VTM and 2047 FIA plots.



Land ownership class	Included Land classes	% of CA	Number of unique plots	Average elevation (m)	Average distance between VTM and FIA plots (km)	Average polygon size (km^2)
NPW: National Parks and Wilderness Areas	Wilderness areas, National Parks, Fish and Wildlife Refuges	17.7%	582	2211	3	23.4
PP: Private Protected	Conservation easements, NGO's	2.4%	28	601	2.5	0.9
SR: State and Regional Parks	State, City, County land	4.4%	94	679	2.5	0.93
NWBT: Non-Wilderness areas, BLM, and Tribal	BLM land and grazing land, Tribal lands	17.7%	59	1290	2.5	13.9
NWNF: Non-Wilderness areas and National Forests	National Forest	15%	1186	1615	2.9	21.6
PT: Private Timberland	Non-Industrial Private Timber, Industrial Private Timber	6.5%	94	1243	2.6	2.95
Total		63.7%	2047			

Table 2 Description of land ownership classes in California (several land owners can be aggregated into a single type), and relevant characteristics

2.3.2 Change over time

To assess if the numbers of trees within each size class (small, medium, large, total) and total basal area changed significantly over time across all land ownership classes we conducted a Wilcox test (also known as the Mann-Whitney U test, or Wilcoxon rank sum test). The Wilcox test is a non-parametric analogue to the t-test often used to detect a significant difference in population means or medians (Fay and Proschan, 2010). This interpretation of the test relies on the assumptions that the two populations are independent and have the same shape and equal variance. However, the Wilcoxon-Mann-Whitney test is still valid in situations where the homogeneity of variance assumption is violated (Fay and Proschan, 2010), does not rely on the assumption that the data follow a normal distribution and can be used on data with extreme outliers

(MacFarland and Yates, 2016), as was apparent in our data. When the variance assumption is violated, rather than testing for a difference in the medians, the hypothesis tests that a randomly selected value from one sample population will be greater than a randomly selected value from the second sample population. Using the two populations (Time1 = VTM and Time2 = FIA), a difference in the population distributions is reflected as a "difference in location" or the median of the differences between a VTM sample and a FIA sample (R Core Team, 2013). To calculate the actual median difference between the two periods we used a bootstrapping approach, generating 5000 iterations of the differences in the population medians from which we took an overall median. We interpret the overall median difference derived from the bootstrapping approach as the median value of change over time in the populations between VTM and FIA within the ownership class. The Wilcox test and resultant p-value < 0.05 describe the change between the FIA and VTM populations as significantly different. All analyses were implemented with the R statistical software base function wilcox.test (R Core Team, 2013), and the boot package (Canty and Ripley, 2017). Finally, we assessed the change in the mean count per plot of the most common species across ownership class.

2.3.3 Change between ownership class

To assess how changes in the number of trees per size class and total basal area vary by land ownership we used the difference between FIA and VTM (i.e. FIA-VTM) variables at the plot level (as distinct from the difference procedure discussed above in 2.3.2, which looks at difference across the entire population using a bootstrapping method) per ownership class directly. The difference or change values for each of the five variables followed a normal distribution, and so we ran an ANOVA to compare the effect of ownership on each variable as well as calculated overall means and 95% confidence intervals. Statistically significant differences between ownership types were determined for each of the five change variables. We then ran post hoc Tukey tests on the significant ANOVAs to compare specific differences between the ownership classes. This permitted us to determine if the change experienced on one ownership class was significantly different from the change on another class.

2.3.4 Spatial pattern of change across ownership class

Additionally, to visualize the spatial distribution of increase, decrease, and extreme change in the five variables we calculated the overall mean of the change values of all plots within 20 km and 5 km resolution grid cells. The 20 km resolution was chosen to be comparable to results from McIntyre et al. (2015), and 5 km resolution represents the maximum allowable distance between the comparison plots. The ownership classes were also aggregated to 20 km and 5 km using the majority method, which does lose granularity as small parcels are subsumed by larger ones. Change values were classed using the Jenks Natural Breaks algorithm, which uses the variance to maximize differences between classes and minimize differences within classes (Jenks, 1977). These classes were reported in terms of increase and decrease, reflecting positive and negative change values. To highlight areas of extreme change, we also classed the values by standard deviations, where extreme change values are defined as change that is three standard deviations from the mean.

3 Results

3.1 Change over time

When viewed holistically (n = 2047 comparison plots), forest structure across California has changed significantly (Table 3). There has been an increase in small trees, medium trees, total trees, and a decline in large trees and TBA. However, critical differences across ownership exist, forests on Private Timberlands, Non-Wilderness National Forests, and National Parks and Wilderness areas (i.e. PT, NWNF, and NPW) showed more pronounced structural changes than do those on State and Regional Parks, Private Protected, and Non-Wilderness BLM and Tribal lands (SR, PP, and NWBT).

PT, NWNF, and NPW represent the largest proportions of managed land in California (Table 2), and these lands have lost significant numbers of large trees (> 61 cm DBH) over the 20th Century (Fig. 2). PT showed the largest declines with a median difference of nearly 50 trees per ha from the 1930s (VTM) to 2010 (FIA), a change of over 83% (Table 3). Rates of large tree declines on Private Timberlands were followed by NPW with a median difference of over 30 trees per ha (71% change) and then NWNF with a 20 tree per ha difference (52%). The median difference on SR, PP, and NWBT land was negligible, as the majority of the plots located on those lands had zero counts of large trees.

Patterns of forest density, primarily driven by significant increases (up to ~137%) in the counts of small trees (10.2–30 cm DBH) per ha were consistent across the state except in SR (Table 3). Private Timberlands (PT) (gains of ~399%), NWNF (gains of ~190%), and NPW (gains of ~85%) had statistically significant increases in small trees (Table 3). Correspondingly, all ownership classes with the exception of State and Regional Parks (SR) experienced increases in total trees per ha (Fig. 2). Significant increases (up to 55%) in total numbers of trees on the landscape were significant on PT with a median difference of 371 trees per ha (139% increase); NWNF with a median increase of 191 trees per ha (88% increase); and NPW with a median increase of nearly 50 trees per ha (16% increase). PP and NWBT showed slight, but not significant gains in trees per ha.

Declines in large trees were generally reflected by corresponding declines in total biomass as represented by total basal area (TBA m^2 / ha) across ownership classes (Table 3 and Fig. 2). Across PT (35% TBA decline), NPW (29% TBA decline), SR (21% TBA decline), and NWNF (6% TBA decline) forests showed statistically significant declines in total basal area over the 20th century. Private Timberlands had the largest median difference between the two-time periods with decreases of 19 m^2 /ha (36%). NPW lost 14 m^2 /ha of basal area (29%), and NWNF showed relatively small overall change losing 2 m^2 /ha (6%). Conversely, NWBT land experienced non-significant gains in basal area (2 m^2 /ha /ha) (Table 3).

With the exception of SR and PP lands, which show contradictory patterns with increase in Pinus species (P.*monticola*,P.*ponderosa*,)and decline in Quercus species most other ownership types experience a decrease in most Pinus species, while Quercus species increase and decrease depending on land owner and species (Figure A1,Table A1). Average counts of Q. *chryolepis* increases across most ownership types except for SR and PP with largest increases on PT. Counts of Q. *lobata* decline across all types, expect for NPW. Mean counts of Q. *wislizeni* increase across all types except for PT. Mean counts of Q. *agrifolia* decline everywhere except for PT. P. *sabinia* decreases on all ownership types, P. *ponderosa* decreases on all except for PP. Most types show increase in average P.*menziesii* counts. Also, increased mean counts of U.*californica*, and L. *densiflorus* occurred across management types.

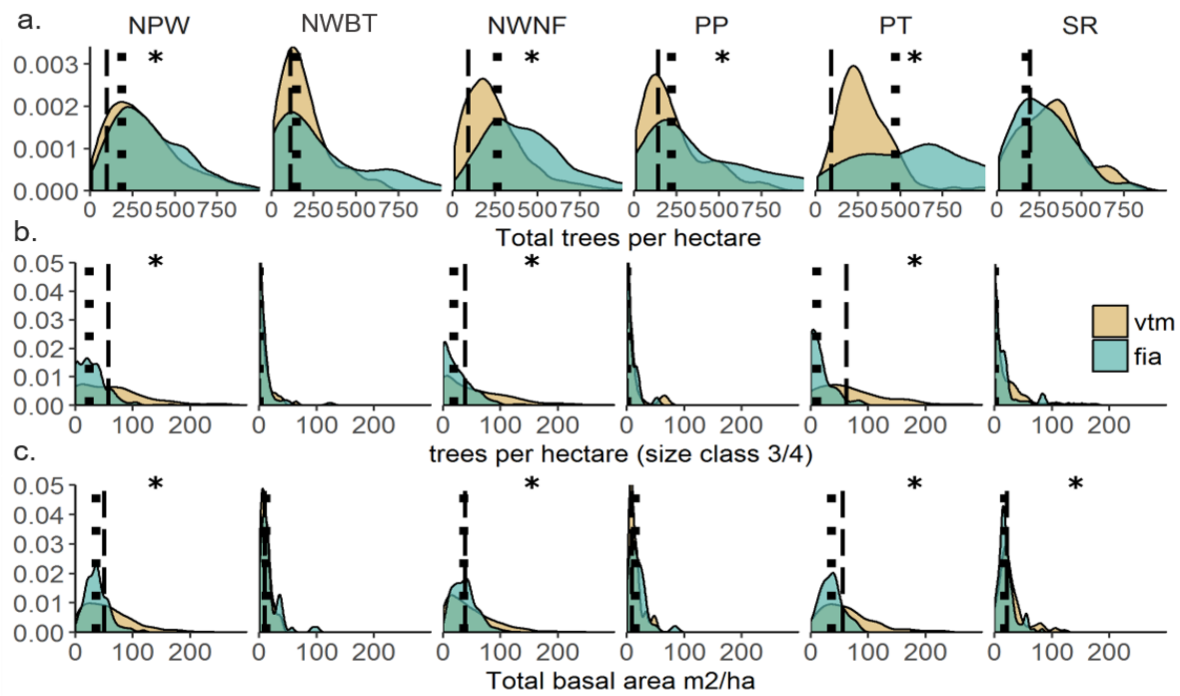


Figure 2 Change in forest structure over time for: (a) total trees per ha; (b) large trees per ha; and (c) total basal area (m2 /ha). Each plot shows the density distribution of each time period for each ownership class. Significant (p-value < 0.05) indicating dissimilar population distributions are shown with an '*'. Median values of VTM distribution are dashed lines, median values of the FIA distribution are shown with dotted lines. Some plots at the ends of the distribution have been removed to show the part of the distribution with the highest densities.

Table 3 Summary statistics for stems/ha for small, medium, large, and total trees and the average m^2/ha of total basal area for each ownership class and the average across the whole dataset. We calculated change between FIA-VTM and report percent change. Median differences from are results from 5,000 iterations comparing a median value from FIA to a median value in VTM and generating a global median. Significant p-values (from Wilcox) < 0.05 are shown in bold.

	Small trees	**Medium trees**	**Big trees**	**Total trees**	**Total basal area**
	stems per hectare				m^2/hectare
All areas (n = 2,047)					
VTM median	99.33	50.94	38.70	235.03	38.96
FIA median	235.43	84.72	18.23	366.60	34.15
% change	137.03	66.32	-52.90	55.98	-12.33
Median Difference (bootstrap)	136.71	33.97	-20.80	131.46	-4.70
P-value (Wilcox)	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
PT: Private Timberlands (n = 98)					
VTM median	93.99	78.54	61.96	262.04	56.18
FIA median	469.81	92.49	10.47	626.40	36.19
% change (relative)	399.86	17.77	-83.11	139.04	-35.58
Median Difference (bootstrap)	375.66	10.63	-49.17	371.35	-18.87
P-value (Wilcox)	**< 0.001**	0.11	**< 0.001**	**< 0.001**	**< 0.001**
NWNF: Non-Wilderness and National Forest (n = 1,186)					
VTM median	90.29	49.48	38.70	216.63	38.25
FIA median	262.56	91.16	18.53	407.55	35.85
% change	190.80	84.23	-52.10	88.14	-6.27
Median Difference (bootstrap)	171.91	42.50	-20.40	191.62	-2.40
P-value (Wilcox)	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**0.002**
NPW: National Parks and Wilderness (n = 582)					
VTM median	100.85	66.53	57.63	273.15	50.24
FIA median	187.01	84.72	16.51	318.35	35.91
% change	85.43	27.34	-71.36	16.55	-28.52
Median Difference (bootstrap)	84.41	19.28	-33.56	48.99	-14.42
P-value (Wilcox)	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
NWBT: Non-Wilderness BLM and Tribal (n = 59)					
VTM median	112.43	25.80	0.00*	152.82	9.94
FIA median	147.24	28.92	0.00*	185.89	11.79
% change	30.96	12.11	NA	21.64	18.60
Median Difference (bootstrap)	28.96	7.34	0.00	36.84	2.29
P-value (Wilcox)	0.08	0.32	NA	0.10	0.15
SR: State and Regional Parks (n = 94)					
VTM median	196.34	63.89	0.00*	321.33	21.94
FIA median	172.12	67.89	0.00*	262.55	17.42
% change	-12.33	6.27	NA	-18.29	-20.59
Median Difference (bootstrap)	-29.11	3.95	0.00	-58.30	-4.49
P-value (Wilcox)	0.29	0.99	NA	0.15	0.02
PP: Private Protected (n = 28)					
VTM median	141.47	39.27	0.00*	173.14	9.16
FIA median	217.93	39.00	0.00*	247.33	14.84
% change	54.04	-0.68	NA	42.85	61.92
Median Difference (bootstrap)	74.90	3.11	0.00	70.08	4.63
P-value (Wilcox)	0.09	0.46	NA	0.05	0.22

*Where the values are equal to zero, the majority of plots had zero stem per hectare counts.
Wilcox tests were not run on zero values.
Mean values for these classes are as follows: SR (VTM: 19.26, FIA: 14.53), PP (VTM: 6.03, FIA: 4.81), NWBT (VTM: 5.56, FIA: 6.19).

3.2 Change between ownership class

The direct change (FIA-VTM) across management type for total trees, large trees, and TBA were significant for PT, NWNF, and NPW (Fig. 3). PT showed the largest average changes with an increase of 310 (95% CI ± 5) total trees/ha and average declines of −54 (95% CI ± 15) in large trees/ha and −26 (95% CI ± 4) TBA (m^2 /ha). NWNF also showed large average increases of 193 (95% CI ± 21) total trees/ ha with declines of −31 (95% CI ± 3) large tree/ha and −12 (95% CI ± 3) TBA (m^2 /ha). Finally, NPW areas showed similar average declines in large trees/ha (−37, 95% CI ± 5) as NWNF, but compared to PT and NWF had lower average increases in total trees/ha (53, 95%CI ± 25) and corresponding declines of −20 (95% CI ± 3) in TBA (m^2 /ha) (Fig. 3).

Despite the constancy in overall patterns of forest densification and shifting tree size class distributions statewide, there are important differences in the amount and direction of change between ownership class (Table 4). Changes in total trees and numbers of small and medium trees per ha on NWNF lands are significantly different from those changes on NPW lands, with NWNF lands experiencing greater increases than NPW. However, NWNF and NPW have changed similarly in terms of declining densities of large trees (Table 4). Changes in NWNF and PT were significantly different from each other in terms of small trees, large trees, and total trees, with PT changes in small trees being significantly different than every other land ownership. Changes on SR were consistently significantly different from other ownership classes, except for PP which due to low samples size is not statistically different than most other ownership types (Table 4). NWBT was most dissimilar in terms of large tree and total basal area decline to NPW, NWNF, and PT (Table 4).

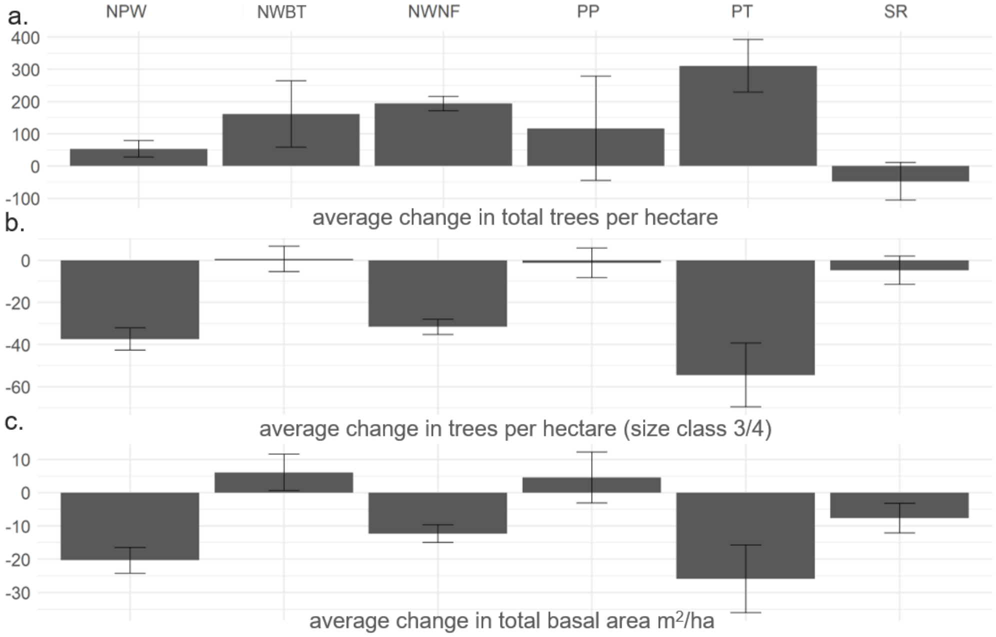


Figure 3 Average direct change (FIA-VTM) in forest structure by plot and by management type for: a) total trees per ha; b) large trees per ha; and c) total basal area (m^2/ha) with 95% confidence intervals on change values.

Table 4 Significant difference in the change experienced by pairs of ownership class for each variable are demarcated by a “ ✓ ” where Tukey Honest Significant Difference test resulted in a P-value < 0.05; pairs that were not significantly different from each other show 0 significant variables.

Management Class Comparison Pair	Small trees	Medium trees	Large trees	Total trees	Total basal area	Number of significant variables
PP-NWBT						0
PP-NWNF						0
SR-PP						0
PP-NPW			✓			1
NWBT-NPW			✓		✓	2
PT-NPW	✓			✓		2
SR-NPW	✓		✓			2
NWNF-NWBT			✓		✓	2
SR-NWBT	✓			✓		2
PT-PP	✓		✓		✓	2
NWNF-NPW	✓	✓		✓	✓	3
PT-NWBT	✓		✓		✓	3
PT-NWNF	✓		✓	✓		3
SR-PT	✓		✓	✓		3
SR-NWNF	✓	✓	✓	✓		4

3.3 Spatial pattern of change across ownership class

Increases in small trees and in total trees were widespread throughout the northern Sierra Nevada region, and more scattered throughout the central coast and southern part of the state (Fig. 4b and e). Patterns of medium tree change were variable across ownership class and regions (Fig. 4c). Declines in large trees were widespread throughout the state and pronounced in the Sierra Nevada region (Fig. 4d). The spatial pattern of large tree declines largely reflected the pattern of decline in total biomass across California (Fig. 4f), however, patterns of TBA exhibited more local heterogeneity particularly in the central coast. The central coast regions depicted strikingly different changes than the Sierra Nevada region, with average decreases in small, medium and total trees in contrast to the increases found in the Sierra Nevada region. These are predominantly SR lands, which showed nonsignificant changes across most measures (Fig. 2, Table 3). Extreme changes (> 2.0 standard deviation decrease or increase in any measure over the 20th century) are shown in Fig. 5. Extreme changes are scattered throughout the state, with some important local trends. The timber production zones (PT) of the North Coast and Klamath Ranges showed extreme increases (> 2.0 standard deviation from the mean) in the number of small trees and total trees (shown in dark green in Fig. 5b and e). Large tree decline was also most extreme in the higher elevation areas of the Sierra Nevada, and in the Non-Wilderness National Forest and Private Timberlands of the northern Klamath Ranges (shown in purple in Fig. 5d). An alternative view of this change along a spectrum of ownership for use or conservation can be found in the appendix (Figure A2 and A3).

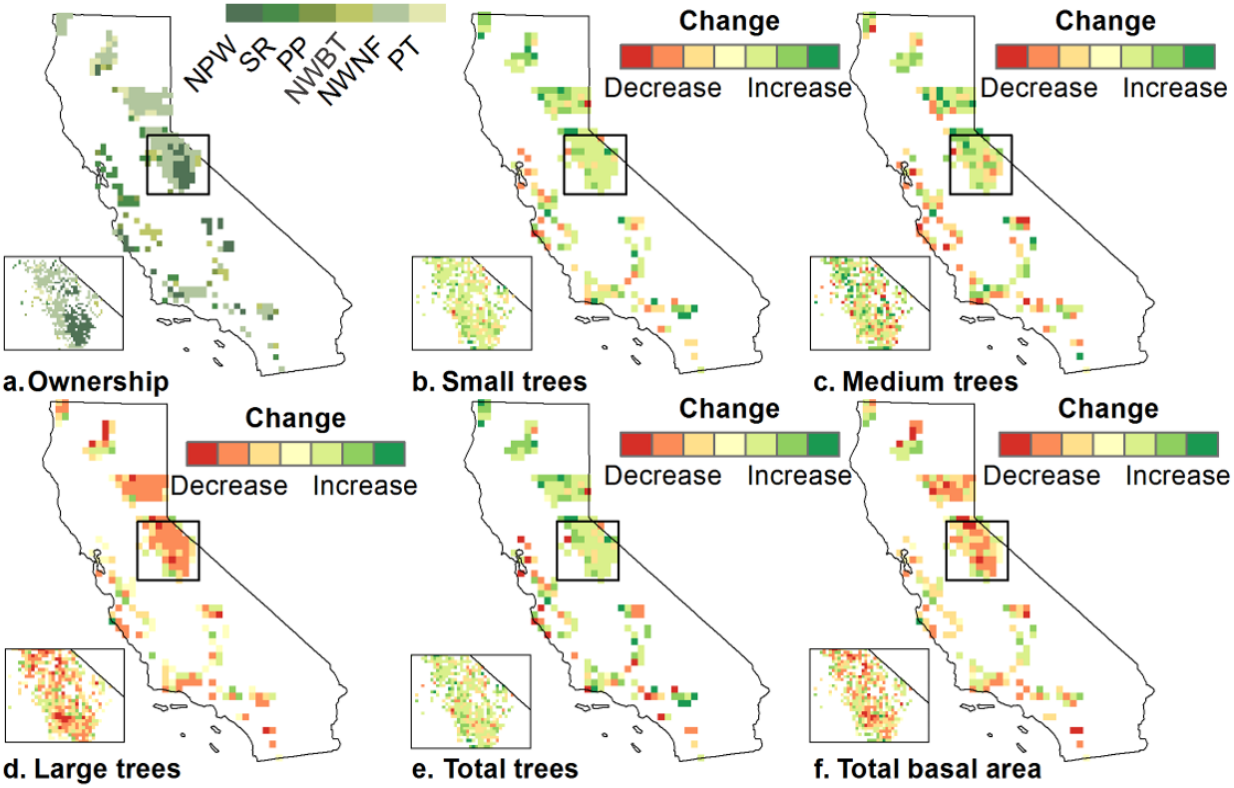


Figure 4 Change in forest structure measures at 20 km resolution for: a) ownership class; b) number of small trees per ha; c) number of medium trees per ha; d) number of large trees per ha; e) number of total trees per ha; and f) total basal area (m^2/ha). The insets of each map show the Sierra Nevada region at 5 km resolution. Change is depicted as increase (> 0) and decrease (< 0).

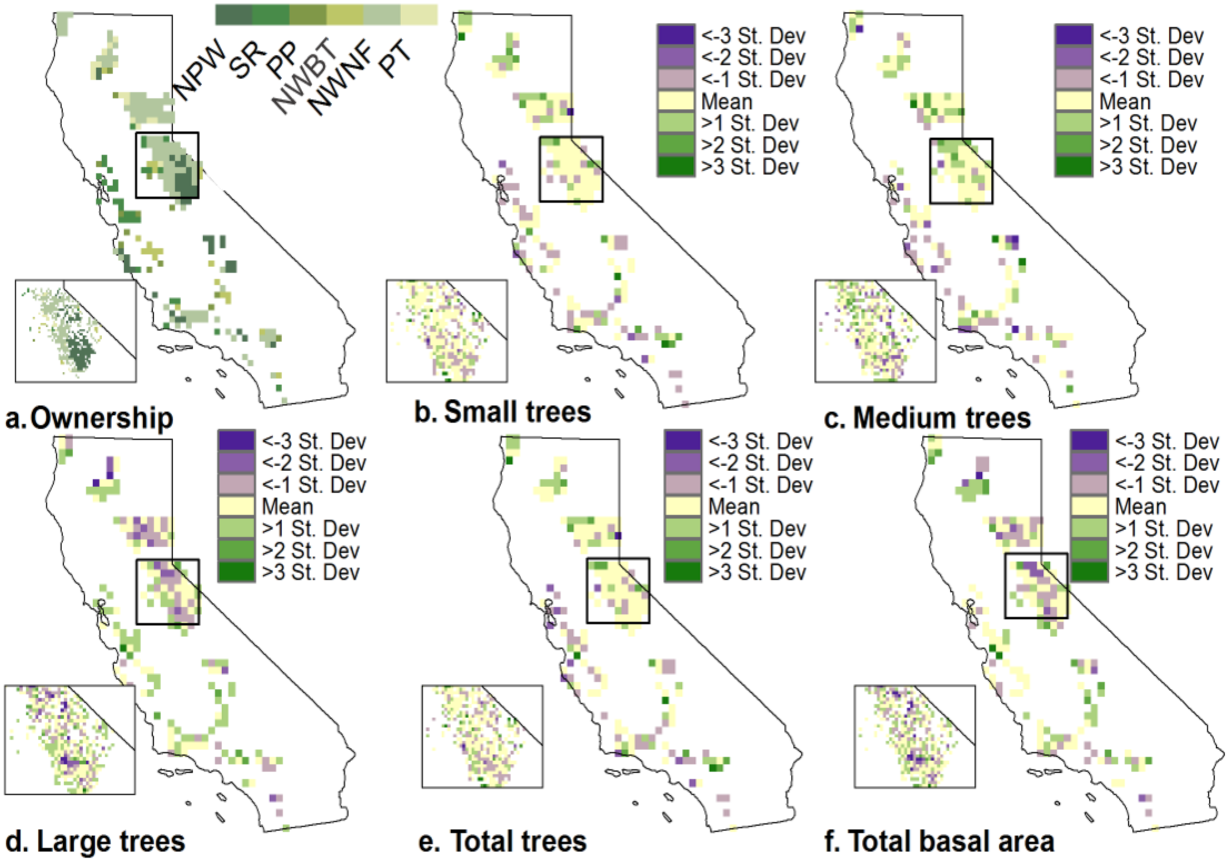


Figure 5 Extreme change in forest structure measures at 20 km resolution for: a) ownership class; b) number of small trees per ha; c) number of medium trees per ha; d) number of large trees per ha; e) number of total trees per ha; and f) total basal area (m^2/ha). The insets of each map show the Sierra Nevada region at 5 km resolution. Change is shown as deviations from the mean cell value, > 2.0 standard deviations from the mean is considered to be extreme change.

4 Discussion

We report overall changes to forests statewide (increases in small trees and total trees and decreases in large trees and total basal area) that are similar to previous results (e.g. McIntyre et al., 2015), however the changes we report uniquely vary across ownership class. For example, large trees declined by 83% on PT, by 71% on NPW, and by 52% on NWNF. The consistency of large tree declines across differing land ownership class aligns with previous research pointing towards systematic influences of climatic drivers such as temperature and climatic water deficit (Das et al., 2013; Lutz et al., 2010; McIntyre et al., 2015) as a primary driver of decline. The lack of variation in large tree decline between PT and NPW is especially surprising given their conflicting logging histories and directly contrast with previous studies that find logging as the primary driver of large tree decline (Knapp et al., 2013). Timber harvesting took place in NWNF and in NPW areas before their establishment, however nearly 95% of the total national timber harvest came from private forests prior to World War II (Hirt, 1994). Increased demand for timber between 1940 and 1960 caused by post-war building led to extensive timber harvesting. Forest stands from private forests and National Forests both contributed to this rapid increase in harvesting. In the late 1940s, timber harvests on National Forests increased dramatically (Winters, 1950). More recently, timber production in California has declined to pre-war averages of about 1.1 billion board feet per year with private lands again producing nearly 95% of California's timber (Stewart et al., 2016). Timber harvesting on PT has been consistent since before WWII, whereas it has been variable on NWNF, and relatively nonexistent on NPW. This consistency in timber harvesting on PT lands may explain why the largest declines in large trees were on private timberlands, but timber harvesting alone does not explain why declines (71% loss) on relatively unlogged NPW are greater than declines on variably logged NWNF (52% loss) and why losses on NPW are not significantly different then losses on PT (Table 4).

Although an active legacy of timber harvesting may not play the primary role in determining statewide patterns of large tree decline, it might help to explain changes to forest density. Specifically, long term studies of post-harvest forest recovery have shown that logged areas have higher densities than corresponding undisturbed areas (GarciaFlorez et al., 2017; Naficy et al., 2010). This is consistent with historical logging practices of targeting large size classes (Bouldin, 1999; Knapp et al., 2013). Large gaps created by the removal of one large tree allow for several small trees to infill the space left behind (Lydersen et al., 2013). PT, which have been the most actively logged of the ownerships presented here, on average have greater densities than their unlogged or variably logged counterparts (i.e. NPW, NWNF) and showed substantially different changes in numbers of small trees than all other land ownership classes. However, depending on the scale of the study, logging history may not be the primary driver of densification (Knapp et al., 2013; Collins et al., 2017, Merschel et al., 2014), and local and regional drivers maybe more explanatory. Regional differences in ownership types may also play an explanatory role, for example patterns of small tree density on PT could be attributed to the fact that they are largely located at lower elevations and are therefore potentially more productive areas than higher elevation NWNF and NPW areas.

Beyond logging histories, different ways in which fire suppression practices have been implemented also likely played a role in the differing magnitude of changes taking place, particularly on NPW lands and NWNF. The National Park Service adopted a perspective of fire as a critical natural process, allowing wildfire to return to the landscape under specific circumstances nearly a decade earlier than did the Forest Service (Miller et al., 2012). Fire managed for natural resource benefits has taken place primarily in Wilderness areas and in many of the National Parks whereas fire suppression is still common on other Forest Service lands (Franklin and Agee, 2003; Stephens et al., 2016). Our data showed smaller changes in the number of small trees on the landscape and significantly lower counts of total trees in NPW when compared to NWNF and PT. On PT and NWNF, where immediate fire suppression is still common, significant increases in small trees have contributed to much denser forests overall. In previous studies, where burned and unburned plots were studied with the goal of explicitly understanding the effect of fire suppression, the largest increases in stem densities occurred in unburned mid-elevation conifer forest where fire suppression is argued to have the greatest impact (e.g. Fellows and Goulden, 2008) and which aligns with our results from the majority of PT and NPW plots.

Understanding how forests have changed over long periods of time is increasingly important for contemporary forest managers. Total biomass and forest density measurements are commonly used and relatively simple proxies for forest function (e.g. productivity and diversity); they serve as key indicators of species habitat (Franklin et al., 2002), and are increasingly used in carbon estimation (Balderas Torres and Lovett, 2013; Brown et al., 1989). The forest structure measures we assessed here (e.g. tree counts by size class and basal area) are simple measures used to target silvicultural practices through existing mechanisms (e.g. thinning, controlled burns, fuel reductions, reseeding).

Furthermore, restoration projects, reforestation efforts, and local policies often use trees per area as a baseline or target for success (Crowther et al., 2015). Many of these efforts use historical estimates as restoration targets (Alagona et al., 2012; Rhemtulla and Mladenoff, 2007) yet in this era of both rapid anthropogenic change and potentially novel climatic regimes historical numbers may not be appropriate baselines (Millar et al., 2007). Therefore, understanding how forest structure changes across space and time, is altered by management, and varies across ownership is important for setting appropriate targets.

Significant increases in forest density and declining stand basal area in California over the 20th century has resulted in forests that are profoundly different than they were 100 years ago. In recent decades resilience has become the overarching framework of forest management, especially within public forests (Churchill et al., 2013; Millar et al., 2007; Stephens et al., 2016). Widespread patterns of increasing density reduce important structural and spatial patterns in forests including the distribution of large individual trees, open spaces, and clusters of trees (Churchill et al., 2013; Lutz et al., 2013). As these features are lost forests become increasingly susceptible and less resilient to catastrophic fire, disease, and drought induced mortality (Larson and Churchill, 2012; Lindenmayer et al., 2012; Lydersen et al., 2013; Stephens et al., 2008).

Considering that patterns of forest densification is consistent across the state and across ownership type, large scale forest management strategies that foster greater horizontal and vertical complexity merit further attention. In heavily managed forests strategizing appropriate post-harvest planting densities and expanding seedling genetic diversity (Millar et al., 2007) could contribute to healthier and potentially more resilient forests. Despite the complexity of forest thinning operations in National Forests, National Parks and Wilderness areas, careful deliberation is needed when deciding to abandon these strategies (including selective harvesting, thinning, and the use of prescribed fire) since this could cause an increase in competition for resources such as water and increase species vulnerability and species stress within a drier climate (Linares et al., 2010). Increasing the amount and scope of management efforts that help to reduce density in forests may help trees survive to become the large trees that we have lost over the last century.

Our work has shown important differences in forest change across land ownership. However, it is important to note that some of this variation could additionally be explained by patterns in regional climate, local differences in tree regeneration, growth, dispersal or disturbance regimes including fire, pests, and disease that may co-vary with ownership. A more nuanced explanation of these changes calls for further investigation of the aggregating spatial unit (biophysical region, county, land owner), a greater emphasis on collecting and analyzing spatial information on past land use legacies, as well as understanding the interactions between all likely explanatory factors.

5 Conclusions

Increased forest density and forest biomass declines over the last century have resulted in profound structural change in the forests of California. Evidence including historical resurveys and contemporary comparisons demonstrate this trend both locally (Dolanc et al., 2013a; Lydersen et al., 2013), and statewide (Fellows and Goulden, 2008; McIntyre et al., 2015). Although these changes are consistent across scales, attributing these patterns to specific drivers is complicated, and a more nuanced understanding requires investigating land use legacies in addition to climate, disturbance, and regional differences. We contribute to this discussion via our investigation into changes in forest structure across differing land ownerships. We compared historical and contemporary forest inventory data and found that contemporary forests in California are denser and have less overall biomass (total basal area) than their 1930s counterparts, and that this pattern is significant at a statewide scale. However, critical differences in forest structural change across ownership exist. Forests on PT, NWNF, and NPW exhibit consistent and more pronounced structural changes (loss of large trees and basal area, increase in small trees and total trees) than those on SR, PP, and NWBT.

Given that the magnitude and directionality of forest structure change differs across ownership class, we argue that land ownership, in part helps explain variations in forest structure. There are also regional differences in forest change (e.g. changes to Sierra Nevada and Central Coast forests are sometimes opposite), and these areas have experienced different management regimes over the 20th century. While we do not explicitly test for such regional differences, we have demonstrated that understanding land ownership and management history is crucial for understanding changing biomass across California, irrespective of region. To further our understanding, predictions, and management of forest ecosystems, consideration of socio-ecological, economic, and biophysical drivers is needed. Our work contributes to the development of a more nuanced understanding of change in California forests that incorporates climate, geomorphology, disturbance and management.

6 References

Alagona, P.S., Sandlos, J., Wiersma, Y.F., 2012. Past imperfect: using historical ecology and baseline data for conservation and restoration projects in North America. Enviorn. Philos. 9, 49–70.

Balderas Torres, A., Lovett, J.C., 2013. Using basal area to estimate aboveground carbon stocks in forests: La Primavera Biosphere’s Reserve, Mexico. Forestry 86, 267–281. http://dx.doi.org/10.1093/forestry/cps084.

Bechtold, W.A., Patterson, P.L., 2005. The Enhanced Forest Inventory and Analysis Program—National Sampling Design and Estimation Procedures.

Beesley, D., 1996. Past Sierra Nevada landscapes. In: Sierra Nevada Ecosystem Project: Final Report to Congress. Assessments and Scientific Basis for Management Options, vol. II, pp. 3–24

Brown, S., Gillespie, A., Lugo, A., 1989. Biomass estimation methods for tropical forests with applications to forest inventory data. For. Sci.

Bouldin, J., 1999. Twentieth-century changes in forests of the Sierra Nevada, California. Ph.D. Dissertation, University of California, Davis.

Canty, A., Ripley, B., 2017. boot: Bootstrap R (S_Plus) Functions [WWW Document]. R Packag. version 1.3-20.

Churchill, D., Larson, A., Dahlgreen, M., Franklin, J., Hessburg, P., Lutz, J., 2013. Restoring forest resilience: From reference spatial patterns to silvicultural prescriptions and monitoring. For. Ecol. Manage. 291, 442–457.

Collins, B.M., Fry, D.L., Lydersen, J.M., Everett, R., Stephens, S.L., 2017. Impacts of different land management histories on forest change. Ecol. Appl. 27, 2475–2486. http://dx.doi.org/10.1002/eap.1622.

Crowther, T.W., Glick, H.B., Covey, K.R., Bettigole, C., Maynard, D.S., Thomas, S.M., Smith, J.R., Hintler, G., Duguid, M.C., Amatulli, G., Tuanmu, M.-N., Jetz, W., Salas, C., Stam, C., Piotto, D., Tavani, R., Green, S., Bruce, G., Williams, S.J., Wiser, S.K., Huber, M.O., Hengeveld, G.M., Nabuurs, G.-J., Tikhonova, E., Borchardt, P., Li, C.-F., Powrie, L.W., Fischer, M., Hemp, A., Homeier, J., Cho, P., Vibrans, A.C., Umunay, P.M., Piao, S.L., Rowe, C.W., Ashton, M.S., Crane, P.R., Bradford, M.A., 2015. Mapping tree density at a global scale. Nature 525, 201–205. http://dx.doi.org/10. 1038/nature14967.

Das, A.J., Stephenson, N.L., Flint, A., Das, T., van Mantgem, P.J., 2013. Climatic correlates of tree mortality in water- and energy-limited forests. PLoS One 8. http://dx.doi.org/10.1371/journal.pone.0069917.

Dolanc, C.R., Safford, H.D., Dobrowski, S.Z., Thorne, J.H., 2013a. Twentieth century shifts in abundance and composition of vegetation types of the Sierra Nevada, CA. US. Appl. Veg. Sci. 1–14. http://dx.doi.org/10.1111/avsc.12079.

Dolanc, C.R., Thorne, J.H., Safford, H.D., 2013b. Widespread shifts in the demographic structure of subalpine forests in the Sierra Nevada, California, 1934 to 2007. Glob. Ecol. Biogeogr. 22, 264– 276. http://dx.doi.org/10.1111/j.1466- 8238.2011.00748.x.

Dolanc, C.R., Safford, H.D., Thorne, J.H., Dobrowski, S.Z., 2014. Changing forest structure across the landscape of the Sierra Nevada, CA, USA, since the 1930s. Ecosphere 5, 1–26.

Fay, M.P., Proschan, M.a., 2010. Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. Stat. Surv. 4, 1–37. http://dx.doi.org/10.1214/09-SS051. Wilcoxon-Mann-Whitney.

Fellows, A.W., Goulden, M.L., 2008. Has fire suppression increased the amount of carbon stored in western U.S. forests? Geophys. Res. Lett. 35, 1–4. http://dx.doi.org/10. 1029/2008GL033965.

Franklin, J.F., Agee, J.K., 2003. Foraging a Science-Based National Forest Fire Policy. Issues Sci. Technol.

Franklin, J.F., Spies, T.A., Pelt, R. Van, Carey, A.B., Thornburgh, D.A., Berg, D.R., Lindenmayer, D.B., Harmon, M.E., Keeton, W.S., Shaw, D.C., Bible, K., Chen, J., 2002. Disturbances and structural development of natural forest ecosystems with silvicultural implications, using Douglas-fir forests as an example. For. Ecol. Manage. 155, 399–423 http://dx.doi.org/10.1016/S0378- 1127(01)00575-8.

Garcia-Florez, L., Vanclay, J.K., Glencross, K., Nichols, J.D., 2017. Understanding 48 years of changes in tree diversity, dynamics and species responses since logging disturbance in a subtropical rainforest. For. Ecol. Manage. 393, 29–39. http://dx.doi. org/10.1016/j.foreco.2017.03.012.

Hirt, P., 1994. A Conspiracy of Optimism: Management of the National Forests Since World War II. University of Nebraska Press.

Jenks, G., 1977. Optimal Data Classification for Choropleth Maps. University of Kansas, Lawrence.

Keeley, J.E., 2004. VTM plots as evidence of historical change: goldmine or landmine? Madrono 51, 372–378.

Kelly, M., Allen-Diaz, B., Kobzina, N., 2005. Digitization of a historic dataset: the Wieslander California vegetation type mapping project. Madrono 52 (3), 191– 201.

Kelly, M., Easterday, K., Koo, M., Thorne, J.H., Mukythar, S., Galey, B., 2017. Geospatial informatics key to recovering and sharing historical ecological data for modern use. Photogramm. Eng. Remote Sens. 83, 779–786. http://dx.doi.org/10.14358/PERS.83. 10.779.

Kelly, M., Easterday, K., Rapicullio, G., Koo, M., McIntyre, P., Thorne, J., 2016. Rescuing and sharing historical vegetation data for ecological analysis: the California vegetation type mapping project. Biodivers. Informatics 11, 40–62.

Knapp, E.E., Skinner, C.N., North, M.P., Estes, B.L., 2013. Long-term overstory and understory change following logging and fire exclusion in a Sierra Nevada mixed-conifer forest. For. Ecol. Manage. 310, 903–914. http://dx.doi.org/10.1016/j.foreco. 2013.09.041.

Larson, A.J., Churchill, D., 2012. Tree spatial patterns in fire-frequent forests of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. For. Ecol. Manage. 267, 74–92. http://dx.doi.org/10.1016/j.foreco.2011.11.038.

Laudenslayer, W.F., Darr, H.H., 1990. Historical effects of logging on the forests of the Cascade and Sierra Nevada ranges of California. Trans. Western Section Wildlife Soc. 26, 12–23.

Linares, J.C., Camarero, J.J., Carriera, J.A., 2010. Competition modulates the adaptation capacity of forests to climatic stress: insights from recent growth decline and death in relict stands of the Mediterranean fir Abies pinsapo. J. Ecol. 98, 592–603. http://dx.doi.org/10.1111/j.13652745.2010.01645.x.

Lindenmayer, D.B., Laurance, W.F., Franklin, J.F., 2012. Global Decline in Large Old Trees. Science. 338, 1305–1306. doi: 10.1126/science.1231070.

Lunt, I.D., Spooner, P.G., 2005. Using historical ecology to understand patterns of biodiversity in fragmented agricultural landscapes. J. Biogeogr. 32, 1859–1873.

Lutz, J.A., Larson, A.J., Freund, J.A., Swanson, M.E., Bible, K.J., 2013. The importance of large-diameter trees to forest structural heterogeneity. PLoS One 8, e82784.

Lutz, J.A., van Wagtendonk, J.W., Franklin, J.F., 2009. Twentieth-century decline of large-diameter trees in Yosemite National Park, California, USA. For. Ecol. Manage. 257, 2296–2307. http://dx.doi.org/10.1016/j.foreco.2009.03.009.

Lutz, J.A., van Wagtendonk, J.W., Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. J. Biogeogr. 37, 936–950. http://dx.doi.org/10.1111/j.1365-2699.2009.02268.x.

Lydersen, J.M., North, M.P., Knapp, E.E., Collins, B.M., 2013. Quantifying spatial patterns of tree groups and gaps in mixed-conifer forests: Reference conditions and long-term changes following fire suppression and logging. For. Ecol. Manage. 304, 370–382. http://dx.doi.org/10.1016/j.foreco.2013.05.023.

MacFarland, T.W., Yates, J.M., 2016. Mann–Whitney U Test. In: Introduction to Nonparametric Statistics for the Biological Sciences Using R. Springer International Publishing, Cham, pp. 103–132. http://doi:10.1007/978-3-319-30634-6_4

Maxwell, R., Taylor, A., Skinner, C., Safford, H., Isaacs, R., Airey, C., Young, A., 2014. Landscape-scale modeling of reference period forest conditions and fire behavior on heavily logged lands. Ecosphere 5, 1–28.

McIntyre, P.J., Thorne, J.H., Dolanc, C.R., Flint, A.L., Flint, L.E., Kelly, M., Ackerly, D.D., 2015. Twentieth-century shifts in forest structure in California: denser forests, smaller trees, and increased dominance of oaks. Proc. Natl. Acad. Sci. 201410186. http://dx.doi.org/10.1073/pnas.1410186112.

McKelvey, K.S., Johnston, J.D., 1992. Historical Perspectives on Forests of the Sierra Nevada and the Transverse Ranges of Southern California: Forest Conditions at the Turn of the Century. USDA Forest Service Gen. Tech. Rep. PSW-GTR-133, 225–246.

McIver, C.P., Meek, J.P., Scudder, M.G., Sorenson, C.B., Morgan, T.A., Christensen, G.A., 2015. California’s Forest Products Industry and Timber Harvest, 2012.

Merschel, A.G., Spies, T.A., Heyerdahl, E.K., 2014. Mixed-conifer forests of central Oregon: effects of logging and fire exclusion vary with environment. Ecol. Appl. 24, 1670–1688.

Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17, 2145–2151. doi.org/10.1890/06-1715.1.

Miller, J., Collins, B., Lutz, J., Stephens, S., van Wagtendonk, J., Yasuda, D., 2012. Differences in wildfires among ecoregions and land management agencies in the Sierra Nevada region, California, USA. Ecosphere 3, 1–20.

Minnich, R.A., Barbour, M.G., Burk, J.H., Fernau, R.F., 1995. Sixty Years of Change in Californian Conifer Forests of the San Bernardino Mountians. Conserv. Biol. 9, 902–914.

Naficy, C., Sala, A., Keeling, E.G., Graham, J., DeLuca, T.H., 2010. Interactive effects of historical logging and fire exclusion on ponderosa pine forest structure in the northern Rockies. Ecol. Appl. 20, 1851–1864. http://dx.doi.org/10.1890/090217.1.

Perring, M.P., De Frenne, P., Baeten, L., Maes, S.L., Depauw, L., Blondeel, H., Carón, M.M., Verheyen, K., 2016. Global environmental change effects on ecosystems: the importance of land-use legacies. Glob. Chang. Biol. 22, 1361–1371.

Rhemtulla, J.M., Mladenoff, D.J., 2007. Why history matters in landscape ecology. Landsc. Ecol. 22, 1–3. http://dx.doi.org/10.1007/s10980-007-9163-x.

Santos, M.J., Watt, T., Pincetl, S., 2014. The push and pull of land use policy: reconstructing 150 years of development and conservation land acquisition. PLoS One.

Smith, W.B., 2002. Forest inventory and analysis: a national inventory and monitoring program. Enviornmental Pollut. 116, 233–242.

Stephens, S.L., Collins, B.M., Biber, E., Fule, P.Z., 2016. U.S. Federal fire and forest policy: emphasizing resilience in dry forests. Ecosphere 7, 1–19. http://dx.doi.org/10.1002/ ecs2.1584.

Stephens, S.L., Fry, D.L., Franco-vizcaíno, E., 2008. Wildfire and Spatial Patterns in Forests in Northwestern Mexico: The United States Wishes It Had Similar Fire Problems.

Stewart, W., Sharma, B., York, R., Diller, L., Hamey, N., Powell, R., Swiers, R., 2016. Forestry. In: Mooney, H., Zavaleta, E. (Eds.), Ecosystems of California. UC Press, pp. 817–834.

Taylor, A., 2000. Fire regimes and forest changes in mid and upper montane forests of the southern Cascades, Lassen Volcanic National Park, California, USA. J. Biogeogr. 1905, 87–104.

Taylor, A., Skinner, C., 2003. Spatial patterns and controls on historical fire regimes and forest structure in the Klamath Mountains. Ecol. Appl. 13, 704–719.

R Core Team, 2013. R: A language and environment for statistical computing.

Thorne, J.H., Le, T.N., 2016. California's historic legacy for landscape change, the Wieslander vegetation type maps. Madrono 63, 293–328.

Turner, M.G., Wear, D.N., Flamm, R.O., 1996. Land Ownership and Land-Cover Change in the Southern Appalachian Highlands and the Olympic Peninsula. Ecol. Appl. 6, 1150–1172.

van Mantgem, P.J., Nesmith, J.C.B., Keifer, M., Knapp, E.E., Flint, A., Flint, L., 2013. Climatic stress increases forest fire severity across the western United States. Ecol. Lett. 16, 1151–1156. http://dx.doi.org/10.1111/ele.12151.

van Mantgem, P.J., Stephenson, N.L., Byrne, J.C., Daniels, L.D., Franklin, J.F., Fulé, P.Z., Harmon, M.E., Larson, A.J., Smith, J.M., Taylor, A.H., Veblen, T.T., 2009. Widespread Increase of Tree Mortality Rates in the Western United States. Science (80-.). 323, 521–524.

Waddell, K.L., 2013. PNW-FIADB Users Manual: A data dictionary and user guide for the PNW-FIADB database.

Wieslander, A.E., 1935. A vegetation type map of California. Madroño 60, 348–352

Wieslander, A., Yates, H.S., Jensen, H.A., Johannsen, P., 1933. Manual of Field Instructions for Vegetation Type Map of California. On File in the Library at Yosemite National Park, Yosemite Valley, CA.

Winters, R., 1950. Fifty years of forestry in the U.S.A.

A Appendix

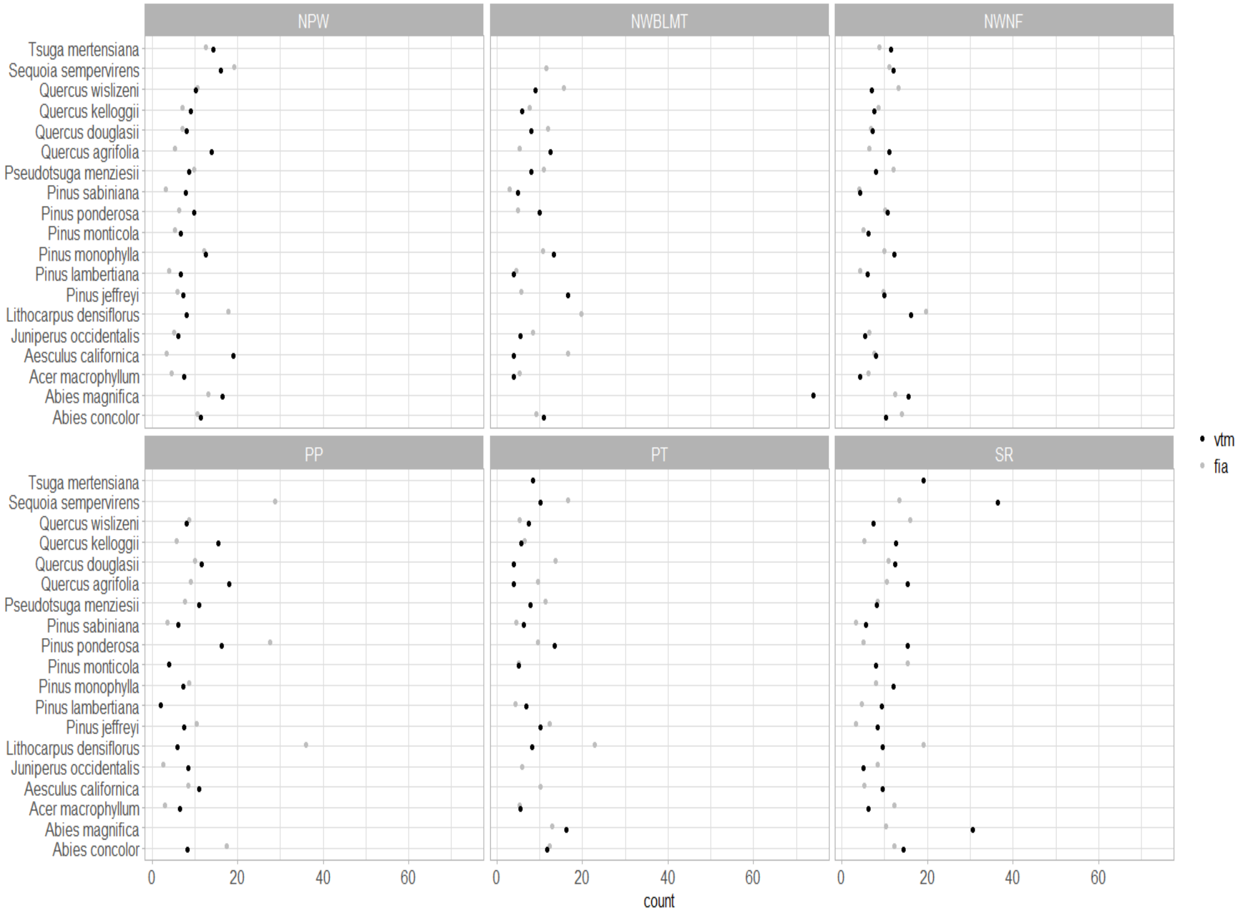


Figure A1 Difference in average frequency count of species per plot by ownership type over time. VTM in dark grey, FIA in light grey.

Species	NPW	PT	NWNF	NWBT	PP	SR
Pinus jeffreyi	-20%	22%	-1%	-65%	40%	-61%
Pinus lambertiana	-40%	-36%	-30%	17%	NA	-50%
Pinus monophylla	-3%	1%	-20%	-19%	20%	-34%
Pinus monticola	-19%	158%	-21%	NA	NA	93%
Pinus ponderosa	-35%	-29%	-5%	-51%	69%	-67%
Pinus sabiniana	-60%	-26%	-7%	-39%	-41%	-40%
Pinus attenuata	-51%	NA	-51%	-76%	NA	0%
Pinus coulteri	-69%	NA	-61%	NA	-20%	-79%
Pseudotsuga menziesii	14%	44%	52%	35%	-32%	3%
Quercus agrifolia	-62%	137%	-43%	-58%	-50%	-32%
Quercus chrysolepis	26%	85%	55%	34%	-25%	-19%
Quercus douglasii	-12%	243%	-5%	47%	-13%	-13%
Quercus kelloggii	-23%	14%	15%	29%	-63%	-59%
Quercus lobata	60%	-35%	-64%	-76%	-71%	-40%
Quercus wislizeni	4%	-30%	90%	71%	8%	118%
Sequoia sempervirens	20%	63%	-9%	NA	NA	-63%
Lithocarpus densiflorus	124%	179%	22%	NA	500%	102%
Umbellularia californica	43%	349%	72%	347%	89%	-39%
Calocedrus decurrens	52%	33%	81%	115%	NA	27%

Table A1 Percent difference in average frequency of species per plot by ownership type, decreased frequency highlighted in red, increased frequency highlighted in green. Increases of S.*sempervirens* on PT and NPW are particularly unexpected.

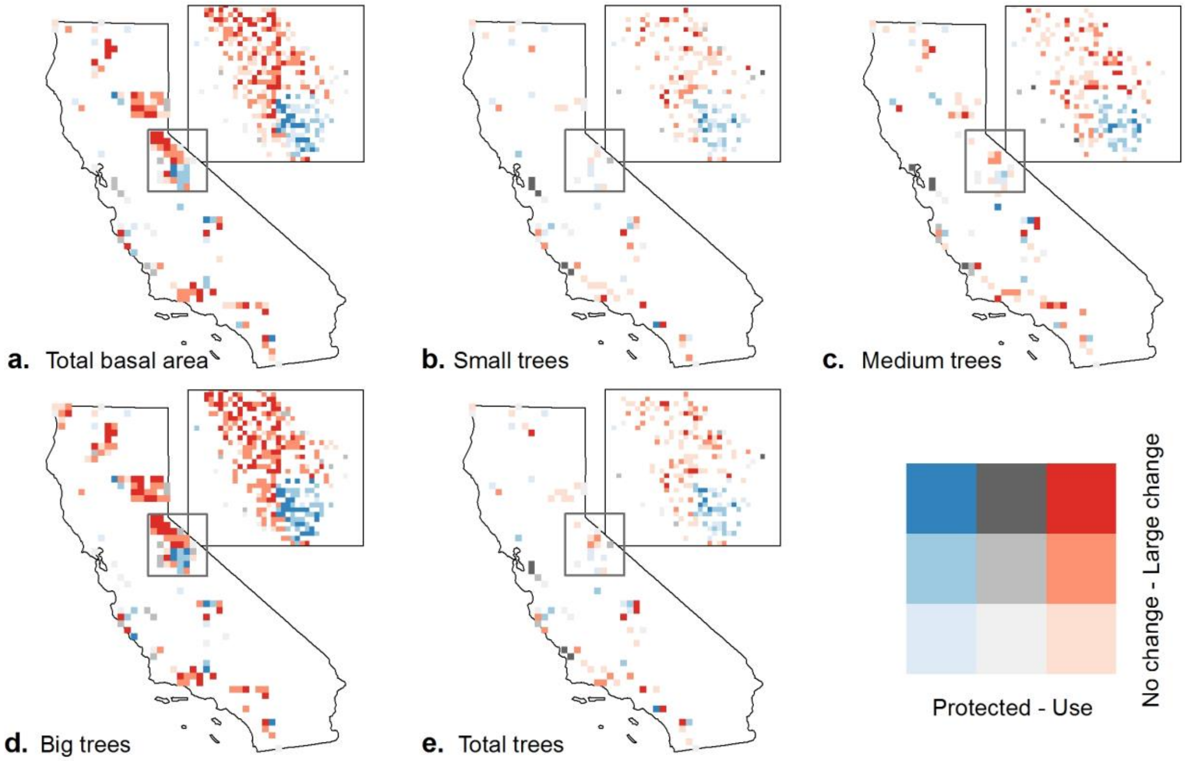


Figure A2. Decreases in forest structure measures at 20 km resolution shown with a 2-dimensional (from protected to intense use and from small to large change) diverging color scheme: a) total basal area; b) number of small trees per ha; c) number of medium trees per ha; d) number of large trees per ha; e) number of total trees per ha. The insets of each map show the Sierra Nevada at 5 km resolution. NPW and PP categories are shown in blue, NWBT and SR are represented in Gray, and PT and NF are shown in red.

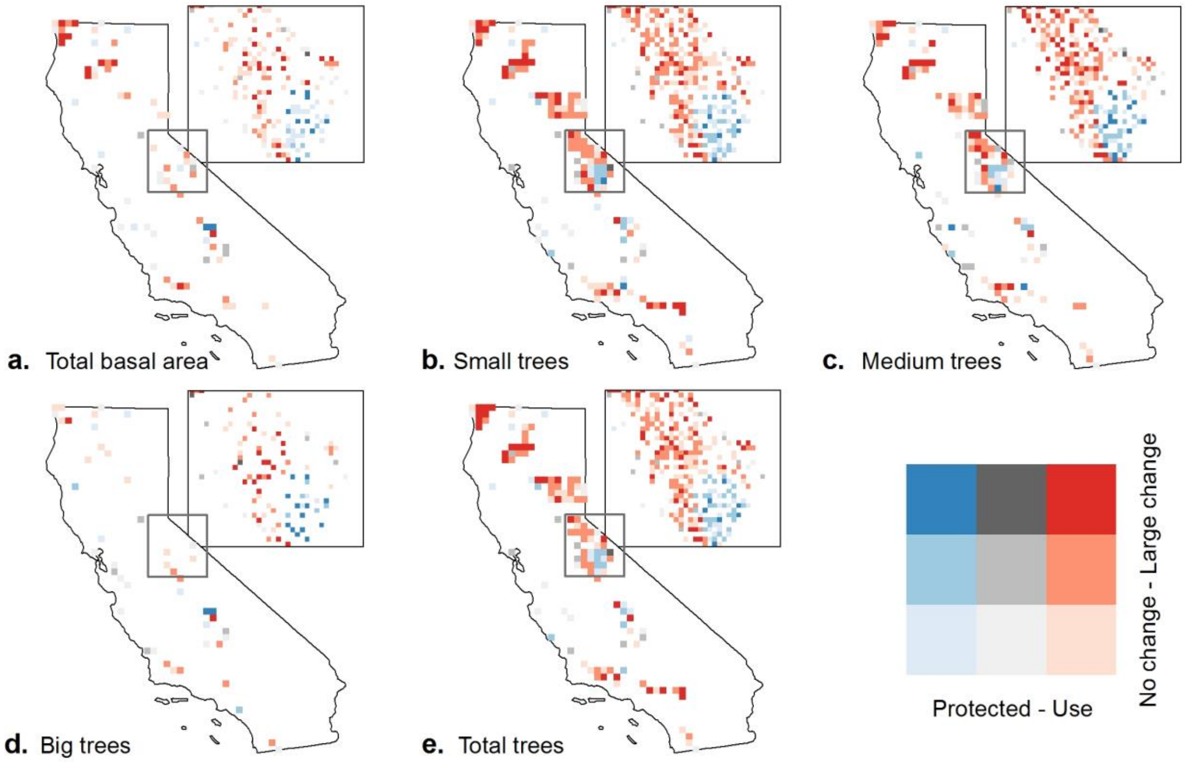


Figure A3. Increases in forest structure measures at 20 km resolution shown with a 2-dimensional (from protected to intense use and from small to large change) diverging color scheme: a) total basal area; b) number of small trees per ha; c) number of medium trees per ha; d) number of large trees per ha; e) number of total trees per ha. The insets of each map focus on the Sierra Nevada at 5 km resolution. NPW and PP categories are shown in blue, NWBT and SR are represented in Gray, and PT and NF are shown in red.

Chapter 3: Assessing threats and conservation status of historical centers of Oak richness in California

This article has previously been published and is reproduced here with kind permission from co-authors and the Graduate Division

Easterday, Kelly Jane, Patrick J. McIntyre, James H. Thorne, Maria J. Santos, and Maggi Kelly. 2016. "Assessing Threats and Conservation Status of Historical Centers of Oak Richness in California." *Urban Planning* 1 (4): 65–78.

Abstract

Oak trees are emblematic of California landscapes, serving as keystone cultural and ecological species and as indicators of natural biological diversity. As historically undeveloped landscapes are increasingly converted to urban environments, endemic oak woodland extent is reduced, which underscores the importance of strategic placement and reintroduction of oaks and woodland landscape for the maintenance of biodiversity and reduction of habitat fragmentation. This paper investigated the effects of human urban development on oak species in California by first modeling historical patterns of richness for eight oak tree species using historical map and plot data from the California Vegetation Type Mapping (VTM) collection. We then examined spatial intersections between hot spots of historical oak richness and modern urban and conservation lands and found that impacts from development and conservation vary by both species and richness. Our findings suggest that the impact of urban development on oaks has been small within the areas of highest oak richness but that areas of highest oak richness are also poorly conserved. Third, we argue that current policy measures are inadequate to conserve oak woodlands and suggest regions to prioritize acquisition of conservation lands as well as examine urban regions where previous centers of oak richness were lost as potential frontiers for oak reintroduction. We argue that urban planning could benefit from the adoption of historical data and modern species distribution modelling techniques primarily used in natural resources and conservation fields to better locate hot spots of species richness, understand where habitats and species have been lost historically and use this evidence as incentive to recover what was lost and preserve what still exists. This adoption of historical data and modern techniques would then serve as a paradigm shift in the way urban planners recognize, quantify, and use landscape history in modern built environments.

Keywords

Quercus; species distribution models; urban planning; vegetation type mapping

1 Introduction

Urban areas serve as important landscapes for a wide range of species. However, the rapid spread of urban development has heightened concern globally over potential losses in biodiversity and ecosystem services generated through landscape conversion. Sustainable planning initiatives in conjunction with ecological knowledge can help sustain biodiversity and reduce landscape fragmentation in urban environments. Calls for the integration of landscape ecology principles, natural resource conservation, and landscape history into urban planning have increased. In conjunction, the types of tools and data normally reserved for ecological analysis have begun to be used in the planning arena. The blending of principles from landscape ecology, urban planning data, and geospatial modelling tools represent a paradigm shift in the way we recognize, quantify, and use landscape history in planning our modern built environments. Current and future sustainable urban planning practices in both developed and undeveloped areas require detailed information on past landscapes. However, historical information is often spatially discontinuous and may require statistical extrapolation to fill in gaps and create regional descriptions. The use of species distribution modeling (SDM), also called environmental niche modeling (ENM), is common in the conservation and ecological restoration communities, but these tools have been underutilized in the urban planning arena. These models generate regional scale descriptions of past vegetation communities or taxa distributions and may offer critical information in sustainable planning processes that seek to reintroduce natural vegetation to already urbanized areas or to avoid substantially altering existing natural environment.

Oaks and oak woodlands are emblematic of California landscapes. They occupy about 13% of the state or 4 million ha in diverse canopy mixtures of eight primary tree species of the genus Quercus: coast live oak (Q. agrifolia), black oak (Q. kelloggii), valley oak (Q. lobata), blue oak (Q. douglasii), Oregon white oak (Q. garryana), Engelmann oak (Q. engelmannii), canyon live oak (Q. chrysolepis), and interior live oak (Q. wislizeni). Oak woodlands are defined by the presence of native oak species within a Mediterranean climate system (Pavlik, Muick, Johnson, & Popper, 1991). In California, tree density and canopy cover vary widely, and woodland appearance ranges from open savanna with widely dispersed trees and understory dominated by Mediterranean annual grasses to dense oak dominated forests (Barbour, Keeler-Wolf, & Schoenherr, 2007). These ecosystems play important roles for wildlife, insects, fungi and lichens (Grivet, Sork, Westfall, & Davis, 2008) while the oaks themselves provide critical ecosystem services, their large canopies creating microclimates and regulating air quality and their root systems providing stability and water filtration (Marañón, Ibáñez, AnayaRomero, Muñoz-Rojas, & Pérez-Ramos, 2012; Standiford & Huntsinger, 2012).

Oaks and oak woodlands are deeply rooted in California’s history. Native Americans used and managed them extensively, deriving food and commodities from oak products (Anderson, 2005). Through the setting of seasonal fires Native Americans retained the quality of oak woodland habitat for game species while curbing pests and disease. Despite the cultural and ecological importance of oaks, the history and practice of converting oak woodlands is lengthy (Bartolome et al., 2002). Lower elevation woodlands, such as the valley oak woodlands of the fertile central valley, were converted to intensive agriculture while the woodlands in the surrounding foothills were historically used for extensive livestock grazing and fire wood production. Since the 1940's it is estimated that California has lost 5,000 km^2 of oak woodland to three main drivers: development, range clearing, and agriculture (Gaman & Firman, 2006, Kueppers, Snyder, Sloan, Zavaleta, & Fulfrost, 2005; Pavlik et al., 1991).

In this paper we focus on one of these drivers, urban development, which is projected to threaten 3,000 km^2 (~one quarter) of the remaining oak woodlands before 2040. California has one of the most rapidly growing human populations and this rate is accelerating (California Department of Finance, 2013; Medvitz & Sokolow, 1995). Over 80% of hardwood lands in California are privately owned (California Fire and Resource Assessment Program, 2010), changing land use in the form of subdivisions has fostered expansion of the urbansuburban footprint (Huntsinger, Buttolph, & Hopkinson, 1997; Huntsinger & Fortmann, 1990). The urban interface with oak woodlands, once confined to the major population centers (San Francisco Bay, Sacramento, the Los Angeles basin), now extends throughout the entire state. Historical ecologists have reconstructed historical distributions and landscapes by extracting mapped and textual data from archives using these products in planning urban and working landscapes (Beller, Downs, Grossinger, Orr, & Salomon, 2015; Grossinger, Striplen, Askevold, Brewster, & Beller, 2007). For example, photographs, maps, and data originally captured for purposes such as taxation or land surveying have become useful data sources in reconstructing historical vegetation conditions (Grossinger et al., 2007; Stein et al., 2010; Whipple, Grossinger, & Davis, 2011).

In addition to mining historical archives, detailed distribution maps of past vegetation conditions are predicted using species distribution modeling (Schussman, Geiger, Mau-Crimmins, & Ward, 2006).SDMs are inferential models that develop relationships between species presence (and sometimes absence) and the key environmental variables that define an environmental niche,and use that relationship to map the niche across space (Graham, Ferrier, Huettman, Moritz, & Peterson, 2004; Keenan, Maria Serra, Lloret, Ninyerola, & Sabate, 2011; Peterson, 2011). The niche, often defined primarily with climatic variables, generates a probability surface of a species occurrence based on the ranges of the climatic variables where a species is known to exist and where those ranges exist in a given space. There are critiques related to these models (e.g. bias in time, assumption of climatic equilibrium, sensitivity to spatial scale), but they can effectively serve regional conservation goals. Given limited species locality information, these models help fill in the gaps of probable species occurrence and generate reasonable regional descriptions of a species distribution based on the input variables. SDMs have traditionally been used in natural resource, conservation, and ecological fields to reconstruct historical habitats and examine climate change impacts (Kueppers et al., 2005; Schussman et al., 2006; Warren, Wright, Seifert, & Shaffer, 2014), to map biotic invasions and disease spread (Kelly, Guo, Liu, & Shaari, 2007; Václavík & Meentemeyer, 2009), to examine bio-richness and speciation mechanisms (Graham et al., 2004; Rushton, Omerod, & Kerby, 2004), and to inform conservation and species management priorities (Kelly, Fonseca, & Whitfield, 2001; Raxworthy et al., 2003; Zhang et al., 2012). Yet their use in urban settings for planning remains limited (Milanovich, Peterman, Barrett, & Hopton, 2012).

In this paper, we argue that urban planning can benefit from a deeper understanding of past distributions of important landscape features, such as vegetation communities and key taxa; the use of historical data and species distribution modeling can aid in protection, guide in planning and management, and lend insight to future distributions given recent climate variability and landscape change.

In this paper, we use a digitized collection of historical vegetation data from a broad-scale California plant community survey from 1920–1930 to map historical oak tree species richness. We then use oak tree occurrence data to model oak richness across California focusing on eight dominant oak species (excluding data on shrub oaks and rare hybrid taxa); coast live oak (Q. agrifolia), black oak (Q. kelloggii), valley oak (Q. lobata), blue oak (Q. douglasii), Oregon white oak (Q. garryana), Engelmann oak (Q engelmannii), canyon live oak (Q. chrysolepis), and interior live oak (Q. wislizenii). We present results in map form for individual species and as overlays conveying oak richness (historical oak "hot spots"). We then analyze how areas of historical oak richness (hot spots) juxtapose current patterns of urban lands and conservation areas and comment on potential opportunities for the reintroduction of lost habitat as well as current areas of potential protection. We use species richness, a known measure of biological diversity—to represent hot spots where several species of oaks overlap. Historical oak richness or oak hot spots describe potential regional biodiversity hot spots that may represent ecological transition zones—areas where species range margins overlap—that constitute a favorable environment for species persistence or adaptation. Regional biodiversity hot spots—as defined in terms of numbers of species—are often conservation priorities that serve as a cost-effective way to preserve the greatest number of species. Using this historical dataset, we are motivated by two questions: (1) where have areas of modeled historical oak richness been lost due to land conversion to urban uses; and (2) to what extent have conservation lands been able to preserve areas of historical oak richness.

1.1 Historical vegetation data: The Vegetation Type Mapping collection

During the 1920 and 1930s, Vegetation Type Mapping (VTM) crews surveyed 16 million ha (40%) of California's wildlands. They collected vegetation information at over 18,000 plots, produced detailed maps of dominant vegetation for over 100,000 km^2, gathered over 23,000 herbarium specimens, and took over 3,000 photographs depicting California vegetation and landscapes (Colwell, 1977; Ertter, 2000; Kelly, Allen-Diaz, & Kobzina, 2005; Kelly, Ueda, & Allen-Diaz, 2008; Wieslander, 1935). Parts of the collection: maps, plot data, and photographs have been used separately, primarily to investigate drivers of change, including climate and fire, and of changes in forest and chaparral communities around the state (Kelly et al., 2016). In this paper we use both the digitized georeferenced plot data (Kelly et al., 2005; Kelly et al., 2008), and digitized georeferenced polygons from the VTM vegetation maps (Thorne, Kelsey, Honig, & Morgan, 2006; Thorne, Santos, & Bjorkman, 2013) to develop distribution models for these oak species. We did not use the VTM georeferenced herbarium specimens to avoid potential duplication. To our knowledge this is the first effort to use both the maps and plot data in conjunction with modern species distribution modelling methods to create a comprehensive historical distribution of taxa. This effort thereby increases the sample size of occurrence records usually gained from the use of georeferenced herbarium specimens alone.

2 Methods

2.1 Historical Oak data

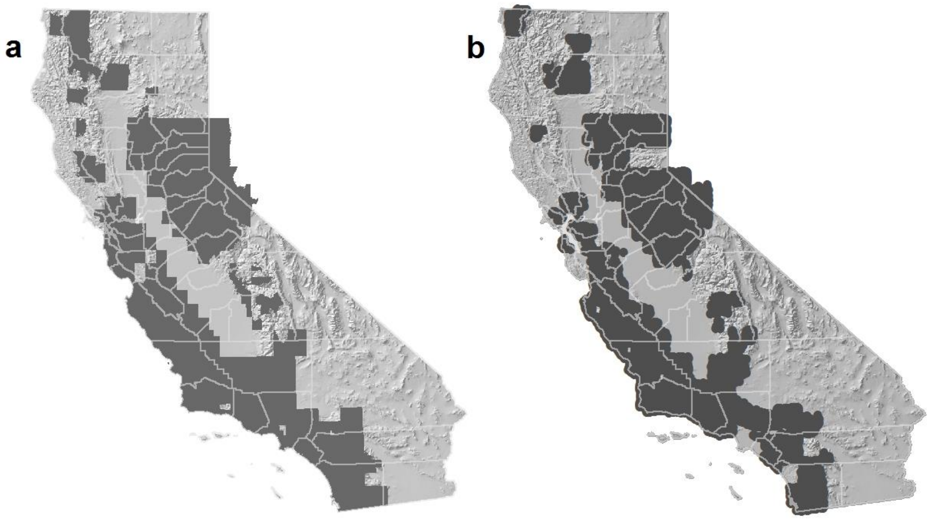


Figure 1 Locations of a) VTM vegetation maps and b) VTM vegetation plots in California.

Location data for eight *Quercus* species was extracted from VTM using digitized vegetation maps and plot data (Kelly et al., 2005). The ~18,000 VTM plots although concentrated primarily along the Sierra Nevada mountain range and the central and southern coastal ranges (Figure 1) were surveyed across a gradient of vegetation types. The records contain data regarding tree stand structure (number per diameter class), percent cover of dominant vegetation by species, soil type, parent material, leaf litter, elevation, slope, aspect, parent material, and other environmental variables. The VTM vegetation map dataset consists of hand drawn polygons covering over 100,000 km^2 in which species comprising 20% or greater of the visual cover of a stand were recorded.

We generated the oak species occurrences used for distribution modeling by obtaining the centroids of polygons in which oaks were recorded as a dominant species. Although the exact extent of the vegetation polygons maybe imprecise as they were hand drawn and distinguished through visual interpretation from nearby vantage points, the use of polygon centroid is likely to reduce the error in the overall sample from inexact locality placement. We removed duplicate localities from map and plot datasets for the same species. We then examined potential outliers and inconsistencies with visual and overlay methods (Hijmans, Schreuder, De la Cruz, & Guarino, 1999). The total sample size for each species is listed in Table 1. It is important to note that these localities were confirmed presence of oak species and do not necessarily constitute the species entire range or environmental niche. The presence data were limited in scope to the extent of the original VTM surveys leaving out large portions of the Central Valley, North Coast, and Mojave. Additionally, the assembled occurrence data may underestimate potential occurrences within mixed stands due to the 20% cover threshold for reporting species. Despite the potential shortcomings of this dataset the VTM survey coverage is the most comprehensive and detailed historical survey of vegetation available for California.

Table 1 VTM dataset sample sizes used in species distribution modeling for eight California oak species.

Species	Common name	Plot locality records	Map locality records	Total
Q. agrifolia	Coast Live Oak	1,653	18,966	20,619
Q. chrysolepis	Canyon Live Oak	1,594	12,484	14,078
Q. douglassii	Blue Oak	1,732	14,826	16,558
Q. engelmannii	Engelmann Oak	61	555	616
Q. garryana	Oregon White Oak	169	952	1,121
Q. kelloggii	California Black Oak	3,126	13,413	16,539
Q. lobata	Valley Oak	601	3,777	4,378
Q. wislizeni	Interior Live Oak	2,677	9,356	12,033
	Total	11,613	74,329	85,942

2.2 Distribution modeling

We use a reduced set of 30-year average (1960–1990) bioclimatic ("Bioclim") (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005) variables at~1km spatial resolution to model the historical distribution of eight oak species. These climatic variables are commonly used to model distributions based on specimens collected from across the 20th century. As this study did not involve predictions across multiple time periods, we opted to use the Bioclim data as it is the most widely used global climate dataset and has benefits in terms of replicability and access. To reduce problems associated with extensive collinearity of predictor variables we examine pairwise correlations among the 19 standard Bioclim variables across California and selected a single variable from pairs with a greater than 0.85 correlation coefficient (Pearson et al., 2006). We used 8 variables: mean diurnal temp range (Bio2), isothermality (Bio3), maximum temperature in the warmest month (Bio5), minimum temperature in the coldest month (Bio6), temperature annual range (Bio7), mean temperature in the wettest quarter (Bio8), annual precipitation (Bio12), and precipitation seasonality (Bio15).

We constructed and assessed the distribution models using Maxent v3.01 called from the R 3.03 statistical environment (R Development Core Team, 2013) using the Dismo package (Hijmans, Phillips, Leathwick, & Elith, 2012). Background (pseudo-absence) data were generated by randomly sampling 10,000 points from the full area of VTM plot and map sampling (Figure 1). We used a k-fold sampling (with k = 4 or 25%) of the occurrence data for each oak species to partition the data into testing and training data, with each round of modeling containing 75% training and 25% testing data. We then assessed model fit using the AUC (area under curve) statistic, which evaluates the performance of model as a series of tradeoffs between true positives and false positives (Fielding & Bell, 1997). AUC values range from 0– 1 with a value of 0.5 representing a model with prediction probabilities close to random, and values greater than 0.5 signify a model with a greater power to predict areas of high suitability in locations of known species presence (Phillips, Anderson, & Schapire, 2006). Using the AUC statistic, we confirmed how well the distribution predicted by our model matched the distribution from a sample of the historical occurrences. We used the maximum sensitivity plus specificity threshold to convert each modeled result from continuous probability scores (e.g. 0–100%) to binary predicted/not-predicted scores (e.g. 0 and 1). This threshold has performed well in a recent evaluation of presence-only threshold methods (Liu, White, & Newell, 2013). We then used this threshold to create individual surfaces that articulated the high probability range of each oak species given the climatic variables. Finally, we summed the eight binary predictions/surfaces for each species to generate a map of modeled historical oak richness for California. Historical oak richness or oak hot spots describes regions where there is spatial coincidence in the modeled ranges of individual oak species. Since these models are based on climatic variables alone the modeled areas of oak richness represent areas of historical climate that were highly suitable for an overlapping number of oak species. Low historical oak richness is represented as single species of oak, moderate represents 2–5 overlapping species ranges, and high represents 6 or more overlapping species ranges.

2.3 Areas of oak threat and conservation

We examined modeled hot spots of historical oak richness as they juxtapose with current urban areas and with protected areas in California using an overlay analysis of the binary maps of modeled historical oak species distributions and statewide spatial layers depicting current urban and protected areas. We used two current statewide products that depict urban footprints and protected areas. The urban footprint, derived from the 2010 decennial census, is useful for analyzing urban growth and associated impacts (U.S. Census Bureau, 2014). The California Protected Areas Database (CPAD, 2013) database tracks public, conservation and trust land ownership representing the most complete publicly available representation of conservation lands for the state of California. Both were provided by the U.S. government data portal: http://www.data.gov.

3 Results

Species distribution model support (AUC) ranged from 0.83 for Q. *chrysolepis* to 0.98 for Q. *engelmannii* (Table 2). The mapped binary results for individual oak species are shown in Figure 2, along with a statewide view of modeled historical oak richness. Areas of high historical oak richness (six or more oak species) include: a) the North Coast Ranges, b) the South Coast Ranges, c) the Sierra Foothill Belt, d) the Transverse Ranges including the Tehachapi Mountains, and e) the Peninsular Ranges (Figure 2).

We overlaid the map of modeled historical oak richness on the current urban footprint and the current conserved lands and found that impacts from development and conservation vary by species and richness. Impacts from urban development have been relatively small (~5.5% of the land) within the areas of high oak richness (Table 3), however 17% of the historical distributions of individual oak species are found in current urban areas. Coast live oaks (Q. *agrifolia*) and Engelmann oaks (Q. *engelmannii*) are the most disproportionately affected; with~19% of each modeled range now under the modern urban footprint. Additionally, the ranges of valley oak (Q. *lobata*), blue oak (Q. *douglasii*), and Oregon white oak (Q. *garryana*) may be underrepresented in these models due to the lack of VTM survey coverage in these species normal ranges which include the Central Valley and the North Coast.

Areas of moderate historical oak richness (2–5 oak species) have some protection on conservation lands ranging from 27 to 39% of their predicted historical distribution. Four oak species have approximately half of their modeled historical range on current protected lands (Q. *chrysolepis*, Q. *garryana*, Q. *kelloggii*, and Q. *wislizeni*).

Table 2 AUC values from each species distribution model of eight California oak species, and threshold values using the Maximum Sensitivity and Specificity method for binary predictions of presence and absence

Species	**Area Under Curve (AUC)**	**Maximum Sensitivity + Specificity Threshold**
Q. agrifolia	0.887	0.42
Q. chrysolepis	0.831	0.43
Q. douglassii	0.842	0.48
Q. engelmannii	0.987	0.17
Q. garryana	0.947	0.33
Q. kelloggii	0.869	0.44
Q. lobata	0.865	0.42
Q. wislizeni	0.853	0.45

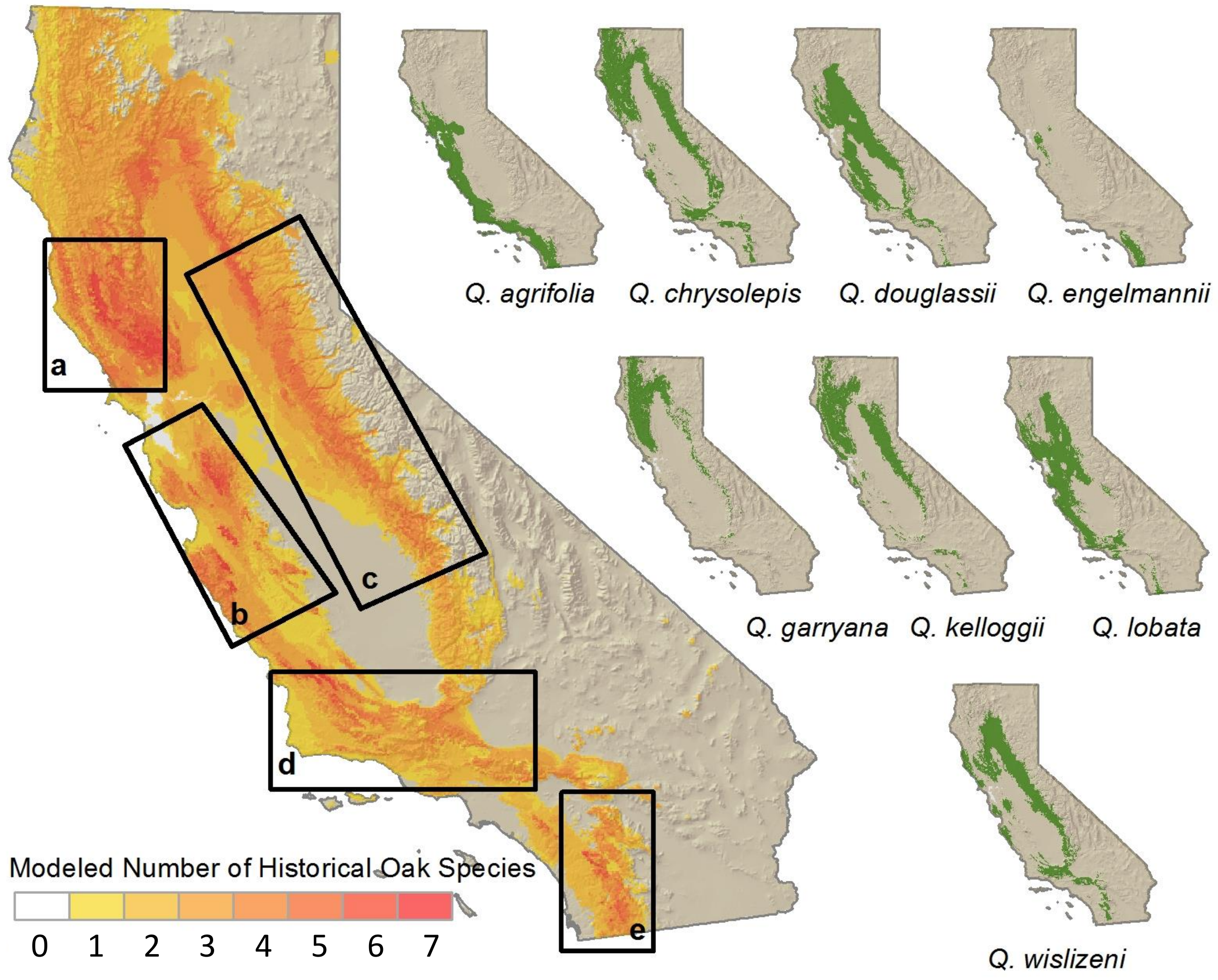


Figure 2 Modeled Number of Historical Oak Species: a) North Coast Range; b) South Central Coast Range, c) Sierra Nevada Foothills, d) Transverse Ranges including the Tehachapi Mountains, and e) Peninsular Ranges. Individual binary maps of eight modeled oak distributions are also shown.

However, hot spots of historical oak richness (6 or more oak species) currently have low representation in conserved lands. Of the mapped areas identified as supporting suitable habitat for seven oak taxa: 4% fall within areas developed since 1930, and 13% fall within lands with current conservation protection. For the conservation of high oak richness these regions would be high priority areas for conservation land acquisition.

A visual comparison of areas of modeled historical oak distribution with urban areas and parks, public, conservation and trust ownership lands are found in Figure 3. We focus on three urbanizing areas of the state: a) the San Francisco Bay Area, b) the Sacramento/Sierra Foothills area, and the c) Los Angeles area; as well as two areas that have high richness and recent conservation: d) the inner Coast Ranges of Napa and Lake Counties, and e) the Tehachapi Mountains. Despite the fact that current urban areas do not occur in areas of high historical oak richness, there is considerable spatial juxtaposition of current urban footprint and areas of moderate historical oak richness (2–5 species) in large urban areas across the state. In the San Francisco Bay Area (Figure 3a), a 3,490 km^2 region covering ten counties, 918 km2 (26.3%) of single species range, 2,556 km^2 (73.3%) of moderate species richness (2–5 oak species), and 10 km^2 (0.3%) of *Quercus* hot spots have been converted to urban areas. This region of the state is a matrix of intermixed parkland and urban area: the cities of Contra Costa and Alameda on the east side of the San Francisco Bay area surround the biologically rich area of Mt. Diablo. In the southern San Francisco Bay area, the rapid expansion between the San Jose urban area and Morgan Hill is encroaching on an area rich in oak species richness.

In the Sacramento and Sierra Foothills area (Figure 3b), a 1,630 km^2 region covering five counties, 241 km^2 (14.8%) of single species range, 1,293 km^2 (79.3%) of moderate species (2–5 oak species) richness, and 2.9 km^2 (0.2%) of Quercus hot spots have been converted to urban areas. The Sierra Foothills are a rich area for oak species and are increasingly threatened with urban and exurban expansion particularly along the Interstate 80 and Highway 50, shown as the twin arms of urbanization located east from the city of Sacramento in Figure 3b. There are few large parks or open space lands in this Foothill Belt (150–900 m in elevation) to help conserve oak richness most federally owned lands in the Sierra Nevada are located in the mixed Conifer belt and higher (above 900 m). In both of these areas urban expansion has affected the moderate (2–5 oak species) richness class the most.

In the Los Angeles area (Figure 3c), a 9,169 km2 region covering five counties, 4,219 km2 (46.0%) of single species range and 1,113 km2 (12.1%) of moderate species (2–5 oak species) richness have been converted to urban areas. No high Quercus richness areas were converted to urban areas. Oak habitat extends south from the Transverse Ranges and rings the mountains surrounding the Los Angeles Basin (Fig. 3c) and Peninsular Ranges to the border with Mexico. This is an area of active urban growth; however, there are considerable large extant open space areas (primarily federal lands) to serve as preserves.

The inner Coast Ranges of Napa and Lake Counties in northern California (Figure 3d) and the Tehachapi mountains of southern California (Figure 3e) are areas of high oak richness that have recently significantly increased their conservation of oak diversity. In 2015 the area identified with high oak richness in Napa and Lake counties was proclaimed as a new National Monument (Berryessa Snow Mountain) and in 2010 the purchase of 62,000 acres of Tejon Ranch, located in the Techachapis was approved.

Table 3 Modeled historical oak richness. Area supporting oaks predicted to occur based on species distribution models, by number of oak tree species richness and individual oak species, and the percentage found within urban or protected areas

Species	Total km^2	% in Urbanized Areas	% in Protected Areas
Q. agrifolia	58,597.8	18.62	26.61
Q. chrysolepis	88,543.9	0.92	54.88
Q. douglassii	83,423.5	5.07	17.18
Q. engelmannii	9,373.6	18.99	32.49
Q. garryana	43,882.6	0.75	50.18
Q. kelloggii	68,182.2	1.57	47.25
Q. lobata	76,616.5	8.27	19.50
Q. wislizeni	46,606.0	4.42	51.80

Number of Species	Description	Total km^2	% in Urbanized Areas	% in Protected Areas
1	Low	39,775.7	16.86	32.91
2		59,748.6	7.57	32.85
3	Moderate	62,484.7	4.02	38.57
4		20,924.4	3.03	39.35
5		9,114.3	2.66	27.81
6	High	2,533.2	1.45	24.66
7		959.5	4.03	13.11

4 Discussion

Reconstructing historical distributions and patterns of richness is critical to our understanding of current landscapes, in addition this knowledge provides the foundation for thoughtful and informed management, protection, restoration, and planning decisions (Rhemtulla & Mladenoff, 2007). The history of a landscape or the historical distribution of a species does not establish a linear path for the future, but rather, provides a foundation of understanding (White & Walker, 1997), and gives context to the trajectories of species and landscapes (Foster et al., 2003). Urban planning principles urge the integration of elements from the surrounding flora, fauna, and topography in building sustainable landscapes (McHarg, 1971; Steiner, 2008). Therefore, integrating historical landscape ecological research with disciplines that investigate and modify the built environment such as planning provides a pathway for directing future landscape change. Understanding and mapping historical distributions of natural vegetation types, as well as using historical data in modern modeling provides opportunity for ecologically and historically based decision making, planning, and policy direction. As human population increases, planning projects increasingly modify current infrastructure and existing structures. Therefore, knowledge of past landscape history could provide critical inspiration for regreening cities and re-connecting them with their past. Many of California's urban areas were constructed in landscapes historically rich in oak woodlands: this disappearance of oaks within the urban landscape has since motivated plans to return oaks even within heavily urbanized areas (Grossinger et al., 2007; Whipple et al., 2011). The utility of historical data to drive environmental niche models, generating past species distributions and reconstructions of vegetation communities is an unexplored theme in urban planning. This study of using a single historical dataset (VTM) to provide historical distributions of one taxon is just one example of the capabilities and value-added information that rich biogeographic data can lend to urban planning. We argue that the lack of understanding of past landscapes and important vegetation communities is a potential oversight within urban planning that is easily remedied through the use of the techniques and data presented in this paper and strengthened with other rich biogeographic datasets available for the state (see Table 4). By linking the past with the present through the use of modeling techniques we carry invaluable ecosystem and human health services into our modern urban environments.

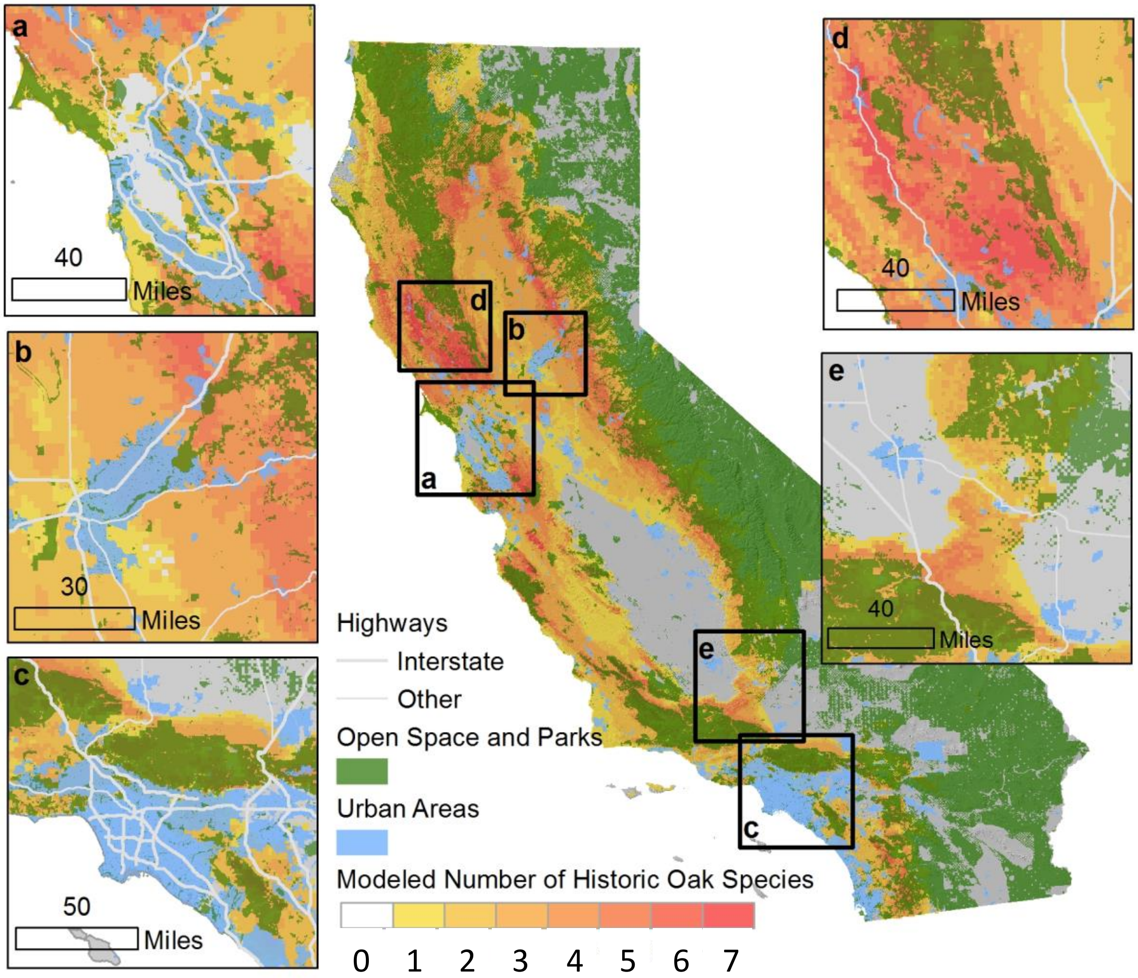


Figure 3 Areas of historical *Quercus* richness mapped with current urban and protected areas, with a focus on juxtaposition of historical richness and urban areas: a) San Francisco Bay Area, b) Sacramento/Sierra Foothills, c) Los Angeles; as well as areas where historical oak richness are not near protected areas: d) Napa/Sonoma/Mendocino Counties and e) Tehachapi mountains.

Through the development of environmental niche models, we have found that California oaks have been greatly impacted by urban development and this is likely to continue. Historical land use change, such as widespread clearing of blue oaks during “rangeland improvement” programs (Bolsinger, 1988), and current and future loss of habitat for urban and ex-urban expansion, will further fragment intact oak woodlands, eroding the sustainability of the oak woodland ecosystem and its associated products and ecological services, including wildlife habitat provision (Hilty & Merenlender, 2004), genetic richness, and evolutionary potential (Grivet et al., 2008). The Sierra Foothills region (Figure 2c) of the state is an example of these complicated interactions with urban and suburban growth predicted to double by 2020 at great consequence to forests and rangelands (Theobald, 2005). Urban growth in this area has extended into rural areas through rapid development of low density housing, increasing competing interests in the urban/wildlife interface, challenging fire management in these arid ecosystems, and illustrating the complex relationship between natural resource management and urban development encountered across the state (Byrd, Rissman, & Merenlender, 2009). Historical species richness and distribution data such as presented may serve to highlight areas where developmental pressures are encroaching upon high oak richness, prompting further investigation.

Table 4 List of the most comprehensive biodiversity databases for California with reference to the type of data they hold, the number of specimens reported at the time (11/2016) and their extent. These databases provide historical and current species occurrence information that can be used to construct species distribution models. Note that some records are redundant and may be housed in multiple databases.

Database	**Data**	**Number of specimens/localities**	**Extent**
Global Biodiversity Information Facility (GBIF) http://www.gbif.org	Plants and Animals	624,423,832	Global
California Natural Diversity Database (CNDDB) https://www.wildlife.ca.gov/Data/CNDDB	Plants and Animals-rare species only	86,000	California
HOLOS-Berkeley Ecoinformatics Engine* https://holos.berkeley.edu	Plants, Animals, Maps	>3 million	Primarily California
GAP http://gapanalysis.usgs.gov/species	Animals only	1,480 species	United States
Vertnet http://vertnet.org	Animals only	80 million	Global
Biodiversity Information Serving Our Nation (BISON) https://bison.usgs.gov	Plants and Animals	>100 million	United States
INaturalist http://www.inaturalist.org	Plants and Animals	3,173,095	Global
CalFlora https://www.calflora.org	Plants only	>1 million	California
Consortium of California Herbaria (CCH) http://ucjeps.berkeley.edu/consortium	Plants only	>2 million	Primarily California
iDigBio https://www.idigbio.org	Plants and Animals	73,192,805	Global

*** locality information used in this paper was sourced from HOLOS-Berkeley Ecoinformatics Engine**

Oaks, in particular are emblematic of California landscapes and serve as keystone cultural and ecological species providing ecosystem services through the provisioning of shade, soil stabilization, air and water quality regulation, food and shelter for animals, as well as providing aesthetics linked to increased property value. As more historical landscapes are being lost to increased urbanization and climatic pressures are projected to reduce species ranges (Kueppers et al., 2005), it is critical to maintain species diversity and reduce habitat fragmentation by making our built and natural environments more cohesive through the strategic placement and reintroduction of important habitats and species, such as oaks. Through the use of historical data and modeling the integration of lost landscape features starts from a more informed position. Current efforts (e.g. Grossinger et al., 2007; Whipple et al., 2011) in the California Bay Area serve as an example of how coordinated efforts between local open space councils, local stewards, and urban planning officials can led to "re-oaking" (Grossinger et al., 2012, Grossinger & Beller, 2011): the reintroduction of oak woodland landscapes and of native oaks to the urban forest canopy. Future efforts in urban planning would also benefit from the use of historical data and modeling to locate hot spots of species richness, understand where habitats and species have been lost historically, and use this evidence as incentive to recover what was lost and preserve what still exists.

Understanding past distributions as we have done in this paper is a critical step in the development of future models that address the impacts of a changing climate. Future climate models for California show trends of increasing temperatures, creating longer summers and shorter, warmer winters, with less snowpack retention and therefore a diminishing water source to last through the longer, drier summers (Cayan, Luers, Hanemann, & Franco, 2006; Luers, Cayan, Franco, Hanemann, & Croes, 2006; Thorne, Boynton, Flint, & Flint, 2015). Expected increasing temperatures will likely exacerbate existing ecological problems from pests and diseases (Cayan et al., 2006; Luers et al., 2006). Diseases such as Armillaria, Hypoxylon (root rot) and Phytophthora ramorum (commonly known as "sudden oak death") are expected to more easily infect drought-stressed trees (California Fire and Resource Assessment Program, 2010; Cayan et al., 2006; Luers et al., 2006). P. ramorum, which can rapidly kill coast live oak (Q. agrifolia) and California black oak (Q. kelloggii), among other species, has already been confirmed in 14 counties in the state of California (California Oak Mortality Task Force).

Policy measures to protect oaks and oak woodlands might be a way to conserve areas of oak richness, but measures are complicated by the fact that the majority of oak-dominated woodlands in the state (>80%) are located on private lands (Davis et al., 1998; Pavlik et al., 1991; Santos & Thorne, 2010; Standiford & Bartolome, 1997). Further, the notion of oak woodlands as a traditional working landscape historically reduced their value in the eyes of the conservation community possibly delaying formalized protection until the 1970's (Cox & Underwood, 2011; Santos, Watt, & Pincetl, 2014). However, following this formalization of protection, the decentralized structure to statewide conservation and protection of oak woodlands, including the lack of statewide information on patterns of oak distribution and richness, has left the responsibility to protect and regulate oaks unclear.

The environmental consequences of inconsistent policy may have detrimental effects on the distribution of oak woodland communities. Since many of the oak hot spots identified span administrative and county boundaries, the need for a statewide mandate and clear delegation of protection and regulation authority is essential in developing a regional approach to conservation of oaks and oak woodland habitat. Although local policies may be inconstant county to county, they are still critical to developing a multi-scalar approach to conservation of oaks from individuals to landscape. Local strategies of conservation such as land acquisition in the form of land trusts and conservation easements (Merenlender, Huntsinger, Guthey, & Fairfax, 2004) and open space designation, would benefit from the mapping of past, current, and future oak distribution and richness. For instance, areas of modeled historical oak richness—the North Coast Ranges, the South Coast Ranges, the Sierra Foothill Belt, the Transverse Ranges, and the Interior Coast Ranges are important repositories for plant species endemism (Grivet et al., 2008; Thorne, Viers, Price, & Stoms, 2009), and are critical conservation areas for oak woodlands that could be looked at more closely for incorporation under conservation easements open space designations, or planning that incorporates oaks and woodland habitat into new communities. Making transparent the locations of hot spots of richness gives strength and reasoning to local initiatives and could potentially initiate consistent statewide policy.

5 Conclusions

In this paper, we combined modeled data from a historical dataset with modern data on urban and protected areas, to provide a base for understanding the pressure of development on the distribution and richness of oak species. Areas of modeled historical oak richness were compared to the current footprint of urban areas and current conserved lands. We found that about a fifth of the area that previously contained a single oak species in the past is now urban with nearly 20% of the modeled historical range of both coast live oaks and Engelmann oaks now under the modern urban footprint. Areas of moderate historical oak richness have some protection on conservation lands but have been disproportionally affected by urban areas. Four oak species (Q. chrysolepis, Q. garryana, Q. kelloggii, and Q. wislizeni) are moderately protected, with around half of their modeled range currently on conservation lands. Hot spots of high oak richness (e.g. six Quercusspecies) currently have low proportional representation in conserved lands with only 13% of the modeled range within current conservation protection. Plans for protecting oak woodlands in California are complicated by policy, which can be local in scale, and fragmented with no uniting statewide mandate. Many of the areas of high historical oak richness span administrative boundaries, and thus are difficult to manage by policy measures alone. We therefore encourage the use of historical data to encourage and guide protection of these landscapes in the form of policy and regulations, and to help in planning for future urban greening efforts resurrecting oak habitat that sits waiting beneath modern sidewalks.

6 References

Alvey, A. A. (2006). Promoting and preserving biodiversity in the urban forest. Urban Forestry and Urban Greening, 5(4), 195–201.

Anderson, K., (2005). Tending the wild: Native American knowledge and the management of California's natural resources. Oakland, CA: University of California Press.

Araújo, M. B. (2002). Biodiversity hotspots and zones of ecological transition. Conservation Biology, 16(6), 1662–1663.

Barbour, M., Keeler-Wolf, T., & Schoenherr, A. A. (2007). Terrestrial vegetation of California. Oakland, CA: University of California Press.

Bartolome, J. W., McClaran, M. P., Allen-Diaz, B. H., Dunne, J., Ford, L. D., Standiford, R. B., Forero, L. C., (2002). Proceedings of the fifth symposium on oak woodlands: Oaks in California's changing landscapes (pp. 22–25), Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture.

Beller, E., Downs, P., Grossinger, R., Orr, B., & Salomon, M. (2015). From past patterns to future potential: using historical ecology to inform river restoration on an intermittent California river. Landscape Ecology, 31(3), 581–600. dx.doi:10.1007/s10980-015-0264-7

Bolsinger, C. L. (1988). The hardwoods of California's timberlands, woodlands, and savannas (Vol. 148). Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture.

Buerki, S., Callmander, M. W., Bachman, S., Moat, J., Labat, J. N., & Forest, F. (2015). Incorporating evolutionary history into conservation planning in biodiversity hotspots. Philosophical Transactions of the Royal Society B, 370(1662), 20140014.

Byrd, K. B., Rissman, A. R., & Merenlender, A. M. (2009). Impacts of conservation easements for threat abatement and fire management in a rural oak woodland landscape. Landscape and Urban Planning, 92(2), 106–116.

California Department of Finance. (2013). Report P-1: State and County Population Projections by County, by Race/Ethnicity, and by Major Age Groups, 2010– 2060 (by decade). Sacramento, CA: California Department of Finance.

California Fire and Resource Assessment Program. (2010). The changing California: Forest and range 2010 assessment. Sacramento, CA: California Department of Forestry and Fire Protection. California Oak Mortality Task Force

CPAD Version 2013b. Oakland, CA: GreenInfo Network. Retrieved from http://www.calands.org

Cavender-Bares, J., Gonzalez-Rodriguez, A., Pahlich, A., Koehler, K., & Deacon, N. (2011). Phylogeography and climatic niche evolution in live oaks (Quercus series Virentes) from the tropics to the temperate zone. Journal of Biogeography, 38(5), 962–981.

Cayan, D., Luers, A. L., Hanemann, M., & Franco, G. (2006). Scenarios of climate change in California: An overview. Sacramento, CA: California Climate Change Center

Colwell, W. L. (1977). The status of vegetation mapping in California today. In M. G. Barbour & J. Major (Eds.), Terrestrial vegetation of California. Sacramento, CA: John Wiley & Sons.

Cooper, W. S. (1922). The broad-sclerophyll vegetation of California: An ecological study of the chaparral and its related communities (No. 319). Washington, DC: Carnegie Institution.

Cox, R. L., & Underwood, E. C. (2011). The importance of conserving biodiversity outside of protected areas in Mediterranean ecosystems. PLoS One, 6(1), e14508.

Davis, F. W., Stoms, D. M., Hollander, A. D., Thomas, K. A., Stine, P. A., Odion, D., Warner, K. (1998). The California Gap Analysis Project—Final Report. Santa Barbara, CA: University of California, Santa Barbara.

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics, 40(1), 677–697.

Ertter, B. (2000). Our undiscovered heritage: Past and future prospects for species-level botanical inventory, Madroño, 47(4), 237–252.

Ferenc, M., Sedláček, O., & Fuchs, R. (2014). How to improve urban greenspace for woodland birds: Site and local-scale determinants of bird species richness. Urban Ecosystems, 17(2), 625–640.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/ absence models. Environmental Conservation, 24(1), 38–49

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., & Knapp, A. (2003). The importance of land-use legacies to ecology and conservation. BioScience, 53(1), 77–88.

Gaman, T., & Firman, J., (2006). Oaks 2040: The status and future of oaks in California. Oakland, CA: California Oak Foundation.

Gordon, A., Simondson, D., White, M., Moilanen, A., & Bekessy, S. A. (2009). Integrating conservation planning and landuse planning in urban landscapes. Landscape and Urban Planning, 91(4), 183–194.

Graham, C. H., Ferrier, S., Huettman, F., Moritz, C., & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. Trends in Ecology and Evolution, 19(9), 497–503.

Grivet, D., Sork, V. L., Westfall, R. D., & Davis, F. W. (2008). Conserving the evolutionary potential of California valley oak (Quercus lobata Née): A muiltivariate genetic approach to conservation planning. Molecular Ecology Notes, 17(1), 139–156.

Grossinger, R. M., Askevold, R. A., Beagle, J., Beller, E. E., Brewster, E., Gardner, S. Striplen, C. J. (2012). Napa Valley historical ecology atlas: Exploring a landscape of transformation and resilience. Berkeley, CA: University of California Press.

Grossinger, R., & Beller, E. (2011). “Oak landscapes in the recent past.” In L. R. Costello, B. W. Hagen, & K. S. Jones (Eds.), Oaks in urban landscapes: Selection, preservation, and care (pp. 222–223). Davis, CA: UC Agriculture and Natural Resources.

Grossinger, R. M., Striplen, C. J., Askevold, R. A., Brewster, E., & Beller, E. E. (2007) Historical landscape ecology of an urbanized California valley: Wetlands and woodlands in the Santa Clara Valley. Landscape Ecology, 22(1), 103–120.

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. Global Ecology and Biogeography, 24(3), 276–292.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. Ecology Letters, 8(9), 993–1009.

Hampe, A. (2004). Bioclimate envelope models: what they detect and what they hide. Global Ecology and Biogeography, 13(5), 469–476.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25(15), 1965–1978.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2012). Dismo: Species distribution modeling. R package version 0.7-17.

Hijmans, R. J., Schreuder, M., De la Cruz, J., & Guarino, L. (1999). Using GIS to check co-ordinates of genebank accessions. Genetic Resources and Crop Evolution, 46(3), 291–296.

Hilty, J. A., & Merenlender, A. M. (2004). Use of riparian corridors and vineyards by mammalian predators in Northern California. Conservation Biology, 18(1), 126–135.

Huntsinger, L., Buttolph, L., & Hopkinson, P. (1997). Ownership and management changes on California hardwood rangelands: 1985 to 1992. Journal of Range Management, 50(4), 423–430.

Huntsinger, L., & Fortmann, L. P. (1990). California's privately-owned oak woodlands: Owners, use, and management. Journal of Range Management, 43(2), 147–152. Jepson, W. L. (1910). The silva of California (Vol. 2). The University Press.

Kark, S., Schneider, C. J., & Smith, T. B. (2001). Biodiversity hotspots and beyond: The need for preserving environmental transitions. Trends in Ecology and Evolution, 16(8), 431.

Keenan, T., Maria Serra, J., Lloret, F., Ninyerola, M., & Sabate, S. (2011). Predicting the future of forests in the Mediterranean under climate change, with niche-and process-based models: CO2 matters! Global Change Biology, 17(1), 565–579.

Kelly, M., Allen-Diaz, B., & Kobzina, N. (2005). Digitization of a historic dataset: The Wieslander California vegetation type mapping project. Madroño, 52(3), 191–201.

Kelly, M., Easterday, K., Rapacciuolo, G., Koo M., McIntyre, P., & Thorne, J. (2016). Rescuing and sharing historic vegetation data for ecological analysis: The California Vegetation Type Mapping project. Biodiversity Informatics, 11, 40–62.

Kelly, M., Guo, Q., Liu, D., & Shaari, D. (2007). Modeling the risk of a new invasive forest disease in the United States: An evaluation of five environmental niche models. Computers, Environment and Urban Systems, 31(6), 689–710.

Kelly, M., Ueda, K., & Allen-Diaz, B. (2008). Considerations for ecological reconstruction of historic vegetation: Analysis of the spatial uncertainties in the California Vegetation Type Mapping dataset. Plant Ecology, 194(1), 37–49.

Kelly, N. M., Fonseca, M., & Whitfield, P. (2001). Predictive mapping for management and conservation of seagrass beds in North Carolina. Aquatic Conservation: Marine and Freshwater Ecosystems, 11(6), 437–451.

Kueppers, L. M., Snyder, M. A., Sloan, L. C., Zavaleta, E. S., & Fulfrost, B. (2005). Modeled regional climate change and California endemic oak ranges. Proceedings of the Academy of Sciences, 102(45), 16281– 16286.

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography, 40(4), 778–789.

Luers, A. L., Cayan, D. R., Franco, G., Hanemann, M., & Croes, B. (2006). Our changing climate: Assessing the risks to California. Sacramento, CA: California Climate Change Center.

Marañón, T., Ibáñez, B., Anaya-Romero, M., MuñozRojas, M., & Pérez-Ramos, I. (2016). Oak Trees and woodlands providing ecosystem services in Southern Spain. In D. Rotherham, C. Handley, M. Agnoletti, & T. Samojlik (Eds.), Trees Beyond the Wood conference proceedings (pp. 291–299). Sheffield, UK: Wildtrack.

Marcucci, D. J. (2000). Landscape history as a planning tool. Landscape and Urban Planning, 49(1–2), 67–81.

McHarg, I. L. (1971). Design with nature. Garden City, NY: Doubleday/Natural History Press.

Medail, F., & Quezel, P. (1999). Biodiversity hotspots in the Mediterranean Basin: Setting global conservation priorities. Conservation Biology, 13(6), 1510–1513.

Medvitz, A. G., & Sokolow, A. D. (1995). Population growth threatens agriculture, open space. California Agriculture, 49(6), 11–17.

Merenlender, A. M., Huntsinger, L., Guthey, G., & Fairfax, S. K. (2004). Land trusts and conservation easements: Who is conserving what for whom? Conservation Biology, 18(1), 65–75.

Milanovich, J., Peterman, W., Barrett, K., & Hopton, M. (2012). Do species distribution models predict species richness in urban and natural green spaces? A case study using amphibians. Landscape and Urban Planning, 107(4), 409–418. doi: 10.1016/j.landurbplan.2012.07.010

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. Nature, 403(6772), 853–858.

Nielsen, A. B., van den Bosch, M., Maruthaveeran, S., & van den Bosch, C. K. (2014). Species richness in urban parks and its drivers: A review of empirical evidence. Urban Ecosystems, 17(1), 305–327.

Niemelä, J. (1999). Ecology and urban planning. Biodiversity & Conservation, 8(1), 119–131.
Pavlik, B. M., Muick, P. C., Johnson, S. G., & Popper, M. (1991). Oaks of California. Los Olivos, CA: Cachuma Press.

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography, 12(5), 361– 371.

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Lees, D. C. (2006). Model-based uncertainty in species range prediction. Journal of Biogeography, 33(10), 1704–1711.

Peterson, A. T. (2011). Ecological niches and geographic distributions (MPB-49). Princeton, NJ: Princeton University Press.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modeling, 190(3), 231–259.

R Development Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Raxworthy, C. J., Martinez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A., & Peterson, A. T. (2003). Predicting distributions of known and unknown reptile species in Madagascar. Nature, 426(6968), 837–841.

Rhemtulla, J. M., & Mladenoff, D. J. (2007). Why history matters in landscape ecology. Landscape Ecology, 22, 1–3.

Rissman, A. R., & Butsic, V. (2011). Land trust defense and enforcement of conserved areas. Conservation Letters, 4(1), 31–37.

Rissman, A. R., Lozier, L., Comendant, T., Kareiva, P., Kiesecker, J. M., Shaw, M. R., & Merenlender, A. M. (2007). Conservation easements: biodiversity protection and private use. Conservation Biology, 21(3), 709–718.

Rudd, H., Vala, J., & Schaefer, V. (2002). Importance of backyard habitat in a comprehensive biodiversity conservation strategy: a connectivity analysis of urban green spaces. Restoration Ecology, 10(2), 368–375.

Rushton, S. P., Omerod, S. J., & Kerby, G. (2004). New paradigms for modelling species distributions? Journal of Applied Ecology, 41(2), 193–200.

Santos, M. J., & Thorne, J. H. (2010). Comparing culture and ecology: Conservation planning of oak woodlands in Mediterranean landscapes of Portugal and California. Environmental Conservation, 37(2), 155–168.

Santos, M. J., Watt, T., & Pincetl, S. (2014). The push and pull of land use policy: Reconstructing 150 years of development and conservation land acquisition. PloS One, 9(7), e103489.

Schussman, H., Geiger, E., Mau-Crimmins, T., & Ward, J. (2006). Spread and current potential distribution of an alien grass, Eragrostis lehmanniana Nees, in the Southwestern USA: Comparing historical data and ecological niche models. Diversity and Distributions, 12(5), 582–592.

Scott, T., & Standiford, R. B. (2001). Value of oak woodlands and open space on private property values in Southern California., 10(1), 137–15.

Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A meta-analysis of global urban land expansion. PloS One, 6(8), e23777.

Standiford, R. B., & Bartolome, J. (1997). The integrated hardwood range management program: Education and research as a conservation strategy (USDA Forest Service General Technical Report PSW-GTR-16). Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture.

Standiford, R., & Huntsinger, L. (2012). Valuing forestland environmental services: A case study for California’s oak woodlands. Proceedings of the International Symposium of Forest Economics 2012 (ISFE2012) (pp. 273–281). Beijing, China.

Stein, E. D., Dark, S., Longcore, T., Grossinger, R., Hall, N., & Beland, M. (2010). Historical ecology as a tool for assessing landscape change and informing wetland restoration priorities. Wetlands, 30(3), 589–601.

Steiner, F. (2008). The living landscape (2nd ed.). Washington, DC: Island Press. Theobald, D. M. (2005). Landscape patterns of exurban growth in the USA from 1980 to 2020, Ecology and Society, 10(1), 32. Thorne, J., Viers, J., Price, J., & Stoms, D. (2009). Spatial patterns of endemic plants in California. Natural Areas Journal, 29(4), 344–366.

Thorne, J. H., Boynton, R. M., Flint, L. E., & Flint., A. L. (2015). The magnitude and spatial patterns of historical and future hydrologic change in California’s watersheds, Ecosphere, 6(2), 1–30.

Thorne, J. H., Kelsey, R., Honig, J., & Morgan, B. (2006). The development of 70-year-old Wieslander vegetation type maps and an assessment of landscape change in the central Sierra Nevada (prepared for the California Energy Commission, Public Interest Energy Research). Davis, CA: University of California Davis.

Thorne, J. H., Santos, M. J., & Bjorkman, J. H. (2013). Regional assessment of urban impacts on landcover and Open space finds a smart urban growth policy performs little better than business as usual. PloS One, 8(6), e65258.

U.S. Census Bureau (2014). National Geospatial Data Asset (NGDA) Urban Area.

Václavík, T., & Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling, 220(23), 3248–3258.

Waddell, K. L. & Barrett, T. M. (2005). Oak woodlands and other hardwood forests of California, 1990s (PNWRB-245).

Warren, D. L., Wright, A. N., Seifert, S. N., & Shaffer, H. B. (2014). Incorporating model complexity and spatial sampling bias into ecological niche models of climate change risks faced by 90 California vertebrate species of concern. Diversity and distributions, 20(3), 334–343.

Whipple, A. A., Grossinger, R. M., & Davis, F. W. (2011). Shifting baselines in a California oak savanna: Nineteenth century data to inform restoration scenarios. Restoration Ecology, 19(101), 88–101.

White, P. S., & Walker, J. L. (1997). Approximating nature's variation: selecting and using reference information in restoration ecology. Restoration Ecology, 5(4), 338–349.

Wieslander, A. E. (1935). A vegetation type map of California. Madroño, 3(3), 140–144.

Wu, J. J. (2008). Making the case for landscape ecology an effective approach to urban sustainability. Landscape Journal, 27(1), 41–50.

Zhang, M. G., Zhou, Z. K., Chen, W. Y., Slik, J. F., Cannon, C. H., & Raes, N. (2012). Using species distribution modeling to improve conservation and land use planning of Yunnan, China. Biological Conservation, 153, 257–264.

Chapter 4: From the field to the cloud: a review of three approaches to sharing historical data from field stations using principles from data science

Abstract

Historical data play an important role in our understanding of environmental change and ecosystem dynamics. By elongating the temporal scale of scientific inquiry, historical data reveal insights into the dynamic nature of ecosystems over long time periods that might otherwise be unavailable. However, most historical data has yet to make a full contribution, remaining ‘dark’ and out of reach to the broader scientific community. This article responds to several calls stressing the importance of historical materials from field research and urges their preservation and accessibility. Despite the importance of historical data collections, few standards have emerged to integrate historical dark data into the larger digital data landscape. To encourage greater use of historical data across scientific disciplines it is vital to make data findable, accessible, interoperable, and reusable (e.g. the FAIR principles). In this paper we discuss the potential of historical dark data to contribute to the modern digital ecological data landscape. We do this by focusing on three cases from the University of California field and research stations and the groups that have worked to make historical dark data discoverable. Despite the common goal of maximizing the potential use of these data collections, each case and the methods employed are unique, and showcase varying levels of success in achieving the FAIR principles and shepherding historical data into the 21st century.

Keywords

dark data, data science, historical data, field stations, open data

1 Introduction

Scientific research increasingly highlights large datasets for their transformative potential in solving enduring and complex problems, leading one recent analysis to declare data the “world's most valuable resource”(Fosso Wamba et al., 2015; Hampton et al., 2013; The Economist, 2017). Large “born digital” data from modern data streams have increased the scope of environmental inquiry. Within the last two decades, advances in computing, databases, sensing technologies, cloud-based services, social media, and mobile data collection (among other things) have ushered in an era of ‘big data’ characterized by a previously unimaginable volume, variety, and velocity of incoming data streams (Gandomi and Haider, 2015). While ‘big data’ have garnered deserved attention, data generated from individual projects in small volumes at local scales (also called the ‘long tail of science’) (Hampton et al., 2013; Heidorn, 2008; Wallis et al., 2013) and ‘dark data’ including both unstructured and unused digital data collected during routine business and research (Ferguson et al., 2014; Hampton et al., 2013; Wallis et al., 2013) as well as analog, unarchived, non-machine readable historical data - also known as legacy, or heritage data (Bürgi and Gimmi, 2007; Salmond et al., 2012), have not. Such datasets are the foundations on which big data is often built (Ferguson et al., 2014) and represent a large portion of the data landscape that is currently underutilized but has recognized potential (Bi et al., 2013; Eitzel et al., 2016; Kelly et al., 2016; Michener and Jones, 2012). This paper responds to the need for new theory and methods to move what we call historical dark data - unarchived, non-digital legacy data - from file drawers to the cloud in order to realize its full value, potential, and become an integral part of the digital data landscape. Historical dark data includes unarchived physical data collections such as accumulated reports, field notes, journals, biological specimens, correspondence, and artifacts.

These materials have three important roles. They: 1) elongate the temporal scale of potential scientific inquiry to include otherwise irretrievable past environments, 2) provide a contextual foundation from which to assess change, and 3) situate the study of the environment in a wider disciplinary context. However, non-digital formats, decentralized physical location and variable condition of the data collections create barriers to productive scientific use and put important data at risk of disposal and loss. Several calls have stressed the importance of these types of materials and their preservation, but few standards have emerged to shed light on historical data. To encourage greater use of historical data across scientific disciplines and ensure a future for our past it is important to make these collections findable, accessible, interoperable, and reusable (FAIR) (Wilkinson et al., 2016).

In this paper, we focus on historical environmental data collected in the field at research stations or research properties. These kinds of physical data collections are common - the result of a century of business-as-usual research and daily operations that focused on forestry, ecology and agriculture. We review three University of California (UC) projects that digitize and share historical data collections and evaluate the collections’ journey out of the dark and into the larger digital data ecosystem with respect to the FAIR principles. These case studies reveal that historical data are complex, requiring diverse approaches to preservation and dissemination, but they also reveal that such efforts can be invaluable to the environmental sciences.

1.1 Data stewardship using FAIR principles

The synthesis of historical and contemporary ecological data with predictive models can be a powerful approach to investigate the complex response of species, communities, and landscapes to changing biophysical conditions in time and space (Kelly et al., 2016). Scientists tackling complex socio-ecological questions regularly deal with large collections of heterogeneous data and recognize that principles from data science can help them in their work (Hampton et al., 2013; Lowndes et al., 2017; Morrison et al., 2017; Peters et al., 2014a; Wilson et al., 2017; Wolkovich et al., 2012). Principles for reproducible data science, such as transparency, reusability, collaboration, and communication (Lowndes et al., 2017; Pedersen et al., 2007) can streamline projects and make science more efficient. Data sharing is a fundamental part of these efforts. Effective data management and sharing are key for data integration, knowledge discovery, and continued use (Wilkinson et al., 2016). The current landscape of scientific data can be fragmented - developed and maintained by individuals or small academic groups, on focused areas, and concentrated in time (Kelly et al., 2016; Michener, 2006; Michener et al., 1997; Waide et al., 2017) - precluding efficient use.

The FAIR (Findable, Accessible, Interoperable, and Reusable) framework is one of several recent efforts to establish best practices and principles for effective data management by the global research community (Wilkinson et al., 2016). Guidelines for long term data stewardship are not new, but their adoption in practice can be ad hoc (Borgman et al., 2015; Michener and Brunt, 2009). FAIR principles serve as umbrella concepts for goal setting and evaluating success that may translate across institutional, educational, disciplinary, and technological barriers. *Findable* requires that data and/or metadata should be uniquely and persistently identified, indexed, and described in detail so that they may be discovered by potential researchers. For data to become *accessible* once found*,* they need to be published using standard, free, and open protocols. Using standard data formats and ontologies in this process makes data *interoperable.* Finally, data need to be *reusable*, so ensuring that data provenance is preserved and well documented for the next user is critical. Data projects that fulfill all FAIR principles are expected to have the most potential for use by new studies, in transdisciplinary research, and are therefore are highly valuable to science.

We believe the FAIR principles promote digital resilience (Wright, 2016) by fostering forward thinking approaches to data archiving, sharing, and use. FAIR can be conceptualized as a road map (Figure 1) with each step elevating the potential and value of data across a spectrum of dark unstructured collections that fulfill no FAIR principles to “open data” that fulfill FAIR (Ferguson et al., 2014). As dark data transition out of file drawers and into digital structures it increases the variety and volume of data that is readily available and for integration into scientific workflows, thereby expanding the temporal data record and increasing its potential reach. Achieving FAIR principles will enable the use of historical data in conjunction with contemporary data (Kelly et al., 2016), in transdisciplinary research (Beller et al., 2017; Michener, 2015), and in synthesis or meta-analysis (Wallis et al., 2013). This framework is useful in context of historical dark data because the principles are flexible, and even partial fulfillment can yield success and contribute towards increased use, potential and value of historical data in science. However, as we discuss in the following case studies, achieving FAIR principles is difficult and costly requiring long-term investment, stewardship, and expertise.

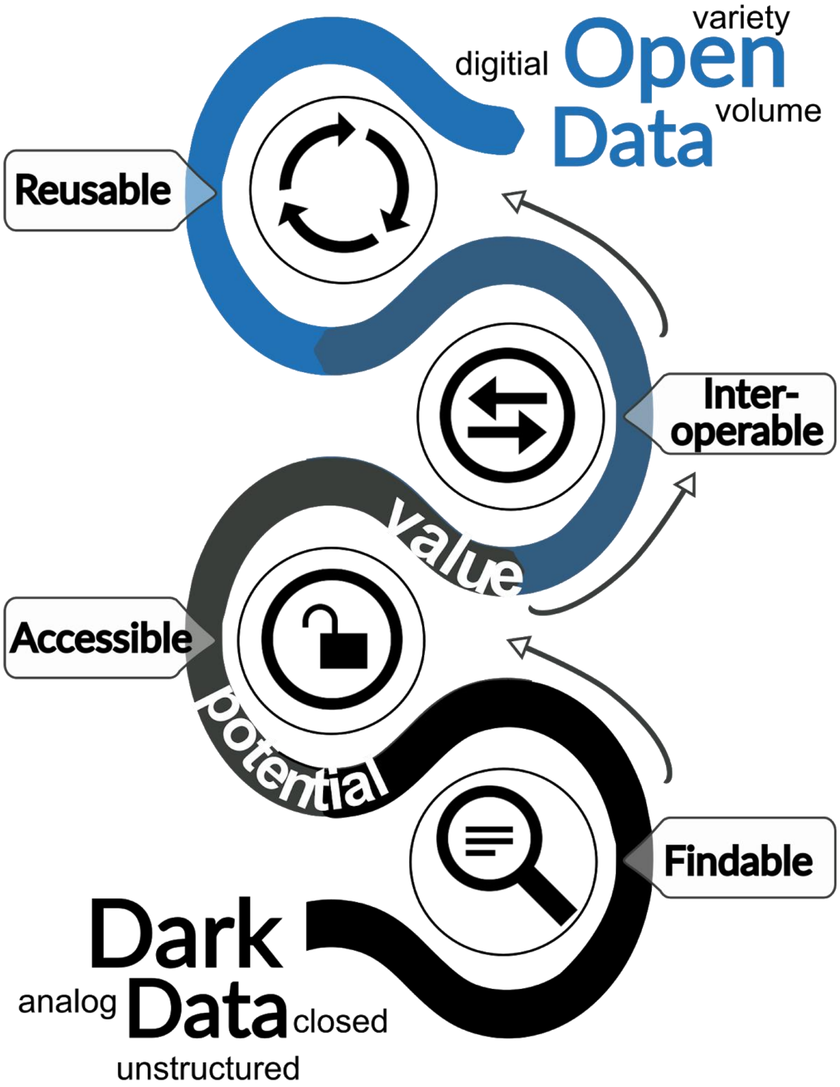


Figure 1 Conceptual diagram of the FAIR roadmap for dark data: each curve represents a step toward increasing the value and potential of dark data for science.

1.2 The current ecological data landscape

The current landscape of ecological data is complex and evolving (key U.S. players are summarized in Figure 2). Within this landscape, ***science and synthesis centers*** (NASA, NCEAS, LTER, NEON, NCALM, and SESYNC) serve as the institutional leads in developing standards in file formats, protocols, tools; and by creating partnerships across institutions or groups that lead to data aggregation or increase the potential for integration (Michener et al., 2011; Rodrigo et al., 2013). These centers work to synthesize and collect heterogeneous environmental data from multiple sources including field observations and experiments as well as sensor networks. Data from these centers have been used to study a wide range of environmental phenomena—

including land use change, invasive species, phenology, aquatic environments, atmospheric processes, and ecosystem dynamics—largely since the 1970s. These centers have largely succeeded at using and re-using diverse datasets by linking them through standardized metadata and centralized repositories, some of which are created by the centers themselves (Jones et al., 2006), but the coverage of their data often misses key historical events that shaped contemporary ecosystems before 1970 (Alagona et al., 2012).

Recent and numerous efforts to make data discoverable and interoperable has resulted in several types of data repositories. First among these are the recent proliferation of ***domain-specific*** data repositories, especially for biological specimens and associated data that use taxonomically-specific language, protocols, and standards. These repositories range from taxon-specific (e.g. VertNet) to taxonomically broad (e.g. GBIF) and include many museum and herbarium records. Data include digitized physical specimens; records of species occurrence, abundance, tolerances; and insight on various other environmental conditions derived from the digitization of field notebooks and journals. Institution-or collection-focused field notebook digitization efforts, such as field notebooks from UC Berkeley libraries (http://ecoreader.berkeley.edu/) or Zooniverse Notes from Nature (https://www.notesfromnature.org/), have shown the potential of crowdsourcing platforms to integrate historical dark data into the digital data landscape through transcription.

Generalist repositories exist on a spectrum of centralized or decentralized models of data aggregation (Franklin et al., 2017), some serve as a data warehouse collating data from disparate institutions and partners while others collect metadata and finding aids to point to the original location of the data but do not store the data itself. Generalist repositories include university-based efforts (e.g. Harvard's Dataverse, Berkeley's HOLOS); government sponsored national spatial data portals or clearinghouses (e.g. National Map, DataOne) (Crompvoets et al., 2004; Maguire and Longley, 2005; Tait, 2005); and proprietary portals (e.g. ESRI's Living Atlas of the World). Generalist repositories are not unique to ecological, biological, and environmental data, and ecological data and materials often exist in generalist repositories that ecologist may be unaware of (e.g. Digital Public Library of America) (Waide et al., 2017). Allied data repositories may establish even greater interconnections using an Application Programming Interface (API), thus, creating gateways to larger data landscapes. APIs are applications that serve machine-readable data and functionality to applications that represent the data to users.

Data registries (e.g. Registry of Research Data Repositories and FairSharing) serve as guides to help users find appropriate data. Registries provide global indexes of research data repositories, allowing users to search, find, or connect with groups that may have similar data (Pampel et al., 2013). Several scientific journals that require data deposition upon submission (e.g. Nature, Science, PNAS) also guide researchers by listing supported discipline-specific and generalist repositories. Registries foster interconnectivity and potentially reduce redundancy in the creation of new repositories, experiments, and data collections.

Finally, we identify emergent ***participatory or citizen science*** data repositories driven by massive public data collection efforts. These include biodiversity databases (e.g. iNaturalist.org) and other distributed and public efforts to document changing climates (e.g. IceWatch). Data from these non-traditional and volunteered collective efforts have already enhanced scientific learning in numerous cases (Connors et al., 2012; Dickinson et al., 2012, 2010; Kearns et al., 2003) and will play a growing role in ecological data collection, sharing, and use. This evolving ecological data landscape (or über network (Michener, 2015; Peters et al., 2014b)) encourages data discovery, integration, and reuse. Sharing data is a public good that is generated by mutual commitment to scientific principles (Hampton et al., 2013; Michener, 2015; Reichman et al., 2011; Wallis et al., 2013). The value of data and metadata repositories is in their capability to help make collections of data FAIR. However, the growth and success of these repositories has tended to overlook vital elements (and indeed the majority) of the data landscape that were not born digital and are not yet FAIR (Jones et al., 2006).

Figure 2 The digital ecological data landscape is comprised of science and synthesis centers, citizen science data repositories, generalist repositories, domain specific repositories, and data repository registries.

1.3 Growing the data landscape with historical collections

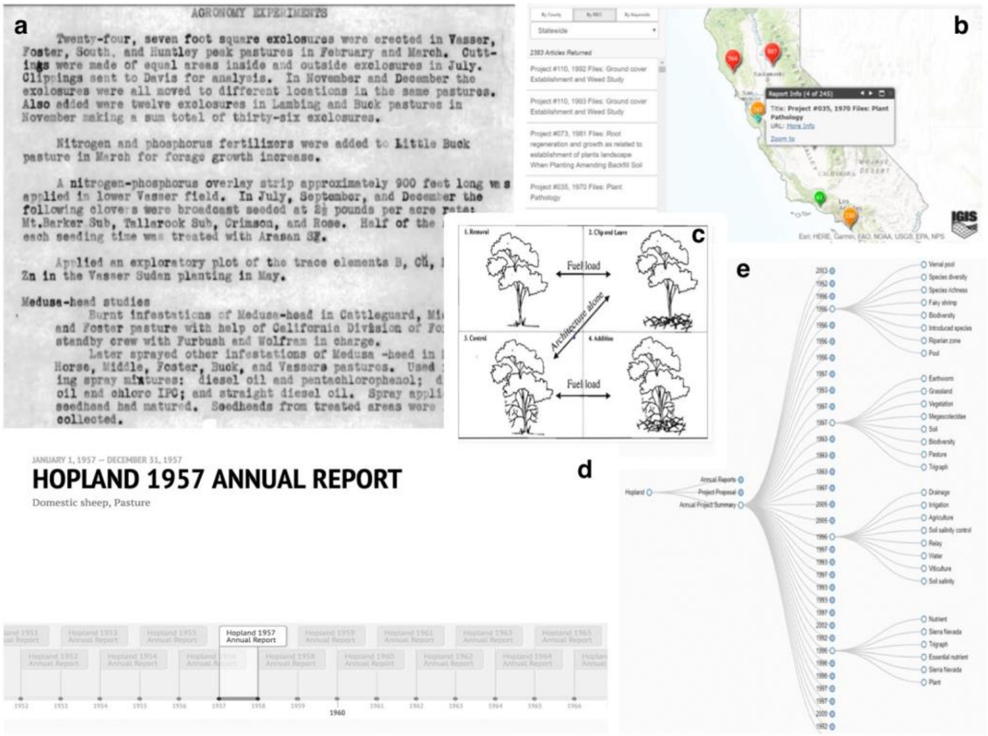
With the exception of domain specific repositories for biodiversity collections and field notebooks, historical data are disproportionately underrepresented in modern ecological repositories leaving temporal gaps in the scientific record (McClenachan et al., 2015; Szabó and Hédl, 2011; Zu Ermgassen et al., 2012). Despite the consensus that historical data are necessary, these types of data are often underutilized in practice (Magurran et al., 2010; Szabó, 2010) due to the difficulty integrating non-digital historical data in routine research. Ecological research using historical data have demonstrated success in modelling the impact of climate change on species abundance and distribution (Lavoie, 2013; Pyke and Ehrlich, 2010; Shaffer et al., 1998; Tingley and Beissinger, 2009), cataloguing drastic changes in forest structure, composition, and distribution (Easterday et al., 2016; Kelly et al., 2016; Petit et al., 2008), contextualizing evolutionary processes (Holmes et al., 2016), documenting the spread of infectious disease (Bradley et al., 2014; DiEuliis et al., 2016; Suarez and Tsutsui, 2004), and extending our knowledge of species lineages (Bi et al., 2013). Growing the reach of studies like these will likely depend on dispersed efforts by the many stewards and potential users of historical data.

Projects hoping to digitize and publish historical data face an overwhelming variety of platforms, technologies, standards, and protocols with few available guidelines. Historical data, due to their analog, unstructured nature, defy classic data deposition methods and require specific approaches that go largely undocumented. The development of protocols to make this type of data FAIR are vitally needed, since any data without redundant and varied storage methods face heightened risk of permanent loss (Elizabeth Griffin, 2015). Historical data emanating from distributed small research collections are often confronted with a lack of logical physical and digital storage options. This conflict forefronts the choice of either fitting the data to the needs of an existing infrastructure (like the repositories above) or developing a structure that fits the needs of the data. This choice is also constrained by current science funding structures that incentivize and value the creation of new repositories and data over the curation and integration of older ones.

2 Case studies of historical data preservation at University of California research centers

The University of California (UC) has been a leader in ecological, natural resource, and agricultural field research since the early 20th century (Chornesky et al., 2015; Rapacciuolo et al., 2017, 2014). We provide three case studies of projects attempting to recover historical dark from the University using different methods and approaches to digitization.

The first data collection comes from nine Research and Extension Centers (RECs) of the UC Division of Agriculture and Natural Resources (ANR) which cover over 5,000 ha of California's Central Valley, Sierra Foothills, and Pacific Coast. Since 1912, ANR RECs have hosted research generating important discoveries across agricultural and ecological disciplines (Downing, 2016; White, 2017) (Figure 3). The second data collection comes from the UC Natural Reserve System (NRS), the largest university-administered network of research reserves and field stations in the world (Fiedler et al., 2013). The earliest NRS site was founded in 1937, and the NRS now manages 39 sites (covering over 303,500 ha) for field research, conservation, teaching, and public outreach. These sites represent nearly every major California bioregion, from the Channel Islands to the High Sierra, and from the Northwest Forest to the Mojave Desert (Figure 4). The third data collection is the California Vegetation Type Map (VTM) Project, which developed from a partnership between UC Berkeley and the U.S. Forest Service (USFS) California Forest and Range Experiment Station (now Southwest Research Station). The VTM Project mapped nearly 40 million hectares of the state's natural areas in the 1930’s (Colwell, 1977; Wieslander, 1961). The full VTM collection includes detailed vegetation maps, floristic and environmental plot data, landscape photographs, maps showing photographer vantage point and record locations, and herbarium specimens for species recorded on vegetation maps and sample plots (Figure 5).


AGRONOMY EXPERIMENTS

Twenty-four, seven foot square exclosures were erected in Vasser, Foster, South, and Huntley peak pastures in February and March. Cuttings were made of equal areas inside and outside exclosures in July. Clippings sent to Davis for analysis. In November and December the exclosures were all moved to different locations in the same pastures. Also added were twelve exclosures in Lambing and Buck pastures in November making a sum total of thirty-six exclosures.

Nitrogen and phosphorus fertilizers were added to Little Buck pasture in March for forage growth increase.

A nitrogen-phosphorus overlay strip approximately 900 feet long was applied in lower Vasser field. In July, September, and December the following clovers were broadcast seeded at 2½ pounds per acre rate: Mt.Barker Sub, Tallarook Sub, Crimson, and Rose. Half of the each seeding time was treated with Arasan SF.

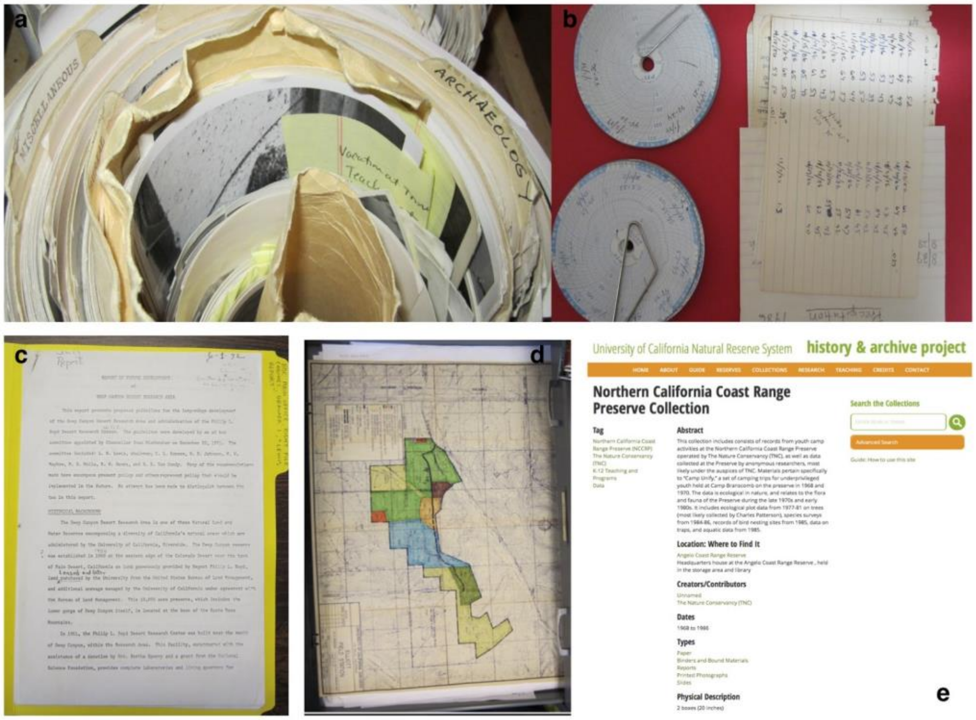
Applied an exploratory plot of the trace elements B, Cu, Zn in the Vasser Sudan planting in May.

Meduse-head studies

Burnt infestations of Medusa-head in Cattleguard, Mi and Foster pasture with help of California Division of Fo standby crew with Furbush and Wolfram in charge.

Later sprayed other infestations of Medusa -head in Horse, Middle, Foster, Buck, and Vassers pastures. Used ing spray mixtures: diesel oil and pentachlorophenol; d oil and chloro IPC; and straight diesel oil. Spray appli seedhead had matured. Seedheads from treated areas were collected.

Figure 3 Panel of images from the REC project: (a) example of an original research report; (b) Web mapping platform providing access to the original scanned PDF documents, searchable by REC, keywords, and date; (c) example of image from an original research report showing a fuel reduction experiment; (d) timeline visualization of researchers and topics conducted on Hopland REC; (e) visualization of keywords extracted from research reports screenshot of the interactive website



Northern California Coast Range Preserve Collection

Tag

Northern California Coast Range Preserve (NCCRP)
The Nature Conservancy (TNC)
K-12 Teaching and Programs
Data

Abstract

This collection includes consists of records from youth camp activities at the Northern California Coast Range Preserve operated by The Nature Conservancy (TNC), as well as data collected at the Preserve by anonymous researchers, most likely under the auspices of TNC. Materials pertain specifically to "Camp Unify," a set of camping trips for underprivileged youth held at Camp Branscomb on the preserve in 1968 and 1970. The data is ecological in nature, and relates to the flora and fauna of the Preserve during the late 1970s and early 1980s. It includes ecological plot data from 1977-81 on trees (most likely collected by Charles Patterson), species surveys from 1984-86, records of bird nesting sites from 1985, data on traps, and aquatic data from 1985.

Location: Where to Find It

Angelo Coast Range Reserve
Headquarters house at the Angelo Coast Range Reserve , held in the storage area and library

Creators/Contributors

Unnamed
The Nature Conservancy (TNC)

Dates

1968 to 1986

Types

Paper
Binders and Bound Materials
Reports
Printed Photographs
Slides

Physical Description

2 boxes (20 inches)

Figure 4 Panel of images from the NRS project: (a) damaged photographic prints at Sweeney Granite Mountains Desert Research Center; (b) historical climate data. Administrative records with land use data from Boyd Deep Canyon Desert Research Center; (e) maps from the UC San Diego NRS campus office; and (f) a sample records collection description on the NRSHAP website.

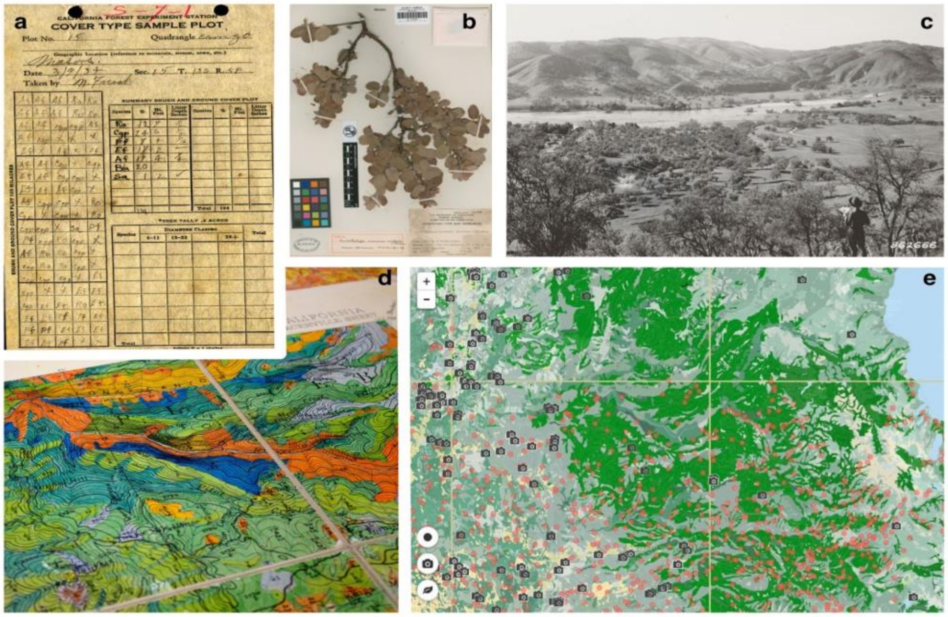


Figure 5 Examples of the components of the VTM collection: (a) plot card recording floristic species; (b) herbarium specimen (Arctostaphylos morroensis, San Luis Obispo Co.); (c) a landscape photograph San Antonio River, Monterey Co. 1938; (d) vegetation type map (Placer Co.); and (e) the digital representation (vtm.berkeley.edu) of maps, plots, and photographs, showing part of Lake Tahoe. In (e) the background colors represent different vegetation types, red dots are plot locations, and black icons are photograph locations.

These three case studies exemplify the data-related problems and opportunities of long-term sites for place-based learning (Alagona and Paulson, 2018). Because of their unique initiatives as centers of science and experimentation, these sites and projects can provide qualitative and quantitative information on human-natural interactions for over a century (Erb et al., 2016; Watson et al., 2014). However, each of these places are sites where valuable data are dark due to lack of infrastructure, incentive, and investment (National Research Council et al., 2014). The three case studies examined take different approaches to digitizing distributed datasets: one data-driven and led by ecologists and geographers, one metadata-focused and led by historians and archivists, and one object-driven and led by administrators and data scientists. While the approaches taken in digitization were different for each project, determined by expertise and project goal, all of the data within these collections were at risk of loss or destruction; at risk of staying dark. In this way, these cases exemplify varying levels of success in achieving data interoperability and moving the data collection out of the dark and into the modern digital data landscape.

2.1 Case Study 1: Creating an object-based digital collection

2.1.1 Background and need

The ANR RECs project originated out of a pressing need to digitize routine research documents (annual reports, project proposals, and annual project summaries) prior to their physical destruction. Each REC had accumulated large volumes of these documents, and the need for space drove a rapid preservation effort, in which all documents were scanned.

2.1.2 Methods

During the project, a single staff member traveled to each REC and used a digital scanner to digitize all available paper documents (total 3,152) as PDF files. Each one of these documents was stored in a SQL relational database (a database that implements a structured query language to manage the data within it) and given a unique article number, title, coordinates associated with the REC it was retrieved from, year, and URL of the digital document. To make these documents findable to the broader research community, an interactive web application using the ESRI (*ESRI ArcGIS Desktop*, 2017) web application stack was created (http://igis.ucanr.edu/Infobase/InfobaseExplorer/). The interactive map-based user interface enabled a spatial representation of the entire document repository and allows for simple queries of information within the database. The web application displayed and made the documents discoverable and allowed users to find and download scanned PDFs.

The documents were scanned using a Fujitsu fi-6140Zdj scanner at 300 dpi (Figure 3a) and run through Adobe Acrobat Professional 11.0 optical character recognition (OCR) tools. The resulting extracted character string was also stored in the database. OCR enables the conversion of images of typed, handwritten, or printed text into machine-readable format text (Holley, 2009) enabling search, storage, display and analysis. The project improved upon the original OCR with an automated workflow by testing several programmatic options on an initial site (Hopland REC, n=564 documents).

The project chose a machine learning (ML) approach, using a series of Natural Language Processing (NLP) tools to extract information on key people, organizations, topics, and scientific keywords from the scanned documents. NLP is an area of computer science focused on training computers to process large collections of written texts. This project developed a machine learning pipeline using NLP techniques to interpret the scanned documents when trained with common subject matter libraries of agricultural, biological and ecological text. We investigated several tools to improve our ability to extract information from the documents including Ocular , a tool that uses unsupervised learning methods to recognize text from scanned historical documents including opaque text (Berg-Kirkpatrick et al., 2013), and Tesseract, a Google-funded open-source OCR system. Tesseract yielded better results compared to baseline OCR text. The output text strings were analyzed with Alchemy, a NLP tool based on IBM's artificial intelligence Watson machine, which uses deep machine learning algorithms to analyze massive amounts of structured and unstructured content. Critical information (e.g. keywords, organizations, people, and scientific topics) from the processed OCR strings were extracted and visualized by year and document type using the D3 javascript library.

2.1.3 Data uses

This project has primarily served internal, custodial goals, and although there has been little external use of the data, the project was able to mine the preserved documents to capture key data on researchers, research projects, scientific concepts, and keywords over a 60-year period at Hopland REC through the ML process described above. For example, we were able to extract a summary of researchers and research conducted at Hopland annually from 1951- present (e.g. “animal science”, “pasture forage”, “spring fertilizer application”, “herbage production”, “biodiversity”). This kind of information, derived automatically from scanned documents, can assist future researchers to find related data for their own projects. This information is also valuable for tracking and understanding the evolution of research and science at the RECs and the intensity, scope, scale, and frequency of management actions taken at each site. Documenting past research and management treatments is needed to understand implications for ongoing and future research projects. Making data findable and accessible to the broader research community would greatly increase the success of this preservation effort and now that it is digitally captured can be ingested into existing repositories with a wider reach such as the Biodiversity Heritage Library.

2.2 Case Study 2: Creating a digital metadata archive

2.2.1 Background and need

The UC Natural Reserve System History and Archive Project (NRSHAP) represents an initial effort to preserve the historical materials of the NRS and promote their use for research and education in both science and history. NRSHAP operated for 7 years (2011-2017) with funding from the National Science Foundation and UC to identify, index, preserve, and promote historical records held on or pertaining to the NRS. NRSHAP was led by historians and archivists who adapted standard archival protocols to non-traditional sites (Society of American Archivists, 2013; The University of Chicago, 2006; Young, 2006). The NRS field station historical datasets come in diverse formats

(Figure 4) therefore, the goal of the project was to provide enough information about the existing records that potential researchers may identify data types and provide for their own use of the materials on site rather than develop individualized digitization workflows. The outcome was a digital metadata archive. Metadata is commonly defined as “data about data” and can be used to locate, describe, and retain data provenance. NRSHAP is a unique document archiving project because of its spatial coverage including 39 field stations and reserves, 8 campus offices, one system-wide administrative office, several independent archives, and personal collections across the state of California.

2.2.2 Methods

NRSHAP developed a multi-step data-preservation method for field stations and other remote or dispersed organizations/sites with potential archives. Initially, NRSHAP distributed questionnaires to all site contacts to assess the scope of potential historical records. this questionnaire was followed by extensive research into the known history of the NRS and its sites. NRSHAP teams travelled to each site and conducted a records inventory following established archival methods (Society of American Archivists, 2013; The University of Chicago, 2006; Young, 2006).

The inventory was divided into either “active records”—still in use for the regular operation of the station— and “inactive records”—no longer used but still of value. The inactive records were then grouped into collections and information on the physical location, the creators, date, physical material types, the arrangement (by subject, chronological) and the physical description of the size of the collection (e.g. linear feet) was tagged and used to create an archival collection description or “finding aid” following established standards (Society of American Archivists, 2013). This information was documented in a database and published online (http://archives.nrs.ucsb.edu/content/research) creating a metadata archive. However, not all records pertaining to the NRS existed on site, and throughout the years various material types and collections were sent to various institutions. NRSHAP researchers identified other collections relating to the NRS history that were held in other archives (e.g. Bancroft Library); affiliate organization offices such as State Parks; and the personal offices and homes of past staff and researchers and linked to these existing collections. Therefore, published descriptions of the field station archives may also sometimes be found on existing archival networks and search engines, such as the Online Archive of California or Archivegrid. NRSHAP made recommendations to station managers regarding the best means of preserving and promoting their historical collections. Preparing for this involved meeting with potential institutions and repositories across California regarding their interest in acquiring and managing NRS materials. Finally, NRSHAP developed a plan for regular review of metadata accuracy, document health (if still held on site), and ongoing off-site research for relevant collections.

2.2.3 Data uses

When NRSHAP began, the project was on the cusp of a broader awakening among scientific researchers and field station managers to the potential of historical documents or dark data. Along the way we encountered lots of support and encouragement from people invested in the NRS system, but many also expected the effort to involve digitization of the records themselves. Historians have used archives as their primary data method for almost two hundred years, but archival research methodologies are mostly project-specific and have never been standardized or fully articulated. Potential data users should not see this as a hindrance, but an opportunity, since archival methods are flexible and can be adapted for inclusion in projects involving other types of data collection and analysis. Morrison et al. (2017) argue that these kinds of connections will be necessary for the future of ecology. NRSHAP bridged epistemological and methodological divides across disciplines to create new opportunities for more robust research and collaborations. Metadata archives hold great potential for data reuse. Projects that reuse dark historical data are effectively collecting new, unknown data because these are not part of existing records, and they are able to collect that data in an environment that we need to understand but which is gone from us—the past. However, the success of field station metadata archives will inevitably rely on targeted and continued efforts promoting use of the archive itself and educating those on best practices once it has been created.

NRSHAP affiliates have promoted use of the metadata archive by speaking at NRS system-wide meetings and academic conferences, using the website as a teaching tool in undergraduate classes, and conducting their own research projects. NRSHAP has attracted interest from researchers across the UC system and is already being used by one, ongoing international collaborative research project.

2.3 Case Study 3: Creating a completely digitized data collection

2.3.1 Background and need

The Wieslander Vegetation Type Mapping (VTM) collection, named after director Albert Wieslander, was an exhaustive and detailed effort to map California land cover in the early 20th century. During the 1920-30s, VTM crews surveyed 16 million ha (40%) of California's wildlands. They collected vegetation information at over 18,000 plots, produced detailed maps of dominant vegetation for over 100,000 km^2, gathered over 23,000 herbarium specimens, and took over 3,000 photographs (Colwell, 1977; Kelly et al., 2016). Until recently, the full collection was distributed throughout libraries and labs statewide. Significant, and partly unknown, portions of the collection were lost to custodial needs and competing collections' demand on space (Kelly et al., 2016). Overt risk of loss, combined with the tremendous depth of content, provided the impetus for many individual digitization efforts across the state, which eventually combined in the early 21st century (Kelly et al., 2016).

2.3.2 Methods

The digitization of the vegetation maps, plots, plot maps, photographs, and locations of herbarium specimens took place over a decade, in several UC labs and libraries including the Marian Koshland Biosciences Library and the Museum of Vertebrate Zoology (MVZ). Linework from the vegetation maps was manually digitized and polygon values linked to a spatial database; plot data were transcribed manually and joined to plot locations which were manually digitized from plot location maps; photographs were scanned and where possible attributed with a geographic location; and herbarium specimens were georeferenced using analog accompanying information (Kelly et al., 2017, 2016, 2008).

All digital VTM data (geographic, ecological, and photographic) were stored using PostgreSQL, a relational database that supports the storage, analysis, and transfer of geospatial vector data through a PostGIS extension (Kelly et al., 2016). The data itself is downloadable as standard text and spatial data formats that can be used in numerous GIS and statistical software packages. An interactive web map interface was built for exploring, searching, aggregating, and downloading the VTM data collection using Leaflet (a JavaScript mapping library for web mapping) and Open Street Map base layers. The VTM website was built using the HOLOS API from which the VTM data is linked to the structured digital database and allows for analysis of raw data, integration with contemporary data, as well as rapid interaction and visualization (Dolanc et al., 2013; Easterday et al., 2016; McIntyre et al., 2015; Thorne et al., 2008).

2.3.3 Data uses

The VTM data has been used since the mid-20th century, and its digitization has increased the scope and scale of the types of analysis performed(Kelly et al., 2016). The vegetation data found in the plot database have been used to develop vegetation classification schemes and to examine changes to chaparral and forest communities around the state enabling prediction of community structure and shifts under a changing climate. The vegetation maps have been used to document regional changes in vegetation communities, to investigate legacies of land use change, and to support land planning.

3 Discussion

3.1 Evaluation of cases with FAIR principles

These three cases provide different protocols for data digitization, and we evaluate them here with respect to the FAIR principles and summarized our findings in Table 1. The REC collections of historical research documents were completely digitized via scanning and OCR and made available via the web. These data “objects” were *findable* as text and text strings through simple database web searches. Several machine learning (ML) algorithms were used to reconstruct context and make the data *accessible*, however, the digitization process did not result in interoperable or reusable data because the data remained unstructured. The major advantage of this approach was speed: the complete collection of physical records can be made digital with limited technical skill and made available to a broader audience.

NRSHAP focused on making physical objects *findable* through metadata, and *accessible* as the metadata contained essential instructions for finding the data. The data itself remained on site in curated and semi-curated collections. The major limitation of the metadata archive approach is the limited access to the data itself. The data can now be discovered but requires further investment to use. The VTM project provides an example of a completely digitized data collection that reaches all the benchmarks of the FAIR standard. Data is *findable and accessible* through links from several data repositories, through an API and as part of a larger data landscape supported by HOLOS; data is *interoperable* as it is stored in standard spatial data file formats that can be used easily in most common spatial analysis and statistical software with updated nomenclature to be readily used in conjunction with contemporary species and vegetation codes; and data is *reusable* because the digitization methodology and data provenance are fully documented.

Table 1 Evaluation of three case studies according to FAIR data management principles; ✔= successful; ✓ = partially successful; and ✗ = not successful.

FAIR Principle	UC REC	NRSHAP	VTM
Findable	✔ Information preserved as digital objects; and available via web.	✔ Archive captures metadata and physical location of data collection.	✔ Findable through links to other data repositories, and API.
Accessible	✓ Not listed in any general repository. ML algorithms used to mine text.	✓ Archive is publicly available online and contains instructions for further research into any of the dispersed collections	✔ Linked to API via Holos, and part of a larger data landscape.
Interoperable	✘ Data remains unstructured. No external access to OCR output.	✘ No effort made to update or migrate data into contemporary digital format. Metadata formats not compatible with other generalist repositories	✔ Data is stored in standard GIS file formats that can be used easily in spatial analysis. Updated to current taxonomy and linked with common standardized vegetation classifications.
Re-usable	✘ Data still unstructured. Captured only objects, not context.	✓ Original order, context, and media format of records is preserved	✔ Data is fully digitized, available for download. Context and data provence preserved.

Analysis of these case studies finds that FAIR is a valuable tool for data preservation planning and evaluation, though not all projects will accomplish FAIR fully. Making datasets *findable* and *accessible,* alone, creates awareness, but is often insufficient to ensure data reuse and longevity. All of these projects faced challenges, yet they all ultimately increased the potential and value of the datasets through their efforts. For example, in our first case, some success was achieved in resurrecting critical components of the historic scientific record at the RECs, and this information was shared via a web application. However, the workflow in extracting value-added information from the documents was not without flaw and most of the information therefore remains unstructured in a non-machine-readable format. Efforts that span the entire FAIR process require diverse skill sets and multidisciplinary teams with some combination of computer scientists, data scientists, ecologists, historians, librarians, land managers, and web developers working together. Indeed, all our cases required input from multidisciplinary teams. When the FAIR process is completed, data can be used in unexpected ways, making valuable, transdisciplinary analysis possible (Kelly et al., 2016). In the case of VTM, there was a documented increase in the scope and scale of research conducted with the dataset once it was made digitally available (Kelly et al., 2016).

However, since there is often a time lag between digitizing, sharing, and use of a data collection, management requires long term stewardship. Each approach dealt with collections of heterogeneous materials that did not readily fit into existing repositories. Rather than separating the collections, individualized databases were created to host and make the materials accessible. In this way dark data from small projects gained recognition and use amongst the immediate research community, but the reach remains limited (Van Noorden, 2013). Potentially mirroring key parts of a collection- such as field notebooks or biological specimens- that have readily recognized repositories can increase this reach but this risks loss of data provenance and potentially reduces the use of other materials from the same collection. Balancing these risks requires careful planning.

Not all collections lend themselves to traditional data digitization. As shown in the case of NRSHAP, the metadata archive approach works well for field stations because it is designed for geographically remote or distributed data collections; it provides for either internal hosting or third-party data-management options as appropriate; it serves both data promotion and long-term data management; and it does not usually require reorganization of documents. The metadata archive approach focuses on the kind of information that both outside researchers and station staff need to be able to reuse existing historical datasets. It does not require the archivist to know or anticipate the character or media format of future scientific reuses. Further, some speculate that physical archives are *the best way to save information*, since physical materials (even under threat of pests, mold, paper acids, and natural disaster) have a much longer shelf life than any known digital forums, and they will remain legible to human eyes long after advancing computer technologies make current digital information obsolete (Clement et al., 2013; Klein, 2008; Scott, 2007; Wright, 2014). However, redundant collections of both physical and digital renditions are best practice. And finally, automated approaches to digitization do not always save time in the long run, since considerable human intelligence might be required to ensure data is fully *interoperable* and *reusable*. As demonstrated in the ANR case, historic documents can be difficult to digitize meaningfully. Uneven typesets, faded ink, and handwriting all pose common and serious obstacles to automated information retrieval.

The FAIR principles provide flexible guidelines for the stewardship of heterogeneous data types, yet do not address the need to first make historical data digitally discoverable. Sharing examples of how historical dark data is made digital, and then FAIR will lead to an exchange of successful protocols that may lead to eventual standards. Developing standards and ontologies is paramount to the interoperability and reuse of all data (Jones et al., 2006), but is largely lacking for historical dark data. Adopting contemporary data science standards, such as FAIR, for historical data will help to integrate historical and contemporary data, but the high standards of “open data” should not preclude preservation of historical data. Primarily, the first two principles -

findable and accessible should have scalar and adaptive rules that are relative to the project's goals, the different stages in which data are created, and to the overarching goal of creating maximum potential use. For example, none of the projects succeeded in assigning the collections persistent identifiers (PIs) including Digital Object Identifiers (DOIs) or Archival Resource Keys (ARKs) that would make them findable the way FAIR is defined. Each of these projects understood that FAIR must be relative to the quality of data, the resources at hand, the projects goal, and the communities' standards. These three case studies made their collections more findable and accessible to their immediate research communities including those presently most interested in using and reusing the data, yet each given the time and resources would open these data collections to a much broader research community and create potential for further discovery.

4 Conclusions

Comprehending the temporal and human dynamics of ecosystems is a central challenge of science in the Anthropocene (Safford et al., 2012; Robin and Steffen, 2007). This requires synthesis and sharing of transdisciplinary, heterogeneous datasets over long time periods and large spatial scales (Kelly et al., 2016; Lowndes et al., 2017). Within the last half century, the increasingly large streams of data from sensor networks, mobile technology, and remote sensing has created both opportunities, "big data" (Gandomi and Haider, 2015), and challenges, "data deluge" (Hampton et al., 2013; Porter et al., 2012), establishing the need for better data science workflows and training across most disciplines. Often overlooked in this discussion are the large majority of scientific data that are created by small research groups with limited resources for data planning and management (Hampton et al., 2013). The majority of scientific data potentially available for future research and synthesis never make it into a discoverable repository and remain inaccessible to the broader community (Heidorn, 2008). Without proper incentive, support, or standards in place to consistently capture data and make it accessible, it often goes "dark" - limiting the scope and potential of scientific research.

Historical data are vital to current ecological research: they provide benchmarks from which to compare change, they can be linked to modern ecological data to create new knowledge, and they can be modeled to help predict future changes and validate models. We argue that these data are "dark" until they are effectively digitized and made discoverable to a wider audience. In the strictest definition, dark data is unstructured, untagged and untapped data that is created through routine activities yet has not been analyzed or processed. Dark data is increasingly recognized in business and economics as vulnerable, underutilized, and valuable (Heidorn, 2008), and we argue that the same is true of historical data for science.

Achieving successful sharing of historical data can be difficult and time consuming, since these collections are often analog, unstructured, and physically distributed. Our review of three novel approaches to digitizing historical field data showcase some of these challenges. We evaluated each approach with respect to the FAIR principles (findable, accessible, interoperable and reusable) (Wilkinson et al., 2016), and revealed both the value of the framework and its limitations in practice. The most effective digitization projects demand lots of human, technical, and capital resources. Making datasets *findable* and *accessible* is a necessary first step to creating demand, but not sufficient to ensure data reuse and longevity. Second, efforts that span the entire FAIR process require diverse skill sets and transdisciplinary work often with some combination of land stewards, ecologists, historians, librarians, archivist, data scientists, computer scientists, and web developers working together.

An encouraging antidote to the challenges facing those working to digitize historical data can be the foresight provided by leaders of early 20th century field data collection. Joseph Grinnell, founder of MVZ and a preeminent field scientist of the day, wrote of his own preservation efforts: “After the lapse of many years, possibly a century, the student of the future will have access to the original record of faunal conditions in California” (Grinnell, 1910). Potential use of Grinnell’s and others’ data only grows as technologies increase to repurpose the data to answer questions unimagined at the time of their collection (Morrison et al., 2017).

We, as a global scientific community, have the responsibility to continue to shed light on historical data through digitization, adding scientific knowledge, strengthening cultural heritage, and increasing public good. Moving data from file drawers to the cloud will require a transdisciplinary exchange of tools, technology, and methodologies at the intersection of data science, history, ecology, and ecoinformatics. Field research and research reserves are not only major producers and repositories of scientific data, but also can be key agents in making data shareable for researchers and the public. Thus, field stations, research reserves, and field data projects are critical nodes in the nexus of big and dark data: enlarging and enriching a growing data landscape. Going forward capturing the intellectual infrastructure from these sites will require systematic investment, strategy, and leadership to preserve and maintain ecological records for future generations. Envisioning a future for historical data will also require an exchange of tools, technology, methodology, and transdisciplinary work at the intersection of data science, history, ecology, and ecoinformatics, and is a vision that if achieved ensures that future generations have access to the past.

5 References

Alagona, P., Paulson, T., 2018. From the Classroom to the Countryside: The University of California’s Natural Reserve System and the Role of Field Stations in American Academic Life, in: Invited Paper for the Dumbarton Oaks Series in Garden & Landscape Studies. Forthcoming Volume Entitled: Landscape and the Academy.

Alagona, P.S., Sandlos, J., Wiersma, Y.F., 2012. Past Imperfect: Using Historical Ecology and Baseline Data for Conservation and Restoration Projects in North America. Environmental Philosophy 9, 49–70.

Beller, E., McClenachan, L., Trant, A., Sanderson, E.W., Rhemtulla, J., Guerrini, A., Grossinger, R., Higgs, E., 2017. Toward principles of historical ecology. Am. J. Bot. 104, 645–648.

Berg-Kirkpatrick, T., Durrett, G., Klein, D., 2013. Unsupervised transcription of historical documents, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 207–217.

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R., Moritz, C., 2013. Unlocking the vault: next-generation museum population genomics. Mol. Ecol. 22, 6018–6032.

Borgman, C.L., Darch, P.T., Sands, A.E., Pasquetto, I.V., Golshan, M.S., Wallis, J.C., Traweek, S., 2015. Knowledge infrastructures in science: data, diversity, and digital libraries. International Journal on Digital Libraries 16, 207–227.

Bradley, R.D., Bradley, L.C., Garner, H.J., Baker, R.J., 2014. Assessing the Value of Natural History Collections and Addressing Issues Regarding Long-Term Growth and Care. Overview Articles BioScience 64. https://doi.org/10.1093/biosci/biu166

Bürgi, M., Gimmi, U., 2007. Three objectives of historical ecology: the case of litter collecting in Central European forests. Landsc. Ecol. 22, 77–87.

Chornesky, E.A., Ackerly, D.D., Beier, P., Davis, F.W., Flint, L.E., Lawler, J.J., Moyle, P.B., Moritz, M.A., Scoonover, M., Byrd, K., Alvarez, P., Heller, N.E., Micheli, E.R., Weiss, S.B., 2015. Adapting California's Ecosystems to a Changing Climate. Bioscience 65, 247–262.

Clement, T., Hagenmaier, W., Knies, J.L., 2013. Toward a Notion of the Archive of the Future: Impressions of Practice by Librarians, Archivists, and Digital Humanities Scholars. Libr. Q. 83, 112–130.

Colwell, W.L., Jr., 1977. The status of vegetation mapping in California today.

Connors, J.P., Lei, S., Kelly, M., 2012. Citizen Science in the Age of Neogeography: Utilizing Volunteered Geographic Information for Environmental Monitoring. Ann. Assoc. Am. Geogr. 102, 1267–1289.

Crompvoets, J., Bregt, A., Rajabifard, A., Williamson, I., 2004. Assessing the worldwide developments of national spatial data clearinghouses. Int. J. Geogr. Inf. Sci. 18, 665–689.

Dickinson, J.L., Shirk, J., Bonter, D., Bonney, R., Crain, R.L., Martin, J., Phillips, T., Purcell, K., 2012. The current state of citizen science as a tool for ecological research and public engagement. Front. Ecol. Environ. 10, 291–297.

Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010. Citizen Science as an Ecological Research Tool: Challenges and Benefits. Annu. Rev. Ecol. Evol. Syst. 41, 149–172.

DiEuliis, D., Johnson, K.R., Morse, S.S., Schindel, D.E., 2016. Specimen collections should have a much bigger role in infectious disease research and response. PNAS 113. https://doi.org/10.1073/pnas.1522680112

Dolanc, C.R., Thorne, J.H., Safford, H.D., 2013. Widespread shifts in the demographic structure of subalpine forests in the Sierra Nevada, California, 1934 to 2007. Glob. Ecol. Biogeogr. 22, 264–276.
Downing, J., 2016. Sierra Foothill REC: Quantifying IPM benefits in rangeland systems. Calif. Agric. 70, 174–174.

Easterday, K.J., McIntyre, P.J., Thorne, J.H., Santos, M.J., Kelly, M., 2016. Assessing Threats and Conservation Status of Historical Centers of Oak Richness in California. Urban Planning 1, 65–78.

Eitzel, M.V., Kelly, M., Dronova, I., Valachovic, Y., Quinn-Davidson, L., Solera, J., de Valpine, P., 2016. Challenges and opportunities in synthesizing historical geospatial data using statistical models. Ecol. Inform. 31, 100–111.

Elizabeth Griffin, R., 2015. When are Old Data New Data? GeoResJ 6, 92–97.

Erb, K.-H., Luyssaert, S., Meyfroidt, P., Pongratz, J., Don, A., Kloster, S., Kuemmerle, T., Fetzel, T., Fuchs, R., Herold, M., Haberl, H., Jones, C.D., Marín Spiotta, E., McCallum, I., Robertson, E., Seufert, V., Fritz, S., Valade, A., Wiltshire, A., Dolman, A.J., 2016. Land management: data availability and process understanding for global change studies. Glob. Chang. Biol. https://doi.org/10.1111/gcb.13443

ESRI ArcGIS Desktop, 2017. Environmental Systems Research Institute, Redlands, CA.

Ferguson, A.R., Nielson, J.L., Cragin, M.H., Bandrowski, A.E., Martone, M.E., 2014. Big data from small data: data-sharing in the "long tail" of neuroscience. Nat. Neurosci. 17, 1442–1447.

Fiedler, P.L., Rumsey, S.G., Wong, K.M., 2013. The Environmental Legacy of the UC Natural Reserve System. University of California Press.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2015. How “big data” can make big impact: Findings from a systematic review and a longitudinal case study. Int. J. Prod. Econ. 165, 234–246.

Franklin, J., Serra-Diaz, J.M., Syphard, A.D., Regan, H.M., 2017. Big data for forecasting the impacts of global change on plant communities. Glob. Ecol. Biogeogr. 26, 6–17.

Gandomi, A., Haider, M., 2015. Beyond the hype: Big data concepts, methods, and analytics. Int. J. Inf. Manage. 35, 137–144.

Grinnell, J., 1910. The Methods and Uses of a Research Museum. Popular Science Monthly 77.

Hampton, S.E., Strasser, C. a., Tewksbury, J.J., Gram, W.K., Budden, A.E., Batcheller, A.L., Duke, C.S., Porter, J.H., 2013. Big data and the future of ecology. Front. Ecol. Environ. 11, 156–162.

Heidorn, P.B., 2008. Shedding Light on the Dark Data in the Long Tail of Science. Libr. Trends 57, 280–299.

Holley, R., 2009. How good can it get? Analysing and improving OCR accuracy in large scale historic newspaper digitisation programs. D-Lib Magazine 15.

Holmes, M.W., Hammond, T.T., Wogan, G.O.U., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., Nachman, M.W., 2016. Natural history collections as windows on evolutionary processes. Mol. Ecol. 25, 864–881.

Hugh Safford, Gregory Hayward, Nicole Heller, John Wiens, 2012. Historical Ecology, Climate Change, and Resource Management: Can the Past Still Inform the Future? in: John Wiens, Gregory Hayward, Hugh Safford, Catherine Giffen (Ed.), Historical Environmental Variation in Conservation and Natural Resource Management. John Wiley & Sons Ltd., pp. 46–62.

Jones, M.B., Schildhauer, M.P., Reichman, O.J., Bowers, S., 2006. The New Bioinformatics: Ecological Data Integrating the Gene to the from Biosphere. Annu. Rev. Ecol. Evol. Syst. 37, 519–544.
Kearns, F.R., Kelly, M., Tuxen, K.A., 2003. Everything Happens Somewhere: Using WebGIS as a Tool for Sustainable Natural Resource Management. Front. Ecol. Environ. 1, 541.

Kelly, M., Easterday, K., Koo, M., Thorne, J.H., Mukythar, S., Galey, B., 2017. Geospatial Informatics Key to Recovering and Sharing Historical Ecological Data for Modern Use. Photogrammetric Engineering & Remote Sensing 83, 779–786.

Kelly, M., Easterday, K., Rapicullio, G., Koo, M., McIntyre, P., Thorne, J., 2016. Rescuing and Sharing Historical Vegetation Data for Ecological Analysis: The California Vegetation Type Mapping Project. Biodivers. Inf. 11, 40–62.

Kelly, M., Ueda, K.-I., Allen-Diaz, B., 2008. Considerations for ecological reconstruction of historic vegetation: Analysis of the spatial uncertainties in the California Vegetation Type Map dataset. Plant Ecol. 194, 37–49.

Klein, D. (Ed.), 2008. Identifying Museum Insect Pest Damage. National Park Service.

Lavoie, C., 2013. Biological collections in an ever-changing world: Herbaria as tools for biogeographical and environmental studies. Journal of PPEES Sources 15, 68–76.

Lowndes, J.S.S., Best, B.D., Scarborough, C., Afflerbach, J.C., Frazier, M.R., O'Hara, C.C., Jiang, N., Halpern, B.S., 2017. Our path to better science in less time using open data science tools. Nat Ecol Evol 1, 160.

Maguire, D.J., Longley, P.A., 2005. The emergence of geoportals and their role in spatial data infrastructures. Comput. Environ. Urban Syst. 29, 3–14.

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D. a., Scott, E.M., Smith, R.I., Somerfield, P.J., Watt, A.D., 2010. Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. Trends Ecol. Evol. 25, 574–582.

McClenachan, L., Cooper, A.B., McKenzie, M.G., Drew, J.A., 2015. The Importance of Surprising Results and Best Practices in Historical Ecology. Bioscience 65, 932–939.

McIntyre, P.J., Thorne, J.H., Dolanc, C.R., Flint, A.L., Flint, L.E., Kelly, M., Ackerly, D.D., 2015. Twentieth-century shifts in forest structure in California: denser forests, smaller trees, and increased dominance of oaks. Proceedings of the National Academy of Sciences 112, 1458–1463.

Michener, W.K., 2015. Ecological data sharing. Ecol. Inform. 29, 33–44.

Michener, W.K., 2006. Meta-information concepts for ecological data management. Ecol. Inform. 1, 3–7.

Michener, W.K., Brunt, J.W., 2009. Ecological Data: Design, Management and Processing. John Wiley & Sons.

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecol. Appl. 7, 330–342.

Michener, W.K., Jones, M.B., 2012. Ecoinformatics: supporting ecology as a data-intensive science. Trends Ecol. Evol. 27, 85–93.

Michener, W.K., Porter, J., Servilla, M., Vanderbilt, K., 2011. Long term ecological research and information management. Ecol. Inform. 6, 13–24.

Morrison, S.A., Sillett, T.S., Funk, W.C., Ghalambor, C.K., Rick, T.C., 2017. Equipping the 22nd-Century Historical Ecologist. Trends Ecol. Evol. 32, 578–588.

National Research Council, Division on Earth and Life Studies, Board on Life Sciences, Committee on Value and Sustainability of Biological Field Stations, Marine Laboratories, and Nature Reserves in the 21st Century Science, Education, and Public Outreach, 2014. Enhancing the Value and Sustainability of Field Stations and Marine Laboratories in the 21st Century. National Academies Press.

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirmbacher, P., Dierolf, U., 2013. Making research data repositories visible: the re3data.org Registry. PLoS One 8, e78080.

Pedersen, B., Kearns, F., Kelly, M., 2007. Methods for facilitating web-based participatory research informatics. Ecol. Inform. 2, 33–42.

Peters, D.P.C., Havstad, K.M., Cushing, J., Tweedie, C., Fuentes, O., Villanueva-Rosales, N., 2014a. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology. Ecosphere 5, 1–15.

Peters, D.P.C., Loescher, H.W., SanClements, M.D., Havstad, K.M., 2014b. Taking the pulse of a continent: expanding site-based research infrastructure for regional- to continental-scale ecology. Ecosphere 5, 1–23.

Petit, R.J., Hu, F.S., Dick, C.W., 2008. Forests of the Past: A Window to Future Changes. Science 320.
Porter, J.H., Hanson, P.C., Lin, C.-C., 2012. Staying afloat in the sensor data deluge. Trends Ecol. Evol. 27, 121–129.

Pyke, G.H., Ehrlich, P.R., 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. Biol. Rev. Camb. Philos. Soc. 85, 247–266.

Rapacciuolo, G., Ball-Damerow, J.E., Zeilinger, A.R., Resh, V.H., 2017. Detecting long-term occupancy changes in Californian odonates from natural history and citizen science records. Biodivers. Conserv. 1–17.

Rapacciuolo, G., Maher, S.P., Schneider, A.C., Hammond, T.T., Jabis, M.D., Walsh, R.E., Iknayan, K.J., Walden, G.K., Oldfather, M.F., Ackerly, D.D., Beissinger, S.R., 2014. Beyond a warming fingerprint: individualistic biogeographic responses to heterogeneous climate change in California. Glob. Chang. Biol. 20, 2841–2855.

Reichman, O.J., Jones, M.B., Schildhauer, M.P., 2011. Challenges and opportunities of open data in ecology. Science 331, 703–705.

Robin, L., Steffen, W., 2007. History for the Anthropocene. History Compass 5, 1694–1719.

Rodrigo, A., Alberts, S., Cranston, K., Kingsolver, J., Lapp, H., McClain, C., Smith, R., Vision, T., Weintraub, J., Wiegmann, B., 2013. Science Incubators: Synthesis Centers and Their Role in the Research Ecosystem. PLoS Biol. 11, e1001468.

Salmond, A., Lythberg, B., Newell, J., 2012. Old objects, new media: Historical collections, digitization and affect. Journal of Material Culture 17, 287–306.

Scott, R.E., 2007. e-Records in health--preserving our future. Int. J. Med. Inform. 76, 427–431.

Shaffer, H.B., Fisher, R.N., Davidson, C., 1998. The role of natural history collections in documenting species declines. Trends Ecol. Evol. 13, 27–30.

Society of American Archivists, 2013. Describing Archives: A Content Standard, second edition. Society of American Archivists, Chicago, IL.

Suarez, A.V., Tsutsui, N.D., 2004. The Value of Museum Collections for Research and Society. Bioscience 54, 66.

Szabó, P., 2010. Why history matters in ecology: an interdisciplinary perspective. Environ. Conserv. 37, 380–387.

Szabó, P., Hédl, R., 2011. Advancing the integration of history and ecology for conservation. Conserv. Biol. 25, 680–687.

Tait, M.G., 2005. Implementing geoportals: applications of distributed GIS. Comput. Environ. Urban Syst. 29, 33–47.

The University of Chicago, 2006. Uncovering New Chicago Archives Project (UNCAP): Models for Library, Faculty, Student Collaboration.

The world’s most valuable resource is no longer oil, but data, 2017. The Economist.

Thorne, J.H., Morgan, B.J., Kennedy, J.A., 2008. Vegetation Change Over Sixty Years in the Central Sierra Nevada, California, USA. Madroño 55, 223–237.

Tingley, M.W., Beissinger, S.R., 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. Trends Ecol. Evol. 24, 625–633.

Van Noorden, R., 2013. Data-sharing: Everything on display. Nature 500, 243–245.

Waide, R.B., Brunt, J.W., Servilla, M.S., 2017. Demystifying the Landscape of Ecological Data Repositories in the United States. Bioscience 67, 1044–1051.

Wallis, J.C., Rolando, E., Borgman, C.L., 2013. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. PLoS One 8, e67332.

Watson, S.J., Luck, G.W., Spooner, P.G., Watson, D.M., 2014. Land-use change: incorporating the frequency, sequence, time span, and magnitude of changes into ecological research. Front. Ecol. Environ. 12, 241–249.

White, H., 2017. Lindcove REC: Developing citrus varieties resistant to huanglongbing disease. Calif. Agric. 71, 18–20.

Wieslander, A.E., 1961. California's vegetation maps: recent advances in botany. University of Toronto Press, Toronto.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018.

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., Teal, T.K., 2017. Good enough practices in scientific computing. PLoS Comput. Biol. 13, e1005510.

Wolkovich, E.M., Regetz, J., O'Connor, M.I., 2012. Advances in global change research require open science by individual researchers. Glob. Chang. Biol. 18, 2102–2110.

Wright, D.J., 2016. Toward a digital resilience. Elem Sci Anth 4.

Wright, J., 2014. Paper vs. Electronic: The Not-So-Final Battle [WWW Document]. Smithsonian Institution Archives. URL https://siarchives.si.edu/blog/paper-vs-electronic-not-so-final-battle (accessed 3.11.18).

Young, S.F., 2006. Don't Throw It Away! Documenting and Preserving Organizational History. Special Collections Department, The University Library, and Jane Addams Hull-House Museum, College of Architecture and the Arts, The University of Illinois at Chicago.

Zu Ermgassen, P.S.E., Spalding, M.D., Blake, B., Coen, L.D., Dumbauld, B., Geiger, S., Grabowski, J.H., Grizzle, R., Luckenbach, M., McGraw, K., Rodney, W., Ruesink, J.L., Powers, S.P., Brumbaugh, R., 2012. Historical ecology with real numbers: past and present extent and biomass of an imperilled estuarine habitat. Proc. Biol. Sci. 279, 3393–3400.

Chapter 5: Conclusions and future research directions

The goal of this dissertation was twofold: 1) to explore the linkages between human activity and landscape change in the context of land ownership, urbanization, and changes to the distribution and structure of California's forests and woodlands, and 2) to understand and evaluate the technical issues of data availability and data aggregation when historical data are used in conjunction with contemporary empirical data in modern ecological analysis. The challenges for natural resource management and planning in the 21st century are complex and require both a long-term perspective and the evaluation of multiple contributing mechanisms across a socio-ecological context to generate robust policy and management mandates. In California, both forest and woodland ecosystems provide a number of functions that are critical in sustaining economic markets, livelihoods, critical ecosystem processes. Thus understanding, modeling, predicting, and managing these ecosystems to be resilient under future change is critical across several fronts and requires an integrated long-term perspective.

Chapter Two is motivated by several recent studies (McIntyre et al. 2015; J. a. Lutz, van Wagtendonk, and Franklin 2010; Flint et al. 2013) that reported climate (i.e. climate water deficit (CWD)) as the primary mechanism of large tree decline and changes in forest structure in California in the 20th century. Reflecting on these studies and other conflicting opinions of primary drivers of change, I found that very few studies have quantified the impact of land ownership and land management on the quantities of large trees and other characteristics of forest structure. Land ownership has been used to understand the long-term effects of and variation in land management practices; especially when spatially explicit data on management practices are unavailable or incomplete. Thus, in Chapter Two I explicitly investigated 20th century changes in forest structure across six ownership classes in California. In comparing historical and contemporary forest structural data I found that declines in large trees and increases in small tree density were consistent across the state, irrespective of ownership boundaries. However, there were important differences in the magnitude of this change. In particular, this pattern is most pronounced on private timberlands which experience up to 400% regional increases in small tree density since 1930. Nearly all land ownership classes experience declines in large trees, however private timberland and National Park/Wilderness areas experience a significant reduction of 83% and 73% respectively.

Understanding how forests have changed over long periods of time is increasingly important for contemporary forest managers. Forests are long lived ecosystems, that to be understood, modeled, predicted, managed, or preserved require spatially coincident long-term data. Significant increases in forest density and declining stand basal area in California over the 20th century has resulted in forests that are profoundly different than they were 100 years ago. In recent decades resilience has become the overarching framework of forest management, especially within public forests ((Churchill et al. 2013; Millar, Stephenson, and Stephens 2007; Scott L. Stephens et al. 2016). Widespread patterns of increasing density reduce important structural and spatial patterns in forests including the distribution of large individual trees, open spaces, and clusters of trees (Churchill et al. 2013; James A. Lutz et al. 2013). As these features are lost forests become increasingly susceptible and less resilient to catastrophic fire, disease, and drought induced mortality(Larson and Churchill 2012; Lindenmayer, Laurance, and Franklin 2012; Lydersen et al. 2013; S. L. Stephens, Fry, and Franco-Vizcaíno 2008) . Comparing nearly a century of forest change across land ownership types that are reflective of differing land management strategies is a significant opportunity to explore how land use legacies result in contemporary forest structural patterns. Understanding this range in variation across ownership types is critical when developing restoration efforts, which are often guided by historical forest conditions. The lines between natural and anthropogenic disturbances are often vague and it is not clear, in terms of forest change, if these disturbances should be managed or treated differently. However, landscape level analyses that can help prioritize areas of extreme change for where forest restoration efforts can be administered at a scale that is directed towards specific entities may help encourage management contributing to large scale forest resilience. In this study I concluded that understanding patterns of change across ownership is essential for targeting federal, state, and locally specific policies that foster healthy and resilient forests for the future.

One of the key challenges for California ecosystems in the 21st century will be encouraging sustainable development and growth. The maintenance of ecosystem structures and processes in urban ecosystems will be critical for biodiversity and encourage ecological resilience. Oaks and oak woodlands have played a foundational role in California's history and ecosystem functioning serving as a keystone ecological species supporting hundreds of terrestrial vertebrate species, insects, plants (George, Roach, and Eastburn, n.d.) and as critical components of urban landscapes sequestering carbon, reducing urban heat island effects, increasing soil and nutrient retention and improving water quality, among others (Whipple, Grossinger, and Davis 2011). In Chapter Three I investigated the effects of urban development on changes to the distribution of oak species in California. First, by modeling historical patterns of richness for eight oak species using historical map and plot data from the California Vegetation Type Mapping (VTM) collection I examined spatial intersections between hot spots of historical oak richness and modern urban and conservation lands. I found that impacts from development and conservation vary by both species and richness. At the state level, impact of urban development on oaks has been small within the areas of the highest oak richness but that areas of highest oak richness are also poorly conserved.

I argue that inconsistencies in current policy measures have led to fragmented conservation efforts and suggest regions to prioritize conservation as well as examine urban regions where previous centers of oak richness were lost as potential frontiers for oak reintroduction. As more historical landscapes are being lost to increased urbanization and climatic pressures are projected to reduce species ranges (Kueppers et al. 2005), it is critical to maintain species diversity and reduce habitat fragmentation by making our built and natural environments more cohesive through the strategic placement and reintroduction of important habitats and species. Using historical data in modern species distribution modeling techniques can lead to robust examinations of where habitats and species have been lost historically and can also be used as evidence to incentivize restoration and recovery in areas of loss and to encourage preservation of what still exists. Adopting perspectives from historical ecology and methods like species distribution modeling that have primarily been using in ecological, natural resource, and conservation fields within sustainable urban planning infrastructures may serve as a paradigm shift in the way urban planners recognize, quantify, and integrate landscape history in modern built environments.

In a time of rapid global change in climate and land use, the era of the Anthropocene has proven to be one without parallel. Concepts like “novel ecosystems” characterize this era placing it firmly outside any historic analog (Williams and Jackson 2007). Despite its uniqueness, quantify the effects of anthropogenic activities in relation to natural ecosystem dynamics can be difficult, even while the velocity, scale, and intensity of our impact is undeniable. This difficulty is driven by time lags in relation to impact and effect that is captured differently across ecosystems (Jackson and Blois 2015). As no-analog conditions persist, the more a long and continuous perspective becomes critical. However, despite their importance, historical data are disproportionately underrepresented in modern ecological repositories, leaving temporal gaps in the scientific record (Szabó and Hédl 2011; McClenachan et al. 2015; Jackson and Blois 2015). In Chapter 4, I review three approaches to sharing historical data from field stations using principles from data science. To encourage greater use of historical data across scientific disciplines it is vital to make data findable, accessible, interoperable, and reusable (e.g. the FAIR principles) (Wilkinson et al. 2016). Future research that encourages and capitalizes on expansion of historical data in the modern digital data landscape is critical. In contemporary ecological research history is still of paramount importance even while some argue that the use of history as a reference can be misleading in a time of rapid change (Simon 2017). As long as we acknowledge the bias and temporality of the historical data we use it can provide explanation, context, and provide point in time measurements to collectively assess long term dynamics even in times of rapid change (Hobbs et al. 2014). This summary of three important data collections emerging from the University of California showcase the potential for their use in research and encourages similar ventures that use common archival, geospatial, and data science practices to shepherd historical data out of file drawers and into the contemporary digital data landscape.

In the first two chapters I have discussed the relationship between social and biophysical drivers of landscape change. This kind of understanding of long term patterns of change requires data availability and the ability to re-use data. Following from these, Chapter 3 discusses preserving history's place in the growing data landscape and the opportunity for transdisciplinary connection and innovation. Preserving history is a challenge that no single discipline can be responsible for because it is fundamental to science writ large and as such will require the synergy of people, data, and tools to be critically evaluated and integrated.

1.1 Future Research Directions

Nearly a century of fire suppression and drier climates have contributed to the increased connectivity and density of fuel loads in several regions throughout the state. These changes in forest structure, composition, and distribution coupled with extended fire seasons have contributed to the occurrence of uncharacteristically large high severity wildfires in recent years (Wimberly and Liu 2014; Keeley and Syphard 2016). In Chapter 2 I found that several papers were focusing on climate related contributions to several of these patterns, and so I focused on management, finding that land ownership does play some role in describing patterns of forest structure but that it is likely only one piece of the puzzle. I suggested that future research directions that explicitly combine climate, fire, and land management and explore their respective contribution to long-term changes in forest distribution, structure, and composition is necessary for the management of complex socio-ecological systems.

In Chapter 3 I focused on sustainable urban planning through the integration of historical data with modern data and using more contemporary methods and tools to analyze historical data. Chapter 4 also contributes to this discussion by calling for increased data digitization and sharing of historical data to increase 1) the temporal scale of potential scientific inquiry to include otherwise irretrievable past environments; 2) provide a contextual foundation from which to assess change; and 3) situate the study of the environment in a wider disciplinary context (Swetnam, Allen, and Betancourt 1999). These foci link back to some of the issues discussed in Chapters 2 & 3: specifically, that historical data and requisite long-term perspectives offer an integrated understanding of change that can illuminate causality and temporality (Szabó 2010). The main historical dataset discussed in this dissertation, VTM, was conducted in the 1930s but was never fully digitized or widely used in ecological research until the early 2000s (Kelly et al. 2016).

The general workflow developed for the VTM project (e.g., scanning analog material, georeferencing, estimation of error, creation of digital database, visualization, and serving of data available on the WEB API) is generalizable to other historical data collections and their use in analysis. In working with historical data, generating robust assessments of change and dynamics can be difficult due to gaps and bias within the data record. Therefore, best practice when analyzing historical data is to identify several overlapping (spatial and temporal) sources to fill in the record and enumerate the variability and spectrum of change dynamics or risk drawing rare, extreme, or incomplete conclusions from limited datasets (Szabó and Hédl 2011; Gimmi and Bugmann 2013). To this end, and especially for understanding the interactions between fire, climate, and land management legacies on contemporary California ecosystems, the VTM dataset is only the beginning. The VTM project was the foundational effort of another statewide vegetation and soil survey, the State Cooperative Soil-Vegetation Survey of California (SVS) (1949-1979) (Colwell 1975), that follows from the scope and time period of the VTM surveys and has to our knowledge not been digitized and may add to the understanding of California flora in the post-war period. Further, timber and vegetation surveys from John Leiberg, George Sudworth, and others pre-date the VTM surveys and extend our knowledge of Sierra Nevada vegetation, timber potential, and forest harvesting operations back to 1890 (Rojas 2004; Leiberg 1902; Sudworth 1900; Keeler-Wolf 2012).

Through some of my dissertation work I used common geospatial techniques (scanning, georeferencing, and digitizing) to preserve these materials and make them available for future research exploring mechanisms of change in California ecosystems. These three datasets, The Leiberg-Sudworth, VTM, and SVS once fully digitized will connect a nearly complete spatio-temporal (1890-1970) record of California's vegetation and timberlands that can then be combined with modern remote sensing imagery and contemporary empirical data, further elongating the spatiotemporal record of data availability for scientific inquiry. These datasets will be increasingly important as new tools and methods become available to capitalize on their potential for asking complex socio-ecological questions through space and time.

In California, as human populations and development expand, and the climate changes, the tension between urban and wildland systems will grow, and we increasingly need a better understanding of the spatio-temporal interactions and feedbacks between climate, fire, and management on natural systems to help plan for future resilient landscapes. Investigations of diverse drivers of change on landscapes are necessarily conducted in a spatial framework, using multi-scale spatial data, and require long-term perspectives. This means that historical spatial ecological data have become increasingly important for analysis and are being used in ecology to elongate the temporal scale of analysis, to provide context from which to assess change, and broaden the disciplinary scope of analysis. My dissertation contributes to this body of work by providing new insights into a century of changes to California's forests and woodlands and provides novel insights into the technical protocols needed when historical data are digitized and used in conjunction with contemporary empirical data in modern ecological analysis.

2 References

Churchill, D., Larson, A., Dahlgreen, M., Franklin, J., Hessburg, P., Lutz, J., 2013. Restoring forest resilience: From reference spatial patterns to silvicultural prescriptions and monitoring. For. Ecol. Manage. 291, 442–457.
Colwell, W.L.J.R., 1975. The status of vegetation mapping in California today, in: Barbour, M.G., Major, J. (Eds.), Terrestrial Vegetation of California.

Flint, L., Flint, A., Thorne, J., Boynton, R., 2013. Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. Ecological Processes 2, 25.

George, M., Roach, L., Eastburn, D.J., n.d. Vegetation Change and Ecosystem Services, in: Ecology and Management of Annual Rangelands. University of California Cooperative Extension.
Gimmi, U., Bugmann, H., 2013. Preface: integrating historical ecology and ecological modeling. Landsc. Ecol. 28, 785–787.

Hobbs, R.J., Higgs, E., Hall, C.M., Bridgewater, P., Chapin, F.S., Ellis, E.C., Ewel, J.J., Hallett, L.M., Harris, J., Hulvey, K.B., Jackson, S.T., Kennedy, P.L., Kueffer, C., Lach, L., Lantz, T.C., Lugo, A.E., Mascaro, J., Murphy, S.D., Nelson, C.R., Perring, M.P., Richardson, D.M., Seastedt, T.R., Standish, R.J., Starzomski, B.M., Suding, K.N., Tognetti, P.M., Yakob, L., Yung, L., 2014. Managing the whole landscape: historical, hybrid, and novel ecosystems. Front. Ecol. Environ. 12, 557–564.

Jackson, S.T., Blois, J.L., 2015. Community ecology in a changing environment: Perspectives from the Quaternary. Proc. Natl. Acad. Sci. U. S. A. 112, 4915–4921.

Keeler-Wolf, T., 2012. The History of Vegetation Classification and Mapping in California, in: Vegetation Classification in California.

Keeley, J.E., Syphard, A.D., 2016. Climate Change and Future Fire Regimes: Examples from California. Geosci. J. 6, 37.

Kelly, M., Easterday, K., Rapicullio, G., Koo, M., McIntyre, P., Thorne, J., 2016. Rescuing and Sharing Historical Vegetation Data for Ecological Analysis: The California Vegetation Type Mapping Project. Biodivers. Inf. 11, 40–62.

Kueppers, L.M., Snyder, M.A., Sloan, L.C., Zavaleta, E.S., Fulfrost, B., 2005. Modeled regional climate change and California endemic oak ranges. Proceedings of the National Academy of Sciences 102, 16281–16286.

Larson, A.J., Churchill, D., 2012. Tree spatial patterns in fire-frequent forests of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. For. Ecol. Manage. 267, 74–92.

Leiberg, J.B., 1902. Forest Conditions of the Northern Sierra Nevada, California. USGS Professional Paper.

Lindenmayer, D.B., Laurance, W.F., Franklin, J.F., 2012. Global Decline in Large Old Trees. Science 338, 1305–1306.

Lutz, J. a., van Wagtendonk, J.W., Franklin, J.F., 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. J. Biogeogr. 37, 936–950.

Lutz, J.A., Larson, A.J., Freund, J.A., Swanson, M.E., Bible, K.J., 2013. The importance of large-diameter trees to forest structural heterogeneity. PLoS One 8, e82784.

Lydersen, J.M., North, M.P., Knapp, E.E., Collins, B.M., 2013. Quantifying spatial patterns of tree groups and gaps in mixed-conifer forests: Reference conditions and long-term changes following fire suppression and logging. For. Ecol. Manage. 304, 370–382.

McClenachan, L., Cooper, A.B., McKenzie, M.G., Drew, J.A., 2015. The Importance of Surprising Results and Best Practices in Historical Ecology. Bioscience 65, 932–939.

McIntyre, P.J., Thorne, J.H., Dolanc, C.R., Flint, A.L., Flint, L.E., Kelly, M., Ackerly, D.D., 2015. Twentieth-century shifts in forest structure in California: denser forests, smaller trees, and increased dominance of oaks. Proceedings of the National Academy of Sciences 201410186.

Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. Ecol. Appl. 17, 2145–2151.

Rojas, R., 2004. Kings River Project historic forest condition. Prather, CA: High Sierra.
Simon, Z.B., 2017. Why the Anthropocene has no history: Facing the unprecedented. The Anthropocene Review 4, 239–245.

Stephens, S.L., Collins, B.M., Biber, E., Fule, P.Z., 2016. U.S. Federal fire and forest policy: Emphasizing resilience in dry forests. Ecosphere 7, 1–19.

Stephens, S.L., Fry, D.L., Franco-Vizcaíno, E., 2008. Wildfire and spatial patterns in forests in northwestern Mexico: The United States wishes it had similar fire problems. Ecol. Soc.

Sudworth, G.B., 1900. The Stanislaus and Lake Tahoe forest reserves and adjacent territory:1862-1927.
Swetnam, T., Allen, C.D., Betancourt, J.L., 1999. Applied Historical Ecology: Using the Past to Manage for the Future. Ecol. Appl. 9, 1189–1206.

Szabó, P., 2010. Why history matters in ecology: an interdisciplinary perspective. Environ. Conserv. 37, 380–387.

Szabó, P.,Hédl, R., 2011. Advancing the integration of history and ecology for conservation. Conserv. Biol. 25, 680–687.

Whipple, A.A., Grossinger, R.M., Davis, F.W., 2011. Shifting Baselines in a California Oak Savanna: Nineteenth Century Data to Inform Restoration Scenarios. Restor. Ecol. 19, 88–101.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., ’t Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018.

Williams, J.W., Jackson, S.T., 2007. Novel climates, no-analog communities, and ecological surprises. Front. Ecol. Environ. 5, 475–482.

Wimberly, M.C., Liu, Z., 2014. Interactions of climate, fire, and management in future forests of the Pacific Northwest. For. Ecol. Manage. 327, 270–279.