

RESCUING AND SHARING HISTORICAL VEGETATION DATA FOR ECOLOGICAL ANALYSIS: THE CALIFORNIA VEGETATION TYPE MAPPING PROJECT

MAGGI KELLY^{1,2,3,4}, KELLY EASTERDAY¹, GIOVANNI RAPACCIUOLO⁵, MICHELLE S. KOO^{4,6},
PATRICK MCINTYRE⁷, JAMES THORNE⁸

¹134 Mulford Hall #3114, Department of Environmental Sciences, Policy and Management, University of California, Berkeley. Berkeley, CA 94720-3114. ²Geospatial Innovation Facility, University of California, Berkeley, CA 94720. ³University of California Division of Agriculture and Natural Resources. ⁴Berkeley Initiative in Global Change Biology, University of California, Berkeley, CA 94720. ⁵Department of Ecology and Evolution, 650 Life Sciences Building, Stony Brook University, Stony Brook, NY 11789. ⁶Museum of Vertebrate Zoology, 3101 Valley Life Sciences, University of California, Berkeley, CA 94720-3160. ⁷Biogeographic Data Branch, California Department of Fish and Wildlife, 1416 9th Street, Suite 1266, Sacramento, CA 95814. ⁸Department of Environmental Science and Policy, University of California, Davis, CA 95616. Maggi Kelly corresponding author: maggi@berkeley.edu.

Abstract.—Research efforts that synthesize historical and contemporary ecological data with modeling approaches improve our understanding of the complex response of species, communities, and landscapes to changing biophysical conditions through time and in space. Historical ecological data are particularly important in this respect. There are remaining barriers that limit such data synthesis, and technological improvements that make multiple diverse datasets more readily available for integration and synthesis are needed. This paper presents one case study of the Wieslander Vegetation Type Mapping project in California and highlights the importance of rescuing, digitizing and sharing historical datasets. We review the varied ecological uses of the historical collection: the vegetation maps have been used to understand legacies of land use change and plan for the future; the plot data have been used to examine changes to chaparral and forest communities around the state and to predict community structure and shifts under a changing climate; the photographs have been used to understand changing vegetation structure; and the voucher specimens in combination with other specimen collections have been used for large scale distribution modeling efforts. The digitization and sharing of the data via the web has broadened the scope and scale of the types of analysis performed. Yet, additional research avenues can be pursued using multiple types of VTM data, and by linking VTM data with contemporary data. The digital VTM collection is an example of a data infrastructure that expands the potential of large scale research through the integration and synthesis of data drawn from numerous data sources; its journey from analog to digital is a cautionary tale of the importance of finding historical data, digitizing it with best practices, linking it with other datasets, and sharing it with the research community.

Keywords.—VTM; vegetation mapping; historical data; California; climate change; application programming interface; API

The grand challenges that human societies face globally—chiefly biodiversity loss, climate change and conflicting land use—are at the intersection of natural and social systems, and will require a synthesis of past, present and projected future data to understand their complex interactions and mitigate the effects of rapid environmental change. The synthesis of historical and contemporary ecological data with predictive modeling approaches is a powerful approach to understand

the complex response of species, communities, and landscapes to changing biophysical conditions through time and in space. Synthesis of historical and current data can create new knowledge through novel combinations of datasets and modeling approaches (Krebs et al. 2001; Schulte and Mladenoff 2001; Mladenoff et al. 2002; Peters 2010; Fox and Hendler 2011); can help guide management and conservation by improving the temporal transferability of predictive models

(Kittinger et al. 2013; Rapacciuolo et al. 2014a); and can provide crucial information for adaptation planning, policy, and management of natural systems (Stein et al. 2010; Whipple et al. 2011; Higgs et al. 2014).

While calls for digital infrastructures that allow for interoperability between multiple ecological data have been heard since at least the mid 1990s (e.g., Michener et al. 1997), there are remaining barriers that limit such data synthesis. For most of the last century, ecological and geographical databases were mostly disparate: developed and maintained by individuals or small academic groups, on focused areas, and concentrated in time (Michener et al. 1997; Frehner and Braendli 2006; Michener 2006). This paradigm, while still prevalent, is poorly suited to answer large multi-scale interdisciplinary questions that require data synthesis, large data files, user interactivity, dynamic and shared updates, and collaboration between scientists and stakeholders across social and ecological domains (Jones et al. 2006; Tenopir et al. 2011; Borgman 2012; Hampton et al. 2013). What is needed are data frameworks that facilitate these aspects of science through novel developments in data curation and integration that rely on web applications, open standards, and application programming interfaces (APIs) that make the interaction between scientist and multiple diverse data archives as well as the building of analytical tools via the web easier. These technological improvements and standardized protocols make multiple diverse and large datasets more readily available for integration, searching, visualization, and analysis by researchers (Frehner and Braendli 2006; Peters 2010; Reichman et al. 2011; Borgman 2012).

By focusing on one valuable and, at times in its history, vulnerable dataset, we highlight the importance of data rescue, digitization and sharing of historical datasets for environmental science. We also make the case that data, open data frameworks, modeling, and scientific collaboration are critical for modern applied ecological exploration in order to understand past flora and land use, to search for mechanisms for decadal-scale vegetation changes, and to predict and plan for the future. We present a brief description of the Wieslander Vegetation Type Mapping (VTM) project and the digitization of its components, and review the ways in which the VTM collection has

been used in ecological and geographical analysis. We close with a discussion of its journey from paper record to digital database and API as an example of many of the critical themes facing ecologists today: data, data sharing, data synthesis, and modeling.

VTM: A CASE STUDY FROM CALIFORNIA

California provides a unique case study for studying the interactions between climate, land use, and the distribution and abundance of populations and communities due to its diversity in climate, biota, topography and land use and its long tradition of natural history recording and collection (Rapacciuolo et al. 2014b; Chornesky et al. 2015). Moreover, the state is at the forefront in a range of climate adaptation measures (Chornesky et al. 2015). These complex and interacting socio-ecological phenomena are evidenced in the current landscape but result from historical processes (Solé and Bascompte 2006). Future planning increasingly requires syntheses of historical and contemporary data as well as modeling approaches to understand the complex nature of environmental processes and the response of species, communities, and landscapes to changing biophysical conditions through time and in space (Beck et al. 2012).

Although many important historical datasets describing California vegetation exist, most exist in analog form and in isolation. Small-scale county maps of forest cover of the more heavily forested portions of the counties in northern California began to be made in the late 19th century (Colwell 1977; Keeler-Wolf 2007) and much of the Sierra Nevada was mapped for forest cover in the first decade of the 20th century by (Leiberg 1902). However, for the most part, California lacked the coverage provided in much of the midwestern and eastern portions of the US by the General Land Office surveys (Galatowitsch 1990; Schulte and Mladenoff 2001). One exception to this is the mapping collection we consider in this paper: the Wieslander Vegetation Type Mapping (VTM) collection, named after its director Albert Wieslander. The VTM collection covered California in the 1920s–1940s and has been described as “the most important and comprehensive botanical map of a large area ever undertaken anywhere on the earth’s surface” (Kuchler 1967; Jepson et al. 2000) and the “most ambitious attempt ever made

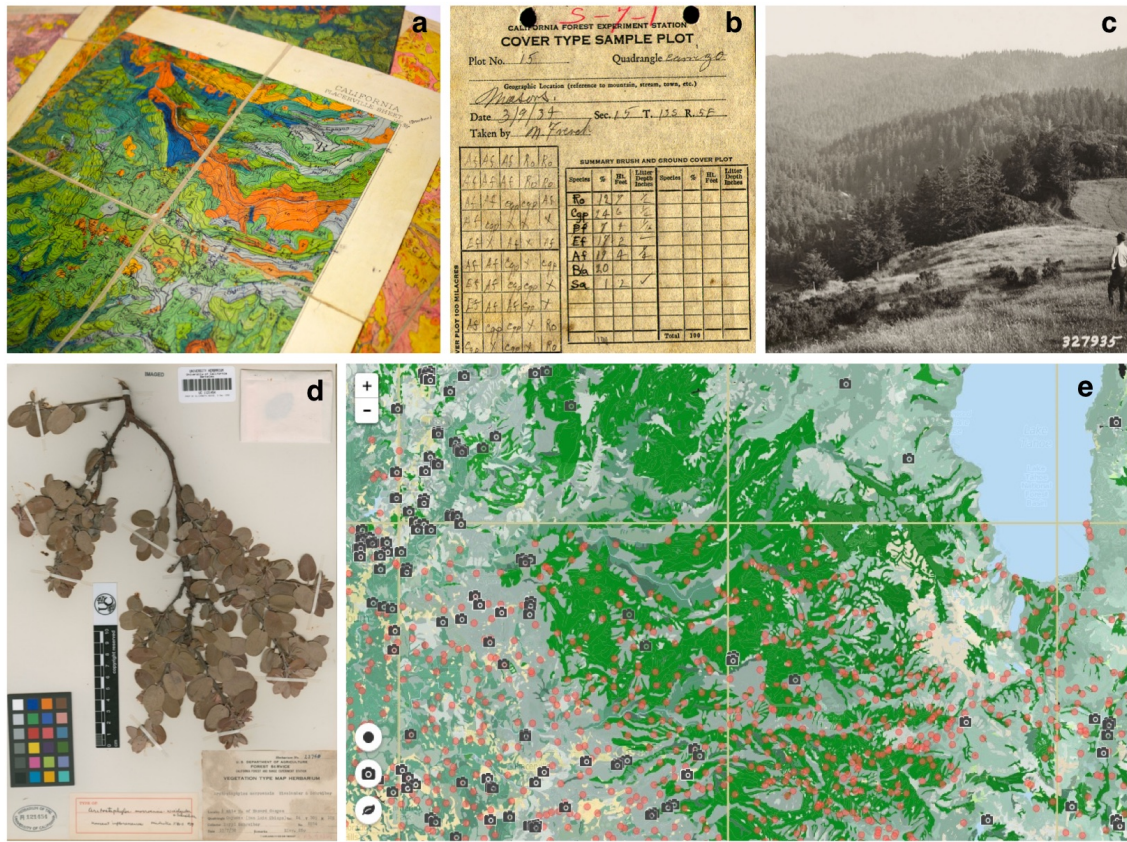


Figure 1. Examples of the components of the original VTM collection: (a) vegetation maps (Placer Co.); (b) a plot card (Imperial Co.), and (c) a landscape photograph (San Mateo Co.), (d) an herbarium specimen (*Arcostaphylos morroensis*, San Luis Obispo Co.), and (e) the digital representation of maps, plots and photographs in the VTM website showing an area covering part of Lake Tahoe and the Tahoe National Forest. Background color = vegetation polygons, red dots = plot locations, black icons = locations of georeferenced photographs.

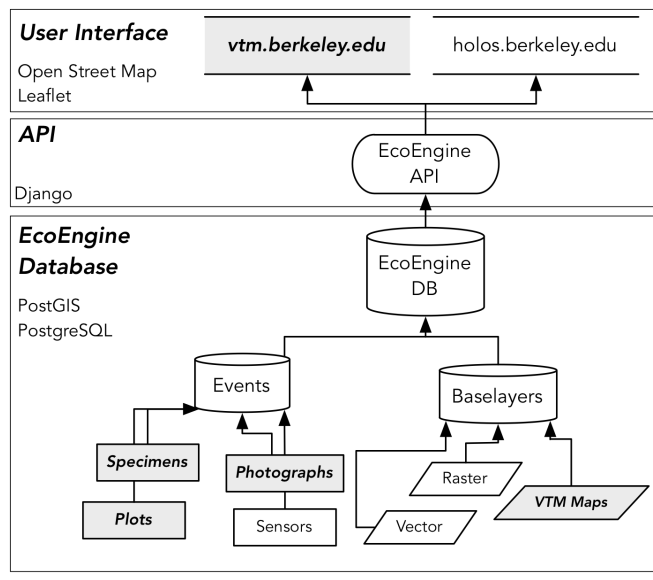


Figure 2. Schematic of the Berkeley EcoEngine API, HOLOS and VTM websites. Elements related to VTM are in bold and italic font.

to describe the complex vegetation of California” (Critchfield 1971). The collection “remains to this date the most exhaustive and detailed effort of mapping vegetation in the state” (Keeler-Wolf 2007). The collection provides a detailed picture of much of California land cover and vegetation in the early 20th century, and it continues to contribute to the study, characterization, and understanding of historical, contemporary, and future California landscapes.

Beginning in the 1920s, Albert Everett Wieslander, an employee of the U.S. Forest Service (USFS) California Forest and Range Experiment Station located in Berkeley, began an effort to map California’s wildlands (Colwell 1977). By the time World War II interrupted the project, 40 million hectares had been visited, covering most of the state’s natural areas exclusive of the deserts and the larger agricultural areas (Wieslander 1961; Colwell 1977). At the end of the war, the project resumed soil-vegetation survey maps (Griffin and Critchfield 1972), but relied on aerial photography techniques. The priority of the original VTM project was to draw detailed vegetation type maps, but the mapping protocol evolved over time with need. It soon added the collection of detailed floristic (trees and shrubs) and environmental data from plots with locations recorded on USGS topographic quadrangles (Wieslander 1986); black and white landscape photographs and maps showing the vantage point of the photographs; and herbarium specimens for species recorded on the vegetation maps or in the sample plots (Ertter 2000; Kelly et al. 2005). The vegetation maps, plots, plot maps, photographs and herbarium specimens are now curated at University of California libraries and herbaria system. However, over the 20th century, parts of the collection have been at risk of loss. In his oral history, Wieslander related how many of the original vegetation maps were thrown away by the University Press, and a few were rescued from destruction by a concerned professor at UC Berkeley (Wieslander 1986). In a second example, in the 1990s the plot data collection was almost discarded to create collection space until a university librarian contacted a professor at Berkeley and together they moved the paper collection into the professor’s office (Norma Kobzina, pers. comm.). Examples of the collection are found in Figure 1; and more details of the

collection can be found several publications (Kelly et al. 2005; Thorne et al. 2008; Thorne and Le in press).

Vegetation maps

Based on direct field observations from vantage points, dominant vegetation types were mapped directly onto U.S. Geological Survey (USGS) topographic quadrangles with a minimum mapping unit (MMU) of 16 ha, and supplemented by sample plots (described below). The vegetation mapping scheme was driven by dominant overstory vegetation and included vegetation mosaics: complex vegetation conditions that resulted from fire or other disturbances, and pure and mixed stand conditions which were associated with “natural plant associations” (Wieslander 1935b, a, 1961; Thorne and Le in press). The mapped products include 215 maps with the major vegetation types shown in different colors and separated by ink lines printed on 7.5’, 15’, and 30’ USGS topographic quadrangles (Colwell 1977; Thorne and Le in press) (Figure 1).

Plot data and plot maps

The VTM crews visited over 18,000 VTM plots statewide; these are concentrated along the central and southern coastal ranges, and in the Sierra Nevada Mountains. These plots were surveyed as a check on the vegetation polygons, and also to provide details on species composition, size and stand density of trees and shrubs and depth of leaf litter. Plots were rectangular (800 m² in forests, 400 m² in shrub and chaparral communities), ran upslope and were divided into milacre sampling units in which dominant species and height characteristics were recorded. Trees greater than 10 cm in diameter breast height within 10 m of either side of the centerline were tallied by species and diameter class. In both forest and understory plots, slope, soil characteristics, and year of last burn were recorded (Wieslander 1935b, a). The plots cover a gradient of vegetation types and include data regarding tree stand structure (number per diameter class), percent cover of dominant overstory and understory vegetation by species, soil type, parent material, leaf litter, elevation, slope, aspect, parent material, and other environmental variables. All plot data were stored on paper data sheets and individual plots were numbered according to map name, quad section

Table 1. Scientific focus of reviewed VTM studies. Asterisks (*) indicate Ph.D. dissertations.

Focus	N	Items	References
Conifer forest / mixed conifer forest / conifer trees	15	Plots: Maps: Photos:	Minnich 1978*; Minnich 1995; Bouldin 1999*; Goforth and Minnich 2008; Fellows and Goulden 2008; Lutz et al. 2009; Lutz et al. 2010; Swanson et al. 2013; Dolanc et al. 2013a; Dolanc et al. 2013b; Dolanc et al. 2014; Maxwell et al. 2014 Walker 2000* Dodge 1975*; Taylor 2000
Conifer and hardwood forest	1	Plots: Maps: Photos:	McIntyre et al. 2015 None None
Shrubland / chaparral	9	Plots: Maps: Plots & Maps: Photos:	Minnich and Dezzani 1998; Franklin 2002; Franklin et al. 2004; Keeley 2004; Taluto and Suding 2008; Syphard and Franklin 2009 Lippitt et al. 2012; Cox et al. 2014 Bradbury 1974* None
Oaks / oak woodland / rangelands	4	Plots: Maps: Photos:	Allen et al. 1991; Allen-Diaz and Holzman 1991; Vayssières et al. 2000; Conlisk et al. 2012 None None
Vascular plants (including both tree and shrub)	5	Plots: Maps: Photos:	Syphard and Franklin 2010; Crimmins et al. 2011; Dobrowski et al. 2011; Crimmins et al. 2013; Crimmins et al. 2014 None None
Land cover	6	Plots: Maps: Photos:	None Thorne et al. 2004; Thorne et al. 2008; Rubidge et al. 2011; Thorne et al. 2013; Santos et al. 2014a; Santos et al. 2014b; None
Grasslands	1	Plots: Maps: Photos:	None Freunenberger et al. 1987 None
Various	3	Plots: Maps:	Kelly et al. 2008 Preston et al. 2012; Davis and Sims 2013

number and plot number. Individual plots locations were denoted by 3.5 mm hollow circles stamped in red ink on USGS topographic maps (editions of 1893–1920, reprinted in the 1930s) that had been cut into sections and mounted on canvas for ease of use in the field. There are about 150 15' (1:62,500 scale) and 30' (1:125,000 scale) plot maps (Figure 1).

Photographs and metadata

The VTM crews also captured about 3100 black and white landscape and stand scale photographs (9.2 x 13.6 cm), taken during 1920–1941. The location of origin of many of these photographs are marked on USGS topographic maps in red pen, with an arrow marking the vantage point and view of the photo. The photograph captions typically includes a description of the location and subject of the photograph including relevant genus and species, timber stand conditions, and examples of cultivation, grazing, logging, mining and fire, and quad name. The photographer, date of the photograph, and occasionally township and range are included (Figure 1).

Herbarium specimens

The VTM crews collected over 23,000 herbarium specimens as part of their efforts. A primary motivation was to voucher specimens for regional identification of taxa observed in plots (Wieslander 1935b, a). The VTM specimens span 3157 taxa, representing 40% of the ~7600 plant taxa recorded from California. Primary holdings are at the University and Jepson Herbaria (UC Berkeley), with duplicates housed at various institutions, most notably at the University of California Santa Cruz where VTM specimens helped form the initial holdings of the herbarium. The specimens are databased and georeferenced but awaiting integration with other VTM data on the vtm.berkeley.edu website.

VTM digitization process

The digitization of the vegetation maps, plots, plot maps, photographs, and locations of herbarium specimens was led by people in several groups in the University of California. The digitizing of the vegetation maps was led by James Thorne at UC Davis over a 10-year period. This involved finding

many of them at various repositories and obtaining the same edition topographic maps as the vegetation maps were drawn on from university libraries. These were digitized and converted to GIS data. The photograph collection was almost completely intact in the 2000s in the Marian Koshland Biosciences Museum at UC Berkeley, and were scanned and uploaded online. In 2014, the efforts to georeference the photographs were led by Michelle Koo of the Museum of Vertebrate Zoology at UC Berkeley. The plot data cards and the majority of the vegetation maps were physically located in the 1990s in the laboratory of Barbara Allen-Diaz at UC Berkeley, and her lab led the digitization of the plot data. The plot maps were physically also located in Dr. Allen-Diaz' lab and Maggi Kelly led the scanning and georeferencing, and the linking of the map data with the plot data. The herbaria specimens are physically located in the Jepson Herbarium at UC Berkeley, and georeferencing of the specimen locations was led by Koo following DarwinCore standards and best practices for georeferencing.

The collection parts were digitized and georeferenced with similar protocols resulting in digital shapefiles of polygons (vegetation maps) or points (plots, photographs and specimens) linked to respective databases. More details on the digitization and georeferencing protocols can be found in (Kelly et al. 2005; Thorne and Le in press). Map georeferencing was done by a suite of analysts using a collection of tie-points gathered from the VTM maps and current USGS topographic quadrangles. The relevant features for each part of the collection (e.g. polygons for the vegetation maps or points for the plots) were transcribed manually from the digital version of the respective map. Vegetation species codes from vegetation maps and vegetation plot cards were transcribed using the *Manual of California Vegetation Types* (Sawyer and Keeler-Wolf 1995), and the California Wildlife Habitat Relationships Models (WHR) (Mayer and Laudenslayer 1988) for land cover classifications. The locations of each photograph depicted on an accompanying map USGS topographic map were georeferenced by measuring the distance and bearing of each marked point from the known southwest corner and calculating its location. Specimens were digitized primarily to township, range and section centroids.

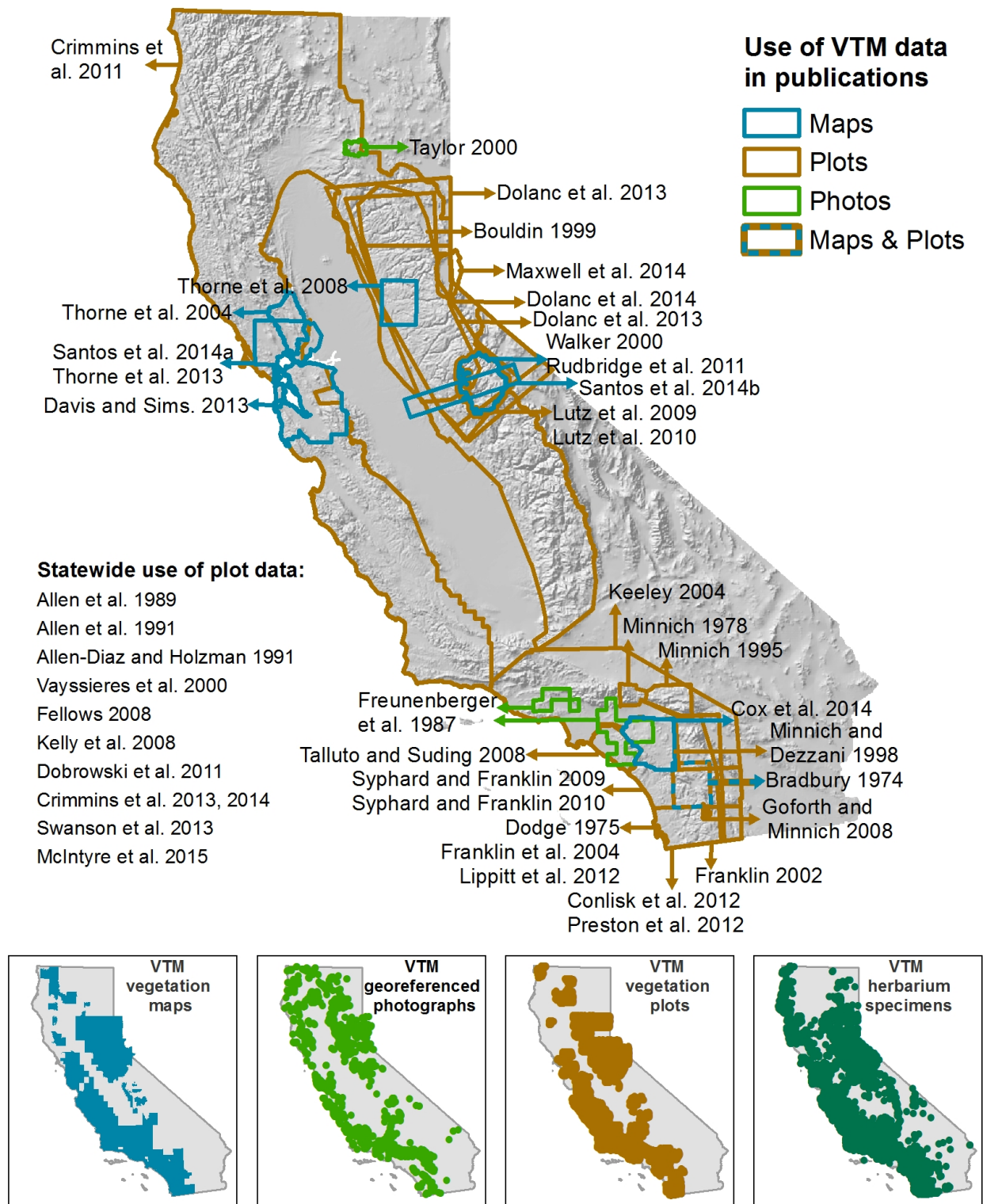


Figure 3. Locations in California of scientific projects that have used VTM data.

Web stack and API development

All digital spatial VTM data are made available in an open source web-mapping application¹ developed with an entirely open source software stack by UC Berkeley's Geospatial Innovation Facility. All VTM data are stored using PostgreSQL, a relational database that supports the storage of and analysis of geospatial vector data through the PostGIS extension. The map interface was built using Leaflet, a lightweight JavaScript mapping library with Open Street Map providing base layers (Figure 2). The website presents a user interface for exploring, searching, aggregating, and downloading the VTM data collection.

The VTM website was built using the Berkeley Ecoinformatics Engine Application Programming Interface (Ecoengine API²). A web API is an application that serves machine-readable data and functionality to applications that represent the data to users. The Ecoengine API is a directory and gateway to many diverse biological and environmental collections at UC Berkeley, part of a recent trend that makes use of digitized museum collections for the study of anthropogenic loss of biodiversity, climate change, and changes to ecosystems (Pyke & Ehrlich 2010). The Ecoengine API serves about 5 million records from museum specimens, soil and pollen data, field station records, sensor readings, as well as biophysical base layers such as climate and land use and the VTM data.

The VTM website accesses the VTM data collection through three main Ecoengine API resources: VTM photos, VTM plots and VTM vegetation (Figure 2). For each resource, the user can explore a map that shows geographic locations of the resource and when the user clicks on a feature, a pop up window displays greater detail about the feature. Querying the Ecoengine API retrieves this detail. The original vegetation maps themselves display vibrant color schemes (Figure 1), but rather than try to replicate this variety (although that might be possible in the future) the polygons are rendered using a modification from the USGS NLCD land cover color scheme (Fry et al. 2008). All original VTM vegetation types were cross-walked with California Wildlife Habitat Relationships (CWHR) information.

¹ <http://vtm.berkeley.edu>.

² <https://ecoengine.berkeley.edu/>.

USES OF THE VTM COLLECTION

We found 39 peer-reviewed journal articles and 5 Ph.D. dissertations that discussed the use of VTM vegetation maps, plots or photographs before 2016 using Google Scholar. Herbarium specimens have been used in combination with other numerous georeferenced specimens available through the Consortia of California Herbaria to assess the responses of different groups of plant species across the entire flora of California to temperature change (Loarie et al. 2008; Wolf et al. 2016). We did not evaluate reports or grey literature. Of the documents reviewed, the majority (n = 31) of studies focused on the VTM plot data, 12 used vegetation maps, and two used photographs. The majority of the studies focused on forested landscapes (conifer or hardwood), but many focused on shrublands or rangelands, or on land use. Research using plot data completed after data digitization in the mid-2000s (Kelly et al. 2005) generally focused on regional or statewide analyses. A summary of the uses of VTM data is provided in Table 1, with their study areas shown in Figure 3.

Use of VTM data to develop vegetation reference systems

The floristic detail found in the plot database has been used since the 1940s to develop Californian vegetation community classification schemes in forests (Maxwell et al. 2014), chaparral (Franklin 2002), oak and rangeland communities (Allen et al. 1991; Allen-Diaz and Holzman 1991), as well as general land cover classification schemes (Thorne et al. 2004). More recently, the plot data has been used to reconstruct past reference conditions or to establish a historical range of variability (HRV) for a particular time period to target ecosystem restoration or landscape management. For example, Maxwell et al. (2014) used 399 VTM plots in conjunction with dendroecological reconstructions in the Lake Tahoe Basin area to reconstruct forest structure, forest fuels, and fire regimes representing pre-settlement conditions as an aid to current forest management. Herbarium specimens helped establish baseline distributions reference material for the distribution and identification of California trees and shrubs (McMinn 1951; Griffin and Critchfield 1972).

20th century changes to vegetation communities

The VTM collection has been used recently to understand the patterns and causes of changes in vegetation structure over decadal scales. Authors have examined shifts in abundance and composition, as well as decadal scale changes in the elevational distribution of vegetation communities (i.e. community shifts upslope or downslope). This type of work is critical for predicting how vegetation communities will respond to a changing climate, as well as planning for community resilience in the face of disturbances.

Changes to structure and composition of California forests.—Comparison of the VTM plots with current vegetation (using both relocated plots and contemporary vegetation maps) revealed consistent evidence of an increase in young-growth and small-diameter trees across the state and decreases in large trees (Minnich et al. 1995; Fellows and Goulden 2008; Goforth and Minnich 2008; Lutz et al. 2009; Dolanc et al. 2013a, 2014; McIntyre et al. 2015), as well as changes in forest composition (Minnich et al. 1995; Dolanc et al. 2013a; McIntyre et al. 2015). Much of this evidence of change in forest age structures derives from resurveyed VTM plots (Table 1). Goforth and Minnich (2008) resurveyed three VTM plots of mixed conifer forest on Cuyamaca Mountain in the Peninsular Range of southern California four years after a stand replacing fire. They replicated measurements at multiple sites around the expected locations of the three original plots and covered similar environmental conditions as described in the original VTM plots. Their analysis showed significant changes in forest composition, increases in tree density, and decreases in stem diameter over a 75-yr period. They augmented these results with an analysis of repeat aerial photographs from 1928 and 1995 that also show significant increase in canopy cover. Dolanc et al. (2013b) resurveyed 139 VTM plots in the subalpine zone of the Sierra Nevada, and compared historical and modern climatic conditions using two high-elevation climate stations nearby. They found fewer larger trees and increases in smaller size classes.

More evidence of increased tree density and change to age structure has been found comparing VTM data with modern plot data, typically provided by the Forest Inventory and Analysis (FIA) program (U.S. Forest Service 2016b) (Table 2). The primary objective of the FIA program is to

determine the extent, condition, volume, growth, and use of trees on US forest land through repeated and widespread sampling. Congress mandated the FIA program in 1928, but it was not until 1999 that collections on a network of plots began to be made annually. Fellows and Goulden (2008) compared 269 VTM forest plots with 260 FIA plots from the 1990s to quantify changes in aboveground biomass for California forests. They found that the size structure of the forests changed dramatically in 70 years, with large increases in stem density driven by increase in number of smaller trees and a net loss of large trees, with a concordant decrease in aboveground carbon stocks, estimated allometrically. They attributed this change to fire suppression. Lutz et al. (2009) compared 655 VTM and 210 modern vegetation plots surveyed by National Park Service field crews from 1988-1999 in Yosemite National Park. They found that large-diameter tree density in Yosemite declined by 24%, and declines were greatest in subalpine and upper montane forests, and least in lower montane forests. Dolanc et al. (2013a) examined changes in abundance and composition across nine vegetation types in northern Sierra Nevada using 4321 VTM plots and 1000 modern FIA plots. Tree density was significantly higher in the modern plots in eight of nine vegetation types. They also found a shift in dominance toward shade-tolerant conifers and evergreen oaks. They attributed some of these changes to fire suppression, although not all, as alpine forests, which do not experience fire, also showed increases in density. Dolanc et al. (2014) conducted a similar study of the entire elevational range of west-facing slopes in the Central and South Sierra Nevada, further confirming significant patterns of change in forest structure and composition. McIntyre et al. (2015) used both VTM plot data (n = 6,572) from the entire state, and compared baselines in forest structure and composition in oaks (*Quercus*) and pines (*Pinus*) with those derived from FIA data (n = 1909). They found declines in large trees in every ecoregion of the state, but also found forest composition has shifted toward increased dominance by oaks relative to pines. They showed that declines in large trees were more severe in areas experiencing greater increases in climatic water deficit since the 1930s.

Loss of shrub and chaparral communities.—Shrublands are vegetation communities dominated

Table 2. Data commonly used to compare with the VTM collection.

Dataset	Date	Agency	Scale	Extent
VTM	1928-1940	U.S. Forest Service	1:125,000, 1:62,500	California
FIA *	1930-Present	VTM plot comparison U.S. Forest Service	Plot Data	United States
CalVeg*	1979-1981	VTM vegetation map comparison U.S Forest Service PSW Forest and Range Experiment Station, U.S Forest Service Remote Sensing Laboratory	MMU 6-800 acres	California (except southern interior)
Hardwood Rangeland Mapping Program	1981-1991	California Dept. of Forestry and Fire Protection	1:24,000– 1:58,000	California hardwood and savannah (<5000 ft)
CAL-GAP*	1990-2008	USGS	1:100,000 (1 km MMU)	California
NLCD*	1992-2001- 2006-2011- 2013	USGS	30m	United States

* FIA = Forest Inventory and Assessment, CalVeg = Classification and Assessment with Landsat of Visible Ecological Groupings; CAL-GAP = California Gap Analysis Program; NLCD = National Land Cover Database.

by short evergreen woody shrubs with sclerophyll leaves. These chaparral and sage scrub plant communities, once dominant in southern California, are now threatened by changing land use, fire regimes, nitrogen deposition, and invasive species. The VTM collection has played a large role in understanding 20th century changes to these communities. The original priority of the VTM project was the vegetation type maps and forest cover, but the mapping protocol developed over time to suit emerging needs. Around 1927, the forest plot protocol was extended to include information on shrub communities in response to requests from southern California-based Forest Service (Wieslander 1986).

The earliest examination of changes to chaparral communities in San Diego County that made use of the VTM plots is Bradbury (1974), who resurveyed VTM plots 40 years after the original VTM crews and found slight changes to chaparral communities, most of which he related to disturbance (Bradbury 1974; Franklin 2002). Freudenberger et al. (1987) used VTM vegetation maps as a guide to help interpret historical aerial photography in a project examining the shifting mosaic of grassland and shrubland in the Los Angeles Basin over 50 years. Franklin et al. (2004) resurveyed 649 plots to investigate the role of fire in changes in species composition and cover in chaparral communities in San Diego County. They report a complex interplay between oak woodlands, chaparral communities and fire response.

More recent uses of the VTM collection explore a range of disturbances driving the loss of Southern California shrublands including fire, invasive species and nitrogen deposition. Talluto and Suding (2008) resampled 54 VTM shrub plots in southern California and found strong evidence of non-native grassland invasion at the expense of coastal sage scrub. Grassland encroachment was positively correlated with increased fire frequency and, in areas with low fire frequencies, air pollution (likely nitrogen deposition). Lippitt et al. (2013) investigated the role of fire in chaparral community resilience in southern California. They used the VTM maps and CALVEG (United States Forest Service 2016a) to identify areas of chaparral that experienced high frequency fires, finding that the number of burns and decreases in mean fire interval over time increased the chance for alteration and type conversion of chemise

chaparral communities. Cox et al. (2014) used a VTM vegetation map to examine the factors leading to dynamics between coastal sage scrub and grassland between 1930 and 2002. Nitrogen deposition and fire were important to understanding and predicting the recovery of the coastal sage scrub system.

With the exception of Cox et al. (2014), these studies relied on the relocation of VTM plots. Plot relocation is known to be difficult (e.g. Minnich et al. 1995), especially in shrublands (Keeley 2004). However, there seems to be ample evidence of successful relocation, particularly in efforts focused over larger scales that incorporate larger numbers of plots which reduce overall spatial uncertainty (Kelly et al. 2008). Additionally, there is evidence that relocation of VTM plots and FIA data (e.g. McIntyre et al. 2015) provide very similar results (Dolanc et al. 2013b).

Range shifts in vegetation communities.—Pioneering work to digitize, georeference and attribute the countless vegetation type polygons from the original VTM collection has enabled analyses of shifts in vegetation classes over time (Thorne and Le in press). For example, Thorne et al. (Thorne et al. 2008) compared the digital VTM vegetation maps to contemporary remote-sensing products such as CALVEG; (Schwind and Gordon 2001) and found significant changes in vegetation types in the Placerville quadrangle on the west slope of the Sierra Nevada. They used the California Wildlife Habitat Relationships (CWHR) classes (Mayer and Laudenslayer 1988) as a cross-walk between the two data types, and found significant shifts in WHR classes. At lower elevations below 700 m, annual grassland expanded, and low elevation hardwoods and conifers, particularly Blue Oak woodland and Blue Oak-Foothill pine, contracted. At higher elevations above 700 m, Ponderosa pine contracted while montane hardwood, montane hardwood-conifer and Douglas fir expanded.

A different approach was taken by Crimmins et al. (2011), who examined the altitudinal distributions of 64 Californian vascular plant species using 13,746 VTM plots and modern FIA plots. They used logistic regression to estimate species' optimum elevations in each period and altitudinal shifts were measured as the difference in optimum elevation between periods. Contrary to the expectation that species from lower elevations will

shift upslope in response to warming (e.g. Kelly and Goulden 2008; Lenoir et al. 2008), they found that climate changes in California have resulted in a significant downward shift in these species' optimum elevations. They explain this downhill shift by regional changes in climatic water balance. Similarly, Dolanc et al. (2013b) who resampled VTM plots in the Sierra Nevada in a study of forest structure changes also found no evidence of upslope shifts in vegetation communities: no species move into or out of the study area when comparing VTM plots with modern resurveys.

Use of VTM photography.—There are only two published examples that highlight the use of the VTM photographs to understand vegetation change, and no recent examples. Wieslander himself discussed the possibilities of using repeat photography (Wieslander 1986). Dodge (1975) used VTM photographs to study historical changes in San Diego County vegetation, with mixed results. He reported that most photographs in the area were stand level rather than landscape view inhibiting his ability to perform exact reshoots. He compared vegetation changes evident from rephotographing the general area of each photograph and reported that vegetation increased in density and large accumulation of dead material on the ground in areas with no evidence of fires, and in areas that experienced wildfires in the intervening years there was near total destruction of coniferous forests and extensive damage to oak woodlands. Taylor (2000) had modest success in relocating four 1925 photographs on Prospect Peak in Lassen National Park; in three of them he found evidence of increased forest density, as a result of fire suppression.

Forecasting future change

The use of past and contemporary data to understand biogeographic responses to recent climate change is fundamental for improving our predictions of likely future responses (Rapacciuolo et al. 2014a). Several recent papers have used the VTM plot data to gather representative samples of species to parameterize current distributions as a precursor to modeling future distributions. Lutz et al. (2010) used data from 655 VTM plots to gather representative tree species data for Yosemite National Park. They correlated climatic water deficit with past and current tree distribution and projected tree distribution into the future using

standard climate scenarios. They suggest that ongoing changes in forest structure and composition can be related to changes in climate water balance. While past increases in temperature since the Little Ice Age have been offset by increasing precipitation, projected future temperature increases coupled with likely decreases in precipitation will increase water deficit, with detrimental impacts on many species. They showed projected future increases in water deficit of 23% across all plots in Yosemite, which may disproportionately affect Western white pine (*Pinus monticola*) and Mountain hemlock (*Tsuga mertensiana*).

Dolanc et al. (2013b) used VTM data for a retrospective examination of climate related changes to subalpine forests, and they commented on the possible future changes to these communities. Their work examining changes in tree abundance and composition in subalpine forests of the Sierra Nevada did not find evidence of change in the direction predicted by vegetation models linked to future climate scenarios. They explain this discrepancy by suggesting a possible lag effect, whereby lower-elevation species may eventually replace higher-elevation species, but only after decades or even centuries.

Conlisk et al. (2012) used VTM plot data and herbarium records with the species distribution modeling algorithm Maxent (Phillips et al. 2006) to find locations and carrying capacities of metapopulation patches of *Quercus engelmannii* in eastern San Diego County. They combined Maxent models with a demographic model to determine the population dynamics within and between these patches in the future. They predicted the dramatic reduction of suitable habitat for *Q. engelmannii* in 2100 under two climate scenarios, with suitable habitat patches predicted to shrink in extent and move to higher elevations.

Model assessment and validation

Species distribution models (SDM) correlate environmental predictors such as climate and topography to the known distribution of a species, and can be used to generate spatial predictions of the suitability or probability of presence for a species given the predictors. The past decade has seen an expansion in the types and sophistication of species distribution modeling approaches that can be used to forecast changes to species and

communities based on locality data and baseline conditions (Graham et al. 2004; Guisan and Thuiller 2005; Guo et al. 2005; Elith et al. 2006; Elith and Leathwick 2009). SDMs have many challenges when forecasting species responses under changing environments, including the potential absence of a species-environment equilibrium, the difficulty to account for dispersal limitations, biotic interactions, phenotypic plasticity and evolutionary changes, and the incidence of novel environments outside the range of conditions used to calibrate the models (Elith and Leathwick 2009). Despite these obstacles, SDMs have become an integral part of conservation planning, resource management, and land decision-making processes. In recent years, the VTM plot data have been increasingly used in SDM studies to provide temporally independent data to validate model predictions and explore some of the methodological nuances of these models. Specifically, VTM data have been used to evaluate model selection and methodology (single vs. ensemble models, or model vs. model) (Vayssières et al. 2000; Crimmins et al. 2013), to assess model uncertainty over time and incorporate spatial autocorrelation (e.g. Swanson et al. 2013), to assess model transferability through time (Syphard and Franklin 2009, 2010; Dobrowski et al. 2011; Crimmins et al. 2014), to test how SDM performance varies with species' traits and environmental predictors (Syphard & Franklin 2009, 2010), and to explore how SDMs can be coupled with demographic models to forecast population responses (Conlisk et al. 2012).

Linkage between vegetation and animal distributions through time

Studies using the VTM data predominantly examine change in vascular plant species. However, a small but growing number of studies have utilized VTM data to understand the impact of long-term vegetation changes on animal distributions. Rubidge et al. (2011) highlighted the promise of using VTM maps with historical and resurveyed Grinnell Resurvey Project transects³ to investigate drivers of change in alpine chipmunk (*Tamias alpinus*) distributions. The detailed descriptions found on the VTM maps allowed vegetation polygons to be matched to California

Wildlife Habitat Relationships (CWHR) classes, generating comparable historical and current habitat maps for use in models of small mammal distribution changes. Though Rubidge et al. (2011) found only a limited effect of vegetation change on small mammal range shifts compared to climate, that study paved the way for the use of VTM maps in this context. As a follow-up to that study, Santos et al. (2014b) investigated the relationship between species' traits and habitat types using VTM maps as historical habitat. The authors found that omnivorous species responded in greater synchrony with shifts in their habitat types than any other diet guild.

Preston et al. (2012) used VTM maps to investigate the drivers of distribution patterns of an endangered butterfly (*Euphydryas editha quino*). The study used VTM maps to aid in the correlation of local scale Quino checkerspot population extinctions with human population growth and land use change, while climate variables determined population distributions at the broader regional scale.

The VTM maps have also been used to understand proximate land use hazards associated with commercially important species. Davis and Sims (2013) analyzed the effect of land use change and precipitation events over time on the prevalence of landslides and gullies, which contribute sediment to the San Pedro Creek watershed in San Mateo County, a key habitat for salmonids (Davis and Sims 2013).

Land use and land cover history and urban planning

Many currently urban areas in California have faced dramatic histories of land use and land cover change as agricultural and natural areas adjacent to cities have been encroached upon. The San Francisco Bay Area exemplifies this complex pattern of land cover change over the previous one hundred and fifty years and is a mosaic of urban, natural and agricultural land. Thorne et al. (2013) examined the dynamics of urban growth and the protection of open spaces—two often contentious interests—using VTM maps and urban growth models to predict how their dynamics under three contrasting urban growth policy scenarios. The authors determine that a historical trend of urban expansion will likely continue into the future, with new development converting an additional 43-53%

³ <http://mvz.berkeley.edu/Grinnell/index.html>.

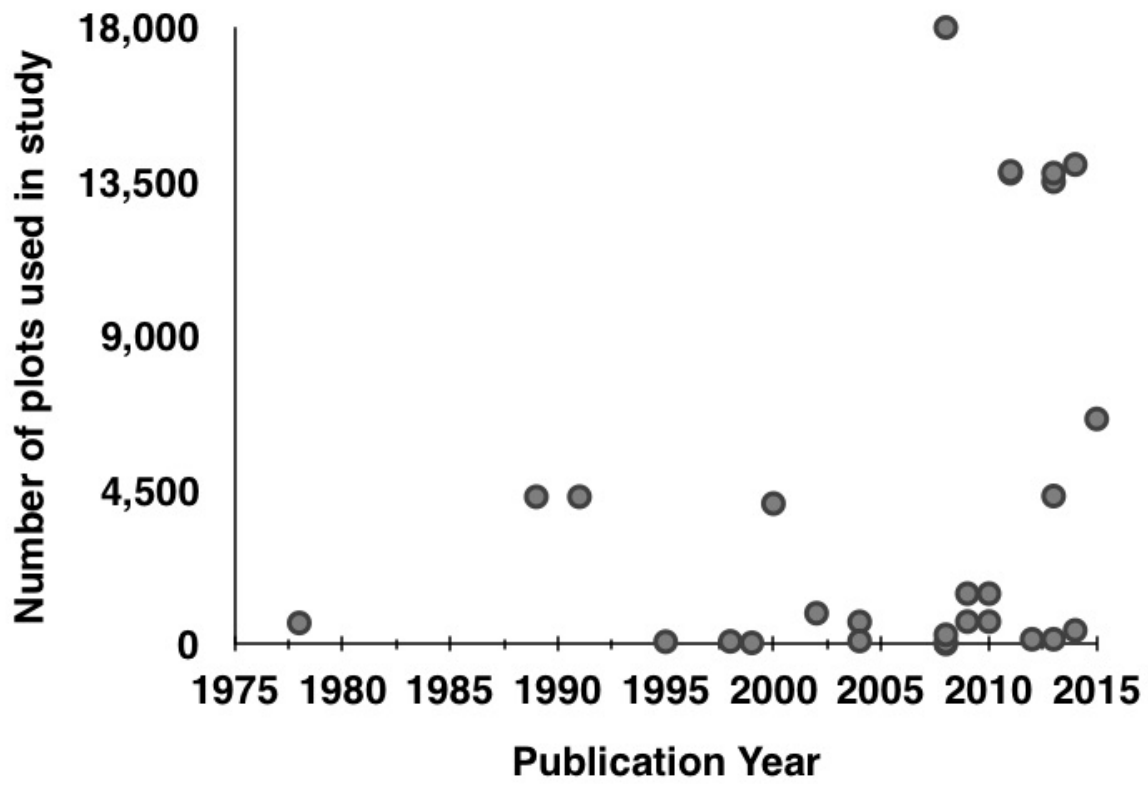


Figure 4. Increase in number of plots used in research over time.

of existing agricultural and grassland areas. However, the growth policy enforced will have a large effect on the overall impact of urban growth on open space. Santos et al. (2014a) used VTM vegetation maps to analyze the timeline of conservation land acquisition in the San Francisco Bay Area from 1850-2010. Describing the evolution of these land acquisitions and the cover types they protected, the authors found that the network of Bay Area conservation lands was highly representative of the diversity of land cover types, encompassing 20% or more of each land cover type in the region. The acquisition of land in the 19th century followed what the authors termed the “fill-in effect”: the purchase of fewer larger properties while the contemporary network was filled in through time by several smaller properties consisting of underrepresented land cover types.

Scale of VTM studies

The geographic scope of study using VTM data has grown over time, particularly after 2005 when much of the data became available in digital form (Kelly et al. 2005) (Figure 4). Studies prior to 2000 were geographically focused and used small numbers of plots (e.g. Minnich and Dezzani 1998; Bouldin 1999). Kelly et al. (2008) were the first to use the complete plot database (n = 18,000) in their study focusing on accuracy assessment. Since the release of the digitized dataset, a multitude of regional scale projects focusing on the central Sierra Nevada forests (Dobrowski et al. 2011; Dolanc et al. 2013a; Dolanc et al. 2013b) and the Sierra Nevada and the coastal ranges (Crimmins et al. 2011; Crimmins et al. 2013; Crimmins et al. 2014) have been published, as well as studies focusing on the state and using the complete plot database (Fellows and Goulden 2008; McIntyre et al. 2015). Another perhaps unforeseen result of the digitization process has been an alteration in the way change is assessed using historical and current data. Many authors have commented on difficulties in re-locating individual plots (e.g. (e.g. Allen-Diaz and Holzman 1991; Minnich et al. 1995; Walker 2000; Franklin 2002; Keeley 2004); and recent efforts have focused less on relocation of individual plots, and more on comparisons with contemporary data (McIntyre et al. 2015).

With some exceptions, a similar trend is seen in the use of vegetation maps: earlier work focused on a single quadrangle (Bradbury 1974), while

more recent work is regional in scope (Thorne et al. 2013; Santos et al. 2014a). This kind of scaling-up of focus is only realistically possible with digital infrastructures to make data discoverable, searchable, and downloadable.

DISCUSSION

Historical ecological data is often required in contemporary ecological analyses for understanding patterns, establishing baselines and past conditions, and detecting changes. These analyses are crucial when we consider planning for possible futures. As a case study, the VTM collection reveals the promise of historical data for ecological analysis. Here, we discuss prospects for increased and integrated usage of the dataset; for novel synergies with other data; and highlight through a discussion of the VTM digitization process the importance of open data frameworks for ecological analysis.

Challenges with the use of historical data

The use of historical ecological data present at least three types of challenges for the modern researcher that result from field methods used at the time of data collection, scientific terminology and taxonomy, as well as errors introduced through the digitization process. First, historical field protocols are often different from contemporary ones. The VTM crew used four size classes to bin the diameter at breast height of trees recorded in the plot data; modern Forest Inventory and Analysis (FIA) data used many more. Comparisons must deal with these differences, and manipulate modern data to best compare with historical data (*sensu* Dolanc et al. 2013a; McIntyre et al. 2015). In contrast, the VTM vegetation maps have a much higher species detail than is provided on modern land cover maps, which requires simplification of the historical vegetation classes and cross-walking of historical and modern classification schemes prior to contemporary analyses (*sensu* Thorne et al. 2008; Thorne et al. 2013). Second, ecological taxonomy has changed since the early 20th century (Barbour et al. 2007), and confusion resulting from species name changes can occur, and careful cross-walking between historical and modern species names is required. Third, as the historical data is scanned, digitized, and georeferenced, error is introduced. Best practices must be used to minimize, estimate, and report the total spatial

error in the final digital data product (Kelly et al. 2008). Such reported measures are critical to guide researchers on the use of the VTM data (Thorne et al. 2008).

Interactions between the academy and the public sector

In addition to its use in academic research over multiple decades, the VTM collection has had a parallel track of use by government resource agencies. The protocols developed by Wieslander and his crew became the foundation for additional surveys of California land including 4.6 million ha of land covered by the State Cooperative Soil-Vegetation surveys from 1947-1977 (Critchfield 1971; Thorne et al. 2007). These early surveys paved the way for many of the contemporary vegetation classification schemes used today in California, including the Manual for California Vegetation, the National Vegetation Classification System, and the California Gap Analysis Program. The CALVEG product, discussed in this paper, had its origins with VTM protocols (Barbour et al. 2007). Wieslander's belief in constructing "an all-purpose map" that "would then be valuable for any kind of management of wildland, for whatever purpose" set the foundation for large scale mapping and biological surveys even serving as the impetus for the Forest Service's "program of mapping all the national forests" (Wieslander 1986). The value of comprehensive and detailed vegetation surveys made apparent through the work of Wieslander and his crew remain a key component of agency mission statements and objectives (e.g. Forest Service, National Park Service, and California Native Plant Society). We hope that the digitization and sharing of the VTM data will continue to expand the reach of this dataset in order to bridge traditional academic and agency silos for advances in research and conservation management practices.

Increased usage of the integrated dataset

The plot data in digital form has been available since 2008, but for the first time are complete (or as complete as possible) vegetation maps and georeferenced photographs are available, and a revival in their integrated use should be encouraged. It is our hope that the combined collection made available via an API and website will facilitate their use in a number of novel ways.

First, the georeferenced photographs are an underused resource. Only two studies focus on the photography collection (Dodge 1975; Taylor 2000), yet repeat photography can provide a clear record of vegetation and landscape change. Only a handful of the georeferenced photographs have been retaken in modern times, yet these photographic pairs provide compelling stories about vegetation change (Figure 5a-f). Additionally, these images provide a window into our past and a sense of place: many of the VTM photographs are not of grand landscapes or vistas, but show the field data collection procedures used by the crews, or forest management practices they encountered in the early 20th century (Figure 5g-h). Thus these images provide unexplored cultural connections to our work and lives today (Higgs et al. 2014).

Second, the plots provide data on structure and composition of vegetation, but they also comment on the terrain, soil types, fire and disturbances, and other environmental variables found at each site. These data might be mined for more nuanced understanding of historical fire behavior (e.g. McIntyre et al. 2015) that complements other early data sources of disturbances. Third, there is great potential in the use of more than one VTM data type at a time. There is only one published example of research using more than one component of the VTM dataset (Bradbury 1974), yet researchers might consider the combined use of plots and maps, or plots and photographs, or all of the components of the collection. There are numerous areas in the state where the three data types coexist including the Sierra Nevada Mountains (e.g. Figure 1), the Central Coastal Ranges, the Transverse Ranges, and the San Bernardino Mountains (Figure 3). The Central Coast Ranges possess all components of VTM collection, and remain under-examined.

New developments in data synthesis

We should also expect and encourage new developments in data synthesis. Data synthesis necessitates an extension of the data beyond its original purpose, increasing the potential to broaden the horizons of scientific inquiry and expand the potential for discovery (Jones et al. 2006; Tenopir et al. 2011). The recent proliferation in publicly available scientific data underscores the need to better integrate and synthesize these

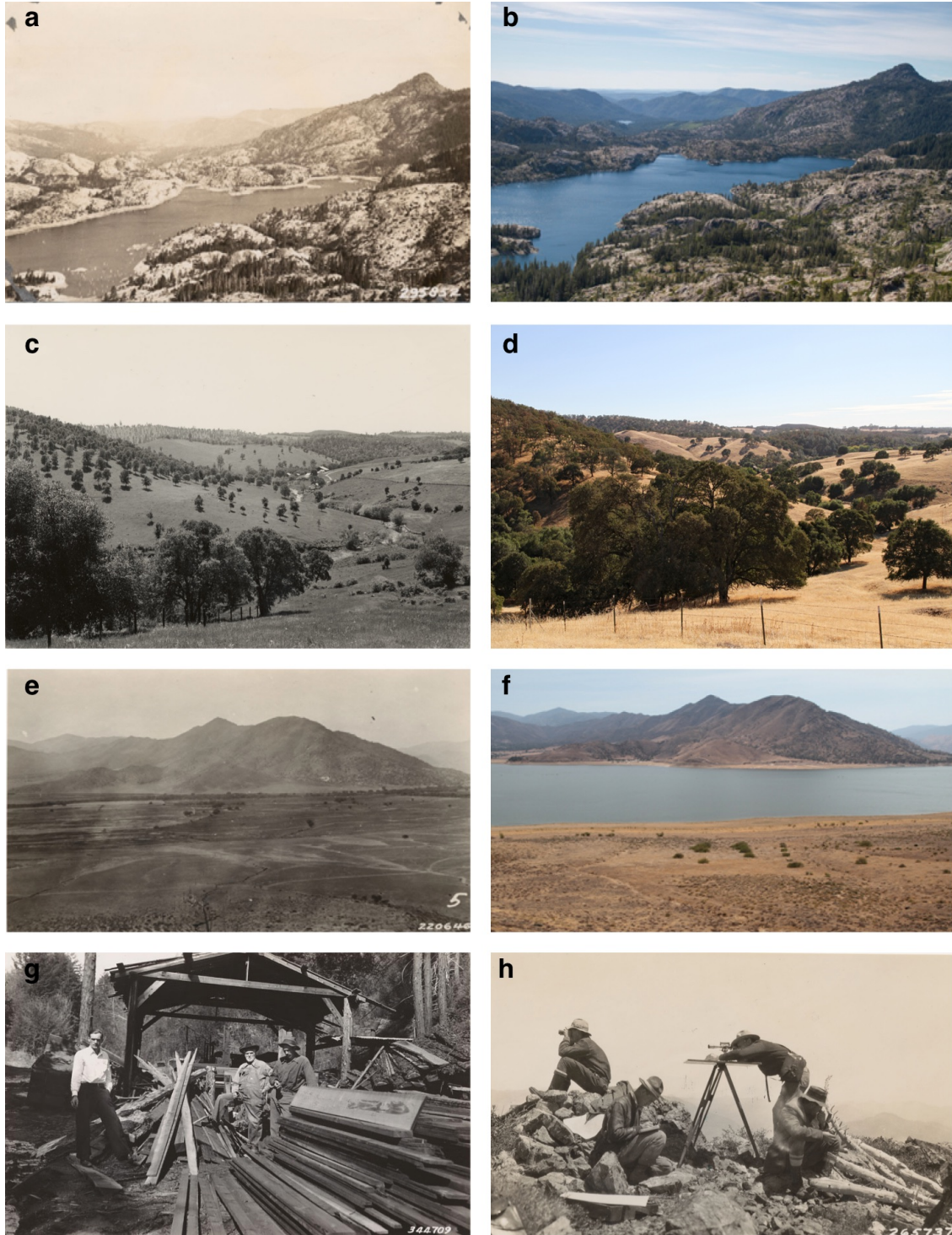


Figure 5. Example of VTM photography. Examples a-f are pairs of VTM and contemporary images: French Lake in the Tahoe National Forest in Nevada Co. taken by (a) N.H. French in June 1934, and (b) Joyce Gross in 2014; Looking southwest from highway between Drytown and Amador in Amador Co. taken by (c) A.E. Wieslander in May 1940, and (d) Joyce Gross in September 2013; and the site of what is now Lake Isabella in Kern Co., taken by (e) L.M. Correll in July 1927, and (f) Joyce Gross in June 2013. Examples g, h are examples of general interest photographs: (g) an example of a local mill and redwood lumber products cut from redwood second growth in Santa Cruz Co.; and (h) topographic mapping crew on Orleans Mountain, in Siskiyou Co.

datasets if they are to be effectively used in addressing issues such as climate and land cover and land use change which both span spatial and temporal dimensions (Tingley and Beissinger 2009). Many of the studies reviewed here employ data synthesis between historical VTM and modern vegetation data (e.g., FIA, CalVeg) to discuss change both qualitatively and quantitatively in vegetation structure (e.g. Goforth and Minnich 2008; Dolanc et al. 2013b), composition (e.g. Minnich et al. 1995), and abundance (e.g. Fellows and Goulden 2008; Table 2). There are additional datasets and approaches that might be used. The VTM database can be linked to much of the ecological data collected at UC Berkeley, among them about 5M records from museum specimen, soil and pollen data, field station records, sensor readings, as well as biophysical base layers such as climate and land use (Figure 2). There is great potential in these data synergies to further expand the scope and scale of VTM analysis. For example, first, synergies between genome sequencing of biological specimens with historical and future vegetation allow us to move beyond single species vegetation type analysis and understand the dynamic nature of ecological communities (Rubidge et al. 2011; Preston et al. 2012; Bi et al. 2013). Second, more can be done using historical vegetation data with current and future climates using new species distribution models. For example, the floristic detail found in both the maps and plot data might be used to look at historical ranges of important taxa. Third, there has been a significant focus on interactions between fire and vegetation in the VTM scholarship for example in conifer forest (Dodge 1975; Minnich 1978; Taylor 2000; Fellows and Goulden 2008; Goforth and Minnich 2008) and shrubland communities (Franklin et al. 2004; Talluto and Suding 2008; Lippitt et al. 2013). McIntyre et al. (2015) used VTM field notes to identify plots with recent fires and were able to detect coarse differences in tree density in multiple size classes between burned and unburned plots within the VTM data and across time in comparisons of burned and unburned FIA plots. More might be done to examine historical vegetation structure and pattern in modern mega fires such as the Rim fire. Conditions influencing fire overlap, fire severity and fire regimes might be explored through synthesis with contemporary fire data (Collins et al. 2007, 2009).

Open data frameworks

The multiple recent discussions in the scientific community around open, transparent, and reproducible science (Jones et al. 2006; Wolkovich et al. 2012) have fostered technological advances in more flexible and user friendly data storage and documentation systems that incentivize researchers to enter, store, and make datasets available to the broader community. Such participation has engendered a need for a more transparent system of data collection and distribution. Contemporary large-scale citizen science efforts promote and alleviate some issues of data transparency and sharing by promoting open data collection, distribution, and assessment (Kearns et al. 2003; Kelly et al. 2012). Technological advances in computer science have increased our ability to query, visualize and use large heterogeneous collections in meaningful ways (Baird 2010; Fox and Hendler 2011; Reichman et al. 2011; Hampton et al. 2013). New web applications such as APIs linked to structured ecological databases allow the rapid generation of maps, charts, timelines, graphs, word clouds, search interfaces, RSS feeds, and many others capabilities (Fox and Hendler 2011).

The voyage from paper to API

Although focused on a single dataset, the VTM collection's voyage from paper archive to API can be seen as a cautionary tale about the importance of finding, archiving, and sharing historical ecological datasets that should resonate more broadly. Despite being well-known and well-documented, in the decades since its creation, the VTM collection has faced the possibility on several occasions of partial destruction (Wieslander 1986). Even today, locations of portions of the collection remain unknown. Thus, its journey from analog to digital is exemplary of a number of important themes facing the scientific community today: the importance of finding and rescuing historical or "dark" data; the need for best practices and standards for data digitization including uncertainty estimation and error control; the value of spatial data visualization and web-based portals for data sharing; the priority in modeling of data fusion and analytical integration; and the critical role of data infrastructures such as APIs for sharing scientific data. Other important detailed records of past biological, ecological, agricultural, and management conditions may exist

across the state of California in paper archives, historical imagery, and/or physical biological specimens. These hidden “dark archives” are currently invisible to researchers, but with the kind of focused work described here, become invaluable.

CONCLUSIONS

There is ample evidence that the rescuing, digitizing, and sharing of historical ecological data is an important scientific endeavor. These data 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. A. Everett Wieslander anticipated many modern uses of the VTM data in 1935. He wrote that it provided: 1) a partial explanation of the current (as of 1930s) distribution of vegetation types and dominant species; 2) a better understanding of vegetation changes that have occurred in the past, those now in progress, or those to be expected to occur in the future; and 3) further contributions to the knowledge of the values of certain plants and vegetation types as indicators of particular soil and climatic conditions (Wieslander 1935b). He was perceptive in this analysis, but did not anticipate all of the uses of the data. The maps, plot data, and photographs have been used in isolation or paired with contemporary data to great effect to study California’s historical flora and land use, for documenting and finding mechanisms for decadal-scale vegetation changes, and for predicting and planning for California’s future.

The digitization and sharing of the VTM collection has expanded the scope and scale of possible analyses. Any analyses larger than a single or few quads were impossible when the analog data were scattered around the state in libraries and research collections; or when plot or map data required laborious digitization. Currently, most papers using VTM plot data explore the full complement of scale and detail. Yet there is more that can be done. Researchers might also use more of the vegetation map data, as well as the georeferenced photographs; they might embark on modeling that fuses data from more than one part of the collection, and synthesizes data from other sources. Additionally, we hope more researchers will explore the connections between vegetation change and other biological signals of change such

as isotopic signatures derived from spatially coincident animal specimens (e.g. Rubidge et al. 2011; Bi et al. 2013).

Finally, we want to highlight the increasingly critical role of data structures that foster scientific sharing and collaboration, such as APIs, especially in their capacity to link at-risk historical data with contemporary ecological data. The digital VTM collection is an example of a web-based data framework that expands the potential of large-scale research through the integration and synthesis of data drawn from numerous data sources. We suggest here that the potential linkages and multi-disciplinary connections waiting to be made with the use of the collection are numerous and important.

Understanding past, present, and future interrelationships between flora, fauna, land use, society, and climate is of paramount importance in ecology. The VTM dataset serves as a valuable and underutilized resource in this regard. The digital and shared data are an expansive historical reference that continues to provide exciting avenues for the modern geographer and ecologist to create connections between diverse datasets to tell the story of a region’s ecological past and better inform the future.

ACKNOWLEDGMENTS

The authors would like to thank several funders, including: the William M. Keck Foundation, Berkeley Institute for Global Change Biology, UC Division of Agriculture and Natural Resources Informatics and Geographic Information Systems Program, and the US Forest Service. Much of the work was conducted in UC Berkeley’s Geospatial Innovation Facility. Several individuals have worked on the digitization process early on; we are indebted to B. Allen-Diaz, K. Ueda, B. Thomas, T. De Chant, E. Zeledon, L. Schile, K. Tuxen-Bettman, and A. Huber, among others. We thank Joyce Gross for use of her photo retakes. We also thank Sarah Hinman and the many undergraduate students who georeferenced the photographs and specimens in the Museum of Vertebrate Zoology and the University and Jepson Herbaria.

REFERENCES

Allen, B. H., C. A. Holzman, and R. R. Evett. 1991. A classification system for California's hardwood rangelands. *Hilgardia* 59:2.

- Allen-Diaz, B. H., and B. A. Holzman. 1991. Blue oak communities in California. *Madrono* 38:80-95.
- Baird, R. C. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodivers. Infor.* 7:130-136.
- Barbour, M. G., T. Keeler-Wolf, and A. A. Schoenherr. 2007. *Terrestrial Vegetation of California*. Univ of California Press.
- Beck, J., L. Ballesteros-Mejia, C. M. Buchmann, J. Dengler, S. A. Fritz, B. Gruber, C. Hof, F. Jansen, S. Knapp, and H. Kreft. 2012. What's on the horizon for macroecology? *Ecography* 35:673-683.
- Bi, K., T. Linderoth, D. Vanderpool, J. M. Good, R. Nielsen, and C. Moritz. 2013. Unlocking the vault: next-generation museum population genomics. *Mol. Ecol.* 22:6018-6032.
- Borgman, C. L. 2012. The conundrum of sharing research data. *J. Am. Soc. Info. Sci. Tech.* 63:1059-1078.
- Bouldin, J. 1999. Twentieth-century changes in forests of the Sierra Nevada, California. *Plant Biol.* University of California, Davis.
- Bradbury, D. 1974. *Vegetation history of the Ramona quadrangle, San Diego County, California (1931-1972)*. University of California, Los Angeles, Los Angeles.
- Chornesky, E. A., D. D. Ackerly, P. Beier, F. W. Davis, L. E. Flint, J. J. Lawler, P. B. Moyle, M. A. Moritz, M. Scoonover, and K. Byrd. 2015. Adapting California's ecosystems to a changing climate. *Bioscience* 65:247-262.
- Collins, B., J. Miller, M. Kelly, J. W. van Wagtenonk, and S. L. Stephens. 2009. Interactions among wildland fires in a long-established Sierra Nevada natural fire area. *Ecosystems* 12:114-128.
- Collins, B. M., M. Kelly, J. W. van Wagtenonk, and S. L. Stephens. 2007. Spatial patterns of large natural fires in Sierra Nevada wilderness areas. *Landscape Ecol.* 22:545-557.
- Colwell, W. L. 1977. *The status of vegetation mapping in California today*. John Wiley & Sons, Sacramento.
- Conlisk, E., D. Lawson, A. D. Syphard, J. Franklin, L. Flint, A. Flint, and H. M. Regan. 2012. The roles of dispersal, fecundity, and predation in the population persistence of an oak (*Quercus engelmannii*) under global change. *PLoS ONE* 7:e36.
- Cox, R. D., K. L. Preston, R. F. Johnson, R. A. Minnich, and E. B. Allen. 2014. Influence of landscape-scale variables on vegetation conversion to exotic annual grassland in southern California, USA. *Global Ecol. Conserv.* 2:190-203.
- Crimmins, S. M., S. Z. Dobrowski, J. A. Greenberg, J. T. Abatzoglou, and A. R. Mynsberge. 2011. Changes in Climatic Water Balance drive downhill shifts in plant species' optimum elevations. *Science* 331:324-327.
- Crimmins, S. M., S. Z. Dobrowski, and A. R. Mynsberge. 2013. Evaluating ensemble forecasts of plant species distributions under climate change. *Ecol. Model.* 266:126-130.
- Crimmins, S. M., S. Z. Dobrowski, A. R. Mynsberge, and H. D. Safford. 2014. Can fire atlas data improve species distribution model projections? *Ecol. Appl.* 24:1057-1069.
- Critchfield, W. B. 1971. *Profiles of California Vegetation*. Pp. 54. Pacific Southwest Forest & Range Experiment Station, Forest Service, U.S. Department of Agriculture, Berkeley, CA.
- Davis, J. D., and S. M. Sims. 2013. Physical and maximum entropy models applied to inventories of hillslope sediment sources. *J. Soils Sed.* 13:1784-1801.
- Dobrowski, S. Z., J. H. Thorne, J. A. Greenberg, H. D. Safford, A. R. Mynsberge, S. M. Crimmins, and A. K. Swanson. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. *Ecol. Monogr.* 81:241-257.
- Dodge, J. M. 1975. *Vegetational changes associated with land use and fire history in San Diego County*. University of California, Riverside.
- Dolanc, C. R., H. D. Safford, S. Z. Dobrowski, and J. H. Thorne. 2013a. Twentieth century shifts in abundance and composition of vegetation types of the Sierra Nevada, CA, US. *Appl. Veg. Sci.* 17:442-455.
- Dolanc, C. R., H. D. Safford, J. H. Thorne, and S. Z. Dobrowski. 2014. Changing forest structure across the landscape of the Sierra Nevada, CA, USA, since the 1930s. *Ecosphere* 5:1-26.
- Dolanc, C. R., J. H. Thorne, and H. D. Safford. 2013b. Widespread shifts in the demographic structure of subalpine forests in the Sierra Nevada, California, 1934 to 2007. *Global Ecol. Biogeogr.* 22:264-276.
- Elith, J., C. H. Graham, R. P. Anderson, M. Dudik, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. M. Overton, A. T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberon, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.
- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol., Evol. Syst.* 40:677.

- Ertter, B. 2000. Our undiscovered heritage: Past and future prospects for species-level botanical inventory. *Madrono* 47:237-252.
- Fellows, A. W., and M. L. Goulden. 2008. Has fire suppression increased the amount of carbon stored in western U.S. forests? *Geophys. Res. Lett.* 35:L12404.
- Fox, P., and J. Hendler. 2011. Changing the equation on scientific data visualization. *Science* 331:705-708.
- Franklin, J. 2002. Enhancing a regional vegetation map with predictive models of dominant plant species in chaparral. *Appl. Veg. Sci.* 5:135-146.
- Franklin, J., C. L. Coulter, and S. J. Rey. 2004. Change over 70 years in a southern California chaparral community related to fire history. *J. Veg. Sci.* 15:701-710.
- Frehner, M., and M. Braendli. 2006. Virtual database: Spatial analysis in a Web-based data management system for distributed ecological data. *Environ. Model. Software* 21:1544-1554.
- Freudenberger, D. O., B. E. Fish, and J. E. Keeley. 1987. Distribution and stability of grasslands in the Los Angeles Basin. *Bull. South. Calif. Acad. Sci.* 86:13-26.
- Galatowitsch, S. M. 1990. Using the original land survey notes to reconstruct presettlement landscapes in the American West. *Great Basin Nat.* 50:181-191.
- Goforth, B. R., and R. A. Minnich. 2008. Densification, stand-replacement wildfire, and extirpation of mixed conifer forest in Cuyamaca Rancho State Park, Southern California. *For. Ecol. Manage.* 256:36-45.
- Graham, C. H., S. Ferrier, F. Huettman, C. Moritz, and A. T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* 19:497-503.
- Griffin, J. R., and W. B. Critchfield. 1972. The distribution of forest trees in California. U.S.D.A. Forest Service, Pacific Southwest Forest and Range Experiment Station, Berkeley, CA.
- Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8:993-1009.
- Guo, Q., M. Kelly, and C. H. Graham. 2005. Support Vector Machines for predicting distribution of sudden oak death in California. *Ecol. Model.* 182:75-90.
- Hampton, S. E., C. A. Strasser, J. J. Tewksbury, W. K. Gram, A. E. Budden, A. L. Batcheller, C. S. Duke, and J. H. Porter. 2013. Big data and the future of ecology. *Front. Ecol. Environ.* 11:156-162.
- Higgs, E., D. A. Falk, A. Guerrini, M. Hall, J. Harris, R. J. Hobbs, S. T. Jackson, J. M. Rhemtulla, and W. Throop. 2014. The changing role of history in restoration ecology. *Front. Ecol. Environ.* 12:499-506.
- Jepson, W. L., R. Beidleman, and B. Ertter. 2000. Willis Linn Jepson's "Mapping in Forest Botany". *Madrono* 47:269-272.
- Jones, M. B., M. P. Schildhauer, O. Reichman, and S. Bowers. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. *Annu. Rev. Ecol., Evol. Syst.* 37:519-544.
- Kearns, F. R., M. Kelly, and K. A. Tuxen. 2003. Everything happens somewhere: using webGIS as a tool for sustainable natural resource management. *Front. Ecol. Environ.* 1:541-548.
- Keeler-Wolf, T. 2007. The history of vegetation classification and mapping in California. *Terrestrial Vegetation of California*. University of California Press, Berkeley:1-42.
- Keeley, J. E. 2004. VTM plots as evidence of historical change: goldmine or landmine? *Madrono* 51:372-378.
- Kelly, A. E., and M. L. Goulden. 2008. Rapid shifts in plant distribution with recent climate change. *Proc. Natl. Acad. Sci.* 105:11823-11826.
- Kelly, M., B. Allen-Diaz, and N. Kobzina. 2005. Digitization of a historic dataset: the Wieslander California vegetation type mapping project. *Madrono* 52:191-201.
- Kelly, M., S. Ferranto, S. Lei, K. Ueda, and L. Huntsinger. 2012. Expanding the table: The web as a tool for participatory adaptive management in California forests. *J. Environ. Manage.* 109:1-11.
- Kelly, M., K. Ueda, and B. Allen-Diaz. 2008. Considerations for ecological reconstruction of historic vegetation: Analysis of the spatial uncertainties in the California Vegetation Type Mapping dataset. *Plant Ecol.* 194:37-49.
- Kittinger, J. N., K. S. V. Houtan, L. E. McClenachan, and A. L. Lawrence. 2013. Using historical data to assess the biogeography of population recovery. *Ecography* 36:868-872.
- Krebs, C. J., R. Boonstra, S. Boutin, and A. R. E. Sinclair. 2001. What drives the 10-year cycle of snowshoe hares? *Bioscience* 51:25-35.
- Kuchler, A. W. 1967. *Vegetation Mapping*. Ronald Press, New York.
- Leiberg, J. B. 1902. *Forest conditions in the northern Sierra Nevada, California*. Department of the Interior, US Geological Survey.
- Lenoir, J., J.-C. Gégout, P. Marquet, P. De Ruffray, and H. Brisse. 2008. A significant upward shift in plant species optimum elevation during the 20th century. *Science* 320:1768-1771.
- Lippitt, C. L., D. A. Stow, J. F. O'Leary, and J. Franklin. 2013. Influence of short-interval fire occurrence on post-fire recovery of fire-prone

- shrublands in California, USA. *Int. J. Wildland Fire* 22:184-193.
- Loarie, S. R., B. E. Carter, K. Hayhoe, S. McMahon, R. Moe, C. A. Knight, and D. D. Ackerly. 2008. Climate Change and the Future of California's Endemic Flora *PLOS One* 3:e2502.
- Lutz, J. A., J. W. van Wagtenonk, and J. F. Franklin. 2009. Twentieth-century decline of large-diameter trees in Yosemite National Park, California, USA. *For. Ecol. Manage.* 257:2296-2307.
- Lutz, J. A., J. W. van Wagtenonk, and J. F. Franklin. 2010. Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. *J. Biogeogr.* 37:936-950.
- Maxwell, R. S., A. H. Taylor, C. N. Skinner, H. D. Safford, R. E. Isaacs, C. Airey, and A. B. Young. 2014. Landscape-scale modeling of reference period forest conditions and fire behavior on heavily logged lands. *Ecosphere* 5:1-28.
- Mayer, K. E., and W. F. Laudenslayer. 1988. A guide to wildlife habitats of California. Pp. 166. State of California, Resources Agency, Department of Fish and Game, Sacramento, CA.
- McIntyre, P. J., J. H. Thorne, C. R. Dolanc, A. L. Flint, L. E. Flint, M. Kelly, and D. D. Ackerly. 2015. Twentieth-century shifts in forest structure in California: Denser forests, smaller trees, and increased dominance of oaks. *Proc. Natl. Acad. Sci.* 112:1458-1463.
- McMinn, H. 1951. An illustrated manual of California shrubs. Univ of California Press.
- Michener, W. K. 2006. Meta-information concepts for ecological data management. *Ecol. Infor.* 1:3-7.
- Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. *Ecol. Appl.* 7:330-342.
- Minnich, R., and R. Dezzani. 1998. Historical decline of coastal sage scrub in the riverside-perris plain, California. *West. Birds* 29:366-391.
- Minnich, R. A. 1978. The geography of fire and conifer forests in the eastern Transverse Ranges, California. University of California, Los Angeles.
- Minnich, R. A., M. G. Barbour, J. H. Burk, and R. F. Fernau. 1995. Sixty years of change in Californian conifer forests of the San Bernardino Mountains. *Conserv. Biol.* 9:902-914.
- Mladenoff, D. J., S. E. Dahir, E. V. Nordheim, L. A. Schulte, and G. G. Guntenspergen. 2002. Narrowing historical uncertainty: probabilistic classification of ambiguously identified tree species in historical forest survey data. *Ecosystems* 5:539-533.
- Peters, D. P. 2010. Accessible ecology: synthesis of the long, deep, and broad. *Trends in Ecology and Evolution* 25:592-601.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190:231-259.
- Preston, K. L., R. A. Redak, M. F. Allen, and J. T. Rotenberry. 2012. Changing distribution patterns of an endangered butterfly: Linking local extinction patterns and variable habitat relationships. *Biol. Conserv.* 152:280-290.
- Rapacciuolo, G., D. B. Roy, S. Gillings, and A. Purvis. 2014a. Temporal validation plots: quantifying how well correlative species distribution models predict species' range changes over time. *Methods Ecol. Evol.* 5:407-420.
- Rapacciuolo G., S.P. Maher, A.C. Schneider, T.T. Hammond, M.D. Jabis, R.E. Walsh, K.J. Iknayan, G.K. Walden, M.F. Oldfather, D.D. Ackerly, and S.R. Beissinger. 2014b. Beyond a warming fingerprint: individualistic biogeographic responses to heterogeneous climate change in California. *Global Change Biology* 20: 2841–2855.
- Reichman, O., M. B. Jones, and M. P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331:703-705.
- Rubidge, E. M., W. B. Monahan, J. L. Parra, S. E. Cameron, and J. S. Brashares. 2011. The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. *Global Change Biol.* 17:696-708.
- Santos, M. J., J. H. Thorne, J. Christensen, and Z. Frank. 2014a. An historical land conservation analysis in the San Francisco Bay Area, USA: 1850–2010. *Landscape Urban Plann.* 127:114-123.
- Santos, M. J., J. H. Thorne, and C. Moritz. 2014b. Synchronicity in elevation range shifts among small mammal and vegetation over the last century is stronger for omnivores. *Ecography* 37:001-013.
- Sawyer, J., and T. Keeler-Wolf. 1995. A Manual of California Vegetation. California Native Plant Society Press, Sacramento, California.
- Schulte, L. A., and D. J. Mladenoff. 2001. The original US Public Land Survey records: Their use and limitations in reconstructing presettlement vegetation. *J. For.* October:5-10.
- Schwind, B., and H. Gordon. 2001. Calveg geobook: a comprehensive information package describing California's wildland vegetation, Version 2. USDA Forest Service, Pacific Southwest Region, Remote Sensing Lab, Sacramento, CA.
- Solé, R. V., and J. Bascompte. 2006. Self-Organization in Complex Ecosystems. Princeton University Press.
- Stein, E. D., S. Dark, T. Longcore, R. Grossinger, N. Hall, and M. Beland. 2010. Historical ecology as a tool for assessing landscape change and informing

- wetland restoration priorities. *Wetlands* 30:589-601.
- Swanson, A. K., S. Z. Dobrowski, A. O. Finley, J. H. Thorne, and M. K. Schwartz. 2013. Spatial regression methods capture prediction uncertainty in species distribution model projections through time. *Global Ecol. Biogeogr.* 22:242-251.
- Syphard, A. D., and J. Franklin. 2009. Differences in spatial predictions among species distribution modeling methods vary with species traits and environmental predictors. *Ecography* 32:907-918.
- Syphard, A. D., and J. Franklin. 2010. Species traits affect the performance of species distribution models for plants in southern California. *J. Veg. Sci.* 21:177-189.
- Talluto, M. V., and K. N. Suding. 2008. Historical change in coastal sage scrub in southern California, USA in relation to fire frequency and air pollution. *Landscape Ecol.* 23:803-815.
- Taylor, A. H. 2000. Fire regimes and forest changes in mid and upper montane forests of the southern Cascades, Lassen Volcanic National Park, California, USA. *J. Biogeogr.* 27:87-104.
- Tenopir, C., S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame. 2011. Data sharing by scientists: practices and perceptions. *PLOS One* 6:e21101.
- Thorne, J., T. R. Kelsey, J. Honig, and B. Morgan. 2007. The Development of 70-Year-Old Wieslander Vegetation Type Maps and an Assessment of Landscape Change in the Central Sierra Nevada. California Energy Commission, Sacramento, CA.
- Thorne, J. H., J. A. Kennedy, J. F. Quinn, M. McCoy, T. Keeler-Wolf, and J. Menke. 2004. A vegetation map of Napa County using the manual of California vegetation classifications and its comparison to other digital vegetation maps. *Madrono* 51:343-363.
- Thorne, J. H., and T. N. g. Le. In press. The Wieslander Vegetation Type Maps, an historic legacy for California landscape change analyses. *Madrono*.
- Thorne, J. H., B. J. Morgan, and J. A. Kennedy. 2008. Vegetation change over sixty years In the Central Sierra Nevada, California, USA. *Madrono* 55:223-237.
- Thorne, J. H., M. J. Santos, and J. H. Bjorkman. 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:e65258.
- Tingley, M. W., and S. R. Beissinger. 2009. Detecting range shifts from historical species occurrences: new perspectives on old data. *Trends in Ecology and Evolution* 24:625-633.
- United States Forest Service. 2016a. CalVeg (Classification and Assessment with Landsat of Visible Ecological Groupings).
- United States Forest Service. 2016b. Pacific Northwest Forest Inventory and Analysis.
- Vayssières, M. P., R. E. Plant, and B. H. Allen-Diaz. 2000. Classification trees: An alternative non-parametric approach for predicting species distributions. *J. Veg. Sci.* 11:679-694.
- Walker, R. E. 2000. Investigations in vegetation map rectification, and the remotely sensed detection and measurement of natural vegetation changes. Pp. 249. *Geography*. University of California, Santa Barbara, Santa Barbara, CA.
- Whipple, A. A., R. M. Grossinger, and F. W. Davis. 2011. Shifting baselines in a California oak savanna: nineteenth century data to inform restoration scenarios. *Restor. Ecol.* 19:88-101.
- Wieslander, A. E. 1935a. The first steps of the forest survey in California. *J. For.* 33:877-884.
- Wieslander, A. E. 1935b. A vegetation type map of California. *Madrono* 3:140-144.
- Wieslander, A. E. 1961. California's vegetation maps: Recent advances in botany. Pp. 4. University of Toronto Press, Toronto, Canada.
- Wieslander, A. E. 1986. A.E. Wieslander, California forester: mapper of wildland vegetation and soils (an oral history conducted in 1985 by Ann Lange). Regional Oral History Office, Bancroft Library, University of California, Berkeley, Berkeley, CA.
- Wolf, A., N. B. Zimmerman, W. R. Anderegg, P. E. Busby, and J. Christensen. 2016. Altitudinal shifts of the native and introduced flora of California in the context of 20th-century warming. *Global Ecol. Biogeogr.*
- Wolkovich, E. M., J. Regetz, and M. I. O'Connor. 2012. Advances in global change research require open science by individual researchers. *Global Change Biol.* 18:2102-2110.