Ecological Informatics 31 (2016) 100–111

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Challenges and opportunities in synthesizing historical geospatial data using statistical models

M.V. Eitzel [a,*], Maggi Kelly [a], Iryna Dronova [b], Yana Valachovic [c], Lenya Quinn-Davidson [c], Jon Solera [d], Perry de Valpine [a]

[a] Department of Environmental Science, Policy, and Management, University of California, Berkeley, 130 Mulford Hall #3114, Berkeley, CA 94720-3114, USA
[b] Department of Landscape Architecture and Environmental Planning, University of California, Berkeley, 230 Wurster Hall #1820, Berkeley, CA 94720-1820, USA
[c] University of California Cooperative Extension, 5630 South Broadway, Eureka, CA 95503, USA
[d] Seven Points Consulting, 900 San Bruno Ave, Brisbane, CA 94005, USA

ARTICLE INFO

Article history:
Received 4 October 2015
Received in revised form 17 November 2015
Accepted 23 November 2015
Available online 2 December 2015

Keywords:
Spatial autocorrelation
Object-Based Image Analysis
Definiens
Black and white images
Scaling
Generalized Additive Models

ABSTRACT

We classified land cover types from 1940s historical aerial imagery using Object Based Image Analysis (OBIA) and compared these maps with data on recent cover. Few studies have used these kinds of maps to model drivers of cover change, partly due to two statistical challenges: 1) appropriately accounting for spatial autocorrelation and 2) appropriately modeling percent cover which is bounded between 0 and 100 and not normally distributed. We studied the change in woody cover at four sites in California's North Coast using historical (1948) and recent (2009) high spatial resolution imagery. We classified the imagery using eCognition Developer and aggregated the resulting maps to the scale of a Digital Elevation Model (DEM) in order to understand topographic drivers of woody cover change. We used Generalized Additive Models (GAMs) with a quasi-binomial probability distribution to account for spatial autocorrelation and the boundedness of the percent woody cover variable. We explored the relative influences on current percent woody cover of topographic variables (grouped using principal component analysis) reflecting water retention capacity, exposure, and within-site context, as well as historical percent woody cover and geographical coordinates. We estimated these models for pixel sizes of 20, 30, 40, 50, 60, 70, 80, 90, and 100 m, reflecting both tree neighborhood scales and stand scales. We found that historical woody cover had a consistent positive effect on current woody cover, and that the spatial autoregressive term in the model was significant even after controlling for historical cover. Specific topographic variables emerged as important for different sites at different scales, but no overall pattern emerged across sites or scales for any of the topographic variables we tested. This GAM framework for modeling historical data is flexible and could be used with more variables, more flexible relationships with predictor variables, and larger scales. Modeling drivers of woody cover change from historical ecology data sources can be a valuable way to plan restoration and enhance ecological insight into landscape change.

1. Introduction

Historical ecology is a flourishing interdisciplinary area of study concerned with the reconstruction of landscapes from decades to centuries ago, often for the purpose of setting restoration targets (Grossinger, 2012; Sanderson, 2009; Swetnam et al., 1999). Increasingly, the field is shifting from descriptions for restoration targets to include quantitative modeling of long term landscape change (Whipple et al., 2011). This development allows historical ecology to contribute to current understandings of the long term effects of global change. Our goal with this study was to explore the challenges and opportunities inherent in the quantitative modeling of historical ecological data using relatively easily available datasets, and to investigate ways to assess the validity of the resulting models in the absence of ground-truth information. We used these methods to investigate vegetation change (specifically, forest densification) in coastal Northern California.

We used quantitative modeling of woody cover change from historical imagery to ask and answer questions regarding the topographic determinants of forest densification. The process of densification has many undesirable consequences for forest ecosystem services, such as increased fuel continuity and subsequent fire hazard, decreased

Abbreviations: OBIA, Object-Based Image Analysis; DEM, Digital Elevation Model; GAM, Generalized Additive Model; NAIP, National Agricultural Imagery Program; NDVI, Normalized Difference Vegetation Index; GLCM, Gray-level Co-occurrence Matrix; MRF, Markovian Random Field; GIS, Geographic Information Systems; RMSE, Root Mean Squared Error.

* Corresponding author at: Science and Justice Research Center, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA. Tel.: +1 831 459 3837; fax: +1 831 459 3518.
E-mail address: mveitzel@ucsc.edu (M.V. Eitzel).

http://dx.doi.org/10.1016/j.ecoinf.2015.11.011

heterogeneity and decreased resiliency, decreases in biodiversity due to decreased light availability on the forest floor, and compromised tree health due to more intense resource competition (Hanberry et al., 2014; Knapp et al., 2013). This process is widespread in California due to fire exclusion policies dating back more than 100 years (Laudenslayer and Darr, 1990). Topographic variables representing water retention capacity and exposure (including slope, aspect, elevation, curvature, solar radiation, and topographic wetness index) have long been used to predict vegetation characteristics (Deng et al., 2007; Franklin, 1995; Jenkins and Coops, 2011). Though they themselves are static, topographic variables can reflect underlying drivers such as solar exposure and moisture accumulation, which could modify the dynamic effects of climate change on cover. Many of these variables are easily calculated from a Digital Elevation Model (DEM) and are often used in GIS modeling studies. Our goal was to model the topographic determinants of current woody cover while using historical woody cover from aerial photographs to account for historical conditions, and to compare the importance of these factors in predicting current woody cover.

Historical aerial imagery, typically dating back to the 1930s and 1940s, is available throughout North America (Morgan and Gergel, 2013) and is often an important data source for historical ecology. Though the imagery can be difficult to find and often requires extensive pre-processing, classification of high spatial resolution black-and-white images has become more common with the commercial availability of Object Based Image Analysis (OBIA) software. OBIA allows analysts to use textural and contextual information in classifying single band images, and the use of OBIA with historical aerial photos has expanded in the last 10 years (Allard et al., 2012; Laliberte et al., 2004; Marignani et al., 2008; Martha et al., 2012; Pringle et al., 2009). Generally speaking, most of the OBIA change detection literature is focused on innovations in mapping techniques and their application to many different systems (Conchedda et al., 2008; de Chant and Kelly, 2009; Desclée et al., 2006; Dronova et al., 2011; Stow et al., 2008). The new goal, however, is not just to map the change but to understand the drivers of the mapped change, an interdisciplinary project involving both modeling and historical ecology (Gimmi and Bugmann, 2013). Among studies using OBIA to classify historical aerial imagery, only a few model the drivers of change (Cserhalmi et al., 2011; Garbarino et al., 2013; Levick and Rogers, 2011; Newman et al., 2014a,2014b; Platt and Schoennagel, 2009).

One issue that arises in combining historical aerial imagery with DEM-derived topographic variables is the problem of scale mismatch between ecological processes and data sources as well as between different data sources. Though geospatial data are becoming available on finer and finer spatial scales, the available data are often at an arbitrary resolution that is more constrained by data acquisition than the process of interest (Deng et al., 2007). Different ecological processes may act at different spatial scales (i.e., raster cell sizes) and different hierarchical organization levels (e.g., individual tree, neighborhood, stand, site, landscape). For instance, a tree may compete for light with other trees in its immediate neighborhood, but moisture accumulation may be a feature of the topography underlying an entire stand of trees. These hierarchical levels may not match spatial scales, and thresholds in the importance of different variables may appear where emergent properties arise (Bissonette, 1997). Ecologically, it would be ideal to explore the effects of densification at multiple levels of the hierarchy, from the tree neighborhood to the forest stand. Recent efforts to study changing forest responses at multiple spatial scales use simulation as a way to achieve this goal (Seidl et al., 2013). Empirical studies on scaling relationships for vegetation patterns so far have only correlated topographic variables with vegetation indices at a range of spatial scales rather than testing multiple variables at once while incorporating spatial autocorrelation (Deng et al., 2007). Methodologically, there is a need for data driven ecosystem modeling using appropriate statistical models and especially for scale sensitivity analysis of these models. Parameters for variables that are clearly important will theoretically be consistent in magnitude, direction, and significance for a range of cell sizes within an ecological scale. Parameter instability over a small range of cell sizes may indicate sensitivity to the particulars of the dataset. We therefore conducted our analysis at a range of cell sizes in order to assess scale-sensitivity.

In this study, we used object based image analysis on high spatial resolution images to map 1948 (historical) and 2009 (recent) woody cover at four sites in northern California, USA. We modeled recent cover as a function of topographic variables and historical cover using a quasi-binomial Generalized Additive Model (GAM) with a nonparametric smooth function of the spatial coordinates. We used these models at a range of raster cell sizes to answer the following questions:

1. Did woody cover increase more at wetter sites (those with higher annual rainfall)?
2. Did variables representing water retention capacity, exposure, and local context within the site demonstrate significant and ecologically reasonable relationships with recent woody cover, after controlling for historical woody cover?
3. Were these relationships stronger at the neighborhood scale or at the stand scale, and was a threshold effect apparent between the two scales (Fig. 1)?
4. Were these results stable over several cell sizes within an ecological scale?

2. Methods

2.1. Study areas

Four research sites were established in Northern California, primarily in Humboldt County. The sites have a Mediterranean climate, with cool, wet winters and hot, dry summers. Oak woodlands at our sites in Humboldt County are characterized by California black oak (*Quercus kelloggii*) and Oregon white oak (*Quercus garryana*) with an understory predominantly composed of grasses and forbs. Densification from woodland (defined as more than 30% cover with 150–300 trees/ha, Agee, 1993) to closed canopy forest (greater than 300 trees/ha, Agee, 1993) can occur when mature oak canopies expand through annual growth, but it more commonly occurs when evergreen species, typically Douglas-fir (*Pseudotsuga menziesii*), encroach into woodlands over time, forming a dense, shaded forest with little to no herbaceous understory. This represents an ecosystem type change with many consequences for biodiversity, forage production, and fire behavior (Engber et al., 2011; Livingston, 2014; Thysell and Carey, 2001). Our sites were chosen to represent several different latitudes and distances from the coast where densification was known to occur: Iaqua Buttes, Bald Hills, Willow Creek, and Blake Mountain (Fig. 2, Table 1). Analysis polygons within

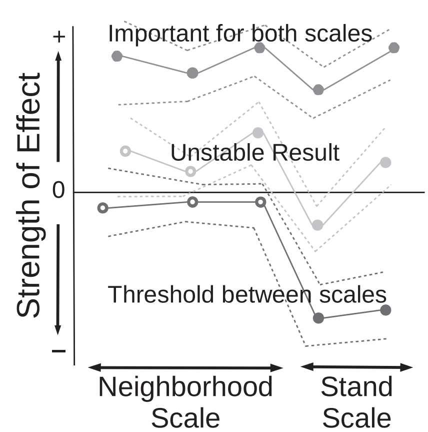

Fig. 1. Diagram of scaling effects. Dashed lines show confidence intervals, solid dots are significant, while hollow dots are not significant (confidence interval overlaps zero). Each dot is a raster cell size. Some parameters may be important for both scales, while others show instability from scale to scale; and still others might indicate a threshold of importance between the two scales indicating the potential for an emergent property.