Ecological Informatics 41 (2017) 116–132

Contents lists available at ScienceDirect

Ecological Informatics

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

Modeling the potential natural vegetation of Minnesota, USA

Seth R. Fore*, Michael J. Hill

Department of Earth System Science and Policy, University of North Dakota, Grand Forks, ND 58201, USA

ARTICLE INFO

Keywords:
Species distribution modeling
Random Forests
Potential natural vegetation
Ecological modeling

ABSTRACT

Assessing the effects of human land use and management decisions requires an understanding of how temporal changes in biodiversity influence the rate of ecosystem functions and subsequent delivery of ecosystem services. In highly modified anthromes, the spatial distribution of natural vegetation types is often unknown or coarsely represented challenging comparative analyses seeking to assess changes in biodiversity and potential downstream effects on ecosystem processes and functions. In this context, the objectives of this study were to construct a multi-resolution representation of potential natural vegetation at four hierarchical classification levels of increasing floristic and physiognomic detail for the state of Minnesota, USA. Using a collection of natural/near-natural vegetation relevés, a series of Random Forest classification models were used to project the potential distribution of natural vegetation types based on their association with a variety of environmental variables.

Model performance varied within and between classification levels with overall accuracy ranging between 64–99% (kappa 0.44–0.99). Model performance tended to decrease and become more variable with increasing floristic complexity at finer classification levels. Classwise performance metrics including precision and sensitivity were also reported. A method for exploring potential class confusion resulting from niche overlap using Random Forest proximities and Nonmetric Multidimensional Scaling is demonstrated.

Collectively, the results presented here provide an analytically supported baseline representation of potential natural vegetation for the state of Minnesota, USA. These data can provide a backdrop to further analyses surrounding the influence of human activity on ecosystem processes and services as well as inform future conservation and restoration efforts.

1. Introduction

There is widespread interest in understanding the consequences of human activity on the biosphere as a means to inform and facilitate optimal land use and conservation decisions among public and private land managers (Bennett et al., 2015; de Groot et al., 2010; DeFries et al., 2004; Fleishman et al., 2011; Fontana et al., 2013; Millennium Ecosystem Assessment, 2005; Polasky et al., 2010). Effective management approaches require understanding how ecosystem processes influence downstream ecosystem services as well as identifying the physical attributes and mechanisms through which biotic and abiotic components interact and regulate these processes (Balvanera et al., 2006; Cardinale et al., 2012; Haines-Young, 2009; Harrison et al., 2014; Mayfield et al., 2005). Plant functional traits have emerged as an effective approach for quantitatively relating the morphological, physiological and phenological attributes of species (collectively referred to as species traits, Violle et al., 2007) to their effect on ecosystem processes and their response to changes in environmental factors and human perturbations (Cornelissen et al., 2003; Díaz et al., 2004; Lavorel, 2012; Lavorel and Garnier, 2002; Weiher et al., 1999). There is widespread agreement that functional diversity, or the identity, value and range of values of functional traits represented by species in a community (Tilman, 2001), is the dominant driver of ecosystem processes and subsequent ecosystem services (Díaz and Cabido, 2001; Díaz et al., 2007; Lavorel, 2012). With respect to land cover and land use change (LCLUC), numerous studies have demonstrated both in concept and in practice the ability to assess the impact of land management alternatives on ecosystem processes and subsequent services by examining their effect on local functional diversity (Díaz et al., 2007; Lavorel, 2012; Lavorel et al., 2010; McIntyre, 2008; McIntyre and Lavorel, 2007; Seppelt et al., 2013; Van Oudenhoven et al., 2012).

While a variety of methods have proven useful in examining the potential consequences of contemporary LCLUC, extending similar concepts to highly developed anthromes with a long-term legacy of LCLUC faces unique challenges. In many cases, an ecologically relevant baseline approximation of vegetation against which changes in

---

Abbreviations: MN, Minnesota; MN DNR, Minnesota Department of Natural Resources; LCLUC, Land cover and land use change; PNV, Potential natural vegetation; RF, Random Forests; NMDS, Non-metric multidimensional scaling; m, Meters; C, Celsius; CV, Cross-validation

* Corresponding author.
E-mail address: seth.fore@icloud.com (S.R. Fore).

http://dx.doi.org/10.1016/j.ecoinf.2017.07.006
Received 14 October 2016; Received in revised form 25 July 2017; Accepted 29 July 2017

landscape structure and ecological processes can be assessed is absent. Potential natural vegetation (PNV) maps have been proposed as useful null model or reference in such scenarios (Ricotta et al., 2002, 2000).

PNV is the vegetation that would become established if all human influence was to cease at the site and its immediate surroundings and if all successional sequences were completed instantaneously under current climatic and environmental conditions (Mueller-Dombois and Ellenberg, 1974; Tüxen, 1956; Westhoff and van der Maarel, 1978). Mapping PNV first involves delineating classes based on existing natural/near-natural vegetation remnants and then projecting their distribution throughout an area of interest based the observed correlation between sites of each class and their environment (Moravec, 1998). The resulting PNV map reflects a hypothesized distribution of natural vegetation for the area (Loidi et al., 2010). In this context, PNV is neither a reflection of pre-anthropic vegetation nor a prediction of future vegetation (Zerbe, 1998). Rather, PNV is a theorized abstraction that, while unobservable in reality, can serve as a useful reference or null model for variety of comparative ecological analyses in the absence of alternate empirical data (Farris et al., 2010; Loidi et al., 2010; Loidi and Fernández-González, 2012; Somodi et al., 2012).

In certain cases, the overall utility of PNV data may be limited due to insufficient spatiotemporal scale, inadequate floristic detail or lack of reproducibility (Zerbe, 1998). Instances where PNV is derived from deterministic models that do not provide a measure of uncertainty or means of assessing possible alternative states may further reduce the usefulness of available datasets (Somodi et al., 2012).

Aside from analytical limitations, the conceptual usefulness of PNV has been debated (Carrión, 2010; Carrión and Fernández, 2009; Chiarucci et al., 2010; Loidi et al., 2010). While widespread consensus has proven elusive, the discussion has prompted useful clarification of terminology and advocated for increased transparency and expanded methodological approaches rather than supporting dismissal of the concept (Farris et al., 2010; Loidi and Fernández-González, 2012; Loidi et al., 2010; Mucina, 2010; Roberts, 2015; Somodi et al., 2012). In spite of challenges, PNV remains a useful conceptual tool for assessing the effect of disturbance on vegetation patterns and ecosystem processes (Ricotta et al., 2002; Somodi et al., 2012).

For the purposes of this study, we focused on Minnesota (MN), USA. MN has a long-term legacy of LCLUC with most changes occurring prior to the early 1900's. By 1920, 42% of MN was characterized as improved cropland (US Census Bureau, 1922) and by the early 1900's the best timber had been harvested (MN DNR, 2008; Wyatt, 1999).

The first comprehensive approximation of MN's PNV was published in 1930 after the majority of conversion had already occurred (Marschner and Heinselman, 1974). While remarkable for its time, Marschner and Heinselman's (1974) methodology is poorly documented (MN DNR, 1988). Kuchler (1964) provided an additional approximation, however, the spatial scale is too coarse for many applications. Moreover, neither approximation is consistent with contemporary vegetation classification standards or compatible with the current classification scheme used for land management decision-making in the study region. Given the absence of empirical data predating conversion and the constraints associated with available maps, a comprehensive approximation of MN's PNV could provide valuable information for a variety of land managers and policy makers across public and private sectors. In this context, our objectives were to:

1) Construct a geospatial dataset based on a network of relevés representative of natural/near-natural plant communities throughout the study region (Aaseng et al., 2011; MN DNR, 2013) documenting the location, classification and environmental characteristics of sites
2) Project the approximated spatial distribution of PNV classes at selected classification levels using a series of Random Forest (RF) classification models (Breiman, 2001)
3) Explore alternate classification scenarios using RF proximities and Non-Metric Multi-Dimensional Scaling
4) Demonstrate an analytical approach for modeling PNV with greater adherence to classification transparency, prediction uncertainty and consideration for alternate classification states (Farris et al., 2010; Loidi and Fernández-González, 2012; Loidi et al., 2010; Somodi et al., 2012).

2. Methods

The following sections provide an overview of the study region, the MN Classification System, the training data used in the analysis and the modeling process.

2.1. Minnesota environment

The State of Minnesota is in the upper-midwest of the USA and is roughly 2.17×10^5 km^2 in total size of which 92% is land and the remainder water. The prevailing climate is continental exhibiting warm summers and cold winters (Borchert, 1950). Mean annual temperatures range between 2 °C in the north and 8 °C in the south while mean annual precipitation increases from 480 mm in the west to 848 mm in the east. Climatic variation throughout the study region supports the convergence of three major biomes including tallgrass prairie, deciduous forest and mixed boreal forest (Bailey, 1995).

Across biomes, the composition, structure and spatial distribution of natural plant communities have been extensively altered. Nearly 54% of the study region is used for agricultural purposes with cropland accounting for 44% and rangeland for 10% of total area (Drotts and Heinzen, 2007). The expansion of agriculture has primarily displaced native grasslands with $< 1\%$ remaining in the study region (Drotts and Heinzen, 2007). While nearly 28% of the study region remains forested, timber harvesting and fire suppression along with other land management practices have led to substantial shifts in composition and structure (Frelich and Reich, 1995; Friedman and Reich, 2005).

2.2. Classification system

In 2003 the MN Department of Natural Resources completed a new classification system for describing regional natural plant communities for conservation, restoration and management purposes (Aaseng et al., 2011, 1993). The updated version was designed similar to other hierarchical classification systems (Kotar, 1986; Kotar et al., 1988) where groups of co-occurring species are successively delineated into smaller subsets based on their mutual association with environmental and biological factors influencing community dynamics at progressively finer spatial scales (Bailey, 2005, 1985; Comer et al., 2003; Grossman et al., 1998; Kotar et al., 1999). Coarser classification levels delineate broad vegetation types (e.g., Forest, Grassland or Peatland) based on the influence of large-scale environmental factors like climate, geology and disturbance. At finer levels, classification units become increasingly descriptive as finer scale variation in landform, topography, soils and resource availability become more influential drivers of community structure.

The MN classification system was numerically constructed using a variety of analytical tools including indicator species analysis in combination with several ordination and clustering techniques (for detailed methodological description, see Aaseng et al., 2011). The vegetation data used to construct the classification system were obtained from a statewide network of surveyed relevés representative of regional natural plant communities (MN DNR, 2013). Using these data, candidate classes were delineated by iteratively clustering species into groups within which indicator species were identified (Aaseng et al., 2011; Dufrêne and Legendre, 1997). Groups were then examined for their association with certain geographic and habitat characteristics by correlating indicator species with site conditions. Final classes reflected floristically and physiognomically distinct groups of species mutually adapted to specific climate, moisture, nutrient and disturbance regimes