Ecological Indicators 79 (2017) 73–82

Contents lists available at ScienceDirect

Ecological Indicators

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

Imputing plant community classification from associated forest inventory and physiographic data in Minnesota, USA

David C. Wilson*, Alan R. Ek

Department of Forest Resources, University of Minnesota, 115 Green Hall, 1530 Cleveland Ave. N., St. Paul, MN 55108, USA,

A R T I C L E I N F O

Keywords:
Forest inventory
Native plant community
Classification
Imputation
Physiography
Maximum likelihood

A B S T R A C T

We use conditions defining growing space (e.g., sum of all relevant biotic and abiotic factors) to systematically associate forested stands with likely native plant community (NPC) classifications. We develop and test imputation methodology to link ecological information with forest inventory records, i.e., to individual stands. These efforts will reduce the cost of obtaining needed ecological data with the detail and site specificity required for regional landscape-scale management. We employ associative techniques similar to those used by the Minnesota Department of Natural Resources native plant sampling program to group sites sharing multiple conditions and plant species. However, the focus here is on linking plant community observations with conditions observed during routine forest stand inventory efforts. Unique associations of herbaceous plants are thought to occur in conjunction with different sets of trees on sites with appropriate soils, physiography, moisture, and disturbance histories. Following this hypothesis, NPC assignments based on field observation of herbaceous and woody plants are used as a source of training data to identify and categorize stands already characterized by routine forest inventory efforts. Additional soils, climate, and physiographic data have also been associated with NPC and forest inventory observations to provide information about shared characteristics influencing vegetation on a site. Imputed classifications provide 4 levels of information about ecological conditions on a stand; ecological system, floristic region, and moisture and nutrient availability. Results suggest that with 13 ecological systems, we might expect an average of 70% ± 10% of imputations to be correct at the system level. Imputation of floristic region, moisture index, and nutrient index ratings resulted in roughly 90% accuracy for each component. The compounding of small error rates in components of the NPC class code leads to a 47% ± 10% rate of success over failure in imputation of the class a stand might later be assigned based on relevé data. Relative correspondence of imputed classes to those assigned through observation (e.g., similarity to the true classification; precision) is roughly 87%. Essentially, this means that even when incorrect, imputed classes are very close to those assigned based on observation of physical and vegetative associations. This level of success will help bio-regional landscape management efforts through savings in NPC mapping and classification efforts, and the ecological understanding provided.

1. Introduction

A driving force behind this research is an interest in using detailed ecological information in the process of developing and selecting appropriate management options for forested stands and ecosystems. The desire to make use of detailed ecological information pre-supposes a knowledge and understanding of the distribution, and successional state of various native plant communities (NPCs) composing the local and bio-regional ecosystems. NPC classifications are based on 4 class code components (Fig. 1) defining the ecological system, floristic region, moisture and nutrient conditions occurring at a site. Unfortunately, such information is only rarely available with the level of detail, coverage, and site specificity needed to make timely decisions for management (e.g., harvest scheduling, wildlife habitat enhancements, water quality protection, maintenance of biodiversity, and other values) at multiple scales. This research seeks to bridge this gap by leveraging what we do know about NPC distribution and landscape associations to assign appropriate NPC classifications to individual forest stands.

Native plant community associations[1] develop on a landscape scale

* Corresponding author.
E-mail addresses: wils0602@umn.edu (D.C. Wilson), aek@umn.edu (A.R. Ek).

[1] Native plant community classes are defined as units of vegetation that generally have uniform soil texture, soil moisture, soil nutrients, topography, and disturbance regimes (MNDNR, 2003). MNDNR currently recognizes approximately 99 distinct NPC classes, with additional types and subtypes defined for many (MNDNR, 2006).

http://dx.doi.org/10.1016/j.ecolind.2017.04.021
Received 12 July 2016; Received in revised form 6 April 2017; Accepted 7 April 2017
Available online 15 April 2017
1470-160X/

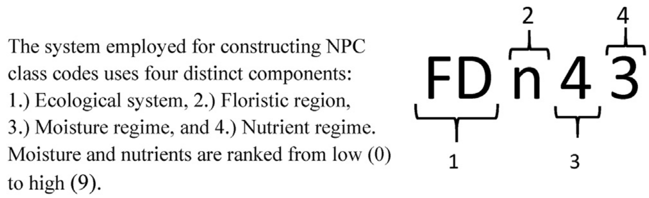

Fig. 1. Native plant community class code example.

over many years (e.g., decades to millennia). Different sites will foster the development of different NPC associations based on physiographic and climatic conditions, the propagule sources prevailing locally, and disturbance history. The time of year a disturbance occurs, its intensity, and the time between disturbances all influence plant community development. Changes in these factors produces great variation in the structure and composition of resulting communities (MNDNR, 2003). Terrain, soil characteristics, and prevailing climate are also related to the frequency, intensity, and timing of fires, drought, windstorms, ice storms, and extreme snowfall events. These relationships are thus suggestive of likely predictor variables related to NPC classification.

Because plant community development is strongly influenced by the combined growing space (e.g., sum of all biotic and abiotic factors affecting plant development) and disturbance history of a site, it is possible to use existing vegetation and abiotic conditions to identify similarities among sites on the landscape (Aaseng et al., 2011; MNDNR, 2013). It is important, however, to note that observed patterns may be related to either physical site characteristics, disturbance history, or both. As discussed above, terrain and climate largely control the frequency and intensity of fires, windstorms, and precipitation, so these factors, combined with vegetative observations, can be used as a surrogate where disturbance history is unknown. We can then group sites according to their physical qualities and the plant associations best able to make use of the growing space represented by that site. The extent to which existing vegetation on a site has been shaped by the interplay of dispersal, competition, succession, and periodic disturbance determines the likely success of using the methodology developed here to identify the same NPC class(es) that would be selected using detailed observation of plant associations.

Growing space is shaped by physical site characteristics, but also by the effects of inter and intra-specific competition for light, nutrients, moisture, and less obvious needs. For these reasons, dominant tree species, successional state, and the extent to which a site has been captured by tree canopy are relevant to the growing space available for both tree and herbaceous species to become established and persist on a bioregional landscape. Herbaceous plant associations provide additional information helpful to differentiating finer NPC classifications (Fig. 1) related to variations in regional to microsite scale conditions (Aaseng et al., 2011; MNDNR, 2003, MNDNR, 2013).

A geographic information system (GIS) is defined as “a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographic data” (Wikipedia.org). The recent proliferation of remote sensing and “big geographic data”, along with the rapid development of GIS capabilities, has provided the opportunity to conduct large-scale predictive analyses of landscape ecological condition. These kinds of analyses typically use a variety of thematic layers describing relevant physical and biotic characteristics of the landscape to make broad prediction of ecological context. Such geographic layers may include themes like land cover, percent forest canopy cover, precipitation, soil drainage, soil fertility, slope, elevation, ecological context, or other relevant factors. Here, we use these broad characterizations combined with selected data layers and field observations to create finer ecological classifications associated with distinct plant communities present on forested stands, or management units, administered by MNDNR.

Existing methods for extending observations to classify very similar nearby entities (e.g., forest inventory units) by nature rely on nonparametric approaches able to accommodate categorical predictor and response variables. The randomForest approach (Breiman 2001; Liaw and Wiener, 2002; Cutler et al., 2007; Packalén et al., 2012), which evaluates a large number of randomly generated classification and regression trees (CARTs) created using different subsets of the observations and predictors, is one such machine learning tool. Withheld, or out of bag, observations (~37% of the data) are used to assess the performance of each CART. Variable importance is assessed through permutation-based mean squared error (MSE) reduction associated with different predictors across the forest of CARTs. Unfortunately, randomForest is limited in the number of categorical predictor levels it can accommodate, and fails to classify entities with incomplete or missing predictor data. Similar tools, including mice (van Buuren and Groothuis-Oudshoorn 2011) and earth (Milborrow 2016), suffer from the same limitations, and also provide incomplete coverage of entities we might wish to classify.

This research explores methodology and performance of more flexible imputation modelling designed to identify most likely NPC class associations for forested stands managed by Minnesota Department of Natural Resources (MNDNR). In doing so, the hypothesis that tree cover and abiotic conditions are highly correlated with the unique associations of plants identified by field ecologists using the relevé method (Minnesota Department of Natural Resources, 2013) is tested. Hence, performance of the imputation model developed here can also be thought of as a further test of the hypothesis that systematically sampled, detailed vegetation lists are truly necessary for classification of forested plant communities at some desired level of detail. Additionally, the hypothesis that inclusion of a detailed ecological predictor (e.g., land type association; Cleland 1997) in the imputation model will enhance classification accuracy is also tested. Implications of the imputation model for relevé survey and forest inventory design, cost, and timing are also explored.

2. Material and methods

2.1. Data and pre-analysis

We begin with a review of available data tracking variables related to forest ecosystem development and growth in Minnesota. This includes soil, landform, temperature, moisture, and nutrient availability, as all are important factors contributing to native plant community development.

2.1.1. Forest inventory

MNDNR Division of Forestry maintains a Forest Inventory Management database (FIM), used to track stand inventory data for the 202,151 stands (5.5 million acres – 2.2 million hectares, 4.17 million forested acres – 1.7 million forested hectares) administered by the agency. Observations detailing major forest cover type, tree species present and order of importance, stand age, basal area, timber volume, physiographic class, topography, and other site characteristics are recorded in this data set (Minnesota Department of Natural Resources, 1981). Most stands have been sampled in the past 10 years. However, many of the slower growing swamp conifer stands have not been visited in up to 30 years.

2.1.2. Native plant community observations/ecological classification system

A total of 16,913 native plant community observations made by MNDNR Division of Ecological and Water Resources and others between 1964 and 2015 (Minnesota Department of Natural Resources, 2015) are employed in this research. The publication *Field Guide to the Native Plant Communities of Minnesota: The Laurentian Mixed Forest Province* (Minnesota Department of Natural Resources, 2003) provides an overview of the classification process. This guide also contains a dichotomous key to help distinguish finer details of NPCs