Forest Ecology and Management 358 (2015) 154–164

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Imputed forest structure uncertainty varies across elevational and longitudinal gradients in the western Cascade Mountains, Oregon, USA

David M. Bell [a,*], Matthew J. Gregory [b], Janet L. Ohmann [b]

[a] Pacific Northwest Research Station, USDA Forest Service, 3200 SW Jefferson Way, Corvallis, OR 97331, USA
[b] Department of Forest Ecosystems and Society, Oregon State University, 321 Richarson Hall, Corvallis, OR 97331, USA

ARTICLE INFO

Article history:
Received 22 May 2015
Received in revised form 26 August 2015
Accepted 3 September 2015
Available online 14 September 2015

Keywords:
Bootstrapping
Forest
Imputation
k-nearest neighbor
Model uncertainty
Vegetation

ABSTRACT

Imputation provides a useful method for mapping forest attributes across broad geographic areas based on field plot measurements and Landsat multi-spectral data, but the resulting map products may be of limited use without corresponding analyses of uncertainties in predictions. In the case of *k*-nearest neighbor (*k*NN) imputation with $k = 1$, such as the Gradient Nearest Neighbor (GNN) approach, where the field plot with the most similar spectral signature is attributed to a given pixel, there has been limited guidance on methods of examining uncertainty. In this study, we use a bootstrapping method to assess the uncertainty associated with the imputation process on predictions of live tree structure (canopy cover, quadratic mean diameter, and aboveground biomass), dead tree structure (snag density and downed wood volume), and community composition (proportion hardwood) for a portion of the Cascade Mountains in Oregon, USA. We performed *k*NN with $k = 1$ imputation with 4000 bootstrap samples of the field plot data and examined three metrics of uncertainty: the width of 95% interpercentile ranges (IPR), the proportion of bootstrap samples with no tally (i.e., forest attribute was imputed as zero), and the imputation deviations (i.e., mean prediction from the bootstrap sample minus baseline GNN prediction [no bootstrapping]). Imputed values of dead tree components and species composition exhibited greater IPR, proportion no tally near 0.5, and greater magnitudes of imputation deviations compared to live tree components, indicating greater uncertainties. Our uncertainty metrics varied spatially with respect to environmental gradients and the variation was not consistent among metrics. Geographic patterns in prediction uncertainties implicated biogeography and disturbance as major factors influencing regional variation in imputation uncertainty. Spatial patterns differed not only by forest attribute, but by uncertainty metric, indicating that no single measure of uncertainty or forest structure provides a full description of imputation performance. Users of imputed map products need to consider the pattern of and the processes that contribute to uncertainty during the early stages of project development and execution.

Published by Elsevier B.V.

1. Introduction

Mapping forest conditions and attributes based on imputation, a method of substituting observed values to replace missing data, is becoming increasingly common (e.g., Ohmann and Gregory, 2002; Tomppo et al., 2008; Wilson et al., 2013) and the utility of the resulting map products in forest management and planning is unknown without estimates of precision, or uncertainty, and accuracy, or bias. While forest inventory programs, such as the USDA Forest Service Forest Inventory and Analysis (FIA) program (Bechtold and Patterson, 2005), provide consistent and extensive sampling of forest conditions appropriate for design-based inference on large areas, their utility in supporting fine-scale decision making (e.g., forest stand management) can be limited by the relative low density of plots (McRoberts and Tomppo, 2007; McRoberts, 2008). For example, at base sampling intensity there is one FIA plot for every 2428 ha of forest land (Bechtold and Patterson, 2005). As a consequence of and in conjunction with increasingly reliable remote sensing products, such as 30-m resolution multi-spectral imagery across the multi-decade life of the Landsat program (Williams et al., 2006; Loveland and Dwyer, 2012), there has been an increasing focus on statistical imputation methods that can produce high-dimensional data products by correlating ground observation of vegetation characteristics with geospatial data products (e.g., Ohmann and Gregory, 2002; Tomppo et al., 2008). While nearest neighbor imputation provides

* Corresponding author.
E-mail addresses: dmbell@fs.fed.us (D.M. Bell), matt.gregory@oregonstate.edu (M.J. Gregory), janet.ohmann@oregonstate.edu (J.L. Ohmann).

http://dx.doi.org/10.1016/j.foreco.2015.09.007
0378-1127/Published by Elsevier B.V.

fine-scale predictions of a variety of vegetation characteristics valued by forest managers, such as descriptions of species composition, live tree structure, and dead wood components, fine-scale measures of uncertainties in mapped ecological predictions are necessary for proper interpretation of results (Wiens et al., 2009).

In addition to traditional model validation and accuracy assessments, where predictions are compared to independent observations to assess a model's ability to predict reality (e.g., Riemann et al., 2010; Wilson et al., 2012), assessing the variation, or uncertainty, in predictions can be a powerful tool for understanding the limitations of statistical imputation and the resulting maps of forest characteristics. Presenting uncertainties in predictions, especially across extensive geographic areas, is a major challenge for the development of useful geospatial predictions (Wiens et al., 2009). Assessing uncertainties can inform future research (e.g., what data and/or processes need to be incorporated; LeBauer et al., 2013) and help identify the limitations of predictions for decision-making (e.g., in what areas are predictions most variable; Beaudoin et al., 2014).

Nearest-neighbor techniques have emerged as useful methods for spatial prediction of forest attributes as combinations of observations (e.g. field plots) that have similar characteristics in a space of mapped auxiliary variables (often from remote sensing) (McRoberts, 2012; McRoberts et al., 2010; Tomppo et al., 2008). Nearest neighbor methods are appealing because they are multivariate and nonparametric, and can be used to map multiple forest characteristics over large areas (Eskelson et al., 2009; McRoberts et al., 2011). Nearest-neighbor techniques based on forest inventory plots and satellite imagery were first implemented operationally in Finland in 1990, but have now been applied in locations spanning the globe (McRoberts, 2012). Recent work on k-nearest neighbor (kNN) techniques has provided valuable insights into estimating prediction uncertainties for imputation methods (McRoberts, 2006; McRoberts et al., 2011). kNN techniques estimate the characteristics of an individual pixel or other areal unit as a function of k observations most similar to the pixel in question based on some set of auxiliary data, such as remote sensing, climate, and soils. When the objective has been to estimate the variance of a prediction, it is often assumed that $k \geqslant 5$ and that each neighbor contributes equally to the estimator to allow for a relatively simple analytical solution (e.g.; McRoberts, 2006; McRoberts et al., 2007). Non-parametric methods of estimating uncertainties, such as bootstrap and jack-knife sampling, have also been employed, both as a way to test the validity of the assumptions used in the analytical solutions as well as estimating forest attributes for management and research (Magnussen et al., 2009, 2010; McRoberts et al., 2015).

While research into kNN variance estimation has received some attention, very little has been done to examine imputation models utilizing small values of k. In particular, $k = 1$ approaches are useful tools for imputation as they, by definition, can only predict combinations of forest attribute values at the pixel-scale that were observed in the field (Ohmann and Gregory, 2002; Henderson et al., 2014). As a result, unrealistic predictions for a 30-m pixel, as might be expected when averaging many nearest neighbors (i.e., large k), are avoided. While there have been attempts to assess model accuracy at the plot- and aggregate-scale for $k = 1$ kNN approaches (Pierce et al., 2009; Riemann et al., 2010; Ohmann et al., 2014), uncertainty characterization for $k = 1$ methods at the pixel level has received less attention (but see McRoberts et al., 2011). When k is small, analytically derived variance estimators are impractical, but non-parametric bootstrapping of the imputation process can provide a method for estimating variability (McRoberts, 2012; McRoberts et al., 2015), such as the width of inter-percentile ranges in bootstrap sample predictions. Bootstrap samples can also be used to examine the likelihood that a forest attribute is present (i.e., >0) by calculating the proportion of bootstrap predictions where the variable of interest equals zero. Similar to zero-truncated models for species abundances (Martin et al., 2005), the quantification of the absence of certain forest attributes provides a deeper understanding of the observed patterns. Finally, the degree to which baseline imputed predictions using $k = 1$ (i.e., no bootstrapping) differ from the mean predictions based on bootstrapped samples (hereafter referred to as imputation deviations) can highlight the influence of extreme data points on prediction. As a result, multifaceted exploration of imputation uncertainties can provide a fuller picture of how the availability of reference plot data modified through non-parametric bootstrapping (i.e., which plots are selected in each bootstrap sample) impacts imputed predictions.

Uncertainties in imputed map predictions arise from a variety of sources. Input data may involve sampling error, due in part to inaccuracies in measurements, such as omission of trees in inventory plots or sensor drift for remotely sensed data. Spatial mismatches and scaling issues are common in geospatial data analysis (Turner et al., 2004; Riemann et al., 2010; Zald et al., 2014). Both of these sources of error could lead to substantial uncertainties in imputed map predictions, especially at the pixel-scale. Statistical models upon which imputation might be based, as with canonical correspondence analysis (CCA; ter Braak, 1986) in the Gradient Nearest Neighbor (GNN; Ohmann and Gregory, 2002) method, are simplifications of reality, contributing to prediction uncertainty. The imputation algorithm assigns predictions to individual pixels based on the model and some set of reference plot data which is itself a sample of forest conditions and is thus an incomplete representation of reality. These differing sources of uncertainty can be manifested in poor predictive performance, often explored through accuracy assessment. For the GNN approach, accuracy assessments demonstrate good agreement between predictions and observations in closed-canopy forests of the Pacific Northwest, especially in the western Cascade Mountains and Oregon Coast Range (Pierce et al., 2009). Even when imputation maps exhibit good accuracy, uncertainties can still manifest themselves as low precision (i.e., high variability) of predictions.

In this paper, we use non-parametric bootstrapping to examine uncertainties in forest attribute imputation for kNN with $k = 1$ methods, because (1) it has direct bearing on the development of fine-scale imputed map products, and (2) focusing only on the contribution of imputation to map uncertainties allows for a general examination of uncertainties associated with kNN methods with $k = 1$ rather than the CCA model underlying the GNN approach. We use bootstrap sampling to explore the sensitivity of $k = 1$ nearest neighbor predictions across 4978 km^2 of forested land in the western Cascade Mountains of Oregon, USA. Specifically, we use one variant of kNN with $k = 1$, the GNN approach, for imputing forest attributes based on 30-m resolution environmental data, including Landsat imagery. Because this implementation of the GNN imputation method relies on data from Landsat imagery as predictor variables, we expect that variability in estimation would depend in part on how closely related a given variable was to the overstory condition being observed by the satellites. For example, since Landsat's TM and ETM+ sensors observe exposed vegetation most directly and do so at a 30-m pixel resolution (Lu, 2006), we would expect live tree forest structure, such as canopy cover and biomass, to be better predicted than dead wood, which can be less abundant and often obscured from view by passive remote sensors. In addition, rare components of the landscape and those strongly influenced by stochastic events may be difficult to predict, such as hardwood contributions to forest communities, because they generally do not dominate in this region (Ohmann and Spies, 1998). Specifically, our objectives were (i) to characterize the imputation uncertainty in predictions for six forest attributes

Download English Version:

https://daneshyari.com/en/article/86068

Download Persian Version:

https://daneshyari.com/article/86068

Daneshyari.com