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

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 (*k*NN) techniques has provided valuable insights into estimating prediction uncertainties for imputation methods (McRoberts, 2006; McRoberts et al., 2011). *k*NN 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 *k*NN 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 $k = 1$ *k*NN 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 *k*NN 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 *k*NN 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 *k*NN 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

Table 1
Names, variable ranges, and descriptions for a subset of forest attributes imputed based on the GNN framework for the western Cascade Mountains.

Variable	Range	Description
BAH_PROP	0–1	Proportion of live tree basal area represented by hardwoods (unitless)
BIOM	0–1.18 × 10^6	Biomass of live trees ⩾ 2.5 cm dbh calculated by the component ratio method (Jenkins et al., 2003) (kg ha^{-1})
CANCOV	0–99	Canopy cover of all live trees (%)
DVPH	0–2003	Volume of down wood ⩾ 25 cm diameter and ⩾ 3 m long (m^3 ha^{-1})
QMD_DOM	0–143	Quadratic mean diameter of all live dominant and codominant trees (cm)
STPH	0–421	Density of snags ⩾ 25 cm dbh and ⩾ 2 m height (snags ha^{-1})

(Table 1) across the region and (ii) to assess how imputation uncertainty varies across regional ecological gradients. Furthermore, we discuss how to use this assessment to explore the potential causes and consequences of imputation uncertainty for users of imputed data products.

2. Methods

2.1. Study area

For this study, we focused on a geographic transect in the western Cascade Mountains which ranges from low elevations (106 m) at the edge of the Willamette Valley to high elevations (3149 m) along the crest of the Cascade Mountains (Fig. 1). Forest ecosystems vary substantially along elevational gradients in the region, with patchy deciduous and coniferous forests located at low elevations, Douglas-fir (*Pseudotsuga menziessii*) forests dominating much of the mid-elevation, montane forests, and true firs (*Abies* species) dominating high-elevation, subalpine forests. This geographic variation in ecosystem distribution provided the opportunity to compare imputation uncertainty across ecological gradients within the study area. Our study focuses on the results for the most recent year for which Landsat imagery was available, meaning that imputed map predictions were produced for 2012.

2.2. Gradient Nearest Neighbor method

Gradient nearest neighbor is one variation of nearest-neighbor imputation which relies on constrained ordination (direct gradient analysis) – specifically CCA – for measuring and weighting distances in nearest-neighbor calculations (Ohmann and Gregory, 2002). Past studies have used GNN to simultaneously map multiple attributes of forest vegetation for a variety of forest ecosystems and objectives for a single date (Ohmann et al., 2007, 2011; Pierce et al., 2009), for two dates (Ohmann et al., 2012), and most recently for a yearly time-series based on the Landsat archive (Ohmann et al., 2014). Even though GNN has often been implemented as a $k = 1$ imputation method, it is important to note that the GNN imputation method does not assume any specific value of k. Still, GNN is often cited as an example of $k = 1$ nearest neighbor method and we focus on the $k = 1$ implementation for the current study.

2.3. Spatial predictors and plot inventory data

Spatial predictors used in GNN are listed in Table 2, are very similar to those used previously, and are discussed in great detail in our previous publications (e.g., Ohmann et al., 2014). Spectral variables were derived from Landsat imagery mosaics developed with the LandTrendr (Landsat Detection of Trends in Disturbance and Recovery) algorithms (Kennedy et al., 2007, 2010). LandTrendr is a trajectory-based change detection method that identifies fitted line segments of consistent trajectory for each pixel that describe sequences of disturbance and growth while minimizing annual variability from differences in sun angle, phenology, and atmospheric effects. Plots were matched to image date to avoid a temporal mismatch between plot measurement and imagery acquisition. We used data from these LandTrendr-derived "temporally smoothed" imagery mosaics, which consisted of an annual time-series of mosaics from 1984 to 2012, as spatial predictors. Temporal smoothing reduces noise in the spectral signal, providing improved vegetation prediction even within a single year (e.g., Ohmann et al., 2012). In addition to the tasseled cap indices (Crist and Cicone, 1984) used in Ohmann et al. (2014), we added the normalized burn ratio (Key and Benson, 2002) as a useful spectral index sensitive to burn severity and other types of change. Climate spatial predictors were derived from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) using 30-year normals (1971–2000) of mean monthly precipitation and temperature data (Daly et al., 2008). The original 800-m rasters were resampled to 30-m using bilinear interpolation.

Response variables for CCA were basal area by tree species and size-class observed on 5124 field plots installed at 3115 locations, measured from 1991 to 2011 in regional and national forest inventories: Forest Inventory and Analysis (FIA) Annual Inventory (Bechtold and Patterson, 2005), periodic inventories by the United States Forest Service (USFS) Pacific Northwest Research Station (PNW) and Region 5 (Waddell and Hiserote, 2005), and Current Vegetation Survey (CVS) by USFS Region 6 and BLM (Max et al., 1996). Detailed descriptions of sample designs and field methods can be found in these references. Different inventories provided slightly different sets of measurements measured over slightly different areas. To address this, we used only those variables shared by all of the plot systems and converted measurements to a per area basis. At each plot location there were as many as three separate field measurements, conducted at different times. We used plots with at least 50% of their area classified as forest (i.e., 10% stocked) or forest-capable (evidence of previous forest and undeveloped for non-forest use) (Bechtold and Patterson, 2005). All plot-level attributes to be imputed were calculated on a per-hectare basis for forested portions of the plot. Pixels identified as non-forest based on the National Gap Analysis Program's land cover data (Kagan et al., 2008) were masked in final vegetation maps.

2.4. Model specification and map imputation

GNN specification, including the fitting of the CCA model and the imputation of plot data to individual pixels, has been discussed at great length in previous literature (Ohmann and Gregory, 2002; Ohmann et al., 2014). In this section, we provide a brief overview of the process, but refrain from extended descriptions. For CCA model fitting, values for spatial predictors were associated with each plot observation using a 3-by-3-pixel block that encompasses the outer extent of the general area sampled by the field plot (Ohmann et al., 2014). The 3-by-3 pixel block roughly matches the extent of the FIA plots, though FIA subplots and macroplots cover only 8% and 50% of the 3-by-3 pixel block area. While some have used individual subplots (e.g., McRoberts et al., 2011), we match the 3-by-3 pixel blocks to the entire plots because (1) geolocation errors for

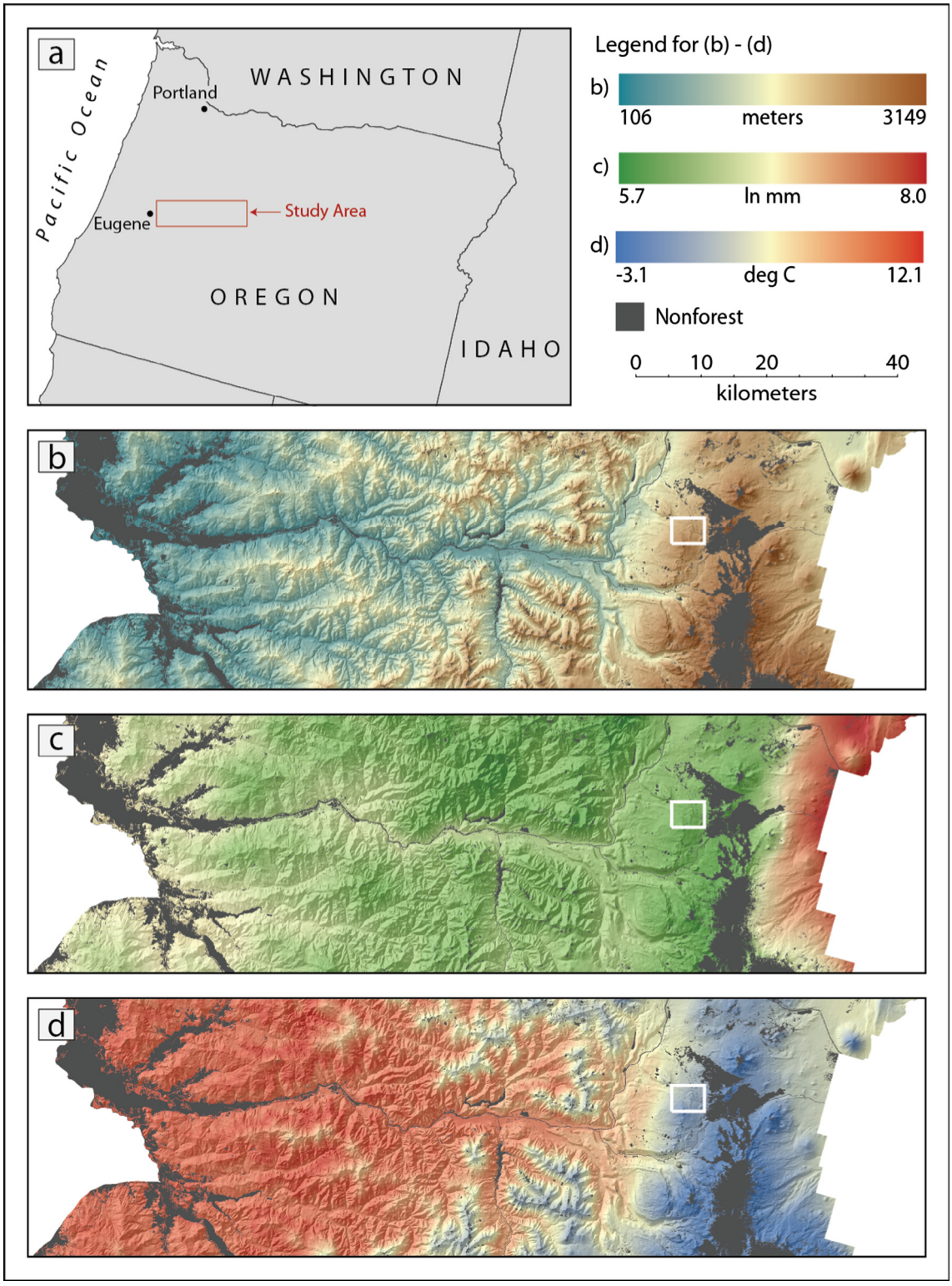

Fig. 1. We present (a) our study region (red rectangle) in western Oregon, USA, as well as (b) elevation (m), (c) mean annual precipitation (ln mm), and (d) mean annual temperature (°C) gradients across forest lands in the western Cascade Mountains. For the representations of the environmental gradients (i.e., panels b–d), dark gray indicates non-forest lands, white indicates landscapes excluded from our analysis, and the white box delineates an area examined in greater detail (Figure 5). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plots in this region can be high (e.g., Zald et al., 2014) which are particularly problematic for smaller sampling units (Stehman and Wickham, 2011), such as subplot-scales, (2) a subplot is substantially smaller than an individual pixel and is not necessarily centered within that pixel, and (3) fine-scale heterogeneity in forest attributes make separate analysis of individual subplots problematic (MacLean, 1980). Spectral data were extracted from the LandTrendr mosaic for the same year as plot measurement. Values for the other explanatory variables were similarly extracted for the plot footprints, but values were assumed constant over the range

Table 2
Spatial predictors used in CCA and GNN imputation. All rasters were 30 m resolution.

Variable subset	Code	Description
Landsat time-series, processed using LandTrendr algorithms (Kennedy et al., 2010)	TC1	Axis 1 (brightness) from tasseled cap transformation (Crist and Cicone, 1984)
	TC2	Axis 2 (greenness) from tasseled cap transformation
	TC3	Axis 3 (wetness) from tasseled cap transformation
	NBR	Normalized burn ratio (using Landsat TM bands 4 and 7)
Climate, from PRISM rasters (Daly et al., 2008), which are 30-year normals	ANNTMP	Mean annual temperature (°C)
	AUGMAXT	Mean maximum temperature in August (°C)
	DECMINT	Mean minimum temperature in December (°C)
	ANNPRE	Mean annual precipitation (natural logarithm, mm)
	CONTPRE	Percentage of annual precipitation falling from June–August, a measure of continentality
	SMRTP	Growing season moisture stress, the ratio of mean temperature (°C) to precipitation (natural logarithm, mm) from May–September
	COASTPROX	Index of coastal proximity, based on penetration of marine air through complex terrain as indicated by minimum and maximum temperatures (index from 0 to 1000)
Topography, developed from 10-m digital elevation model (DEM) and rescaled to 30 m	ELEV	Elevation (m)
	ASP	Cosine transformation of aspect (degrees)
	SLP	Slope (%)
	SOLAR	Cumulative potential relative radiation during the growing season (Pierce et al., 2005)
	TPI	Topographic position index, calculated as the difference between a cell's elevation and the mean elevation of cells within a 450-m-radius window
Location	LAT	Geographic latitude (degrees)
	LONG	Geographic longitude (degrees)

of plot measurement dates. Heterogeneous plots that encompassed strongly contrasting land cover classes or forest conditions were identified through an outlier detection process and excluded from the models. We also screened plots for disturbances occurring between the plot measurement date and the imagery date, such as harvest or fire, where the spectral signal would not be representative of the plot data. This was quite rare because of the yearly temporal matching. Plots of all measurement years were used in the CCA models. The validity of this approach relies on the assumption that LandTrendr effectively normalizes the spectral values among image dates, i.e. that spectral values representing similar forest conditions were equivalent across the time-series. Imputed map products and accuracy assessment reports for California, Oregon, and Washington are freely available to users and are provided in standard GIS formats (http://lemma.forestry.oregonstate.edu).

To impute forest attributes to 30-m pixels, imputation was implemented as described in Ohmann et al. (2014), using $k = 1$, a single-pixel imputation grain, and a multi-pixel accuracy assessment grain. For *k*NN approaches using $k = 1$, the single-pixel imputation and multi-pixel accuracy assessment grain was superior to other combinations of single and multi-pixel imputation and assessment in terms of RMSE across numerous variables and vegetation classification accuracy. Neighbor-finding was based on weighted Euclidean distance within multivariate gradient (CCA) space, with axis scores weighted by their eigenvalues. As a result, the median and mean number of times each field, or reference, observation was imputed to a pixel (i.e., pixel assignments) in our study area were 214 and 1162 pixels, respectively. If all plots were attributed with equal frequency to pixels, we would expect 863 pixels assignments per plot. The difference between the mean and median number of pixel assignments indicates a skewed distribution, implying that the frequency of certain forest types differs between the networks of field plots and the complete forest landscape in our study area. The difference in frequencies likely arises from the fact that the plot data are distributed across the entire western Cascade Mountains, where as the study area for this research was not: The plot network represents a larger region than our study area. For the purposes of clarity, we refer to the imputed map based on all reference plots (i.e., no bootstrapping) as the baseline GNN predictions.

For each variable, we carried out accuracy assessments for the study region using a modified leave-one-out methodology (e.g., Ohmann et al., 2014). In this accuracy assessment, the nearest neighbor with a plot identity differing from the actual plot identity was attributed to the plot, predicted forest attributes were extracted using a multi-pixel (3×3) window, and mean predictions were compared with observations. For each forest attribute, we calculated the Pearson correlation coefficient, root mean squared error (RMSE), relative RMSE, percent bias, and coefficient of determination (R^2). Accuracy metrics were calculated for the study region and the entire western Cascade Mountains (Fig. 1).

2.5. Bootstrap sampling and uncertainty estimation

Because of the impacts of small sample sizes on variance estimators and the inability to calculate variance when $k = 1$, non-parametric bootstrap sampling has been suggested as a means of characterizing variability in predictions from *k*NN imputation for small values of k (McRoberts et al., 2011; McRoberts, 2012). To characterize uncertainties in forest attribute predictions based on *k*NN imputation where $k = 1$, we developed a series of vegetation maps using non-parametric bootstrap sampling of the reference data with 4000 iterations. For each iteration of the non-parametric bootstrap, we sampled the reference plots with replacement ($n = 5124$) to be used in the imputation process. Thus, imputations for each bootstrap contain different, but overlapping, lists of reference plots with which vegetation maps can be estimated. The bootstrapping does not impact the CCA model fitting, which uses all plots and is the basis for the nearest neighbor distances. By limiting the bootstrap sampling to the imputation step of the analysis our results can be viewed as a general characterization of *k*NN imputation uncertainty rather than an artifact of the methodology specific to GNN (i.e., distances based on weighted CCA scores).

At the pixel-scale, the behavior of the *k*NN imputation approach with $k = 1$ can be understood probabilistically by taking into account how the bootstrapping algorithm samples the reference plot data (Appendix A). Because of the *k*NN imputation for $k = 1$ strategy assigns the nearest neighbor within the set of reference plots to a given pixel, one can show that, on average, 63.2% of bootstrap values for a given pixel should arise from the first nearest neighbor in the complete list of reference plots and that 99.91% should arise from the first seven nearest neighbors (Table A.1). As a result, the distribution of imputed forest attributes for a pixel

across the 4000 samples will not be normally distributed, indicating that results should be summarized using non-parametric statistics.

To summarize the results of the bootstrapping at the pixel-level, we chose three metrics for imputation uncertainty. Similar to zero-truncated models (e.g., Martin et al., 2005), we separated uncertainty metrics into two classes: (1) uncertainties in the imputation of the presence of the imputed forest attribute (e.g., whether or not snags were imputed to be present) and (2) uncertainties in the predictions given the forest attribute was imputed to be present (e.g., variation in the density of snags given at least one snag was imputed to be present). To examine imputation uncertainties in the predictions given the forest attribute was present, we calculated the relative (to the median) 95% interpercentile range (IPR), a robust nonparametric measure of variability in the response variable, as the difference between the 97.5th and 2.5th percentiles of non-zero values from the bootstrap sample divided by the median prediction of the forest attribute across all pixels. To examine variation in the presence of a forest attribute, we calculated the proportion of bootstrap samples with no tally of the forest attribute (i.e., predicted attribute equal zero). In this case, estimates of the proportion with no tally equal to 0.5 indicate large uncertainties (e.g., flipping a coin), while estimates near zero or one indicate greater certainty in the presence and absence of a forest attribute, respectively. In addition, we calculated the imputation deviations as the difference between the bootstrap sampling mean and the baseline GNN prediction. Large values of the absolute value, or magnitude, of the imputation deviations indicate that the prediction is sensitive to the removal of the nearest neighbor, and is therefore unstable. Bootstrap sampling appeared to produce stable estimates of uncertainty metrics after 4000 samples for nearly all combinations of variables and uncertainty metrics (Appendix B).

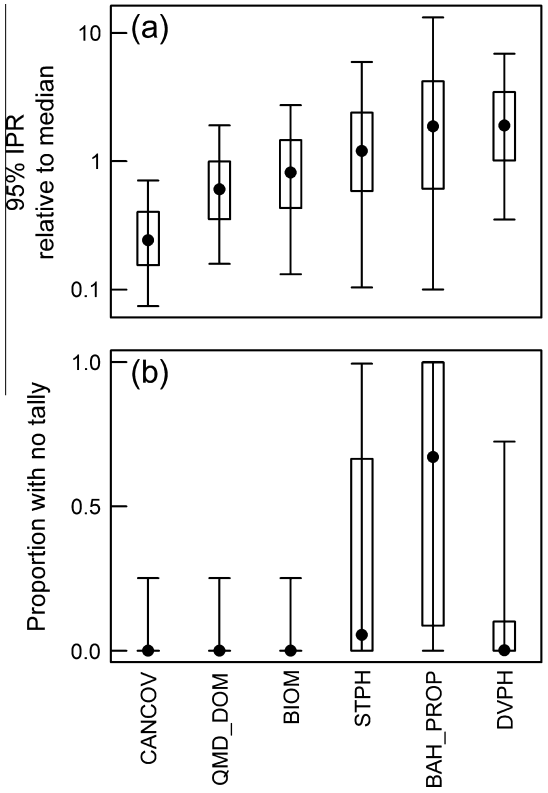

Fig. 2. Variation in imputed bootstrapped values in the (a) 95% IPR and (b) proportion no tally for each forest variable (see Table 1 for attribute descriptions). The closed circles denote the medians, the bounds of the rectangle denote the 25% and 75% percentiles, and whiskers denote the 5% and 95% percentiles.

3. Results

Variability in imputations from bootstrapping was greatest for dead wood and community composition attributes and lowest for live tree forest structure attributes. Accuracy assessments indicated that live tree components of the forest attributes (CANCOV, BIOM, and QMD_DOM) performed well in terms of Pearson correlations ($\geqslant$0.74), R^2 ($\geqslant$0.54), and RMSE (normalized RMSE $\leqslant$0.61) (Table 3). With the exception of BAH_PROP and STPH, forest attribute predictions exhibited mean bias between −2% and 2%. Similar to accuracy assessments, CANCOV, QMD_DOM, and BIOM exhibited the lowest relative IPR (median across pixels equal to 0.24, 0.60, and 0.82, respectively) while STPH, DVPH, and BAH_PROP exhibited the greatest relative IPR (median across pixels equal to 1.20, 1.90, and 1.87, respectively) (Fig. 2a). At each pixel, the proportion of the bootstrap samples with no tally was generally greater than zero for STPH, DVPH, and BAH_PROP (Fig. 2b).

Regional variation in IPR was high for three of the attributes, with patterns and magnitudes varying based on the attribute in question (Fig. 3). For example, the variability in BAH_PROP decreased from west to east. As elevations increased toward the crest of the Cascade Mountains (eastern portion of study region; Fig. 1), variability in DVPH decreased while variability in STPH increased. Regional patterns for CANCOV, QMD_DOM, and BIOM were less obvious. The proportion of bootstrap samples with no tally for live tree forest structure attributes (i.e., CANCOV, QMD_DOM, and BIOM) was generally equal to zero across most of the study area (Fig. C.2a–c). The proportion of bootstrap samples with no tallies of the other three forest attributes exhibited elevational (STPH and DVPH) and longitudinal (BAH_PROP) patterns (Fig. C.2d–f), with large proportion no tally associated with locations where median predictions from the bootstrap sample equaled zero (Fig. C.1). Regional variation in the magnitude of the imputation deviations (Fig. C.3) followed similar patterns to those observed for 95% IPR (Fig. 3). The primary difference between them is that imputation deviations appear to be more variable at fine spatial scales, as indicated by the relative rarity of large blocks of large imputation deviations. Maximum observed imputation deviations scaled with the 95% IPR: large deviations were generally associated with high 95% IPR, but high 95% IPR could be associated with either large or small deviations (Fig. C.4).

When pixels were grouped by Level IV Ecoregion (McMahon et al., 2001), proportion of predictions with no tally, 95% IPR, and bootstrap deviations indicated ecoregional difference in model uncertainty (Fig. 4). For STPH and DVPH, proportion of the

Table 3
Summary of accuracy assessment statistics for six variables (Table 1) based on a modified leave-one-out procedure for 409 plots in the study region.

	BAH_PROP	BIOM	CANCOV	DVPH	QMD_DOM	STPH
Pearson correlation	0.49	0.76	0.80	0.31	0.74	0.63
RMSE	0.12	141,337	13.98	160.39	17.51	33.02
Normalized RMSE	2.39	0.61	0.21	1.30	0.42	1.12
percent bias	−18.16	1.95	1.34	0.60	0.67	−4.25
Pseudo-R^2	0.21	0.55	0.63	−0.03	0.54	0.38

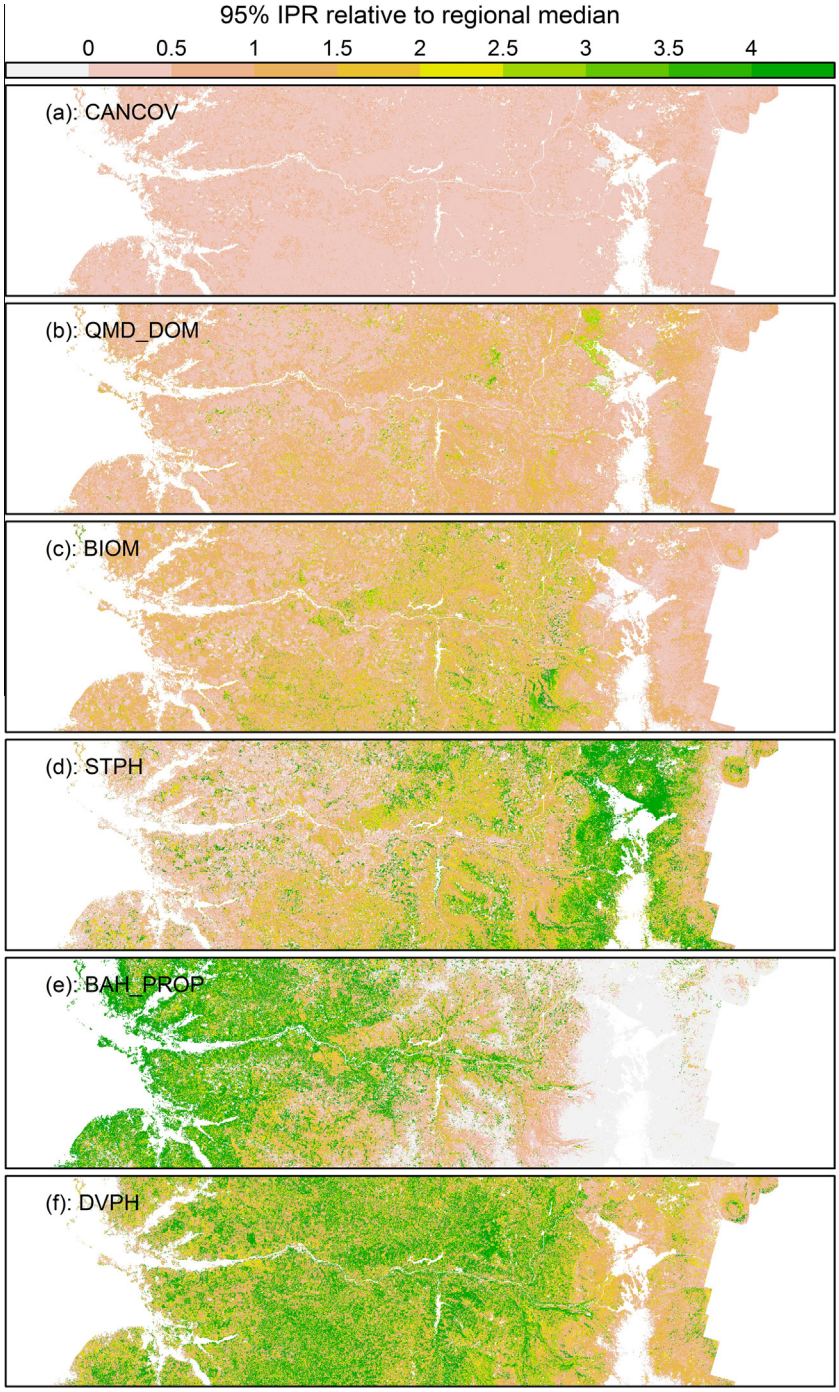

Fig. 3. Regional patterns in the 95% IPR relative to the regional medians. Attributes are sorted by the regional IPR medians (Fig. 2a). Gray areas indicate zero. Note that the maximum category includes all values greater than 4.

bootstraps with no tally were similar to 0.5 (i.e., high uncertainty) in western (particularly WV) and eastern (particularly GF and PP) ecoregions. In addition, the proportion with no tally for DVPH was also similar to 0.5 at the highest elevation ecoregion (CS). Proportion of the bootstrap with no tally for BAH_PROP generally increased from west to east. Pixels in high-elevation forests (CC and CS) tended to exhibit greater IPR in STPH and lower IPR in BAH_PROP compared to lower-elevation forests (WV, CL, CM, GF, and PP). For DVPH and BIOM, 95% IPR tended to be greater in western ecoregions (WV, CL, CM, and CC). There were few trends for CANCOV and QMD_DOM. The magnitudes of bootstrap deviations were larger in regions also exhibiting high 95% IPR.

4. Discussion

The utility of imputed maps of forest attributes depends of the uncertainty of the predictions, but assessing uncertainty can be difficult for some modeling approaches. In the case of *k*NN with $k = 1$, we employed non-parametric bootstrapping approach to examine imputation uncertainties across a diverse mosaic of forest landscapes covering several ecoregions and stretching across the Cascade Mountains of Oregon. Our results indicated that there were both clear differences between forest attributes in which imputed attributes carry substantial uncertainties (Fig. 2) and that geographic patterns in those uncertainties imply broad-scale

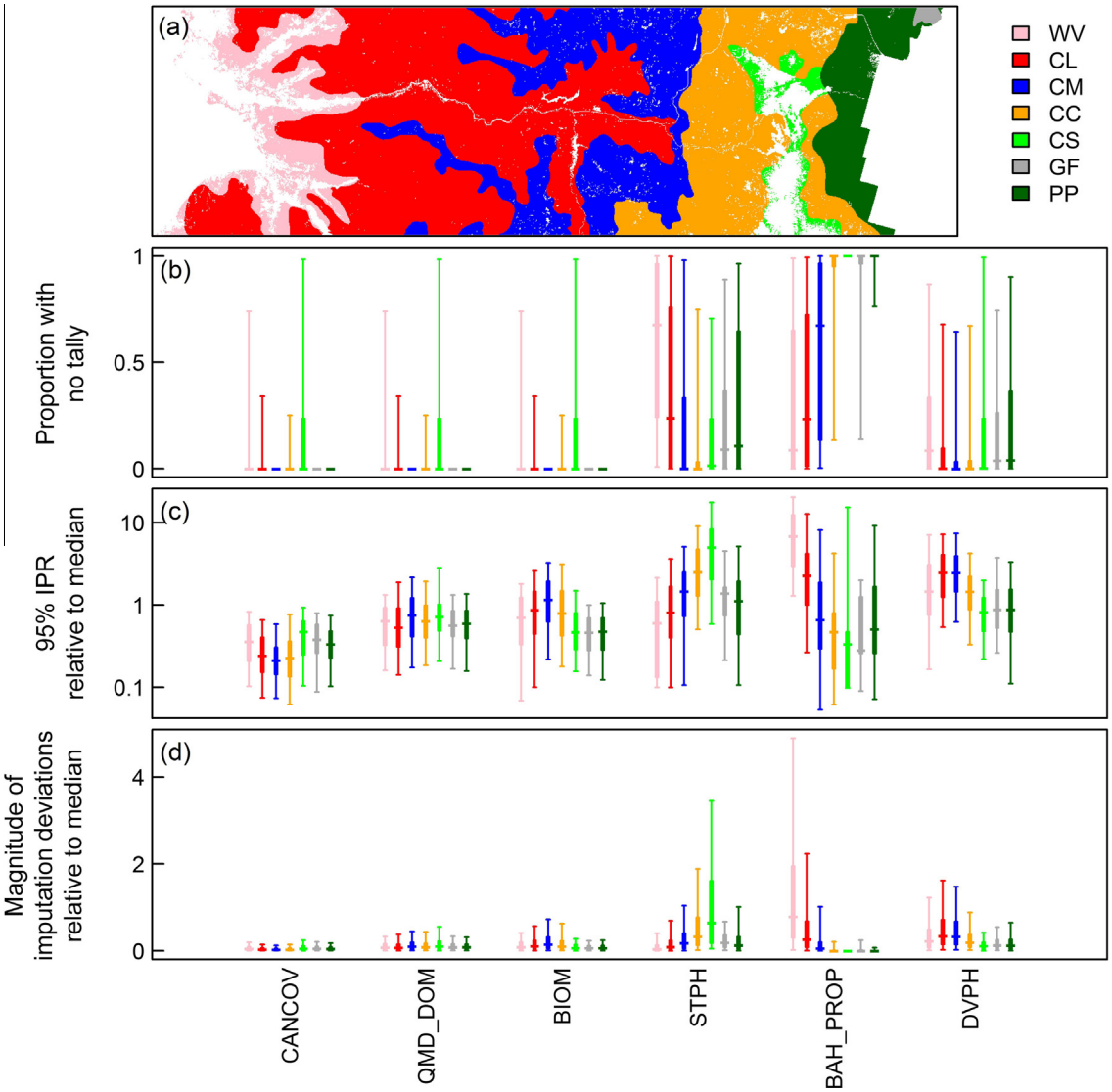

Fig. 4. Bootstrap imputation variability with respect to ecoregion. (a) Nine Level IV ecoregions (McMahon et al., 2001): WV = Willamette River and Tributaries Gallery Forest, Prairie Terraces, and Valley Foothills; CL = Western Cascades Lowlands and Valleys; CM = Western Cascades Montane Highlands; CC = Cascade Crest Montane Forest; CS = Cascade Subalpine/Alpine; GF = Grand Fir Mixed Forest; PP = Ponderosa Pine/Bitterbrush Woodlands. For imputation variation in (b) the proportion of no tally, (c) the 95% IPR, and (d) the magnitude of deviations between baseline GNN and bootstrap imputations, predictions are presented as a function of ecoregion, the horizontal bars denote the medians, the bounds of the rectangle denote the 25% and 75% percentiles, and whiskers denote the 5% and 95% percentiles.

factors influence variability in imputed forest attributes (Figs. 3 and 4). As expected, components of forest structure most closely linked to the Landsat observations upon which the maps were based (e.g., live tree forest structure) exhibited magnitudes of variability in bootstrap sample estimates (i.e., 95% IPR) 20% to 80% less than the regional median of the estimates themselves. Conversely, components less directly linked to spectral reflectance and more associated with disturbance history (e.g., STPH and DVPH) or representing rare components of the landscape (BAH_PROP) did not perform as well. To our surprise, tree size (QMD_DOM) had the second lowest variability in estimates even though Landsat imagery does not observe tree stems which are often obscured in the same fashion as dead wood. Low variability in tree size predictions could be due to correlations between stand height and diameter of the dominant trees.

Geographic patterns in bootstrap results emphasized the roles of biogeography, land-use, and landscape features, both large and small, in influencing imputation predictions. The decline in variability (IPR) in the hardwood component (BAH_PROP) from west to east (Figs. 3e and 4c) results from the nearly complete loss of hardwood species from these forests across the same gradient (Fig. 4b). The elevational trends in standing dead, with greater 95% IPR in high-elevation forests (Fig. 4c), may, in part, be a reflection of the prevalence of private ownership in low-elevation, western portions of the study region and wilderness areas in the high-elevation, eastern portions of the study region. For example, many stands in the Western Cascades Lowlands and Valleys ecoregion (CL) are managed as Douglas-fir plantations and are cut before stands age enough to develop substantial stocks of standing dead biomass nor do they retain legacy dead wood structures from prior to initial disturbance (Hansen et al., 1991). This is supported by larger estimated proportion of no tally for STPH and DVPH in the two western most ecoregions (WV and CL) compared to mid- and high-elevation forests (CM and CC, respectively; Fig. 4b). Similarly, large proportion no tally for STPH and DVPH on the eastern edge of the study region (Fig. 4b) may reflect the historical importance of fire in consuming dead wood (Agee, 2003). The spatial variation in the proportion of bootstraps with no tally of CANCOV, QMD_DOM,

and BIOM reflect the impacts of greater rates of disturbance in managed forests of the western portion of the study area, most notably timber harvest. The low no tally proportion for CANCOV, QMD_DOM, and BIOM was not surprising given these are basic attributes of all forests and should generally be greater than zero in all but the most recently disturbed forests. Conversely, variation in the proportion of no tally in the northeastern portion of the study area may reflect the distribution of lava flows in various states of invasion by tree species or recent wildfires.

4.1. Using uncertainties to understand imputation limitations

Given that biogeography, land ownership, and landscape features are likely to covary spatially (e.g., lava flows are not generally found at lower elevations nor on private lands devoted to timber production), parsing out the drivers of local uncertainties in imputed map predictions will be a time-consuming process requiring knowledge of the landscapes, ecosystems, and species under consideration. While maps of prediction uncertainties are becoming more common (e.g., Beaudoin et al., 2014), little direction has been provided concerning how to use these uncertainties to understand the limits of the imputed maps for answering specific questions. For example, 95% IPR and magnitudes of imputation deviations show similar regional patterns, but imputation deviations appear substantially more variable at a fine scale (Figs. 3, C.3 and C.4). How do users reconcile the differences at fine spatial scales and how can they use these different pieces of information to understand the limitations of the data? Here we present an example of how a scientist or manager could explore imputation predictions and uncertainties to gain a fuller understanding of the appropriate use and interpretation of the data.

In contrast to much of the study area, QMD_DOM was relatively poorly predicted in the northeastern portion of the study area (Fig. 3b), particularly in a small, 20 km^2 area of the map (Fig. 1). In this area, 95% IPR for QMD_DOM were often above 40 cm (Fig. 5a) and baseline GNN predictions and predictions from the bootstrap sample often disagreed (Fig. 5b). After examining aerial photos, it became clear that a major disturbance had occurred on the landscape (Fig. 5c and d). In fact, the Scott Mountain Fire of 2010 burned 1360 ha with high severity occurring in 52% of the burned area (http://www.mtbs.gov). After identifying this disturbance, we examined the QMD_DOM predictions from GNN and found that much of the burned area was predicted to have QMD_DOM near 0 cm (Fig. 5e) and that the predictions over a large continuous area (i.e., within the burned area) were largely imputed from a single forest inventory plot. Furthermore, the mean predictions from the bootstrap sample were greater than the baseline GNN predictions in all portions of the fire except the western edge which was characterized by lower fire severities. Because variation and deviations in QMD_DOM arose from bootstrap samples lacking the nearest neighbor for these pixels, these results highlight a lack of data for QMD_DOM imputation in similar high-elevation, post-disturbance ecosystems. More generally, landscape features poorly represented by the data are likely to be prone to prediction uncertainties and biases and would be challenging to account for, regardless of estimation approach.

The degree to which this uncertainty impacts the use of imputed results depends upon the question being addressed. If precise estimates of QMD_DOM are central to a given project, as might be expected when characterizing habitat suitability for large, cavity-nesting birds, the imputed maps may offer a relatively poor data source within the disturbed landscape of the Scott Mountain Fire. However, if the central goal had been to estimate forest biomass, variations in imputed biomass in the intact forest (northwest corner; Fig. 5f), which are orders of magnitude greater than in the neighboring burned area, may be of much greater concern. As a result, the limitations of imputed maps and the drivers of uncertainties in their predictions will depend on the forest attribute of interest and the geographic setting being examined. Spatial patterns differed not only by forest attribute, but by uncertainty

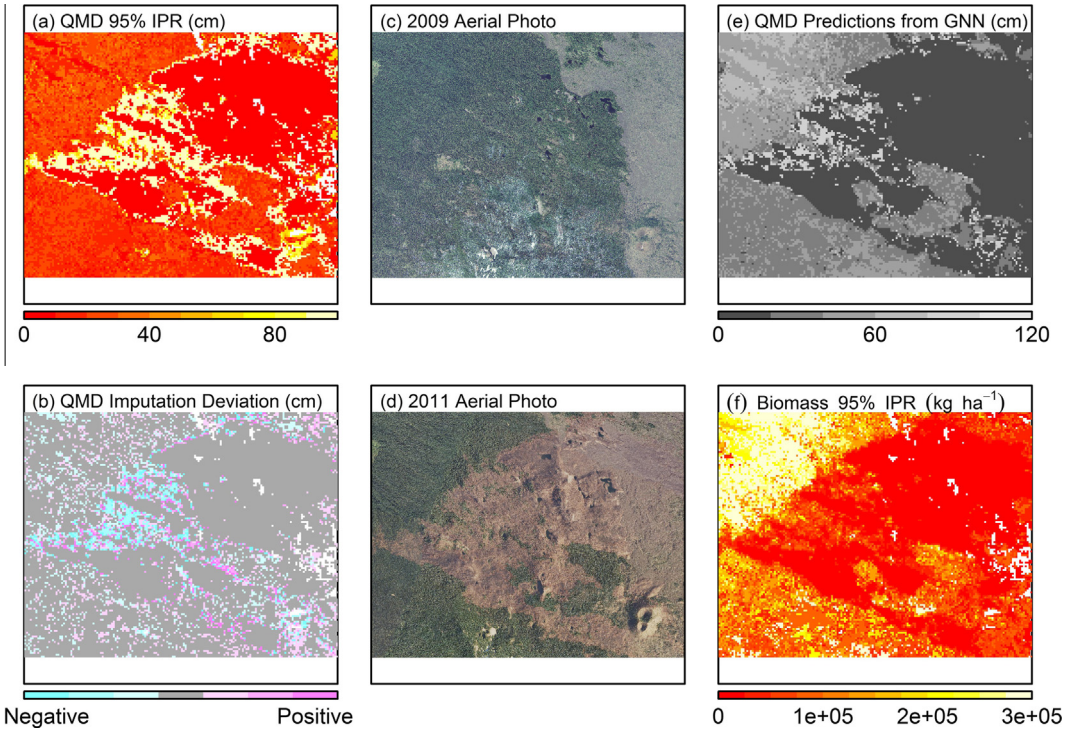

Fig. 5. Representation of the impacts of the Scott Mountain Fire west of Sisters, OR, on QMD predictions. We present spatial patterns in (a–b) QMD 95% IPR and imputation deviations in predictions from the bootstrap sample, (c–d) evidence of wildfire in the area, as shown by aerial photos taken in 2009 and 2011, and spatial patterns in (e) QMD predictions from GNN and (f) 95% IPR of biomass.

metric (Figs. 3 and 4), indicating that no single measure of uncertainty of forest structure provides a full description of imputation performance. This complexity is likely to be challenging for managers and decision-makers working across landscapes and regions, whether examining a recently disturbed area or not. Furthermore, the use of imputed results to initialize other analyses, such as state-and-transition modeling (Halofsky et al., 2013) and forest dynamics modeling (e.g., Falkowski et al., 2010), will almost certainly be impacted by imputation uncertainties. 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.

5. Conclusions

As opposed to previous work involving GNN and other *k*NN with $k = 1$ approaches, the results of the current study explicitly examine geographic variation in imputation uncertainties among a suite of forest attributes to understand the fundamental limitations of Landsat-based forest vegetation mapping using *k*NN with $k = 1$. Components of forest structure most closely linked to the Landsat observations tended to have lower variability whereas components linked to disturbance history and rare components had higher variability. Geographic patterns in bootstrap sample manifested themselves in uncertainty estimates from local (harvest units) to regional (climate gradients) scales. However, uncertainties did not necessarily covary across variables. As a result, the production of uncertainty maps could aid both managers and researchers in understanding the strengths and limitations of Landsat-based vegetation data products at local to regional scales. Further research is needed to determine how uncertainties change over the course of the multi-decadal Landsat record.

Acknowledgements

We thank Warren Cohen, Raymond Davis, Heather Roberts, and Zhiqiang Yang for discussion of concepts and results associated with this research. We thank Barry (Ty) Wilson, Ronald McRoberts, and one anonymous reviewer for comments that improved the quality of the manuscript. Funding for this work was provided by Joint Venture Agreement between Oregon State University and the USDA Forest Service Pacific Northwest Research Station.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.foreco.2015.09.007.

References

Agee, J.K., 2003. Historical range of variability in eastern Cascades forests, Washington, USA. Landscape Ecol. 18, 725–740.

Beaudoin, A., Bernier, P.Y., Guindon, L., Villemaire, P., Guo, X.J., Stinson, G., Bergeron, T., Magnussen, S., Hall, R.J., 2014. Mapping attributes of Canada's forests at moderate resolution through *k*NN and MODIS imagery. Can. J. For. Res. 44, 521–532.

Bechtold, W.A., Patterson, P.L. (Eds.), 2005. The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures. Gen. Tech. Rep. SRS-80. U.S. Department of Agriculture, Forest Service, Southern Research Station, Asheville, NC, 85 p.

Crist, E.P., Cicone, R.C., 1984. A physically-based transformation of Thematic Mapper data – the TM Tasseled Cap. IEEE Trans. Geosci. Remote Sens. 3, 256–263.

Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28, 2031–2064.

Eskelson, B.N.I., Temesgen, H., Lemay, V., Barrett, T.M., Crookston, N.L., Hudak, A.T., 2009. The roles of nearest neighbor methods in imputing missing data in forest inventory and monitoring databases. Scand. J. For. Res. 24, 235–246.

Falkowski, M.J., Hudak, A.T., Crookston, N.L., Gessler, P.E., Uebler, E.H., Smith, A.M.S., 2010. Landscape-scale parameterization of a tree-level forest growth model: a *k*-nearest neighbor imputation approach incorporating LiDAR data. Can. J. For. Res. 40, 184–199.

Hansen, A.J., Spies, T.A., Swanson, F.J., Ohmann, J.L., 1991. Conserving biodiversity in managed forests. Bioscience 41, 382–392.

Halofsky, J.E., Hemstrom, M.A., Conklin, D.R., Halofsky, J.S., Kerns, B.K., Bachelet, D., 2013. Assessing potential climate change effects on vegetation using a linked model approach. Ecol. Model. 266, 131–143.

Henderson, E.B., Ohmann, J.L., Gregory, M.J., Roberts, H.M., Zald, H., 2014. Species distribution modeling for plant communities: stacked single species or multivariate modeling approaches? Appl. Veg. Sci. 17, 516–527.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., Birdsey, R.A., 2003. National-scale biomass estimators for United States tree species. For. Sci. 49, 12–35.

Kagan, J., Ohmann, J.L., Gregory, M.J., Tobalske, C., 2008. Land Cover Map for Map Zones 8 and 9 Developed from SAGEMAP, GNN, and SWReGAP: A Pilot for NWGAP. USGS Gap Analysis Bulletin, No. 15, February 2008, pp 15–19.

Kennedy, R.E., Cohen, W.B., Schroeder, T.A., 2007. Trajectory-based change detection for automated characterization of forest disturbance dynamics. Remote Sens. Environ. 110, 370–386.

Kennedy, R.E., Yang, Z., Cohen, W.B., 2010. Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr – temporal segmentation algorithms. Remote Sens. Environ. 114, 2897–2910.

Key, C.H., Benson, N.C., 2002. Measuring and Remote Sensing of Burn Severity. US Geological Survey Wildland Fire Workshop, October 31 to November 3. US Geological Survey Open-File Report 02–11, Los Alamos, NM, USA.

LeBauer, D.S., Wang, D., Richter, K.T., Cavidson, C.C., Dietze, M.C., 2013. Facilitating feedbacks between field measurements and ecosystem models. Ecol. Monogr. 83, 133–154.

Loveland, T.R., Dwyer, J.L., 2012. Landsat: building a strong future. Remote Sens. Environ. 122, 22–29.

Lu, D., 2006. The potential and challenge of remote sensing-based biomass estimation. Int. J. Remote Sens. 27, 1297–1328.

MacLean, C.D., 1980. A Technique for Identifying Treatment Opportunities from Western Oregon and Washington Forest Survey Plots. Gen. Tech. Rep. PNW-102. USDA Forest Service Pacific Northwest Research Station, 16 p.

Magnussen, S., McRoberts, R.E., Tomppo, E.O., 2009. Model-based mean square error estimators for *k*-nearest neighbor predictions and applications using remotely sensed data for forest inventories. Remote Sens. Environ. 113, 476–488.

Magnussen, S., Tomppo, E.O., McRoberts, R.E., 2010. A model-assisted *k*-nearest neighbor approach to remove extrapolation bias. Scand. J. For. Res. 25, 174–184.

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P., 2005. Zero tolerance ecology: improving ecological inference by modeling the source of zero observations. Ecol. Lett. 8, 1235–1246.

Max, T.A., Schreuder, H.T., Hazard, J.W., Oswald, D.D., Tepley, J., Alergia, J., 1996. The Pacific Northwest Region Vegetation and Inventory Monitoring System. Research Paper PNW-RP-492. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

McMahon, G., Gregonis, S.M., Waltman, S.W., Omernik, J.M., Thorson, T.D., Freeouf, J. A., Rorick, A.H., Keys, J.E., 2001. Developing a spatial framework of common ecological regions for the conterminous United States. Environ. Manage. 28, 293–316.

McRoberts, R.E., 2006. A model-based approach to estimating forest area. Remote Sens. Environ. 103, 56–66.

McRoberts, R.E., 2008. Using satellite imagery and the k-nearest neighbor technique as a bridge between strategic and management forest inventories. Remote Sens. Environ. 112, 2212–2221.

McRoberts, R.E., 2012. Estimating forest attribute parameters for small areas using nearest neighbor techniques. For. Ecol. Manage. 272, 3–12.

McRoberts, R.E., Magnussen, S., Tomppo, E.O., Chirici, G., 2011. Parametric, bootstrap, and jackknife variance estimators for the k-nearest neighbors technique with illustrations using forest inventory and satellite image data. Remote Sens. Environ. 115, 3165–3174.

McRoberts, R.E., Næsset, E., Gobakken, T., 2015. Optimizing the k-nearest neighbor technique for estimating forest aboveground biomass using airborne laser scanning. Remote Sens. Environ. 163, 13–22.

McRoberts, R.E., Tomppo, E.O., 2007. Remote sensing support for national forest inventories. Remote Sens. Environ. 110, 412–419.

McRoberts, R.E., Tomppo, E.O., Finley, A.O., Heikkinen, J., 2007. Estimating areal means and variances of forest attributes using the k-nearest neighbors technique and satellite imagery. Remote Sens. Environ. 111, 466–480.

McRoberts, R.E., Tomppo, E.O., Naesset, E., 2010. Advances and emerging issues in national forest inventories. Scand. J. For. Res. 25, 368–381.

Ohmann, J.L., Gregory, M.J., 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest-neighbor imputation in coastal Oregon, USA. Can. J. For. Res. 32, 725–741.

Ohmann, J.L., Gregory, M.J., Henderson, E.B., Roberts, H.M., 2011. Mapping gradients of community composition with nearest-neighbour imputation: extending plot data for landscape analysis. J. Veg. Sci. 22, 660–676.

Ohmann, J.L., Gregory, M.J., Roberts, H.M., Cohen, W.B., Kennedy, R.E., Yang, Z., 2012. Mapping change of older forest with nearest-neighbor imputation and Landsat time-series. For. Ecol. Manage. 272, 13–25.

Ohmann, J.L., Gregory, M.J., Roberts, H.M., 2014. Scale considerations for integrating forest inventory plot data and satellite image data for regional forest mapping. Remote Sens. Environ. 151, 3–15.
Ohmann, J.L., Gregory, M.J., Spies, T.A., 2007. Influence of environment, disturbance, and ownership on forest vegetation of Coastal Oregon. Ecol. Appl. 17, 18–33.
Ohmann, J.L., Spies, T.A., 1998. Regional gradient analysis and spatial pattern of woody plant communities of Oregon forests. Ecol. Monogr. 68, 151–182.
Pierce Jr., K.B., Lookingbill, T., Urban, D., 2005. A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. Landscape Ecol. 20, 137–147.
Pierce Jr., K.B., Ohmann, L.J., Wimberly, M.C., Gregory, M.J., Fried, J.S., 2009. Mapping wildland fuels and forest structure for land management: a comparison of nearest-neighbor imputation and other methods. Can. J. For. Res. 39, 1901–1916.
Riemann, R., Wilson, B.T., Lister, A., Parks, S., 2010. An effective assessment protocol for continuous geospatial datasets of forest characteristics using USFS forest inventory and analysis (FIA) data. Remote Sens. Environ. 114, 2337–2352.
Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: choosing a spatial unit for thematic accuracy assessment. Remote Sens. Environ. 115, 3044–3055.
ter Braak, C.J.F., 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 67, 1167–1179.
Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sens. Environ. 112, 1982–1999.
Turner, D.P., Ollinger, S., Smith, M.L., Krankina, O., Gregory, M., 2004. Scaling net primary production to a MODIS footprint in support of earth observing system product validation. Int. J. Remote Sens. 25, 1961–1979.
Waddell, K., Hiserote, B., 2005. The PNW-FIA Integrated Database User Guide: A Database of Forest Inventory Information for California, Oregon, and Washington. Forest Inventory and Analysis Program, Pacific Northwest Research Station, Portland, Oregon, USA.
Wiens, J.A., Stralberg, D., Jongsomjit, D., Howell, C.A., Snyder, M.A., 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proc. Natl. Acad. Sci. 106, 19729–19736.
Williams, D.L., Goward, S., Arvidson, T., 2006. Landsat: yesterday, today, and tomorrow. Photogramm. Eng. Remote Sens. 72, 1171–1178.
Wilson, B.T., Lister, A.J., Riemann, R.I., 2012. A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. For. Ecol. Manage. 271, 182–198.
Wilson, B.T., Woodall, C.W., Griffith, D.M., 2013. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage. Carbon Balance Manage. 8, 1.
Zald, H.S.J., Ohmann, J.L., Roberts, H.M., Gregory, M.J., Henderson, E.B., McGaughey, R.J., Braaten, J., 2014. Influence of lidar, Landsat imagery, disturbance history, plot location accuracy, and plot size on accuracy of imputation maps of forest composition and structure. Remote Sens. Environ. 143, 26–38.