Forest Ecology and Management 358 (2015) 361–364

Contents lists available at ScienceDirect

Forest Ecology and Management

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

Correspondence

How sampling and scale limit accuracy assessment of vegetation maps: A comment on Loehle et al. (2015)

David M. Bell [a,*], Matthew J. Gregory [b], Heather M. Roberts [b], Raymond J. Davis [c], Janet L. Ohmann [a]

[a] USDA Forest Service Pacific Northwest Research Station, United States
[b] Department of Forest Ecosystems & Society, Oregon State University, United States
[c] USDA Forest Service Pacific Northwest Region, United States

ARTICLE INFO

Article history:
Received 17 June 2015
Received in revised form 9 July 2015
Accepted 20 July 2015

Keywords:
Accuracy assessment
Imputation mapping
Sampling bias
Spatial resolution

ABSTRACT

Accuracy assessments of remote sensing products are necessary for identifying map strengths and weaknesses in scientific and management applications. However, not all accuracy assessments are created equal. Motivated by a recent study published in Forest Ecology and Management (Volume 342, pages 8–20), we explored the potential limitations of accuracy assessments related to characteristics of the field data: sampling bias and spatial resolution. The authors of the previous paper used data from variable radius plots near northern spotted owl nest sites to assess the predictive accuracy of gradient nearest neighbor (GNN) maps in portions of Oregon and Washington, USA. The field plots used for accuracy assessment (1) potentially biased the accuracy assessment toward older forests and (2) examined accuracy at finer scales than the imputation map predictions under consideration. To examine both the impacts of bias and scale in accuracy assessment, we assessed the predictive accuracy of GNN maps in western and southern Oregon. We found correlation coefficients between predicted (900 m^2) and observed forest attributes for small plots (506 m^2) were consistently lower than accuracy assessments using larger plots (4048 m^2). Similarly, correlation coefficients based only on field plots near nest sites were lower than correlations based on all field plots. These results imply that sampling bias and small plot areas result in accuracy assessments that underestimate map predictive performance. In particular, assessing accuracy at spatial scales below the resolution of the map products are overly pessimistic (i.e., low correlation coefficients). While accuracy assessment is important, care needs to be taken to ensure that the sampling design for field data does not limit inference on map accuracy.

Published by Elsevier B.V.

1. Introduction

In a recent paper (Loehle et al., 2015), the authors (Loehle, Irwin, Manly, and Merrill, hereafter LIMM), as part of a larger analysis of recent northern spotted owl (NSO) habitat modeling (USFWS, 2011), performed an accuracy assessment of a forest vegetation map product, the gradient nearest neighbor (GNN) map, used as input data in the habitat modeling. Briefly, the GNN map is an imputed data product based on a canonical correspondence analysis and a nearest neighbor imputation approach using Landsat imagery and forest inventory data (Ohmann and Gregory, 2002). Imputed maps use statistical models to predict forest attributes for areas where no field observations are available, most commonly using remotely sensed data (Tomppo et al., 2008). An essential component of map production and use is the validation of predictions, allowing both the producer and the user to understand the limitations of map products (Olofsson et al., 2014). While validations of map products are needed, inference on map performance depends on the limitations of the maps themselves and the data used for accuracy assessment. The primary objective of this comment is to discuss the influence of some elements of sample design on the assessment of imputed map products, like the GNN maps, using the work of LIMM as a motivating example. Therefore, we discuss the work by LIMM in the context of (1) past assessments of GNN predictions and (2) influences of sampling bias and scale on accuracy assessment. In addition, we examined GNN map performance in the same general area as the results reported by LIMM.

1.1. Past assessments of GNN predictions

LIMM's paper reported that GNN maps provided little or no predictive accuracy based on their field data (see Section 1.2), but they overstated the support for this claim based on existing

* Corresponding author at: 3200 SW Jefferson Way, Corvallis, OR 97331, United States.

E-mail address: dmbell@fs.fed.us (D.M. Bell).

http://dx.doi.org/10.1016/j.foreco.2015.07.017
0378-1127/Published by Elsevier B.V.

assessments of the GNN product and the limitations of imputation mapping in general. As LIMM point out, the accuracy of GNN predictions range substantially in quality (Pierce et al., 2009). For example, live tree components of forest structure (e.g., canopy cover, basal area, quadratic mean diameter, and stand height) were poorly correlated to GNN data in structurally diverse forests of eastern Washington (squared correlation 0.07–0.17). However, the landscapes of western Oregon, which are dominated by closed canopy forests, exhibited relatively high squared correlation coefficients (0.56–0.70). For other variables, like snag density or down wood volume, no method is yet available that can achieve high predictive accuracy (Pierce et al., 2009). Still, this previous assessment (i.e., Pierce et al., 2009) does not incorporate large areas of southern Oregon most comparable to the results of LIMM. Accuracy assessments like this one have motivated continued improvements to the GNN mapping methodology, resulting in improved performance across forested landscapes of the Pacific Northwest, USA (e.g., Zald et al., 2014).

In addition to previously published examinations, GNN map products are distributed with an accuracy report for each modeling region and year being examined based on a modified leave-one-out (LOO) methodology (Osborne and Tiger, 1991; Ohmann and Gregory, 2002). Examples of these accuracy assessments can be found online (http://lemma.forestry.oregonstate.edu/data/structure-maps). The LOO comparisons provide information to users concerning the quality of the data. It is important to note that the LOO method compares GNN predictions to field plots that (1) represent the entire breadth of ecological variation represented by the model itself and (2) are the same spatial scale as the input data. As opposed to the work by Pierce et al. (2009), the LOO comparisons are available across all of California, Oregon, and Washington (14 forested modeling regions), providing an opportunity for comparison with the results reported by LIMM. For the Klamath (KLE) and Oregon Coast Range (OCR) ecoregions, LOO-based correlation coefficients using all inventory plots within each modeling region indicate a difference in predictive performance ranging from 1.5 to 31 times higher compared to results reported by LIMM (Table 1). This raises the question: What is causing such a large discrepancy?

At least three explanations seem reasonable. First, GNN users attempting to validate GNN predictions with independent data often fall victim to a common pitfall: differences in calculations of forest attributes. Attribute definitions can differ because of eccentricities of the forest inventory data used as GNN inputs. For example, some attributes distributed with the GNN product, such as the quadratic mean diameter of dominant trees, depend on forest inventory field crew designation of trees as dominant, making them difficult or impossible to reproduce based on independent data. However, for the variables reported by LIMM (basal area of all trees, conifers, and hardwoods, basal area of trees between 2.5 cm and 25 cm dbh, and trees per hectare), this issue is unlikely. The second potential explanation is that the sample used to validate GNN predictions may not have been representative of the entire range of forest conditions in the region. Finally, it could be that the field data used by LIMM characterizes a finer spatial resolution (i.e., smaller plot size) than GNN predictions (Fig. 1). We discuss the latter two possibilities in greater detail in the following section.

1.2. Influence of sampling bias and scale on accuracy assessment

Because forest measurements can be expensive and time-consuming to collect, acquiring independent data for accuracy assessments can be a major challenge. Scientists tend to rely on existing field measurements, thus assuming that those previous samples provide an unbiased representation of the population of interest. LIMM used plots from within NSO home ranges. As a result, the data may be biased toward forests characteristic of owl habitat (e.g., older forest; USFWS, 2011). Therefore, the assessment may not examine GNN accuracy in general, but the ability of GNN to predict variation in forest attributes within older, complex forests. Because input data for habitat modeling is informative when it helps to distinguish between places at which species might be present or absent (e.g., Phillips et al., 2009), an accuracy assessment that is biased toward areas where the species is present is of limited use: one needs information characterizing all types of habitats, not just the good habitat, to make conclusions about how map performance might influence habitat modeling.

Another consideration in model evaluation is the appropriate spatial resolution of inference. In particular, it would be inadvisable to apply predictions to finer scales than the original source data. In the case of GNN, input data from a single plot represents several subplots (e.g., four subplots totaling 672 m^2 nested within four macroplots covering 4048 m^2 for FIA) and is predicted at 30-m resolution (900 m^2; Fig. 1a). The 30-m resolution reflects the resolution of the remote sensing data, but users generally aggregate to coarser scales to avoid making inference at finer scales than the field data (e.g., 200 ha). In contrast, LIMM compared individual variable radius plots with BAF 20–40 (English units) with GNN predictions, which implies a scale mismatch as a potential source of poor accuracy. Sampling area for variable radius plots increases with tree size, making the definition of the spatial scale of observations for each plot unclear. For example, with BAF 20 and 40, sampling area for 30 cm trees is 154 and 77 m^2, respectively, while

Table 1
Comparison of GNN map performance in the Klamath (KLE) and Oregon Coast Range (OCR) ecoregions for all plots in the region (A) and only plots within 800-m of northern spotted owl nest sites (approximately 200 ha; *N*) based on Pearson correlation coefficients between predictions and plot data based on a modified leave-one-out validation, LIMM plots, and LiDAR calibration plots. Because LiDAR plots were distributed across both OCR and KLE lands, no region is provided. Note that LIMM provide two assessments for KLE.

	Region	All plots (A) or near nest only (N)	Sample size	Basal area (m^2 ha^{-1})	Conifer basal area (m^2 ha^{-1})	Hardwood basal area (m^2 ha^{-1})	Small tree basal area (m^2 ha^{-1})[a]	Tree density (trees ha^{-1})
LOO validation	KLE	A	3703	0.70	0.73	0.62	0.51	0.47
		N	173	0.59	0.69	0.32	0.31	0.21
	OCR	A	2024	0.75	0.80	0.62	0.60	0.42
		N	81	0.70	0.73	0.38	0.53	0.33
LIMM Plots	KLE	N	410	0.21	0.25	0.02	0.09	0.08
	KLE	N	2092	0.20	0.30	0.16	0.10	0.08
	OCR	N	1779	0.48	0.51	0.06	0.02	0.05
LiDAR 1 × 1	–	A	862	0.36	0.42	0.33	0.31	0.23
	–	N	66	0.40	0.45	0.15	0.35	0.24
LiDAR 3 × 3	–	A	899	0.46	0.50	0.45	0.39	0.40
	–	N	71	0.42	0.47	0.24	0.48	0.31

[a] Small trees are defined as those with diameter less than 25 cm at breast height (1.37 m height).