Ecological Informatics 18 (2013) 20–34

Contents lists available at ScienceDirect

Ecological Informatics

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

Effect of inventory method on niche models: Random *versus* systematic error

Heather E. Lintz [a,*], Andrew N. Gray [b], Bruce McCune [c]

[a] Oregon Climate Change Research Institute, College of Oceanic and Atmospheric Science, Oregon State University, Corvallis, OR, United States
[b] USDA Forest Service, PNW Research Station, Corvallis, OR, United States
[c] Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, United States

ARTICLE INFO

Article history:
Received 3 February 2013
Received in revised form 30 April 2013
Accepted 3 May 2013
Available online 24 May 2013

Keywords:
Niche model
Forest Inventory
Sample design
Uncertainty
Non-parametric Multiplicative Regression

ABSTRACT

Data from large-scale biological inventories are essential for understanding and managing Earth's ecosystems. The Forest Inventory and Analysis Program (FIA) of the U.S. Forest Service is the largest biological inventory in North America; however, the FIA inventory recently changed from an amalgam of different approaches to a nationally-standardized approach in 2000. Full use of both data sets is clearly warranted to target many pressing research questions including those related to climate change and forest resources. However, full use requires lumping FIA data from different regionally-based designs (pre-2000) and/or lumping the data across the temporal changeover. Combining data from different inventory types must be approached with caution as inventory types represent different probabilities of detecting trees per sample unit, which can ultimately confound temporal and spatial patterns found in the data. Consequently, the main goal of this study is to evaluate the effect of inventory on a common analysis in ecology, modeling of climatic niches (or species-climate relations). We use non-parametric multiplicative regression (NPMR) to build and compare niche models for 41 tree species from the old and new FIA design in the Pacific coastal United States. We discover two likely effects of inventory on niche models and their predictions. First, there is an increase from 4 to 6% in random error for modeled predictions from the different inventories when compared to modeled predictions from two samples of the same inventory. Second, systematic error (or directional disagreement among modeled predictions) is detectable for 4 out of 41 species among the different inventories: *Calocedrus decurrens*, *Pseudotsuga menziesii*, and *Pinus ponderosa*, and *Abies concolor*. Hence, at least 90% of niche models and predictions of probability of occurrence demonstrate no obvious effect from the change in inventory design. Further, niche models built from sub-samples of the same data set can yield systematic error that rivals systematic error in predictions for models built from two separate data sets. This work corroborates the pervasive and pressing need to quantify different types of error in niche modeling to address issues associated with data quality and large-scale data integration.

1. Introduction

Survey data collected *in situ* across space and time are indispensable for understanding and managing earth's ecosystems. Biological inventories occur worldwide as repositories of ecological data that can accommodate diverse stakeholders and research (European Commission, 1997; Rudis, 2003a, 2003b). The Forest Inventory and Analysis Program (FIA) of the U.S. Forest Service conducts the largest *in situ* forest data collection effort in North America. However, the current nationally-standard inventory resulted from modifications of regional inventories beginning in 2000. Before 2000, FIA used varied plot sizes, densities, sampling extents, sampling periods, and protocols. While these differences do not affect the ability to provide statistical summaries at local to national scales for many attributes of interest (Barrett, 2004; Bechtold and Patterson, 2005), the comparison of plot-level attributes across inventories can be affected (Gray, 2003). To use historical data with new standardized data, we need to confront an important question that can pervade other biological inventories and large scale data (*e.g.* Hijmans et al., 2000; National Research Council, 2000; Nelson et al., 1990): how do differences among inventories, when combined, confound patterns found in the data? While one study concluded that lumping FIA sample designs is hazardous for the assessment of tree migration through time (Woodall et al., 2009), data from the different FIA sample designs are often lumped across regions (before 2000) or across sample designs (pre- and post-2000) without knowledge of the consequences.

One important use of forest inventory data is the development of ecological niche models (*e.g.* Evans and Cushman, 2009; Iverson

Abbreviations: FIA, Forest Inventory and Analysis Program; NPMR, Non-parametric multiplicative regression.

* Corresponding author. Tel.: +1 541 737 2996 (office); fax: +1 541 737 2540.
E-mail address: hlintz@coas.oregonstate.edu (H.E. Lintz).

1574-9541/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ecoinf.2013.05.001

and Prasad, 1998; McKenzie et al., 2003; Rehfeldt et al., 2006, 2008; Svenning and Skov, 2004). The relationship between a species and its environment is part of a 'species niche' or an '*n*-dimensional hypervolume' that describes conditions where a species can persist (Hutchinson, 1957). Currently, niche models are used for many research purposes including species' conservation (*e.g.* Hannah et al., 2002; Marini et al., 2009), species' re-introduction (Yanez and Floater, 2000), species' migration and invasion (*e.g.* Crossman et al., 2011; Woodall et al., 2009), biodiversity conservation (*e.g.* Newbold and Eadie, 2004), specimen collection (*e.g.* Jarvis et al., 2005), the discovery of new species (Raxworthy et al., 2003), and estimating potential effects of climate change on species (*e.g.* Iverson et al., 2008). Niche models are also used to investigate basic scientific questions on various topics (*e.g.* Engelbrecht et al., 2007; Graham et al., 2004; Hugall et al., 2002; Kelly et al., 2008; Svenning and Skov, 2004).

The effect of combining data from different inventories can be problematic in niche modeling largely due to different plot-level probabilities of detecting species (Azuma and Monleon, 2011; Grosenbaugh and Stover, 1957). This can bias niche models, particularly if different sample designs are used in different portions of a species' niche and/or geographic range. In addition, the amount of resources and effort devoted to sampling often increases with tree size in forest inventories, which could bias models if tree size varies substantially in different portions of a species' niche or range. A change in forest inventory methods clearly has the potential to affect niche models, and to our knowledge, this effect has not yet been examined.

Consequently, we ask, does the change in forest inventory type affect niche models and their predictions for tree species across the Pacific coastal United States? In so asking, we determine how results from an amalgam of approaches (from the old inventory) compare to results from a single, large-scale, standardized approach (the new inventory).

We assume little change in species' probability of occurrence across two sequential time periods of study. We also treat all aspects of sampling that changed in 2000 (further described below) as a single source of potential error. We compare two large samples from the old and new inventory, and we compare two large samples from the new inventory. We then juxtapose these two comparisons, and we examine both random error and systematic error from these comparisons. We define random error as differences between modeled predictions that follow a random frequency distribution and can be due to chance. We define systematic error as error between modeled predictions that demonstrate bias or directional disagreement due to causes likely not associated with chance.

Specifically, we address the following questions:

- How does the change in FIA inventory affect random prediction error for models built from each inventory?
- How does the change in FIA inventory affect systematic prediction error for models built from each inventory?
- How do prediction error types shown across inventories compare to prediction error types for samples within the same inventory?
- How do models and mapped prediction errors differ for within-inventory samples and across-inventory samples?

We answer these questions and also investigate evidence of bias in the data themselves to assist our interpretation of its potential effect on the models and their predictions.

2. Methods

2.1. FIA inventories

The U.S. Forest Service Research branch was mandated in 1928 to report on the status and trends of forest resources on all lands, with a focus on timber production (Frayer and Furnival, 1999). Since the 1930s, surveys have differed by state and region, and inventories were conducted all-at-once for a state (known as the periodic inventory). The periodicity of state inventories ranged from 4 to 18 or more years, and by the early 1990s, most states completed a third inventory cycle with some states re-inventoried as many as six times (Frayer and Furnival, 1999; Hiserote and Waddell, 2003). Starting in 1990, the Forest Service initiated surveys through the Forest Health Monitoring Program (FHM), which used a fixed sampling grid across the U.S. with a quarter of the grid measured each year (Scott et al., 1993). A decade later, the periodic inventory merged with FHM to form the nationally-consistent annual inventory. Instead of periodic sampling that differed by state and rotated by state, the annual inventory samples 10% of permanent plots yearly in the west on a common national grid irrespective of state boundaries.

Key differences among the periodic and the annual inventories for our study region include periodicity, grid density, sampling extent, plot size, and protocol (Tables 1 and 2). Each inventory represents a systematic sample with variable density, sampling dates, and types of lands included. In the old periodic inventory, some management regions only included land capable of timber production (or sites capable of producing 1.4 cubic meters per hectare per year at their peak of mean annual increment). Not all inventories included lands protected from timber production (*e.g.* State and National Parks). The greatest difference in sample population between old and new inventories was the large National Parks in California and Washington that were measured in the new inventory. Some inventories relocated subplots from the fixed design if a plot happened to straddle more than one condition class. Condition class classifies variation in a sampled area with respect to land use, forest type, and stand size class. Important differences between inventories at the plot level include the area sampled and the sampling methods used; the new design relies on a standard set of nested, fixed-radius plots centered on four points for sampling trees of different sizes (see Appendix A, Fig. A1). However, the old inventory often used a variable-radius method to sample most trees >12.7 cm DBH. This was done using a wedge

Table 1
Data source definition, inventory dates, and sampling density by inventory units in California, Oregon, and Washington (Hiserote and Waddell, 2003).

Code for source name	Source name	States	Dates of inventory	Distance between points of sample grid
Old, periodic inventory				
WWA	FIA, Western Washington	WA	1988–1990	3.9 km
EWA	FIA, Eastern Washington	WA	1990–1991	5.5 km
CA	FIA, California	CA	1991–1994	5.5 km (7.7 km in oak woodland)
WOR	FIA, Western Oregon	OR	1995–1997	5.5 km
EOR	FIA, Eastern Oregon	OR	1998–1999	5.5 km
R6	Forest Service, Region 6, Pacific Northwest	OR, WA	1993–1997	2.7 km (5.5 km in wilderness)
R5	Forest Service, Region 5, Pacific Southwest	CA	1993–2000	Numerous (5.5 km base grid)
BLM	Bureau of Land Management	Western OR	1997	5.5 km
RMRS	FIA, Rocky Mountain Research Station	Eastern WA, Eastern CA	2001, 1997	5.5 km
New, annual inventory—all states				
PNW annual	*FIA*	*CA, OR, WA*	*CA and OR began in 2001;WA in 2002*	5 km

Table 2
Differences in plot-level protocols by inventory units for the "large tree" size class. Variable radius plots were sampled using a wedge prism. Each inventory sampled smaller trees with a variety of fixed-radius plot sizes.

Code for source name	Plot radius (m)	BAF* (m^2/ha)	Sub-plots (#)	Total area (m^2)	Tree tally size criteria (DBH, cm)
Periodic inventory					
WWA	Variable	6.9	5	Variable	>17.8**
EWA	Variable	9.2	5	Variable	>12.7
CA	Variable	6.9	5	Variable	>17.8
WOR	Variable	6.9	5	Variable	>12.7
EOR	Variable	4.6 or 6.9	5	Variable	>12.7
R6 + BLM	8.016	N/A	5	1009.4	>7.6
R6 + BLM	15.575	N/A	5	3810.6	>33
R5	Variable	4.6 or 9.2	5	Variable	>12.7
Annual inventory					
PNW annual	7.32	N/A	4	672.5	>12.7
PNW annual	15.575	N/A	4	4050.1	>76.2 west; >61 east

"West" and "east" refer to locations relative to Cascade Mountains in Oregon and Washington. In California, the same minimum diameter as "east" was used.
* BAF stands for basal area factor.
** DBH stands for tree diameter at breast height.

prism projecting a fixed angle from a central point (Bitterlich, 1948; Grosenbaugh, 1952). The prism angle (or basal area factor, BAF) varied among regions to maximize efficiency (Table 2; also see Appendix A; Fig. A1), and this affected the probability of sampling trees of different diameters. The probability of tree capture is proportional to tree basal area for the variable-radius method (Grosenbaugh and Stover, 1957); whereas, the probability of tree capture is proportional to tree density for fixed-radius plots (Grosenbaugh and Stover, 1957). Also, the subplots of the new design, in general, are closer together and fewer compared to the old design. The tree tally size criteria differ among and within designs (Table 2; Fig. A1 of Appendix A).

The data we used from the old design sampled 100% of the plots on a 1 to 4 year cycle but not synchronously among regions (Hiserote and Waddell, 2003). We used data from the new design that sampled 10% of plots across regions per year with a full round occurring every decade. Data from the new design spanned 2001 to 2007, and the data from the old design spanned 1988 to 2000 (with the exception of one ownership, which ended in 2001).

2.2. Study area and data preparation

The area of study contained the conterminous states within the Pacific Coast unit for the FIA program. The included states, California, Oregon, and Washington, are topographically diverse with numerous mountain ranges. Maritime influence combines with complex orographic effects and a wide span of latitude to create a variety of different climatic zones and vegetation types. The region has the broadest range of average annual precipitation (from 25 to 4600 mm/year) found in the lower 48 United States.

Plot grid density and sample size were equalized among model-building data sets. Plots were selected from the old design that was on the shared base grid with 5.5 km spacing, except for the R5 inventory where this was not possible, which resulted in a subset of plots with mean spacing of 3.9 km. Plots with a tree species recorded more than 100 km outside the species' range (as determined by existing flora and current herbarium records) were examined and removed from the analysis because they were presumed to be errors in identification. The data removed for this work subsequently went through an FIA internal review process (while this paper was in review) and were confirmed to be incorrectly recorded through plot re-visits. A total of 27 plots from the old design and 36 from the new design were removed. A certain error rate in identification should be assumed among con-generics in the FIA database, and most outlier populations in our region would be otherwise documented in herbaria. Also, only plots that were at least 50% forested were selected from each dataset, resulting in pools for the old and new designs of 10,831 and 6950 plots respectively. The larger sample from the old design was randomly sampled to obtain a sample size equivalent to that of the new design (N = 6950). The sample size of 6950 is still large, and for this reason, the two samples are statistically similar. For example, the number of presences for all species studied essentially fall on the 1-to-1 line (when compared among samples). A linear regression between numbers of presences in each sample yielded R^2 of 0.9992.

Each data set (one for the old and one for the new inventory) was randomly sampled without replacement to split into two halves of equal size, referred to as the "training" and "testing" data sets. Species' occurrences were summed for the training data. Species with more than 25 occurrences in both training sets were retained for a total of 41 species (Table 3). The training data were used to build models and the testing data were used for model selection and evaluation. Fig. 1 shows geographic comparisons of training data set locations among designs.

We extracted climate data corresponding to FIA plot locations to serve as climatic variables or predictors. We started with 11 variables derived by Ohmann and Gregory (2002) from Daymet grids of the western United States (Thornton et al., 1997) (Table 4). We reduced the number of grids or climate variables as input for our analyses by performing a Principal Components Analysis (PCA) (PC-ORD version 5.2, McCune and Mefford, 2006). We first gathered data for the PCA by taking a random sample comprising 8000 points across Washington, Oregon, and California. These random points were not associated with FIA plot locations. Instead, they were a random sample across the entirety of California, Oregon, and Washington. We choose 8000 points to ensure high density coverage at the resolution of the climate grids. The climate grids resolved data to 4 km. (The authors had past success building strong niche models for tree species using climate grids at the scale of 4 km, so this scale was maintained for this work.) We extracted the climate data from the 11 Daymet grids corresponding to our random sample of 8000 points. We based our PCA on a matrix of correlation coefficients among the data. We selected the first four components that represented 97% of variability in the data (Table 5). The four corresponding eigenvectors were then used to generate the new grids representing PCA scores across Washington, Oregon, and California (Table 5). These PCA scores were extracted from the PCA grids to corresponding FIA pot locations. The use of PCA scores as predictors in our niche models ensured statistical independence among the predictors. The PCA scores also simplified the comparison among inventories by reducing the number of predictors.

While the fourth PCA component has a low eigenvalue, we are not testing a hypothesis concerning how much of the variance is represented by the principal components. Rather, we are interested in whether the principal components serve as useful predictors for external variables. The amount of variance explained in the original matrix has no bearing on whether the component represents a phenomenon with biological importance in subsequent regression (Joliffe, 1982).

2.3. Model building

Non-parametric multiplicative regression (NPMR) was used to build the climate niche models using HyperNiche 2.0 (McCune, 2006; McCune and Mefford, 2008). This technique was chosen among numerous empirical techniques in the literature for species-habitat models (e.g. Elith et al., 2006; Guisan et al., 2007; Kampichler et al., 2010; Pino-Mejías et al., 2010) as it captures the nature of biological response to multiple interacting factors (McCune, 2006). The technique has been compared with linear regression, logistic regression, General Additive Models (GAMs), Classification and Regression Trees (CART), and Random Forest (Lintz et al., 2011; McCune, 2006). Current theory supports that species' response patterns take non-linear, complex shapes (Austin, 2002; Oksanen and Minchin, 2002), and when tested against other species distribution modeling methods, NPMR proved to

Table 3
Forty-one species were studied. Species were retained for study if they contained more than 25 occurrences in the primary training data sets from the old and the new design.

Species code	Latin name	Common name	Prevalence (old)	Prevalence (new)
ABAM	*Abies amabilis*	Pacific silver fir	243	213
ABCO	*Abies concolor*	White fir	801	579
ABGR	*Abies grandis*	Grand fir	333	351
ABLA	*Abies lasiocarpa*	Subalpine fir	189	164
ABMA	*Abies magnifica*	California red fir	214	121
ABPR	*Abies procera*	Noble fir	70	80
ABSH	*Abies shastensis*	Shasta red fir	31	46
ACMA	*Acer macrophyllum*	Bigleaf maple	291	261
AECA	*Aesculus californica*	California buckeye	26	37
ALRU	*Alnus rubra*	Red alder	334	358
ARME	*Arbutus menziesii*	Pacific madrone	305	301
CADE	*Calocedrus decurrens*	Incense-cedar	573	398
CHNO	*Chamaecyparis nootkatensis*	Alaska yellow-cedar	32	35
CONU	*Cornus nuttallii*	Pacific dogwood	100	72
JUOC	*Juniperus occidentalis*	Western juniper	274	356
LAOC	*Larix occidentalis*	Western larch	210	212
LIDE	*Lithocarpus densiflorus*	Tanoak	197	271
PIAL	*Pinus albicaulis*	Whitebark pine	57	34
PICO	*Pinus contorta*	Lodgepole pine	475	450
PIEN	*Picea engelmannii*	Engelmann spruce	178	145
PIJE	*Pinus jeffreyi*	Jeffrey pine	389	213
PILA	*Pinus lambertiana*	Sugar pine	515	324
PIMO	*Pinus monophylla*	Singleleaf pinyon	58	76
PIMONT	*Pinus monticola*	Western white pine	255	184
PIPO	*Pinus ponderosa*	Ponderosa pine	1133	958
PISA	*Pinus sabiniana*	California foothill pine	48	73
PISI	*Picea sitchensis*	Sitka spruce	76	73
POBAT	*Populus balsamifera*	Black cottonwood	60	35
PSME	*Pseudotsuga menziesii*	Douglas-fir	1830	1920
QUAG	*Quercus agrifolia*	California live oak	48	31
QUCH	*Quercus chrysolepis*	Canyon live oak	407	374
QUDO	*Quercus douglasii*	Blue oak	56	97
QUGA	*Quercus garryana*	Oregon white oak	101	106
QUKE	*Quercus kelloggii*	California black oak	463	317
QUWI	*Quercus wislizenii*	Interior live oak	84	94
SESE	*Sequoia sempervirens*	Redwood	44	76
TABR	*Taxus brevifolia*	Pacific yew	116	81
THPL	*Thuja plicata*	Western redcedar	332	321
TSHE	*Tsuga heterophylla*	Western hemlock	567	613
TSME	*Tsuga mertensiana*	Mountain hemlock	180	161
UMCA	*Umbellularia californica*	California-laurel	97	98

be most tractable with any shape of underlying data structure (Lintz et al., 2011). Further, recent tests with MaxEnt against NPMR demonstrated that NPMR had the highest mean externally-validated prediction accuracy compared to MaxEnt for 48 different shapes of simulated data structures (Yost and Lintz, 2013, unpublished data). NPMR has advantages over other methods through reduced model bias and improved prediction accuracy (Lintz et al., 2011). It is designed to automatically accommodate complex interactions and increase user understanding of the nature of complex interactions. NPMR is an iterative algorithm that can take more time than other methods especially with large data sets.

NPMR objectively optimized kernel width by maximizing fit. To do this, a local window centered on a target point within the predictor space (as with other kernel smoothers). A weight function was applied within the window to assign numeric weight to points surrounding the target that decrease with distance from the target. A local model (the local mean) predicted the dependent variable at the target point. NPMR repeated this procedure for all target points to generate a prediction curve or surface. NPMR omitted the target point when predicting the response at that point and multiplied weights from individual predictors. The multiplicative combination of weights can accommodate any type of interactions including additive or multiplicative.

To adequately gauge the effect of inventory on niche models, we compared statistics from models built from the different data. We also compared qualitative attributes of the models and geographic maps of predicted probabilities of occurrence generated from the different models. We juxtaposed comparisons of the old and new inventories (referred to as 'old–new') with comparisons of the two subsamples of the new inventory (referred to as 'new–new'). The old–new comparison used the training sets from each inventory for model building. The new–new comparison used both the testing and training data sets within the new design for model building. Each half of the data set from the new inventory was a sample for model building with the other half used for model testing. In this way, we examined the type of error that can arise in NPMR from taking different samples of the same data set where the difference among data sets was maximized. The data sets were so large that bootstrapped confidence bands were both unnecessary and too time consuming. We reasoned that if the same magnitude and type of deviations from the predicted 1:1 line across species in the 'old–new' comparison are also seen in the 'new–new' comparison, then error arising from sources other than sample design or environmental change was likely the cause.

2.4. Model evaluation and selection

Two measures of model performance for binary data were used for different purposes, model selection and model evaluation. First, we used a measure called the $\overline{LogB}$, which is based in the common likelihood ratio. To derive this measure, we take the log of the likelihood ratio and divide it by the number of sample units (or FIA plots used for model building). In so doing, we obtain the average contribution of a sample unit to the log likelihood ratio for the purpose of model selection (McCune, 2006). Popular statistics for model selection with classifiers are often derived from the likelihood ratio such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Hastie and Tibshirani, 2001); however, the AIC and BIC approximate the optimization curve used in model selection for sample sizes too small to generate the curve empirically (Hastie and Tibshirani, 2001). The empirical optimization curve represents the loss in externally-validated fit that occurs as model complexity increases. Theoretically, the model with the greatest externally-validated likelihood ratio is the most robust and accurate. Given our large sample size, instead of relying on an index to approximate this curve such as an AIC, we rely on external validation itself using $\overline{LogB}$. This is considered to be the ideal scenario for model selection that can occur only when sufficient data are available (Hastie and Tibshirani, 2001).

Second, we use the Area under the Receiver Operating Characteristic (ROC) curve or AUC for model evaluation (Hanley and McNeil, 1982). The $\overline{LogB}$ is not suited to compare the fit of models among species with different numbers of presences because the $\overline{LogB}$ varies with number of presences. Instead, we use the AUC for model evaluation among species as it is much less sensitive to number of presences (Fawcett, 2006). An AUC of 0.5 represents a model fit no better achieved by chance alone. The maximum value of the AUC is 1. See Fig. A2 for a comparison of AUC with $\overline{LogB}$ for candidate models of three species.

Climate and species' occurrence data from plots withheld from model building were used to calculate externally-validated measures of model performance. The withheld climate data were supplied as input to the candidate models (after the models were built from the training data sets), and the corresponding withheld species' presence/absence data were compared to the resulting predictions.

2.5. Inventory effect on models and data

We compared models based on the two inventory types using the following measures: the fit as externally-validated AUC, the quality of the predictors or the type of predictors chosen for a model, random

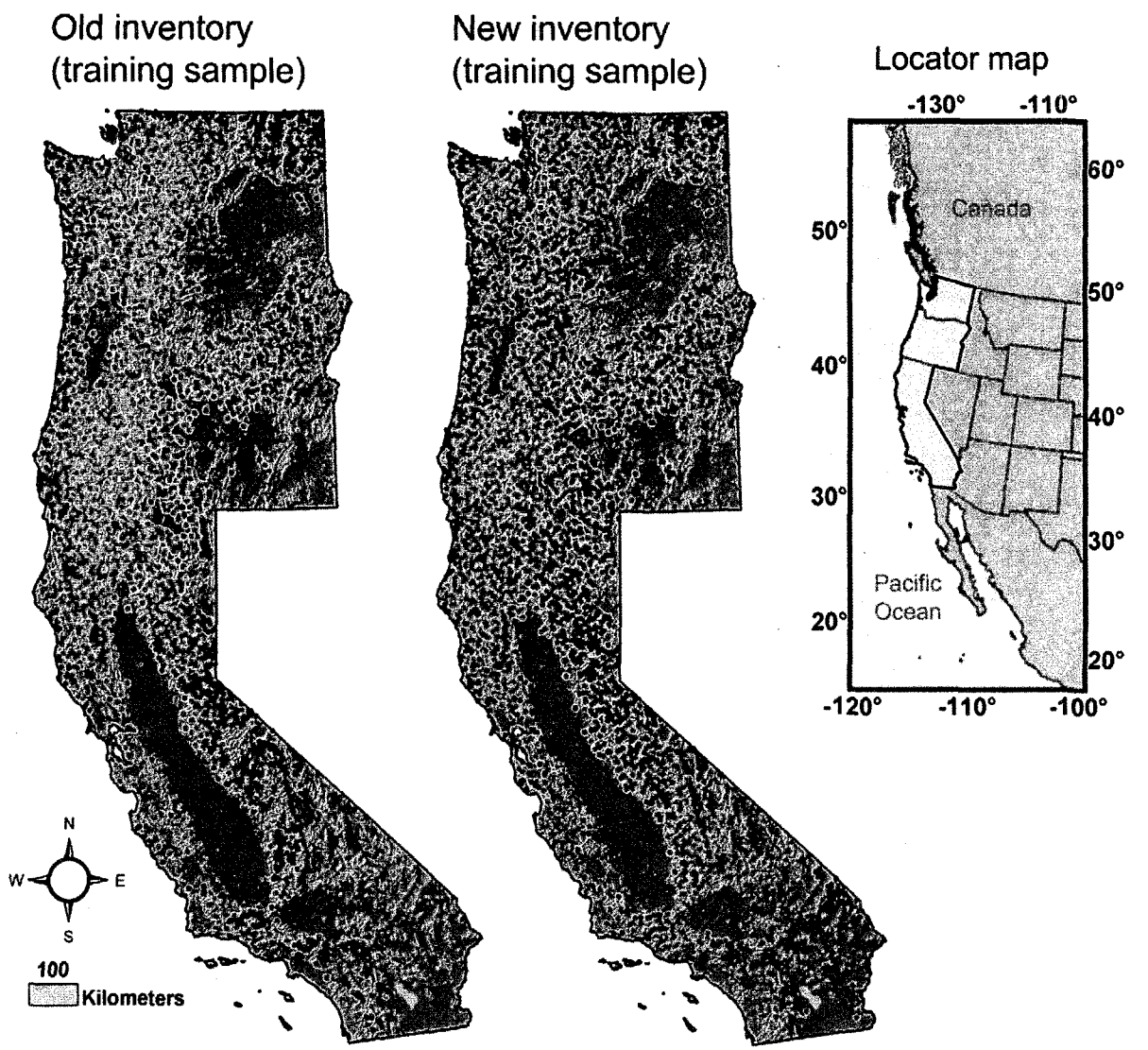

Fig. 1. Study area. Locations of FIA plots for training samples (N = 3475) are shown for the old and new design.

deviations from the 1:1 line for predicted values among models of the same species, and systematic deviations from the 1:1 line (or whether models from one data set tended to overestimate or underestimate probability of occurrence, relative to the other data set). Model comparisons were performed by obtaining predictions from models based on the same random sample of 3475 points and their corresponding

Table 4
Definitions for climate variables derived by Ohmann and Gregory (2002) from Daymet values (Thornton et al., 1997) used in the Principal Components Analysis.

Variable	Definition
ANNPRE	Natural logarithm of mean annual precipitation (mm)
ANNSWRAD	Annual average of the total daily incident shortwave radiative flux (MJ m^{-2} day^{-1})
SMRPRE	Natural logarithm of mean precipitation from May through September (mm)
CVPRE	Coefficient of variation of mean monthly precipitation for wet and dry months (December and July)
ANNGDD	Average number of growing degree days where daily air temperatures exceed 0.0 °C
SMRTP	Moisture stress during the growing season; a ratio of mean summer temperature (SMRTMP) over mean summer precipitation (SMRPRE)
ANNVP	Annual mean of the daily average of partial pressure of water vapor in the air near the surface (Pa)
AUGMAXT	Mean maximum temperature in August (°C)
DECMINT	Mean minimum temperature in December (°C)
DIFTMP	Difference between AUGMAXT and DECMINT (°C)
CONTPRE	Percentage of mean annual precipitation falling June through August

Table 5
PCA loadings corresponding to the Daymet climate variables for each component or eigenvector. We used (and show) the first four eigenvectors or V vectors as climatic predictors where each is scaled to its standard deviation. The PCA was based on a matrix of correlation coefficients among the data. Each respective eigenvalue is shown in the second to last row, and the cumulative variance explained with the addition of each eigenvector is shown in the last row.

Variable	PCA1	PCA2	PCA3	PCA4
ANNGDD	0.9228	0.2589	0.2695	0.0319
ANNPRE	−0.8082	0.5282	−0.0299	0.1217
ANNSWRAD	0.7613	−0.1694	−0.4915	−0.2046
ANNVP	0.2223	0.7232	0.6341	0.1060
AUGMAXT	0.9291	−0.1634	0.1353	0.2793
CONTPRE	−0.1127	−0.7740	0.6006	−0.0894
CVPRE	−0.0056	0.7378	−0.6187	0.2054
DECMINT	0.6329	0.6928	0.3103	−0.0735
DIFTMP	0.3694	−0.8390	−0.1563	0.3644
SMRPRE	−0.9135	−0.0095	0.3162	0.1599
SMRTP	0.9703	0.0903	0.0506	−0.0096
Eigenvalue	5.323	3.260	1.702	0.361
Cumulative % variance	48.39	78.03	93.49	96.78

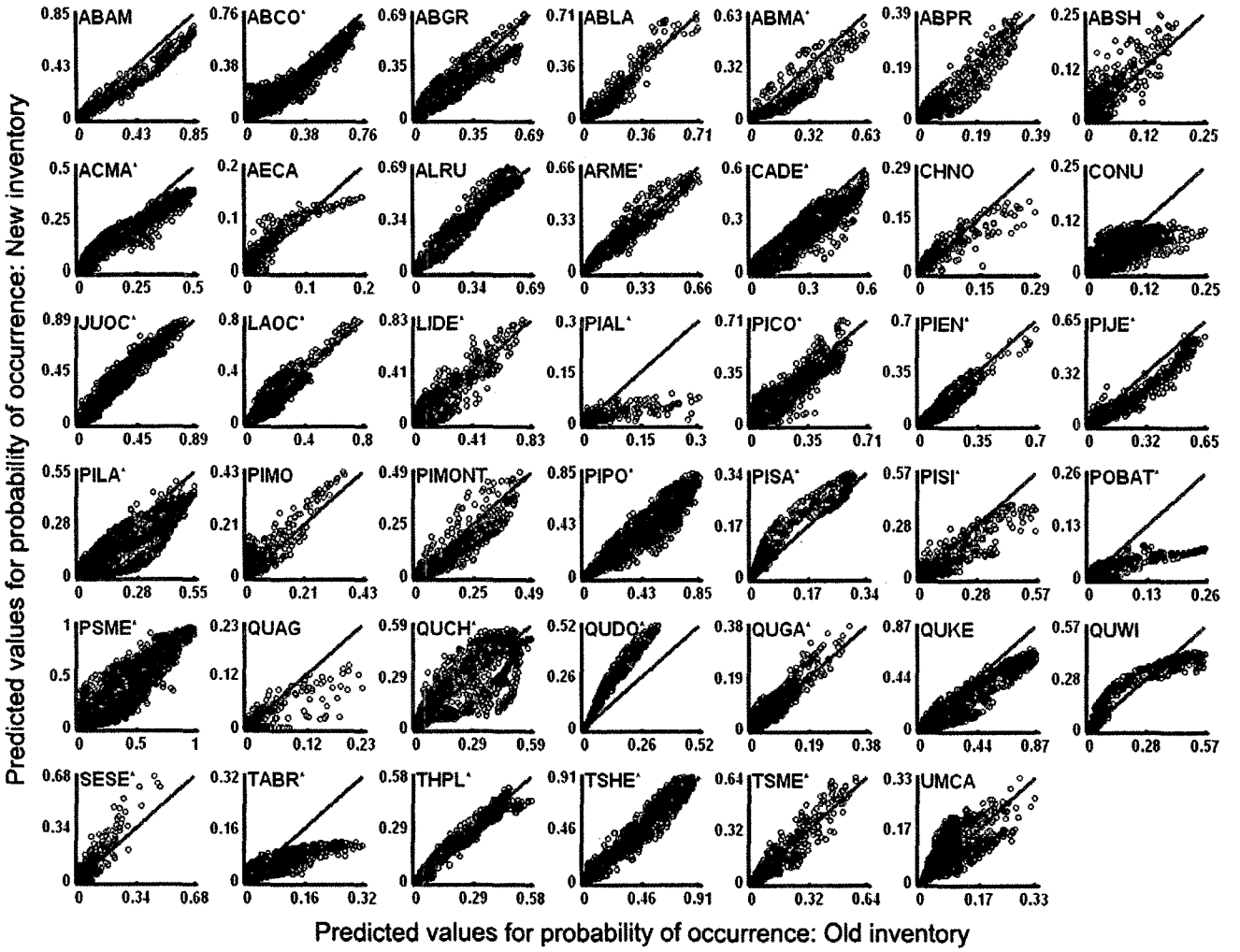

Fig. 2. Predicted probability of occurrence by species for models built from new *versus* old sample designs. Red line represents the ideal 1:1 line. Predicted values were generated from a random sample of unseen climate data (N = 3475). Species codes with asterisks* denote that the compared models for that set of axes had the same model functional form or number and type of predictors.

climate values within the study area. Predictions were compared for old–new and new–new comparisons. The random error or non-directional deviation from the 1:1 line was derived using Root Mean Squared Prediction Error (RMSE). However, the RMSE is a function of the maximum in predicted probability of occurrence for a species, which varied with a species number of presences. Hence, the RMSE was standardized to compare across species with different maxima in probability of occurrence. The differences from the 1:1 line were divided by the range of the data or maximum probability of occurrence before squaring and summing to yield the 'normalized RMSE' or NRMSE. The NRMSE is often expressed as a percentage. The NRMSE represents the degree of deviation from the 1:1 line as a percentage of the axis length. The Wilcoxon signed-rank test evaluated the null hypothesis that the median in NRMSEs from the new–new and old–new comparisons was equal (Wilcoxon, 1945).

The systematic deviations from the 1:1 line were also compared. The residuals were standardized by their range, and three quartiles, the 25th, 50th, and 75th, were plotted and compared for the old–new and new–new comparisons. These indicate the central tendency of the residuals and whether one data set tends to model and predict a greater probability of occurrence compared to another. The median standardized residual tracks the median *non-zero* standardized residual closely and linearly except for values very near to zero, which were not meaningful.

Raw differences among data sets were investigated to aid the interpretation of modeled comparisons. We compared several metrics from the data: 'climatic bias,' probability of tree capture, and number of presences for a species. Climatic bias was defined as the disagreement among two histograms where each histogram represents the frequency of climatic data values corresponding to locations where a species' was found (Kadmon et al., 2003). This characterized and compared species-specific structure of climatic data from two different samples of presence/absence data. Instead of calculating a statistic to characterize climate bias, we relied on data visualization for the following reason. We considered using a measure of effect size, the Kolmogorov–Smirnov two-sample test statistic (d) (Massey, 1951), to compare climatic bias across different sample sizes (or species with different numbers of presences). This statistic assumes no particular form between the compared distributions, it measures the maximum absolute difference among empirical cumulative distribution functions, and it accommodates differences in both shape and central tendency. However, we checked the immunity of this statistic, d, to sample size, and we discovered d to depend on sample size using simulated data (which was surprising for a measure of effect size) (Appendix A, Fig. A3; Appendix B).

Also, we derived gross measures of probability of tree capture across sample designs using the average number of trees per FIA plot by management region. Before the switch from the old to new design, approaches to sampling differed in numerous aspects by management

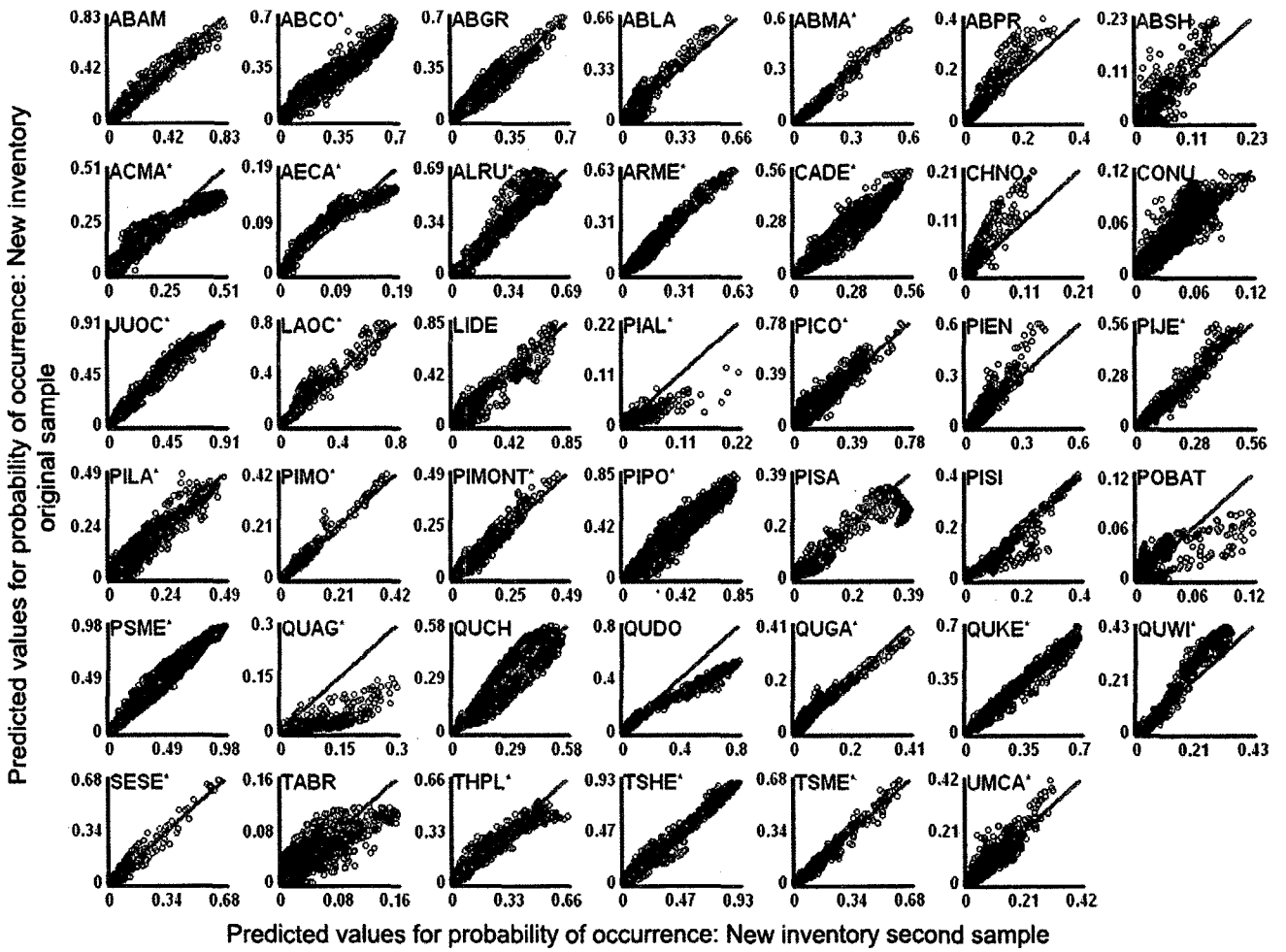

Fig. 3. Predicted probability of occurrence by species for models built from 'new *versus* new' sample designs where a random sample from one design is compared to a second random sample from the same design. Red line represents the ideal 1:1 line. Predicted values were generated from a random sample of unseen climate data (N = 3475). Species' codes with asterisks* denote that the compared models for that set of axes had the same model functional form or number and type of predictors.

region; hence, we used management regions as sample units to examine probability of tree capture across designs.

2.6. Sample design effect on maps

We used geographic grids of PCA scores of climate variables across the study area as input to make maps of probability of species' occurrence for niche models (using HyperNiche 2.0 and ArcGis 9.2). We mapped differences among predictions for new–new and old–new comparisons. This was done for three species, *Arbutus menziesii*, *Tsuga heterophylla*, and *Pinus ponderosa*. The species were chosen arbitrarily to span a range of numbers of presences in the data (N = 301, 613, 958 respectively). We extrapolated conservatively or little beyond the existing range of the data using the default setting in HyperNiche 2.0 where the minimum neighborhood size (in environmental space) for an estimate (in geographic space) is equal to or greater than a quarter of the average neighborhood size for a model. The average neighborhood size is the average amount of data bearing on the estimate of the response variable at each point.

3. Results

3.1. Sample design effect on models

Most species from the old and the new FIA sample designs yielded similar predicted values for models of species' probability of occurrence (Fig. 2). The mean NRMSE in the old–new comparison was 6% (95% quantiles: 3%, 12%) compared to a mean of 4% (1%, 10%) in the new–new comparison for the 41 species. The difference among median NRMSEs (across species for old–new and new–new comparisons) was not likely to be due to chance alone (Wilcoxon signed-rank *two-tailed*, $p < 0.001$). Additionally, many species that showed strong systematic deviation from the 1:1 line among designs (detectable by eye) also showed this type of deviation for models built from different samples of the same design (AECA, CHNO, PIAL, POBAT, QUAG, QUDO, QUWI; Figs. 2, 3). Others showed greater deviation from the 1:1 line in the old–new comparison than in the new–new (ABAM, ABMA, CONU, PIJE, PILA, PIMO, PIMONT, PIPO, PISA, QUKE, SESE). Many of these species represented a subset of 16 species with numbers of presences below 100 among training data sets (where N = 3475; Table 3).

The median standardized residual remained close to zero for most species. Medians ranged from 0.02 to −0.06 among old–new and new–new comparisons (Fig. 4). For most species, the inter-quartile range was smaller and medians were closer to zero for the new–new comparison than for the old–new comparison. Four species, ABCO, CADE, PIPO, and PSME, had greater probabilities of occurrence predicted for the old design compared to the new (Fig. 4). The medians of the deviations were, however, small (absolute systematic deviation <0.06). Systematic error was present but substantially weaker in the new–new comparison than in the old–new comparison for CADE, PIPO, and PSME (Fig. 4).

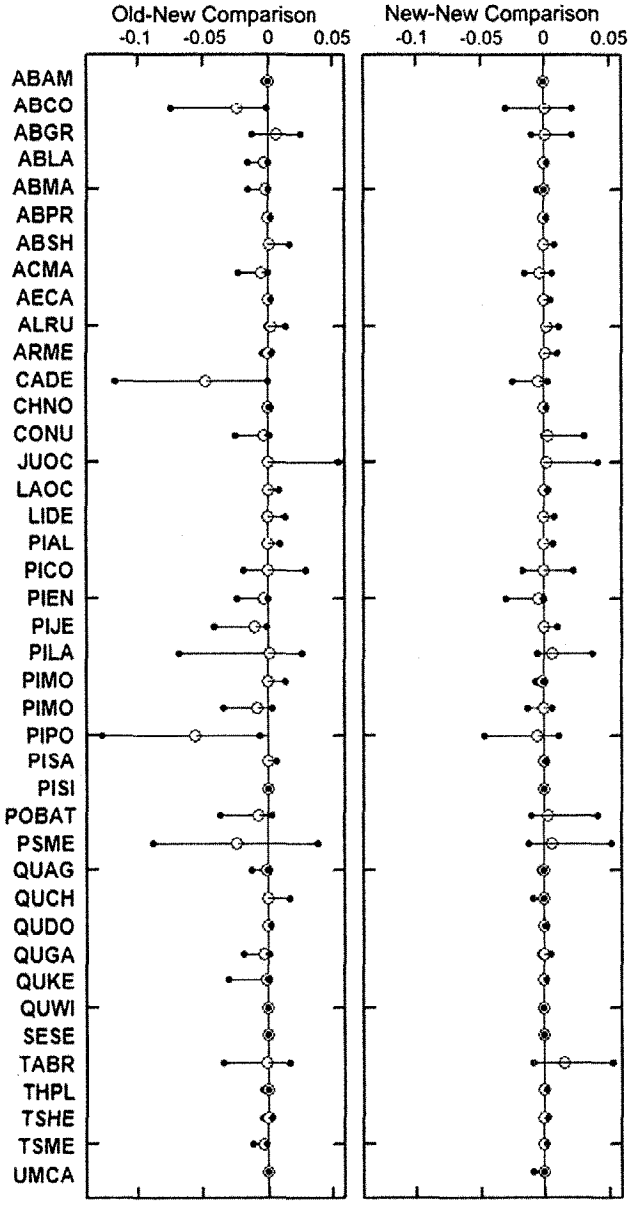

Fig. 4. Stem plots show the central trend in the standardized residuals with respect to the 1:1 lines in Figs. 2 and 3. The mean (white dot), the 25th quantile (black dot), and the 75th quantile (black dot), are shown by species (rows) and by comparison, old–new (left) and new–new (right).

The differences among models built with old and new data did not appear to be caused by model selection of different predictor variables. The same number and type of predictors were chosen among models for 26 species in the new–new and 26 species in the old–new comparisons (see species codes with asterisks in Figs. 2 and 3). Most models selected four predictors out of the four available (an overall mean of 26 species among models built from different data pools). Only four instances occurred with models containing two predictors, and the rest contained three.

All models contained the first PCA component. Also, most predictions fell near the 1:1 line even for models without the same number and/or type of predictors (*e.g.* ABAM, ABGR, ABLA in Figs. 2 and 3). Despite the differences among predictors for many of the compared models, agreement in model fit for old–new was strong and similar for new–new (NRMSE of AUC = 8% for new–new and 13% for old–new; Fig. 5). The minimum model fits (or the minimum externally-validated AUCs) among the comparisons were 0.766 (new–new) and 0.835 (old–new), and the maximum models fits (or the maximum externally-validated AUCs) were essentially the same, 0.995 and 0.996 (new–new and old–new respectively). The mean externally-validated AUC across all models was 0.935. Four species consistently had fits above 0.975 across data sets: *Pinus monophylla, Sequoia sempervirens, Quercus douglasii,* and *Quercus agrifolia.* These species have presences under 100 and occur mostly in California. Most species with strong fits tended toward increased agreement among predictions (Fig. 5). However, the agreement between compared models in type and number of predictors did not play a role in these relationships (see symbol coding in all subplots of Fig. 5). Species with high numbers of presences had slightly lower fit than species with low numbers of presences, albeit very weakly (nonparametric regression, cross-validated $r^2 = 0.08$) and nonlinearly (results not shown). Model fit did not vary with sample design across all models.

3.2. Sample design effect on data

The mean number of trees per plot by size class (where the mean is used as a proxy for probability of tree capture) deviated strongly from the expected 1:1 line (RMSE = 8 or NRMSE = 25%). The deviation was pronounced for all but the largest trees (Fig. 6). The number of trees per plot was higher for the R6 and BLM areas for all tree sizes in the old design compared to the new design. For the regions using variable-radius sampling in the old inventory, numbers of small trees (12.7–38.1 cm dbh) per plot were lower in the old inventory, while numbers of larger trees (38.1–76.2 cm) were higher in the old inventory (Fig. 6; tree selection probabilities for each design are shown in Fig. A1).

Differences among the ECDFs (empirical cumulative distribution functions) from different sample designs for each species were more visually pronounced for species with presences less than 100 (*e.g.* CONU, PIMO, QUAG, and POBAT) (Fig. 7). However, the difference among two ECDFs drawn from the same population is always more pronounced for smaller samples (see Appendix A; Fig. A4). Further, one would expect that species with the most pronounced systematic error in the modeled predictions would demonstrate greatest evidence of climatic bias in the data (which may be attributed to sampling differences). This was not clearly the case as ABCO, CADE, PIPO, and PSME did not have climatic bias greater than that shown for species without systematic error among predictions (*e.g.* ALRU, QUCH, TSHE, LAOC, CHNO, ABSH)(Figs. 4, 7). However, for species with greater numbers of presences like PSME, a small gap has more ecological consequence and meaning compared to the same gap in a species with low numbers of presences (see Fig. A4). To zoom in on ABCO, CADE, PIPO, and PSME, we used Quantile–Quantile plots (QQ-plots). QQ-plots among two samples will be linear if two samples come from the same distribution. The ABCO QQ-plots showed the strongest exception to linearity for PCA2, and the tail of the QQ-plot for the old–new comparison deviated compared to the new–new comparison (Fig. 8).

3.3. Sample design effect on maps

Coarse patterns in predicted probability of occurrence in geographic space appeared similar between sample designs within species, except for the aerial extent where predictions were made (Fig. 9). Also, maps of the differences among predictions within the old–new and new–new comparisons revealed spatial clustering of strong differences at different scales (Fig. 10). These patterns (*e.g.* in the highest magnitude residuals per map) tended to follow broad gradients in topography and climate rather than cluster by management region (Fig. 10). A map of FIA management regions in the old inventory is provided (Appendix A; Fig. A5). Spatial clustering at scales much smaller then FIA management units was evident across new–new and old–new comparisons (*e.g.* the contiguous blue patch in the upper left portion of the map of "ARME new–new," Fig. 10). The source of that error is unknown

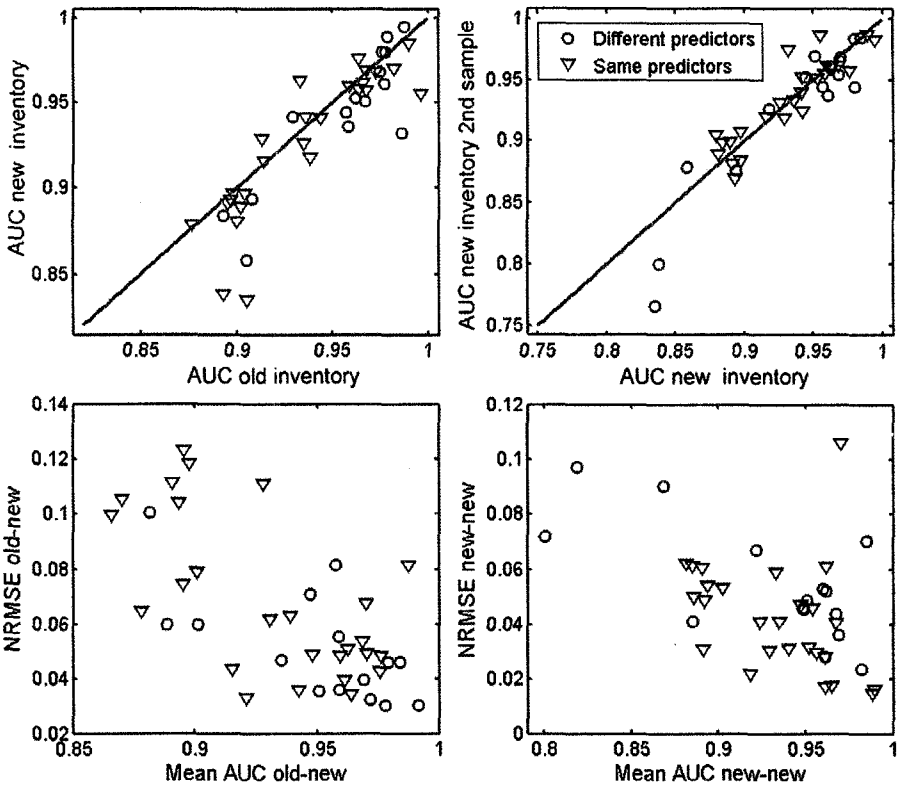

Fig. 5. Top row: Externally-validated area under the receiver-operator characteristic curve (AUC) compared from old–new comparison (left) and new–new comparison (right). Bottom row: Standardized RMSE (NRMSE) *versus* mean AUC among models for old–new (right) and new–new (left) comparisons. Points represent species. Paired models where qualitatively different predictors were chosen for the same species are shown with circles. Paired models where the same predictors were chosen for a species are shown with triangles.

but could be due to sub-sampling or to a contributing variable not included in models such as fire or competitive exclusion.

The maximum absolute differences in predictions differed among species (Fig. 10). These magnitudes depended on the difference among predictions for a location and the range in probability of occurrence predicted for a species. In all three examples the old–new had the highest absolute differences among the comparisons.

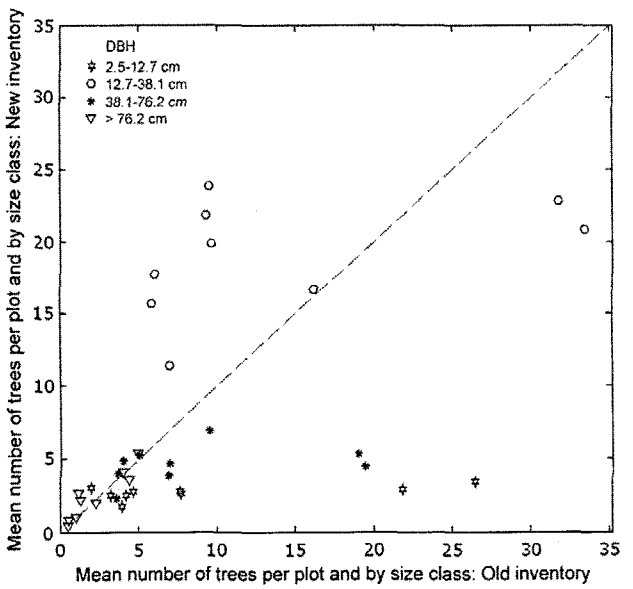

Fig. 6. Mean number of trees per plot and by age class. The ideal 1:1 line is shown. Each point represents a size class in a particular management region from the old design (see Table 1).

4. Discussion

We found two likely effects of inventory method on niche models and their predictions. First, there is a 2% increase in random error among modeled predictions when using one design to predict occurrences in the other design (an average 2% more than within-design sample error for a single within-design comparison). Second, small quantifiable systematic error is present for 4 out of 41 species, yet the error for only one species, ABCO, showed the strongest evidence for a link to inventory.

Although it is possible that the 2% increase in error derived from climate change or other potential factors, we believe the 2% increase in random error is due largely to the change in inventory methods for the following reason: probability of detection and resulting estimates of occurrence are a function of plot size, and plot sizes differed across management regions in the old inventory. Larger plots were used for all tree sizes on R6 and BLM lands while other management regions used variable radius plots with plot size proportional to tree diameter. At least with respect to tree frequency, the varied effects of different regional inventories on tree frequency are suggested (Fig. 6). Hence, when the data from the old design were pooled across regions with different protocols, the cumulative effect (of different types of bias imposed by different sampling methods) likely manifested as random error that propagated to modeled predictions. The difference in plot size that changes a presence record to an absence record at a local scale will depend on the species' density and size distribution for each of species. Our work suggests that the variation in size of FIA plots within and among inventories probably has a small effect

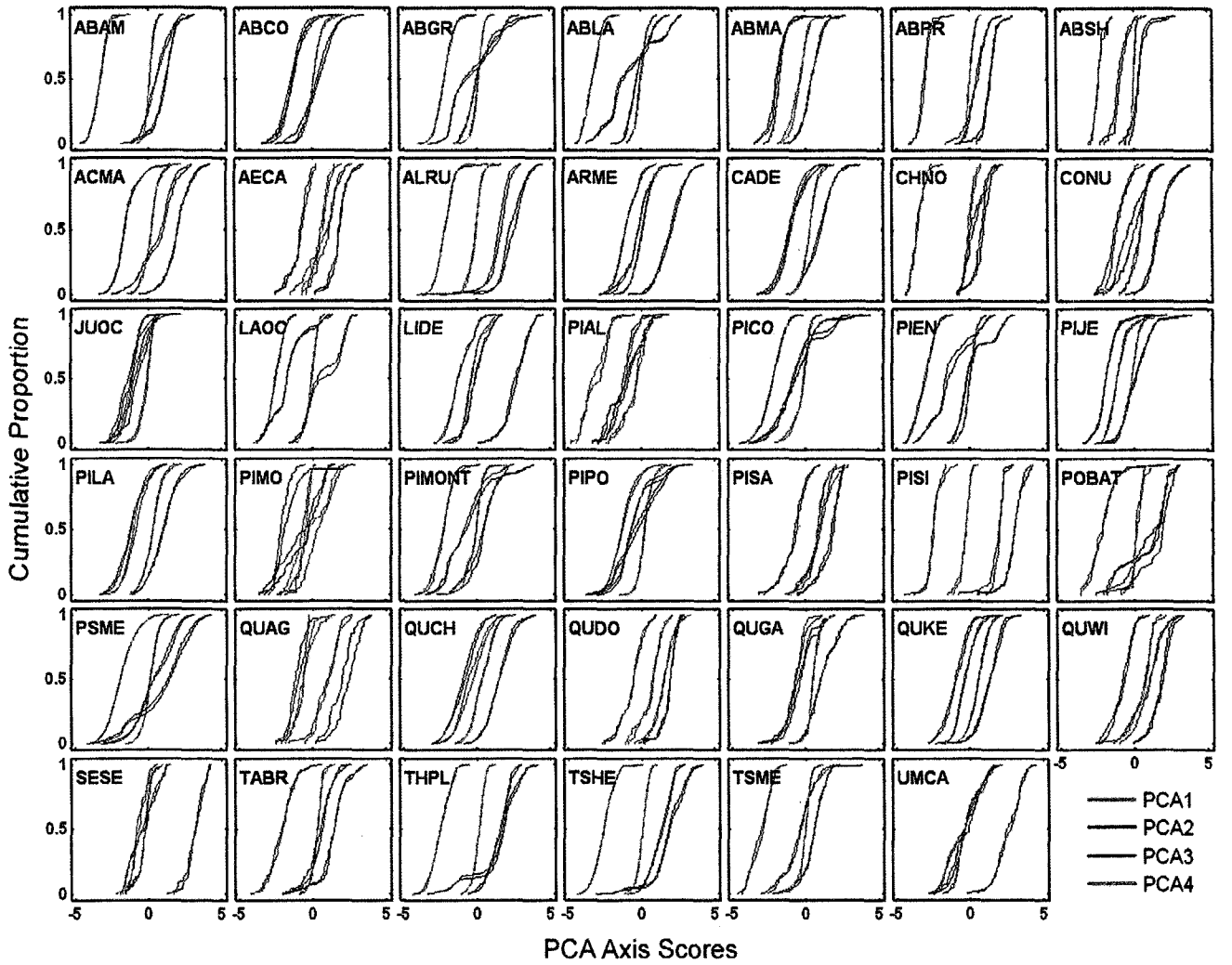

Fig. 7. The ECDF (empirical cumulative distribution function) for each climate variable (or PCA component) as it corresponds to species' occurrence for each design (old and new). Climate variables are color coded, magenta shows PCA1 from both designs, black shows PCA2 from both designs, blue shows PCA3 from both designs, and red shows PCA4 from both designs.

on presence/absence data patterns. However, the future study of the effects of inventory on precision could benefit from not only from old–new and new–new comparisons, but also old–old comparisons. Variations within the old inventories could be directly compared to variations within the new inventories to better understand the 2% difference in random error between modeled predictions.

Inventory change does not seem to be the cause for the systematic error associated with *Calocedrus decurrens*, *Pseudotsuga menziesii*, and *P. ponderosa*. We reach this conclusion given the assumption that the effect of the inventory would first manifest as climatic bias. We define climate bias as the difference between histograms of climatic data corresponding to presence locations. Climate bias was not clearly evident for these species (Fig. 8).

Conversely, stronger evidence of climatic bias existed for *Abies concolor* (Figs. 4, 7, 8). *A. concolor* is part of an intergrading complex of species with *Abies grandis* (Critchfield, 1988). Field discernment of the two species can be difficult where they hybridize yet the species are ecologically distinct (Ferrell and Smith, 1976; Zobel, 1974). Protocols for identification of both species changed in some areas from the old to new design. Old data from R6 national forests tended to either use one species code or the other in areas where research crews have used both, which would affect the tails of the frequency distributions in climate (see Figs. 7 and 8). This difference in inventory procedures shows up as climatic bias because it changes the geographic distribution of presences. The change manifests as difference between frequency distributions of a climate variable.

Spatial patterns in the mapped differences between predictions occurred across scales for the old–new as well as the new–new comparisons. To further examine whether differences among inventories may be the cause for these patterns, we clipped the maps shown in Fig. 10 by management regions (see Appendix A; Fig. A5) and compared distributions of the differences for old–new and new–new by region. Although the disparity in median differences (among mapped predictions) was greater for some management regions in old–new compared to new–new, the patterns did not appear to correspond with differences in inventory (not shown). Apparently, the climate predictors did not segregate by management region enough to affect the distribution of residuals within them. The overlap in climate values among regions with different plot designs probably explains why, despite greater error, the accuracy of the species' predictions was high and similar among designs (*e.g.* revealed by externally-validated AUCs).

4.1. Spatial autocorrelation

Maps of the differences between predictions (Fig. 10) show spatial autocorrelation. Spatial autocorrelation occurs when observations

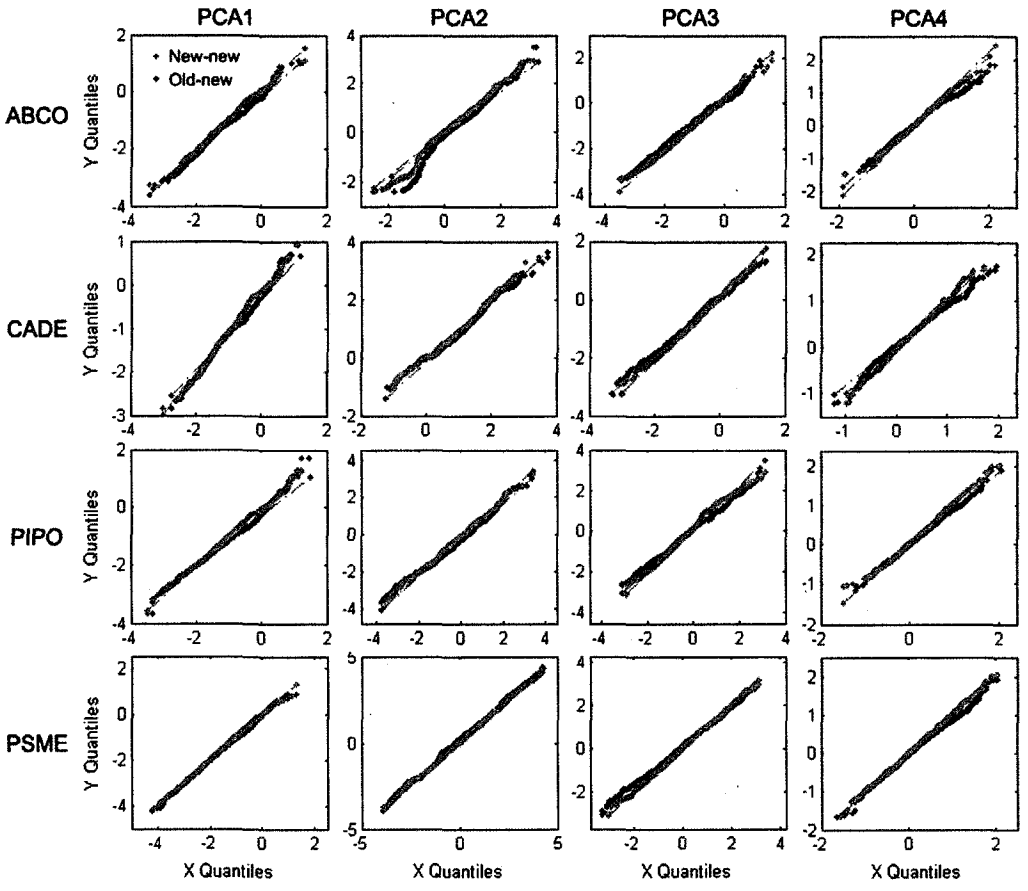

Fig. 8. QQ-plots are shown comparing the frequency distributions of climate variables corresponding to species' presence records across four species (rows) and climate variables (columns). QQ-plots for the old–new comparisons are shown in black and QQ-plots from the new–new comparisons are overlaid in pink. Confidence bounds on QQ-plots that compare two unknown distributions are not possible due to issues regarding multiple comparisons and resulting uncertainty (Chambers et al., 1983).

in close spatial proximity tend to be more similar than expected for observations more spatially separated (Schabenberger and Gotway, 2005). The treatment of spatial autocorrelation in niche modeling is a topic of active research (Dormann et al., 2007; Fortin and Dale, 2009; Miller et al., 2007; Segurado et al., 2006). In our work, we used a kernel smoother in climate space. The climate variables were spatially auto-correlated in geographic space. We did not account separately for the autocorrelation structure among the data in climate space, and this is known as the "working independence" method in kernel regression, which has theoretical basis (Lin and Carroll, 2000; Ruckstuhl et al., 2000). Kernel smoothers are designed to accommodate spatially dependent data (*e.g.* Hutchinson and Gessler, 1994; Wahba, 1990; Yakowitz and Szidarovszky, 1985). Spatial aggregation at variable scales should be expected in the maps of differences among predictions (our results). This is because local climate has spatial autocorrelation. Additionally, spatial aggregation in residuals can result from failing to account for other potential drivers. Here, our main interest lies in the climate signal; thus, we don't attempt to account for the autocorrelation in climate. However, it is possible to pursue other local processes governing spatial autocorrelation using NPMR by examining the residuals with respect to spatial coordinates.

4.2. Cross-validation methods

We found noteworthy variation in modeled output due to within-design sub-sampling. While the noteworthy variations could be due to random chance (through taking a random sub-sample), the sub-sample sizes were large enough to warrant exploring whether something else might exacerbate this such as the modeling process. Thus, we further examined whether model selection by external validation (as opposed to cross validation) could be the reason behind the noteworthy variation in modeled output due to within-design sub-sampling. Model selection by external validation (the method we used here) is a variation of k-fold cross-validation where $k = 2$. Model selection with greater than 2-fold cross-validation uses models built from repeated sub-sampling of the data. In cross-validation, the sample is randomly divided into $k > 2$ sub-samples. One sub-sample is retained (out of the k sub-samples) as the testing or validation data, and the other k-1 subsamples are used each to train or build models. Results of model performance for iteration through k subsamples are combined for optimization. The model that is most robust to external data and accurate is selected.

We examined the subsets of models which used different predictors for a species and among sub-samples. We found that model selection using cross-validation with NPMR may result in greater precision (without compromising robustness) compared to model selection using external validation. This interpretation is based on preliminary results (not shown) that were stimulated from the modeling results of this work. New–new predictions generated from models selected by external validation were compared with new–new predictions generated from models selected by cross-validation. The agreement among predictions improved and model similarity increased for model selection using cross-validation. This is an interesting although preliminary finding given that the statistical paradigm considers external validation as the gold standard in model selection (Hastie and Tibshirani, 2001). More research is needed in this regard with respect to kernel regression, NPMR, and model selection.

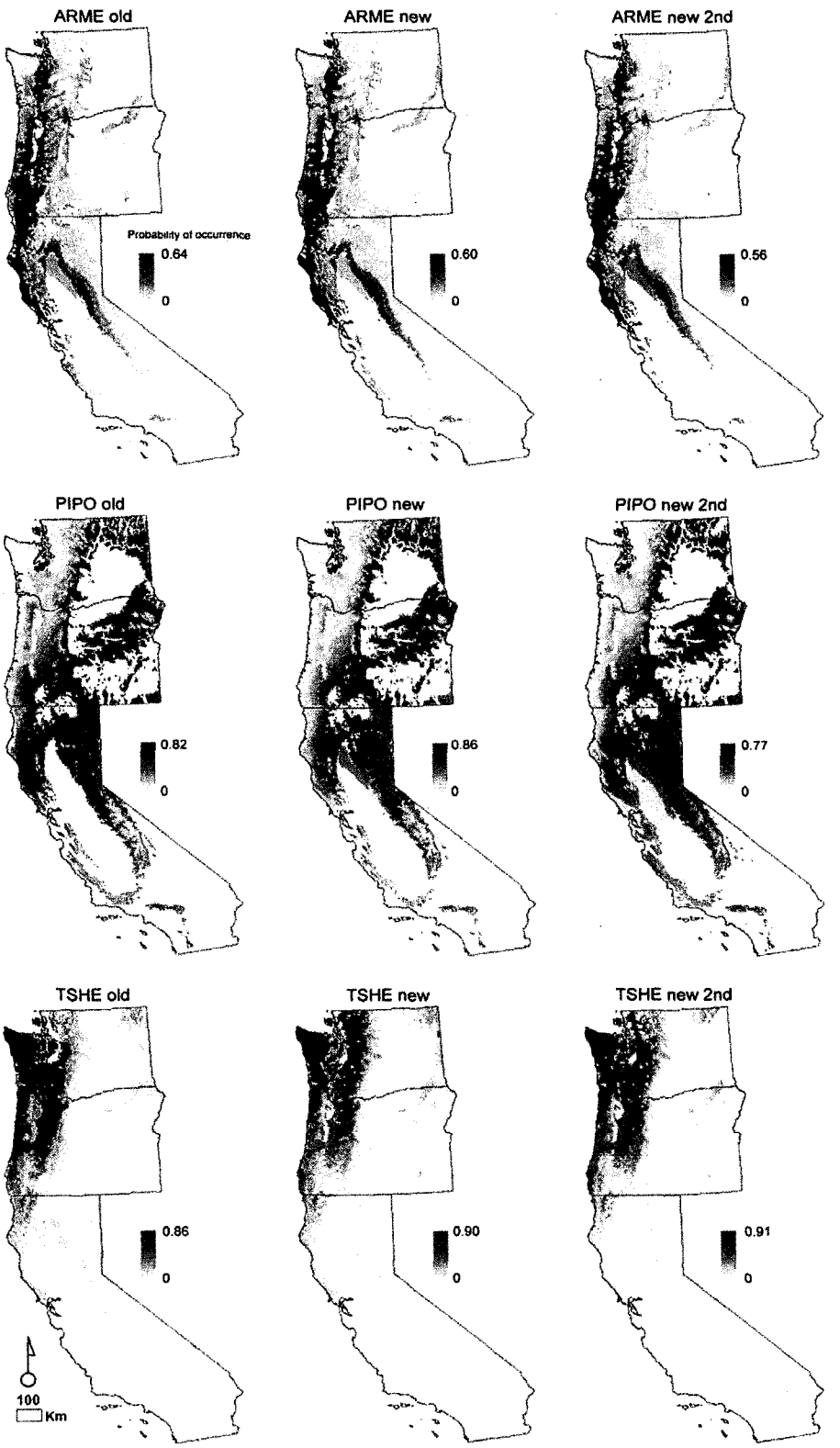

Fig. 9. Maps of probability of occurrence are shown for three species *Arbutus menziesii* (top row), *Pinus ponderosa* (middle row), and *Tsuga heterophylla* (bottom row). Maps correspond to prediction from each data set compared (see Fig. 2).

4.3. Sampling effects

Sampling has been identified as an area requiring further investigation in niche modeling especially given the haphazard nature of some data sets (Araújo and Guisan, 2006; Elith et al., 2006; Guisan and Zimmerman, 2000; Hampe, 2004; Heikkinen et al., 2006). Recent papers that address sampling issues with respect to niche modeling emphasize presence-only data and explore sample size (*e.g.* Elith et al., 2006; Hernandez et al., 2006; Pearce and Ferrier, 2000; Stockwell and Peterson, 2002), sample completeness (*e.g.* Kadmon et al., 2003),

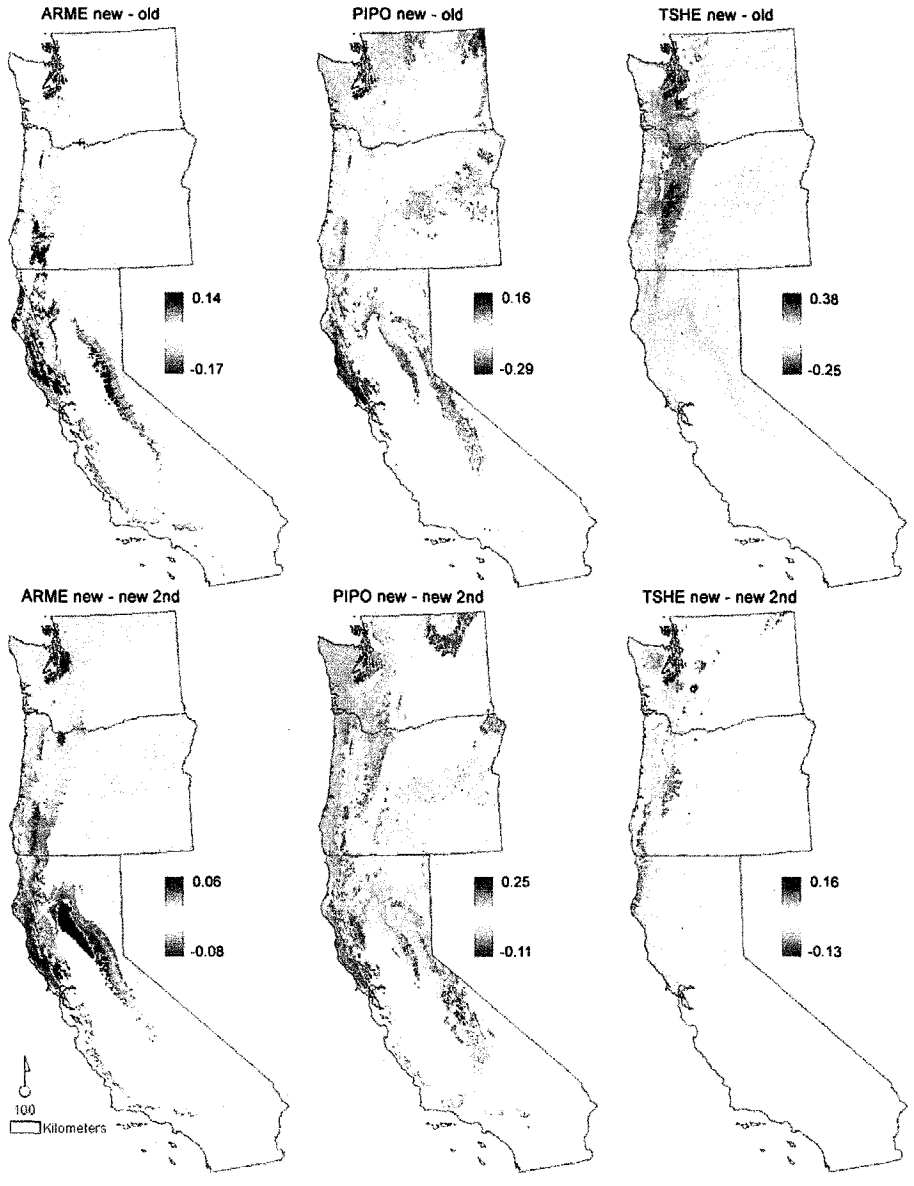

Fig. 10. Maps of differences among probabilities of occurrence show regions of uncertainty among the maps of probability of occurrence (see Fig. 9). The difference for the old–new comparison equals the new minus the old (maps shown in Fig. 9 were subtracted). The difference for the new–new comparison equals the second sample of the new minus the primary sample of the new. Differences are shown for the old–new comparison (top row) and the new–new comparison (bottom row) for the three species in Fig. 9, *Arbutus menziesii* (left column), *Pinus ponderosa* (middle column), and *Tsuga heterophylla* (right column).

sample attributes for optimal model validation (*e.g.* Araújo et al., 2005), idealized sampling strategy for niche models (*e.g.* Reese et al., 2005; Wessels et al., 1998), and sample bias due to site accessibility and/or presence-only data (Kadmon et al., 2003, 2004; Reese et al., 2005). Sampling bias can add extraneous error to ecological signals especially with presence-only data, which can mislead model development, spuriously increase or decrease fit, and affect spatial predictions (Araújo and Guisan, 2006; Barry and Elith, 2006; Guisan et al., 2006). Our work suggests that a minimal effect occurs on predictions from niche models due to a large change in a biological inventory. The inventory measures presences and absences of plant species across the landscape, and various important aspects of the sample design changed with the inventory including plot size.

5. Conclusion

The many features that changed with the overhaul of sample design for the Forest Inventory Analysis Program had a small cumulative impact on niche models and maps of probability of occurrence based on tree species' presence/absence data. The niche relations between tree species and climate are mostly unchanged across the two decades straddling the year 2000 for climate variables estimated at 4 km spatial resolution. The fit or accuracy of all models developed for species' occurrence based on climate was high for both data sets (mean externally-validated AUC = 0.935). Additionally, this work corroborates the pervasive and pressing need to quantify different types of error in niche modeling to address issues associated with large-scale data integration (Barry and Elith, 2006; Elith et al., 2002). The isolation of error types can help to ascertain the cause and uncertainty related to observed patterns.

Acknowledgments

This research was supported in part by funds provided by the USDA Forest Service, PNW Research Station, Forest Inventory and Analysis Program.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ecoinf.2013.05.001.

References

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modeling. Journal of Biogeography 33, 1677–1688.

Araújo, M.B., Pearson, R.G., Thuiller, W., Erhard, M., 2005. Validation of species–climate impact models under climate change. Global Change Biology 11, 1504–1513.

Austin, M.P., 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modeling. Ecological Modelling 157, 101–118.

Azuma, D., Monleon, V.J., 2011. Differences in forest area classification based on tree tally from variable- and fixed-radius plots. Canadian Journal of Forest Research 41, 211–214.

Barrett, T.M., 2004. Estimation procedures for the combined 1990s periodic forest inventories of California, Oregon, and Washington. USDA Forest Service, Pacific Northwest Research Station General Technical Report PNW-GTR-597. Portland, OR.

Barry, S., Elith, J., 2006. Error and uncertainty in habitat models. Journal of Applied Ecology 43, 413–423.

Bechtold, W.A., Patterson, P.L., 2005. The enhanced Forest Inventory and Analysis Program—national sampling design and estimation procedures. USDA Forest Service, Southern Research Station General Technical Report SRS-80. Asheville, North Carolina.

Bitterlich, W., 1948. Die Winkelzählprobe. Allgemeine Forst und Holzwirtschaftliche Zeitung 59, 4–5.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 1983. Graphical Methods for Data Analysis. Wadsworth Publishing Company, Belmont, CA.

Critchfield, W.B., 1988. Hybridization of the California firs. Forest Science 34, 139–145.

Crossman, N.D., Bryan, B.A., Cooke, D.A., 2011. An invasive plant and climate change threat index for weed risk management: integrating habitat distribution pattern and dispersal process. Ecological Indicators 183–198.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kuhn, I., Ohlemuller, R., Peres-Neto, P., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography 30, 609–628.

Elith, J., Burgman, M.A., Regan, H.M., 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecological Modelling 157, 313–329.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettman, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S., Richardson, K., Schachetti Pereira, R., Schapire, R.E., Soberón, J., Williams, S.E., Wisz, M., Zimmermann, N.E., 2006. Novel methods improve predictions of species' distributions from occurrence data. Ecography 29, 129–151.

Engelbrecht, B.M.J., Comita, L.S., Condit, R., Kursar, T.A., Tyree, M.T., Turner, B.L., Hubbell, S.P., 2007. Drought sensitivity shapes species distribution patterns in tropical forests. Nature 447, 80–82.

European Commission, 1997. Study on European Forestry Information and Communication System: Report on Forest Inventory and Survey Systems (Luxembourg).

Evans, J.S., Cushman, S.A., 2009. Gradient modeling of conifer species using random forests. Landscape Ecology 24, 673–683.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognition Letters 27, 861–874.

Ferrell, G.T., Smith, R.S., 1976. Indicators of *Fomes annosus* root decay and bark beetle susceptibility in sapling white fir. Forest Science 22, 365–369.

Fortin, M., Dale, M.R.T., 2009. Spatial autocorrelation in ecological studies: a legacy of solutions and myths. Geographical Analysis 41, 392–397.

Frayer, W.E., Furnival, G.M., 1999. Forest survey sampling designs: a history. Journal of Forestry 97, 4–10.

Graham, C., Ron, S.R., Santos, J.C., Schneider, C.J., Mortiz, C., 2004. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. Evolution 58, 1781–1793.

Gray, A.N., 2003. Monitoring stand structure in mature coastal Douglas-fir forests: effect of plot size. Forest Ecology and Management 175, 1–16.

Grosenbaugh, L.R., 1952. Plotless timber estimates—new, fast, easy. Journal of Forestry 50, 32–37.

Grosenbaugh, L.R., Stover, W.F., 1957. Point sampling compared with plot sampling in southeast Texas. Forest Science 3, 2–14.

Guisan, A., Zimmerman, M.E., 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135, 147–186.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overont, J., Aspinall, R., Hastie, T., 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology 43, 386–392.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S.J., Townsend Peterson, A., 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecological Monographs 77, 615–630.

Hampe, A., 2004. Bioclimate envelope models: what they detect and what they hide. Global Ecology and Biogeography 13, 469–471.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a Receiver Operating Characteristic (ROC) curve. Radiology 143, 29–36.

Hannah, L., Midgley, G.F., Millar, D., 2002. Climate change-integrated conservation strategies. Global Ecology and Biogeography 11, 485–495.

Hastie, T., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer, New York.

Heikkinen, R.K., Luoto, L., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modeling under climate change. Progress in Physical Geography 30, 751–777.

Hernandez, P.A., Graham, C.H., Master, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29, 773–785.

Hijmans, R.J., Garrett, K.A., Huaman, Z., Zhang, D.P., Schreuder, N., Bonierbale, N., 2000. Assessing the geographic representativeness of genebank collections: the case of the Bolivian wild potatoes. Conservation Biology 14, 1755–1765.

Hiserote, B., Waddell, K., 2003. The PNW-FIA integrated database user guide. A Database of Forest Inventory Information for California, Oregon and Washington, Version 2.0. FIA PNW Station, Portland, Oregon.

Hugall, A., Moritz, C., Moussalli, A., Stanisic, J., 2002. Reconciling paleodistribution models and comparative phylogeography in the wet tropics rainforest land snail *Gnarosophia bellendenkerensis* (Brazier 1875). Proceedings of the National Academy of Sciences of the United States of America 99, 6112–6117.

Hutchinson, G.E., 1957. A Treatise on Limnology, Vol. 1. John Wiley and Sons, New York.

Hutchinson, M.F., Gessler, F.R., 1994. Splines: more than just a smooth interpolator. Geoderma 62, 45–67.

Iverson, L.R., Prasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. Ecological Monographs 68, 465–485.

Iverson, L.R., Prasad, A.M., Matthews, S.N., Peters, M., 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. Forest Ecology and Management 254, 390–406.

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P.J., Mottram, G., 2005. Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum* Sendtn.) in Paraguay. Genetic Resources and Crop Evolution 52, 671–682.

Joliffe, I.T., 1982. A note on the use of principal components in regression. Journal of the Royal Statistical Society, Series C 31, 300–303.

Kadmon, R., Farber, O., Danin, A., 2003. A systematic analysis of factors affecting the performance of climatic envelope models. Ecological Applications 13, 853–867.

Kadmon, R., Farber, O., Danin, A., 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecological Applications 14, 401–404.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: a comparison of five machine-learning methods. Ecological Informatics 5, 441–450.

Kelly, C.K., Bowler, M.G., Pybus, O.G., Harvey, P.H., 2008. Phylogeny, niches and relative abundance in natural communities. Ecology 89, 962–970.

Lin, X., Carroll, R.J., 2000. Nonparametric function estimation for clustered data when the predictor is measured without/with error. Journal of the American Statistical Association 95, 520–534.

Lintz, H., McCune, B., Gray, A., McCulloh, K., 2011. Quantifying ecological thresholds from response surfaces. Ecological Modelling 222, 427–436.

Marini, M.A., Barbet-Massin, M., Lopes, L.E., Jiguet, F., 2009. Predicted climate-driven bird distribution changes and forecasted conservation conflicts in a neotropical savanna. Conservation Biology 23, 1558–1567.

Massey Jr., F.J., 1951. The Kolmogorov–Smirnov test of goodness of fit. Journal of the American Statistical Association 46, 68–78.

McCune, B., 2006. Non-parametric habitat models with automatic interactions. Journal of Vegetation Science 17, 819–830.

McCune, B., Mefford, M.J., 2006. PC-ORD. Multivariate Analysis of Ecological Data, Version 5.2. MjM Software Design, Gleneden Beach, Oregon, USA.

McCune, B., Mefford, M.J., 2008. HyperNiche. Multiplicative Habitat Modeling. Version 2.0. MjM Software, Gleneden Beach, Oregon.

McKenzie, D., Peterson, D.W., Peterson, D.L., Thornton., P.E., 2003. Climatic and biophysical controls on conifer species distributions in mountain forests of Washington State, USA. Journal of Biogeography 30, 1093–1110.

Miller, J., Franklin, J., Aspinall, R., 2007. Incorporating spatial dependence in predictive vegetation models. Ecological Modelling 202, 225–242.

National Research Council, 2000. Ecological Indicators for the Nation. National Academy Press, Washington D.C.

Nelson, B.W., Ferreira, C.A.C., da Silva, M.F., Kawasaki, M.L., 1990. Endemism centres, refugia and botanical collection density in Brazilian Amazonia. Nature 345, 714–716.

Newbold, S., Eadie, J.M., 2004. Using species-habitat models to target conservation: a case study with breeding mallards. Ecological Applications 14, 1384–1393.

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. U.S.A. Canadian Journal of Forest Research 32, 725–741.

Oksanen, J., Minchin, P.R., 2002. Continuum theory revisited: what shape are species responses along ecological gradients? Ecological Modelling 157, 119–129.

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133, 225–245.

Pino-Mejías, R., Cubiles-de-la-Vega, M.D., Anaya-Romero, M., Pascual-Acosta, A., Jordán-López, A., Bellinfante-Crocci, N., 2010. Predicting the potential habitat of oaks with data mining models and the R system. Environmental Modelling & Software 25, 826–836.

Raxworthy, C.J., Martinez-Meyer, E., Horning, E., Nussbaum, R.A., Schneider, G.E., Ortega-Huerta, M.A., Townsend-Peterson, A., 2003. Predicting distributions of known and unknown reptile species in Madagascar. Nature 426, 837–841.

Reese, G.C., Wilson, K.R., Hoeting, J.A., Flather, C.H., 2005. Factors affecting species distribution predictions: a simulation modeling experiment. Ecological Applications 15, 554–564.

Rehfeldt, G.E., Crookston, N.L., Warwell, M., Evans., J.S., 2006. Empirical analyses of plant–climate relationships for the western United States. International Journal of Plant Sciences 167, 1123–1150.

Rehfeldt, G.E., Ferguson, D.E., Crookston, N.L., 2008. Quantifying the abundance of co-occurring conifers along the inland northwest (USA) climate gradients. Ecology 89, 2127–2139.
Ruckstuhl, A., Welsh, A.H., Carroll, R.J., 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statistica Sinica 10, 51–71.
Rudis, V.A., 2003a. Comprehensive regional resource assessments and multipurpose uses of Forest Inventory and Analysis data, 1976 to 2001: a review. General Technical Report SRS-70. U.S. Department of Agriculture, Forest Service, Southern Research Station, Ashville, North Carolina.
Rudis, V.A., 2003b. Fresh ideas, perspectives, and protocols associated with Forest Inventory and Analysis surveys: graduate reports, 1974 to July 2001. General Technical Report SRS-61. U.S. Department of Agriculture, Forest Service, Southern Research Station, Ashville, North Carolina.
Schabenberger, O., Gotway, C.A., 2005. Statistical Methods for Spatial Data Analysis. Chapman & Hall/CRC, Boca Raton, Fla.
Scott, C.T., Cassell, D.L., Hazard, J.W., 1993. Sampling design of the U.S. National Forest Health Monitoring Program. In: Nyyssonen, A., Poso, S., Rautala, J. (Eds.), Ilvessalo symposium on national forest inventories: Finnish Forest Research Institute, Research Paper, 444, pp. 150–157.
Segurado, P., Araújo, M.B., Kunin, W.E., 2006. Consequences of spatial autocorrelation for niche-based models. Journal of Applied Ecology 43, 433–444.
Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecological Modelling 148, 1–13.
Svenning, J., Skov, F., 2004. Limited filling of the potential range in European tree species. Ecology Letters 7, 565–573.
Thornton, P.E., Running, S.W., White, M.A., 1997. Generating surfaces of daily meteorological variables over large regions of complex terrain. Journal of Hydrology 190, 214–251.
Wahba, G., 1990. Comment on Cressie. The American Statistician 44, 255–256.
Wessels, K.J., Van Jaarsveld, A.S., Grimbeek, J.D., Van der Linde, M.J., 1998. An evaluation of the gradsect biological survey method. Biodiversity and Conservation 7, 1093–1121.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. Biometrics 1, 80–83.
Woodall, C/.W., Oswalt, C.M., Westfall, J.A., Perry, C.H., Nelson, M.D., 2009. Tree migration detection through comparisons of historic and current forest inventories. In: McWilliams, W., Moisen, G., Czaplewski, R. (Eds.), Forest Inventory and Analysis Symposium 2008, Park City, UT. Proc. RMRS-P-56CD. Fort Collins, Colorado.
Yakowitz, S.J., Szidarovszky, F., 1985. A comparison of kriging with nonparametric regression methods. Journal of Multivariate Analysis 16, 31–53.
Yanez, M., Floater, G., 2000. Spatial distribution and habitat preference of the endangered tarantula *Brachypelma klassi* (Araneae: Theraphosidae) in Mexico. Biodiversity and Conservation 9, 795–810.
Yost, A., Lintz, H.E., 2013. A comparison of the prediction accuracy between MaxEnt and non-parametric multiplicative regression (NPMR). Unpublished data analysis; in preparation for submission.
Zobel, D.B., 1974. Local variation in intergrading *Abies grandis–Abies concolor* populations in the central Oregon Cascades. II. Stomatal reaction to moisture stress. Botanical Gazette 135, 200–210.