Ecosystems (2014) 17: 43–53
DOI: 10.1007/s10021-013-9703-y

ECOSYSTEMS

Using Publicly Available Forest Inventory Data in Climate-Based Models of Tree Species Distribution: Examining Effects of True Versus Altered Location Coordinates

Jacob Gibson,[1]* Gretchen Moisen,[2] Tracey Frescino,[2] and Thomas C. Edwards Jr.[3]

[1]*Department of Wildland Resources and Ecology Center, Utah State University, Logan, Utah 84321, USA;* [2]*USDA Forest Service, Rocky Mountain Research Station, Ogden, Utah 84401, USA;* [3]*U.S. Geological Survey, Utah Cooperative Fish and Wildlife Research Unit, Department of Wildland Resources, Utah State University, Logan, Utah 84321, USA*

Abstract

Species distribution models (SDMs) were built with US Forest Inventory and Analysis (FIA) publicly available plot coordinates, which are altered for plot security purposes, and compared with SDMs built with true plot coordinates. Six species endemic to the western US, including four junipers (*Juniperus deppeana* var. *deppeana*, *J. monosperma*, *J. occidentalis*, *J. osteosperma*) and two piñons (*Pinus edulis*, *P. monophylla*), were analyzed. The presence–absence models based on current climatic variables were generated over a series of species-specific modeling extents using Random Forests and applied to forecast climatic conditions. The distributions of predictor variables sampled with public coordinates were compared to those sampled with true coordinates using *t* tests with a Bonferroni adjustment for multiple comparisons. Public- and true-based models were compared using metrics of classification accuracy. The modeled current and forecast distributions were compared in terms of their overall areal agreement and their geographic mean centroids. Comparison of the underlying distributions of predictor variables sampled with true versus public coordinates did not indicate a significant difference for any species at any extent. Both the public- and true-based models had comparable classification accuracies across extent for each species, with the exception of one species, *J. occidentalis*. True-based models produced geographic distributions with smaller areas under current and future scenarios. The greatest areal difference occurred in the species with the lowest modeled accuracies (*J. occidentalis*), and had a forecast distribution which diverged severely. The other species had forecast distributions with similar magnitudes of modeled distribution shifts.

Key words: forest inventory and analysis; ''fuzz-swap'' plot coordinates; perturbed coordinates; species distribution models; junipers; piñon pine.

Received 25 January 2013; accepted 28 July 2013;
published online 4 September 2013

Electronic supplementary material: The online version of this article (doi:10.1007/s10021-013-9703-y) contains supplementary material, which is available to authorized users.

Author Contributions: GM and TF conceived of or designed study. JG performed research. JG and TF analyzed data. GM and TE contributed new methods or models. JG, GM, and TE wrote the paper.

**Corresponding author; e-mail:* j.gibson@aggiemail.usu.edu

Introduction

The US Forest Service Forest Inventory and Analysis Program (FIA) maintains a systematic array of survey plots which collectively provide a spatially comprehensive database of tree species and their condition across the US (McRoberts and others 2005). This database is public and provides an unprecedented resource for a range of research [for example, quantifying empirical tree species–climate relationships (Rehfeldt and others 2008); detecting regional demographic trends of tree species (Woodall and others 2010)] and management applications [for example, estimating regional mortality (Shaw and others 2005); mapping national landcover (Homer and others 2007)]. However, to protect plot integrity and sensitive information from privately owned lands, the true locations of FIA sample plots are altered before they are made public to comply with Federal security requirements [Food Security Act of 1985 (7 U.S.C. 2276(d)], as amended). These altered plot coordinates are often referred to as fuzzswap plot coordinates. Plots are ''fuzzed'' by randomly perturbing their location an average of 0.8 km away from their true locations. In addition, locations on up to 20% of plots on private lands are ''swapped'' with other private land plots within the same county having similar forest characteristics.

The effect of altering plot coordinates on research and management questions is contingent upon the resolution, extent, and nature of the application with respect to the underlying FIA sample design. The sample design underlying true plot locations inherently limits the resolution and extent of any application. Altered coordinates increase these limits in theoretically predictable ways for several applications given the algorithm employed to alter the coordinates (McRoberts and others 2005). These predicted effects have been corroborated by investigations into several applications [for example, summaries of plot attributes over an area of interest (McRoberts and others 2005); modeled relationship between plot attributes and environmental variables (Coulsten and others 2006; Pricely and others 2009)], which collectively provide basic guidelines for selecting an appropriate extent and resolution. Here, we extend these investigations to the use of FIA data for developing species distribution models (SDMs) (Guisan and Zimmermann 2000), an increasingly common application, especially under scenarios of forecast global change (Zimmerman and others 2009).

SDMs portray a species' environmental niche (for example, Pearman and others 2008; Weins and others 2009) by empirically relating the probability of a species current presence (and absence) to environmental variables considered relevant to the species' distribution. Although SDMs have been used in a wide range of ecological studies [for example, lichen distributions (Edwards and others 2006); goshawk nesting habitat (Zarnetske and others 2007); rare tree species distributions (Zimmerman and others 2007)], they are increasingly used for exploring the potential response of tree species to climate change via extrapolating modeled distributions to forecast climate change scenarios (Iverson and others 2008; Zimmerman and others 2009; Cole and others 2008a, b; Refheldt and others 2009). The FIA data, with its design-based observations of true presence and absence, therefore provide an ideal set of observations on tree species from which SDMs can be developed (Iverson and others 2008; Rehfeldt and others 2009). However, selecting an appropriate resolution and extent to model species distributions introduces technical and conceptual challenges notwithstanding those arising from the potential effects of the altered public coordinates.

To investigate the effect of using public FIA coordinates in SDMs we compared SDMs built with true coordinates to those built with public coordinates. We focus on two conceptually important elements of SDMs potentially affected by altered coordinates: (i) extent, considered in terms of ensuring appropriate geographic and environmental sample coverage of the species being modeled (for example, Sánchez-Fernández and others 2011); and (ii) resolution, which was simply held constant at 1 km^2, the resolution at which altered coordinates are theoretically predicted to be negligible. All predictor variables where acquired at a resolution of 1 km^2 and this resolution was maintained for all modeled distributions. All modeling was done on six species of conifers which are endemic to the Western U.S., and that have strong, relatively well understood relationships to climate. All species geographic ranges were completely encompassed by the FIA sampling grid, insuring sampling coverage of each species geographic and environmental space.

Complete geographic (sampling) coverage allowed us to vary modeling extents, and observe cascading effects of true versus public plot coordinates into environmental space. No definitive technical or conceptual guidelines exist for selecting a modeling extent given known geographic extents and so we built models across a range of buffered extents. The extents are species specific, centering on the FIA-based sample presences of the

respective species, and increase at 50 km increments from 50 to 200 km, forming buffered distances from presences, the effect resulting in an increase in the absences as extent increased.

The underlying sampled variables were compared between true and public coordinates using simple *t* tests adjusted for multiple comparisons. Models were built using Random Forests (Briemann 2001; Cutler and others 2007) across the range of extents (with resolution held constant at 1 km^2) and compared in terms of classification accuracy, including model sensitivity, specificity (after Fielding and Bell 1997), area under the curve (AUC) (Hanley and McNeil 1982), and the true skill statistic (TSS) (Allouche and others 2006). Models from the 100-km buffered extent were applied to forecast climates and compared spatially in terms of overall areal agreement and distribution centroids to determine if use of true versus public coordinates had any effect on estimated distribution size and geographic location, respectively.

Methods

Choice of Species

Four junipers (*Juniperus deppeana* var. *deppeana*; *J. monosperma*; *J. occidentalis*; *J. osteosperma*) and two piñons (*Pinus edulis*; *P. monophylla*) dominate the arid woodlands of the western US, occupying 10–20% of regional land cover (Figure 1). These six species are endemic to the US and therefore their entire distribution is covered by the FIA sample frame, allowing for their entire realized distribution to be used in SDMs. Furthermore, the relationship between climate and piñons and junipers is relatively well understood across multiple levels of biotic organization (for example, physiology: McDowell and others 2008; West and others 2008, coevolution: Lanner 1998; Robinson and others 2010; populations: Chambers and others 1999; Martens and others 2001, continental distributions: Betancourt and others 1990; Cole and others 2008a, b).

Species Observation Data

The FIA collects information on the number and condition of trees on plots approximately 0.4 ha in area with approximately one plot per 2,400 ha. However, the spatial intensity of annual FIA plots collected to date varies throughout the western US due to different start dates of annual inventory practices and land tenure, with grid resolutions of 10,000, 5,000 and 2,500 m commonly found. To create a sample of plots with equal probability of selection, thereby avoiding sample weight issues, we extracted one randomly chosen FIA plot per 10 km^2, which is the lowest sampling intensity in the area covered by our study species. For the

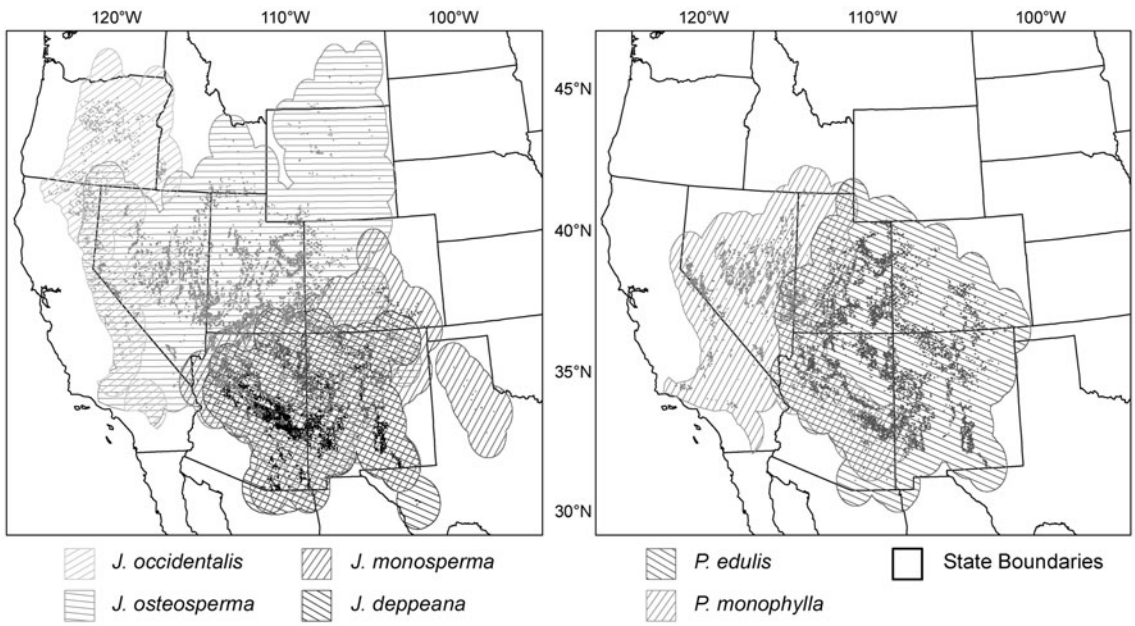

Figure 1. Distribution of FIA plots with observations of the presence, by the study species, and their 100-km modeling extent. The modeling extent is simply a buffer around the presences of each respective species which serves as a basis to select absences and subsequently limit model extrapolation.

purpose of this analysis, we collapsed this random subset of FIA plot observations into simple presence or absence for each species. Although statistical methods exist for modeling abundance (count) data, we collapsed plot abundance data to a nominal response (that is, count ≥ 1 is presence, count = 0 is absence) so we could employ a commonly used distribution modeling tool, Random Forests (see ''Modeling Current Distributions'' section). These responses were related to environmental variables to produce species-specific models of climatic suitability. Subsequently, modeled climatic suitability at FIA plots was related to observations of establishment and mortality (at their respective plots) to set thresholds corresponding to distribution expansion and contraction, respectively (see Gibson 2011).

Modeling Extent

The selection of presences for our modeled species is based on entire species distributions and thereby minimizes the risk of under-sampling the breadth of environmental space occupied by each species. However, the selection of absences, dictated by the size and shape of the modeling extent, lacks obvious guidelines yet has important implications on model building. With respect to sessile species, the underlying biological meaning of an absence (for example, Zarnetske and others 2007) represents a set of conditions in which individuals of the species cannot persist but must cautiously be distinguished from areas in which the species has simply never had the opportunity to establish (for example, Barve and others 2011). This is typified by piñons and junipers of the western US because their continental distributions are not considered in equilibrium with the current climate (for example, Betancourt and others 1991; Gray and others 2006). Although the challenges introduced by the climatic disequilibrium of distributions remain a confounding problem for developing SDMs (for example, Araújo and Pearson 2005; Stokland and others 2011; García-Valdés 2013), it is beyond the scope of this analysis to address. Rather, we isolate the effect of altered coordinates on SDMs by comparing true- and public-based models developed at four species-specific modeling extents. The extents are generated by buffering presences, for each species, at 50 km increments from 50 to 200 km. The species buffers are simply the dissolved composite of round buffers centered on each presence. Each of these four extents are used to define a set of absences (described in Species Observation Data, above) which are then used to compare sampled environmental variables and to build models. The 100 km species-specific modeling extents are illustrated in Figure 1. This is done for both the public coordinates and the true coordinates. Thus, for both the public and true coordinates, there are four sets of extents/absences and four corresponding models built for each species which are then compared.

Predictor Layers and Future Climate Scenarios

Eight climate variables, in combination with a topographic roughness variable, constitute the predictor variables used for all species models (Table 1). This selection was made based on a review of physiological and demographic studies, which indicate piñons and junipers respond to these basic climatic variables (for example, Martens and others 2001; Floyd and others 1982). Specifically, piñons and junipers are partitioned along a common set of climatic gradients within shared climatic bounds. They are collectively bounded in

Table 1. Climate and Topographic Variables Used to Evaluate Effect of True Versus Public FIA Plot Coordinates on SDMs for Four Juniper and Two Piñon Pine Species, Western North America

Variable class/code	Description	Units
Climate		
Dwi	Sum of maximum temperatures for Dec, Jan, Feb	°C × 10
Dsp	Sum of maximum temperatures for Mar, Apr, May	°C × 10
Dsu	Sum of maximum temperatures for Jun, Jul, Aug	°C × 10
Dau	Sum of maximum temperatures for Sep, Oct, Nov	°C × 10
Pwi	Sum of precipitation for Dec, Jan, Feb	mm
Psp	Sum of precipitation for Mar, Apr, May	mm
Psu	Sum of precipitation for Jun, Jul, Aug	mm
Pau	Sum of precipitation for Sep, Oct, Nov	mm
Topography		
Zro	Standard deviation from 3 × 3 km neighborhood	n/a

the north and south within a continental thermal belt (~30–45 N) associated with the positioning of the polar air mass (for example, Neilson 1987; Cole and others 2008a, b; Romme and others 2009). On the east and west they are bound within a negative annual moisture balance (~120–105 W). Within this common arid continental interior, piñons and junipers are partitioned along seasonal moisture gradients, principally the balance between summer and winter precipitation. We examined correlations among the variables prior to analysis

Several standardized emissions scenarios have been developed and run with numerous global circulation models (IPCC 2007). Uncertainty associated with the various GCMs and emission scenarios poses a significant challenge for modeling the response of species distributions to climate change. We limit our analysis to one scenario as our emphasis is on the effect of modeling with true coordinates versus public coordinates, not on climate change effects per se. We chose the a2a scenario, which is a high emissions scenario, run by the Hadley center coupled ocean–atmosphere GCM as part of the CMIP3 (IPCC 2001). We derived current and forecast climate grids from the WorldClim dataset (Hijmans and others 2005). Current climatic conditions represent the average monthly values over the last half of the 20th century and have a spatial resolution of 1 km^2. Forecast conditions represent monthly averages over three 30-year increments centered on 2020, 2050, and 2080.

Modeling Current Distributions

The presence and absences within the four modeling extents were modeled as functions of the predictor variables using Random Forests (RF, Breiman 2001; Cutler and others 2007). All of the predictive models in these analyses were fit using the ''randomForest'' library and extrapolated using the ''raster'' library in R (Liaw and Wiener 2002; Hijmans and Elith 2011, respectively). Spatial extrapolation of the species models to current climatic conditions was constrained to the species modeling extent. SDMs are sensitive to classification thresholds, which are related to prevalence of the species (presence) (Manel and others 2001) relative to absence. Given the systematic design of the FIA plots, our ratio of the presence to absence (prevalence) is a true sample estimate of the species prevalence on the landscape. We consequently used these species-specific sample prevalence values as the classification thresholds for our SDMs within the ''PresenceAbsence'' library in R (Freeman and Moisen 2008).

Projecting Future Distributions

The models developed for the current climate were applied to each interval of forecast climate conditions resulting in a sequence of probability surfaces, that is, modeled climatic suitability, under the current and three forecast climate intervals. Linking this sequence of probability surfaces to produce a single categorical distribution forecast map takes the form of conditional queries combining changes in modeled climatic suitability with the distance from current distributions. Changes in modeled climatic suitability are related by life-stage specific thresholds which scale up to distribution contraction, persistence, and expansion. For contraction, this is simply a decrease in climatic suitability below a threshold. For expansion, this is an increase in climatic suitability conditional to the distance from modeled presence in the previous time step. For the purpose of this study, we set the expansion rate for all species at 30-km/30-year time step. Climatic suitability thresholds for expansion and contraction were generated by relating plot-based counts of seedlings and of percent mortality, respectively, to each species current modeled climatic suitability (Gibson 2011). We acknowledge the uncertainty surrounding the forecast rates and magnitudes of species distribution expansion and contraction, and use these thresholds here simply to compare models built with FIA data using true coordinates to those built with altered coordinates.

Comparative Analyses

Differences between models built with true versus public coordinates originate with differences in the values of sampled predictor variables. Sample distributions for each predictor variable were compared by species for each modeling extent. Adjustments for multiple comparisons were made using the Bonferroni correction (Rice 1989; Cabin and Mitchel 2000) within the ''p.adjust'' function in R (R Development Core Team 2008).

Comparisons of the resulting models were based first on classification accuracies and second on the areal discordance of spatial extrapolation under current and forecast distributions. Classification accuracies which require a threshold, including sensitivity, and specificity (after Fielding and Bell 1997), were based on sample prevalence. In addition to these, the threshold-independent area under the receiver operating characteristic curve (AUC) (Hanley and McNeil 1982) was also compared. We used the true skill statistic as a measure of off-diagonal error (Allouche and others 2006).

All accuracies were generated using out-of-bag bootstrap samples. Areal discordance of modeled distributions was calculated between all model sets. Temporal projections were assessed, in terms of areal and geographic distribution centroid discordance, for models produced with the 100-km-buffered extent (Figure 1). Areal discordance was measured as total area with true- and fuzz-based model agreement/disagreement and summarized by combining all species into a richness map. Geographic centroids of distributions were also generated for each time step to compare divergence in forecast distributions.

Results

Adjusted *P* values from the *t* test comparison of sampled predictor variables indicated only one significant difference for any of the variables by species by modeling extent ($P = 0.03$, Z_ROUGH, *P. edulis*, 50-km buffer) (see Supplemental S1 for all comparisons). Thus, with the exception of 1 of 270 possible variable combinations, values of variables extracted from the underlying GIS predictor set did not differ irrespective of whether true or public plot coordinates were used (Figure 2). This indicates that the environmental space used to model the six species was, for all intent and purpose, virtually identical among true- and public-based models.

Accuracies for both true and public models are shown in Table 2. True-based models were generally 1–4% more accurate than public-based models, suggesting the impacts of using public versus true were largely negligible. Of 96 public versus true comparisons, 54 true-based models were 1–4% greater, 39 were equal, and only three were lower than the public-based models. Differences in model accuracies for public- and true-based models, by species and extent, are illustrated in Figure 3. This difference primarily comes from higher sensitivities of true-based models. Differences in specificity generally decreased with increasing modeling extents.

Absolute areal agreement between public- and true-based modeled distributions, for each time step, are illustrated in Figure 4. In general, public-based models produced geographic distributions with greater total areas than the true-based models. The difference in modeled current distribution was approximately 1% for two species, *J. monosperma* (1.1%) and *J. osteosperma* (1.3%), and around 5% for *P. edulis* (6.9%) and *P. monophylla* (4.4%). This difference was greatest for *J. deppeanna* (17.2%) and *J. occidentalis* (38.7%).

The larger total areas of current distributions produced with public-based models increased with forecast distributions with the exception of *J. osteosperma* whose public-based forecast distribution

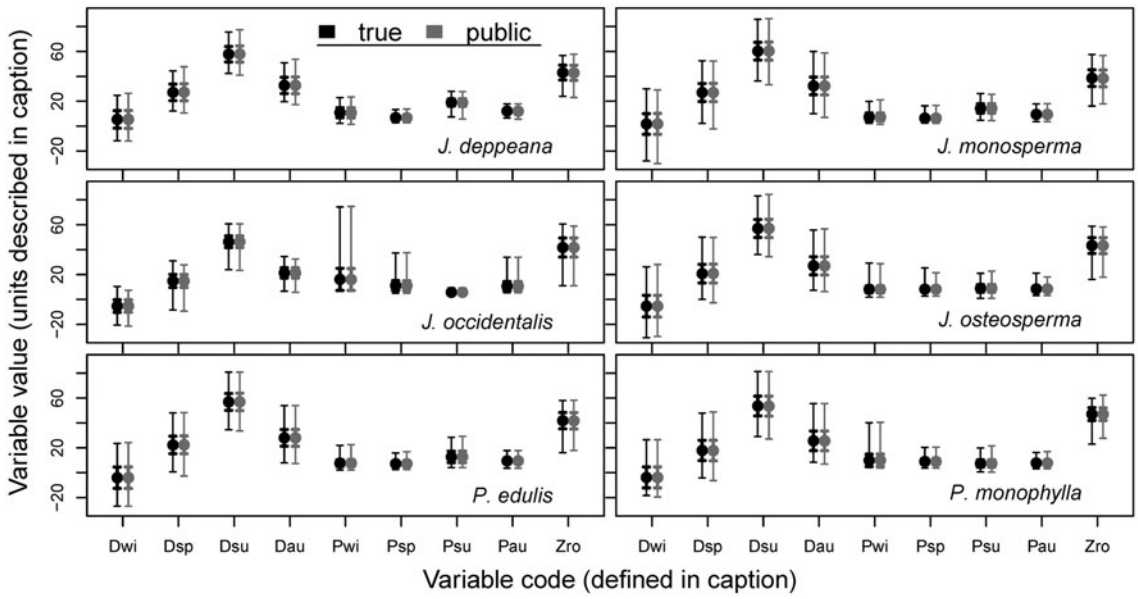

Figure 2. Distribution (mean, first standard deviation, and entire range) of true- and public-based sampled variables for species presence plots: *Dwi* degree sum of winter °C, *Dsp* degree sum of spring °C, *Dsu* degree sum of summer °C, *Dau* degree sum of autumn °C, *Pwi* precipitation sum of winter cm, *Psp* precipitation sum of spring cm, *Psu* precipitation sum of summer cm, *Pau* precipitation sum of autumn cm, *Zro* topographic roughness m.

Table 2. Four Measures of Accuracy Applied to Models Built Using Public Versus True FIA Plot Coordinates for Four Species of Juniper and Two Species of Piñon Pine, Western North America

Extent	Species = JUDE				Species = JUMO			
	SENS	SPEC	AUC	TSS	SENS	SPEC	AUC	TSS
50	0.75/**0.78**	0.95/**0.96**	0.95/**0.97**	0.71/**0.75**	0.61/0.61	0.92/0.92	0.89/0.89	0.53/0.53
100	0.76/**0.78**	0.97/**0.98**	0.96/**0.97**	0.73/**0.76**	0.61/**0.62**	**0.95**/0.94	0.92/0.92	**0.56**/0.55
150	0.75/**0.78**	0.98/0.98	0.97/**0.98**	0.73/**0.76**	0.61/0.61	**0.96**/0.95	0.93/**0.94**	0.56/0.56
200	0.75/**0.78**	0.98/0.98	0.97/**0.98**	0.73/**0.77**	0.61/0.61	0.96/0.96	0.94/0.94	0.57/**0.58**
Extent	Species = JUOC				Species = JUOS			
	SENS	SPEC	AUC	TSS	SENS	SPEC	AUC	TSS
50	0.70/**0.71**	0.93/**0.94**	0.93/0.93	0.63/**0.65**	0.49/0.49	0.92/0.92	0.85/**0.86**	0.41/0.41
100	0.70/**0.71**	0.95/**0.96**	0.95/0.95	0.65/**0.67**	0.48/**0.50**	0.95/0.95	0.90/**0.91**	0.43/**0.45**
150	0.69/**0.71**	0.96/0.96	0.95/**0.96**	0.65/**0.67**	0.49/**0.50**	0.96/0.96	0.92/0.92	0.45/**0.46**
200	0.69/**0.71**	0.96/**0.97**	0.95/**0.96**	0.66/**0.67**	0.49/**0.52**	0.97/0.97	0.93/0.93	0.45/**0.49**
Extent	Species = PIED				Species = PIMO			
	SENS	SPEC	AUC	TSS	SENS	SPEC	AUC	TSS
50	0.73/**0.74**	0.90/**0.91**	0.92/0.92	0.63/**0.65**	0.70/0.70	0.95/0.95	0.94/0.94	0.65/0.65
100	0.73/**0.74**	0.93/0.93	0.94/0.94	0.66/**0.67**	0.69/**0.71**	0.96/0.96	0.95/**0.96**	0.65/**0.68**
150	0.73/**0.74**	0.94/**0.95**	0.95/0.95	0.67/**0.69**	0.70/0.70	0.97/0.97	0.96/0.96	0.67/0.67
200	0.73/**0.74**	0.95/0.95	0.95/**0.96**	0.68/**0.69**	0.69/**0.71**	0.97/0.97	0.97/0.97	0.66/**0.68**

SENS, sensitivity; SPEC, specificity; AUC, area under curve; and TSS, true skill statistic. See ''Methods'' section for each reference. JUDE, Juniperus deppeana; JUMO, J. monosperma; JUOC, J. osteosperma; JUOS, J. occidentalis; PIED, Pinus edulis; PIMO, P. monophylla.
Accuracy measures are read as public/true; bold indicates comparisons where either public- or true-based model accuracy is higher.

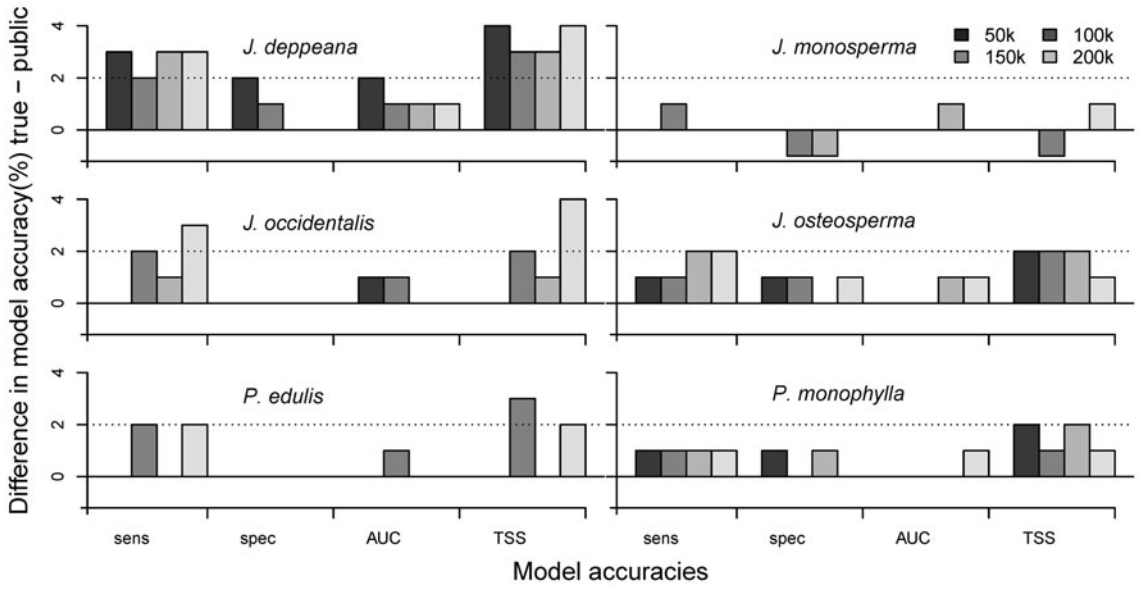

Figure 3. Percent difference (true models –public models) in model accuracies: *sens* sensitivity, *spec* specificity, *AUC* area under curve, and *TSS* true skill statistic for each modeling extent: 50, 100, 150, and 200 k buffers.

was 0.7% lower than its true-based forecast distribution. The relatively larger total areas of public-based models increased with the forecast by 0.2% for *P. edulis*, 5.3% for *P. monophylla*, 5.4% for *J. monosperma*, and 9.2% for *J. deppeana*. For *J. occidentalis* the true-based model had the greatest

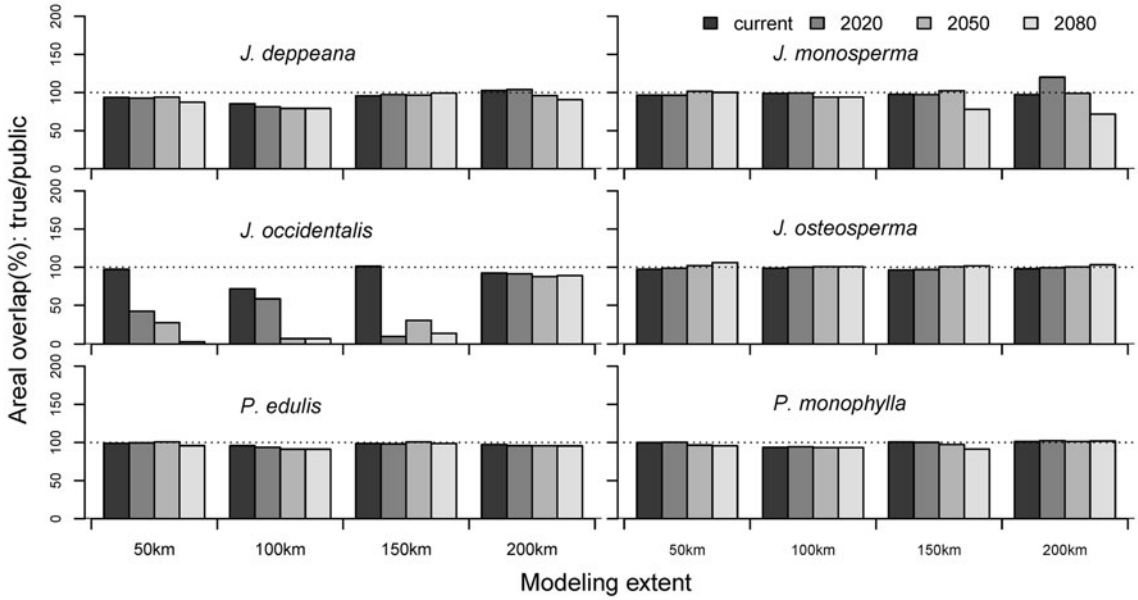

Figure 4. Areal overlap of true- and public-based modeled distribution forecasts for each modeling extent, by species.

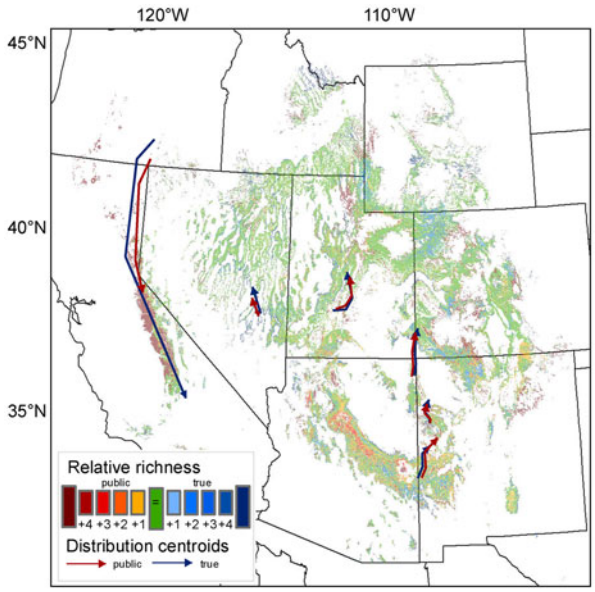

Figure 5. Relative richness of piñon-juniper species distributions projected with true- versus public-based models and their centroids. Areas where public-based models project more species are shown in *red* and areas with more species projected by true-based models are shown in *blue*. Public-based distribution centroids are shown in *red* and true-based centroids shown in *blue*.

discordance of an astonishing 1,265% as a result of the true-based model forecasting an enormous decrease in its distribution (Figure 4).

The collective discordance between all true- and public-based species forecast distributions, for models produced at the 100-km extent, are summarized as differences in species richness (Figure 5). Shifts in mean geographic distribution centroids reflect shifts in total area and are indicative of the similarity in overall magnitude and direction of true- and public-based forecast distribution shifts (Figure 5). Distance between centroids of current distribution models built from true versus public plot coordinates ranged from a low of 6 km for *J. osteosperma* to a maximum of 64 km for *J. occidentalis* (Table 3). By the end of the climate projection period, 2080, these distances ranged from 14 to 368 km for *J. monosperma* and *J. occidentalis*, respectively. Given the scale of the maximum axes of the distributions of the six modeled species was 1,200–2,000 km, these differences in distribution centroids were minor (1–3%), with the exception of *J. occidentalis*.

Discussion

The objective of this analysis is to isolate the effect of using altered coordinates in SDMs. This study does not explicitly address the conceptual challenges facing their development or interpretation of their forecast. The results of this analysis indicate the effects of using public FIA plot coordinates to develop and extrapolate SDMs at a resolution of 1 km^2 are influenced more by the particulars of SDMs than by the errors introduced with altered coordinates. We found no significant differences between the underlying distributions of predictor

Table 3. Distance (km) Between Current and Projected 2080 Distribution Centroids by Species Models Built Using Public versus True FIA Plot Coordinates for Four Species of Juniper and Two Species of Piñon Pine, Western North America

Species	Distance (km) by model		
	Current	2080	Maximum axis
JUDE	12	34	1,200
JUMO	7	14	1,300
JUOC	64	368	1,600
JUOS	6	18	2,000
PIED	14	15	1,300
PIMO	9	35	1,000

Maximum axis is the longest length of the current distribution, and is shown to provide a measure of the scale of differences in the true- versus public-based models.
JUDE, Juniperus deppeana; JUMO, J. monosperma; JUOC, J. osteosperma; JUOS, J. occidentalis; PIED, Pinus edulis; PIMO, P. monophylla.

variables sampled with true versus public coordinates at any modeling extent. This finding supports previous work which indicates the effects of using public coordinates to sample gridded variables with a resolution of 1 km^2 are largely negligible (for example, McRoberts and others 2005; Coulsten and others 2006; Pricely and others 2009).

Although the underlying distributions of sampled variables were not found to be significantly different at this resolution, classification accuracies did vary slightly between and within public- and true-based models as a function of extent. Within the two sets there was a ubiquitous trend of increasing model specificities in proportion to the accumulation of absences associated with increasing extents. In short, the addition of absences from outside the known geographic distribution of the species added greater variability to the environmental space and enhanced the discrimination capability of the RF models. Note, however, that this is a statistical rather than ecological consequence, and may actually lead to misleading models as the number of absences are increasingly selected from locations that have geographic isolation but potentially suitable climatic conditions.

In addition, the difference between accuracy specificities of public- and true-based models decreased as modeling extent increased. Differences in sensitivity, however, generally remained constant within and between the two sets of models across the range of modeling extents, with the true-based models slightly higher (1–5%) than the public-based models. The higher sensitivity of models developed with true coordinates translated into geographic distributions with smaller total areas than the public-based models. For all species but *J. osteosperma* the difference increased with temporal extrapolation. Similarly, for all species but *J. occidentalis*, the negligible differences in distribution centroids also support a general conclusion that no real effect is apparent whether true or public plot coordinates are used in SDMs.

The greatest difference between public- and true-based models occurred for the SDM with the worst classification accuracies, *J. occidentalis*. The comparison of sampled environmental variables for this species indicated there was no statistically significant difference between public and true coordinates. Instead, we reason this difference stems from the fact that the distribution of this species is difficult to model for several reasons which often beset distribution modeling. Principle among these reasons is (i) the myriad of land-use legacies impacting this species distribution in addition to (ii) the complex terrain and relationship with fire-adapted grasses that this species inhabits, and (iii) is compounded by relatively few presences to model within an area where environmental variables, that is, climatic conditions, are similarly difficult to model. The challenges inherent in modeling species distributions, exemplified in part by *J. occidentalis*, are of more concern than the effects of using altered coordinates.

The FIA database is an unprecedented geographic database of tree presence and condition, ideal for applications of SDMs. Our research indicates there are only negligible effects for using public FIA coordinates for developing SDMs at regional scales. Researchers working on such a scale may, therefore, not be concerned with the effects of using public coordinates and instead focus on the challenges inherent in the process of developing SDMs relevant to broader research objectives.

ACKNOWLEDGMENTS

We thank J. DeRose, T. Sharik, and J. MacMahon for their reviews of this and earlier drafts. Funding for this research was provided to Edwards and Gibson by the US Forest Service, Rocky Mountain Research Station, Forest Inventory and Analysis Program. Mention of any product by name does not constitute endorsement by the U.S. Geological Survey or the Federal Government.

REFERENCES

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). J Appl Ecol 43:1365–2664.

Araújo MB, Pearson RG. 2005. Equilibrium of species' distributions with climate. Ecography 28:693–5.

Barve N, Barve V, Jiménez-Valverde A, Lira-Noriega A, Maher SP, Peterson AT, Soberón J, Villalobos F. 2011. The crucial role of the accessible area in ecological niche modeling and species distribution modeling. Ecol Model 222:1819–20.

Betancourt JL, Van Devender TR, Martin PS. 1990. Packrat middens: the last 40,000 years of biotic change. Tucson: University of Arizona Press.

Betancourt JL, Schuster WS, Mitton JB, Anderson RS. 1991. Fossil and genetic history of a pinyon pine (*Pinus edulis*) isolate. Ecology 72:1685–97.

Breiman L. 2001. Random forests. Mach Learn 45:5–32.

Cabin RJ, Mitchel RJ. 2000. To Bonferroni or not to Bonferroni: when and how are the questions. Bul Ecol Soc Am 81:246–8.

Chambers JC, Vander Wall SB, Schupp EW. 1999. Seed and seedling ecology of piñon and juniper species in the pygmy woodlands of western North America. Bot Rev 65:1–38.

Cole KL, Ironside K, Arundel S, Duffy P, Shaw J. 2008a. Modeling future plant distributions on the Colorado Plateau: an example using *Pinus edulis*. In: van Riper C, Sogge IIIM, Eds. The Colorado Plateau III; integrating research and resources management for effective conservation. Tucson, AZ: The University of Arizona Press. p 319–30.

Cole KL, Fisher J, Arundel ST, Cannella J, Swift S. 2008b. Geographical and climatic limits of needle types of one- and two-needled pinyon pines. J Biogeogr 35:257–69.

Coulston JW, Riitters KH, McRoberts RE, Reams GA, Smith WD. 2006. True versus perturbed forest inventory plot locations for modeling: a simulation study. Can J Forest Res 36:801–7.

Cutler RCTC, Edwards KH Jr, Beard A, Cutler KT, Gibson Hess J, Lawler JJ. 2007. Random forests for classification in ecology. Ecology 88:2783–92.

Edwards TC Jr, Cutler DR, Zimmermann NE, Geiser L, Moisen GG. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecol Model 199:132–41.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ Conserv 24:38–49.

Floyd ME. 1982. The interaction of piñon pines and gambel oak in plant succession near Dolores, Colorado. Southwest Nat 27:143–7.

Freeman EA, Moisen GG. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol Model 217:48–58.

García-Valdés R, Zavala MA, Araújo MB, Purves DW. 2013. Chasing a moving target: projecting climate change-induced shifts in non-equilibrium tree species distributions. J Ecol 101:441–53.

Gibson J. 2011. Individualistic responses of piñon and juniper distributions to projected climate change. Unpublished M.S. Thesis, Utah State University, Logan, UT, USA.

Gray ST, Betancourt JL, Jackson ST, Eddy RG. 2006. Role of multidecadal climate variability in a range extension of pinyon pine. Ecology 87:1124–30.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. Ecol Model 135:147–86.

Hanley JA, McNeil BJ. 1982. The meaning and use of the area under the receiver operating characteristic (ROC) curve. Radiology 143:29–36.

Hijmans RJ, van Etten J. 2011. raster: Geographic analysis and modeling with raster data. R package version 1.9-5. http://CRAN.R-project.org/package=raster.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25:1965–78.

Homer C, Dewitz J, Fry J, Coan M, Hossain N, Larson C, Herold N, McKerrow A, VanDriel JN, Wickham J. 2007. Completion of the 2001 National Land Cover Database for the Conterminous United States. Photogramm Eng Rem S 73:337–41.

Iverson LR, Prasad AM, Matthews SN, Peters M. 2008. Estimating potential habitat for 134 eastern US tree species under six climate change scenarios. Forest Ecol Manag 254:390–406.

Lanner RM. 1998. Seed dispersal in *Pinus*. pp 281–293. In: Richardson DM, Ed. Ecology and biogeography of *Pinus*. Cambridge University Press, Cambridge, UK.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. R News 2:18–22.

Manel S, Williams HC, Ormerod SJ. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–31.

Martens SN, Breshears DD, Barnes FJ. 2001. Development of species dominance along an elevational gradient: population dynamics of *Pinus edulis* and *Juniperus monosperma*. Int J Plant Sci 162:777–83.

McDowell N, Pockman WT, Allen CD, Breshears DD, Cobb N, Kolb T, Plaut J, Sperry J, West A, Williams DG, Yepez EA. 2008. Mechanisms of plant survival and mortality during drought: why do some plants survive while others succumb to drought? New Phytol 178:719–39.

McRoberts RE, Holden GR, Nelson MD, Liknes GC, Moser WK, Lister AJ, King SL, LaPoint EB, Coulston JW, Smith WB, Reams GA. 2005. Estimating and circumventing the effects of perturbing and swapping inventory plot locations. J For 103:275–9.

Neilson RP. 1987. Biotic regionalization and climatic controls in western North America. Vegetatio 70:135–47.

Pearman PB, Guisan A, Broennimann O, Randin C. 2008. Niche dynamics in space and time. Trends Ecol Evol 23:149–58.

Pricely SP, Wang H, Radtke PJ, Coulston J. 2009. Combining FIA plot data with topographic variables: are precise locations needed? USDA For Serv Proc: RMRS-P-56.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org.

Rehfeldt GE, Ferguson DE, Crookston NL. 2008. Quantifying the abundance of co-occurring conifers along the inland northwest (USA) climate gradients. Ecology 89:2127–39.

Rehfeldt GE, Ferguson DE, Crookston NL. 2009. Aspen, climate, and sudden decline in western USA. Forest Ecol Manag 258:2353–64.

Rice WR. 1989. Analyzing tables of statistical test. Evolution 43:223–5.

Romme WH, Allen CD, Bailey JD, Baker WL, Bestelmeyer BT, Brown PM, Eisenhart KS, Floyd ML, Huffmann DW, Jacobs BF, Miller RF, Muldavin EH, Swetnam TW, Tausch RJ, Weisberg PJ. 2009. Historical and modern disturbance regimes, stand structures, and landscape dynamics in piñon-juniper vegetation of the western United States. Rang Ecol Manag 62:203–22.

Sánchez-Fernández D, Lobo JM, Hernández-Manrique OL. 2011. Species distribution models that do not incorporate global data misrepresent potential distributions: a case study using Iberian diving beetles. Divers Distrib 17:163–71.

Shaw JD, Brytten BE, DeBlander LT. 2005. Forest inventory and analysis (FIA) annual inventory answer the question: what is happening to Pinyon-Juniper woodlands? J Forest 103:280–5.

Stokland JN, Rune H, Stoa B. 2011. Species distribution modeling: effect of design and sample size of pseudo-absence observations. Ecol Model 222:1800–9.

West AG, Hultine KR, Sperry JS, Bush SE, Ehleringer JR. 2008. Transpiration and hydraulic strategies in a piñon-juniper woodland. Ecol Appl 18:911–27.

Wiens JA, Stralberg D, Jongsomjit D, Howell CA, Snyder MA. 2009. Niches, models, and climate change: assessing the assumptions and uncertainties. Proc Nat Acad Sci USA 106:19729–36.

Woodall CW, Oswalt CM, Westfall JA, Perry CH, Nelson MD, Finley AO. 2010. Selecting tree species for testing climate change migration hypotheses using forest inventory data. Forest Ecol Manag 259:778–85.

Zarnetske PL, Edwards TC Jr, Moisen GG. 2007. Habitat classification modeling with incomplete data: pushing the habitat envelope. Ecol Appl 17:1714–29.

Zimmermann NE, Edwards TC Jr, Moisen GG, Frescino TS, Blackard JA. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. Ecol Model 44:1057–67.

Zimmermann NE, Yoccoz NG, Edwards TC Jr, Meier ES, Thuiller W, Guisan A, Schmatz DR, Pearman PB. 2009. Climatic extremes improve predictions of spatial patterns of tree species. Proc Natl Acad Sci USA 106:19723–8.