RESEARCH PAPER

Using a dynamic forest model to predict tree species distributions

Alvaro G. Gutiérrez[1,2]*, Rebecca S. Snell[1] and Harald Bugmann[1]

[1]*Forest Ecology, Institute of Terrestrial Ecosystems, Department of Environmental Sciences, ETH Zürich, 8092 Zürich, Switzerland,* [2]*Departamento de Ciencias Ambientales y Recursos Naturales Renovables, Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago, Chile*

ABSTRACT

Aim It has been suggested that predicting species distributions requires a process-based and preferably dynamic approach. If dynamic models are to contribute towards understanding species distributions, uncertainties related to their spatial extrapolation and bioclimatic parameters need to be addressed. Here, we analyse the potential of a forest gap model for predicting species distributions.

Location Pacific Northwest of North America (PNW).

Methods We used the dynamic forest gap model ForClim, which includes climate, competition and demographic processes, to simulate the distribution of 18 tree species outside the domain of the data used for fitting. We explored model accuracy for species distributions at the regional scale by: (1) estimating species climatic tolerances so as to maximize agreement with regional distribution maps versus (2) employing a bioclimatic parameter set that produces high accuracy at the local scale. We then performed the opposite tests and simulated local forest composition in a small area in the PNW, using (3) the local bioclimatic parameters and (4) the bioclimatic parameters that produced the highest accuracy at the regional scale. We also compared the ForClim results with predictions from a standard correlative species distribution model (SDM).

Results ForClim produced regional species distributions with fair to very good agreement for 12 tree species. The optimized bioclimatic parameters consistently improved the accuracy of regional predictions compared with simulations run with the local parameters, and were consistent with SDM results. At the local scale, predictions using the local parameters conformed to descriptions of forest composition, but accuracy decreased strongly when using the regionally calibrated parameters.

Main conclusions Forest gap models can predict regional species distributions, but at the cost of reduced accuracy at the local scale. Future applications of gap models to understand regional species distributions should include robust parameterization schemes and additional ecological processes that are important at large spatial scales (e.g. dispersal, disturbances).

Keywords
Dynamic biogeography, dynamic vegetation models, forest gap models, inverse modelling, model parameterization, prediction accuracy, sensitivity analysis, temperate rainforests.

*Correspondence: Alvaro G. Gutierrez, Departamento de Ciencias Ambientales y Recursos Naturales Renovables, Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago, Chile.
E-mail: bosqueciencia@gmail.com*

INTRODUCTION

Predicting species distributions is challenging due to the dynamic processes that influence species ranges (MacArthur, 1972; Gaston, 2003) and the various spatial and temporal scales at which these process operate (Levin, 1992; Chave, 2013). A variety of statistical species distribution models have been developed (Guisan & Zimmermann, 2000; Thuiller *et al.*, 2008), but there is an increasing recognition that understanding species distributions is likely to require a process-based, dynamic

modelling approach (Dormann *et al.*, 2012; García-Valdés *et al.*, 2013). Dynamic vegetation models (DVMs) have the potential to improve our understanding and prediction of species distributions, but there have been only a few such applications to date (cf. Snell *et al.*, 2014, for a review). As most DVMs were not designed to address questions of species range dynamics, an important first step is to test the limitations and assumptions of a particular model at a scale that is appropriate for the prediction of species distributions.

DVMs simulate changes through time in the occurrence, abundance and productivity of plant species. They typically include processes that are important for simulating range dynamics, such as competition and demography, and the influences of climate on these processes. Forest gap models (Bugmann, 2001) are DVMs that are widely used to predict forest community composition at the local scale. These models appear to be suitable for simulating species distributions because they incorporate biotic interactions and abiotic constraints at the level of individual trees based on dynamic calculations of growth, establishment and mortality (Bugmann, 2001). To date, forest gap models have not been used for species distribution analyses, for two main reasons. First, gap models often require a large number of parameters to be estimated for each species, thus limiting their application to a comparatively small set of species (Guisan & Zimmermann, 2000; Guisan & Thuiller, 2005; Dormann *et al.*, 2012). Second, the bioclimatic parameters in gap models are typically estimated using somewhat arbitrary approaches (e.g. qualitative estimates of climatic conditions) or by using coarse map resolutions that fail to capture the influence of topography on local climate (Loehle & LeBlanc, 1996; Schenk, 1996). While the uncertainty about the quantification of species climatic tolerances may influence the ability of a gap model to simulate tree species distributions, reservations regarding the number of species needed to run simulations are no longer valid. For example, the ForClim model currently has about 140 tree species parameterized, allowing the simulation of forest composition over large regions (Bugmann & Solomon, 2000; Busing *et al.*, 2007).

Inverse modelling techniques (Grimm & Railsback, 2005) offer a transparent and robust method for estimating the climatic tolerances of species. In such an approach, data on species distributions and climate can be used for calibrating species parameters in forest gap models, as is traditionally done in correlative species distribution modelling (Hartig *et al.*, 2012). However, parameters denoting species climatic tolerances in DVMs are rarely changed without: (1) a reformulation of the model (with respect to ecological processes) or (2) an explicit up-scaling of the model to gain efficiency (Lischke *et al.*, 2006).

Here, we propose to use an inverse modelling approach to test the applicability of a forest gap model at the regional scale, and to address uncertainties related to their spatial extrapolation and bioclimatic parameters. Specifically, we assess whether the gap model framework is an appropriate representation of those processes that determine species ranges; if so, the model should be able to predict regional species distributions. Further, we assess

the ability of a forest gap model to predict species distributions across spatial scales by employing a bottom-up versus top-down parameterization approach. The bottom-up approach uses a parameter set that has high accuracy at the local scale for simulating regional-scale distributions, whereas in the top-down approach we evaluate model accuracy at the local scale when using a parameter set that has high accuracy at the regional scale. Ultimately, these analyses serve to reveal whether there is a trade-off between model generality and specificity when using DVMs to predict regional species distribution (Thuiller *et al.*, 2008).

METHODS

The forest gap model – ForClim

ForClim (Bugmann, 1996; Bugmann & Solomon, 2000) is a generalized forest succession model that can be used for the temperate zone (Bugmann *et al.*, 2001). ForClim incorporates simple yet reliable formulations of biotic (i.e. competition for light) and abiotic (i.e. climatic) influences on ecological processes, while using a limited number of ecological assumptions (Bugmann, 1996). ForClim simulates independent forest patches (usually 200 patches of 800 m^2), where trees establish and compete for light following the gap model approach (Bugmann *et al.*, 2001). Stand properties are derived by averaging the properties across the replicated patches (Bugmann, 1996). Tree growth is modelled using an empirical growth equation, which is regulated by species-specific responses to the availability of light across the canopy, growing degree-days and soil moisture (Bugmann, 1996; Bugmann & Solomon, 2000). Mean monthly temperatures are used to calculate growing degree-days and winter temperature. A monthly drought index is calculated based on soil water holding capacity, monthly precipitation sums and mean monthly temperatures (Bugmann & Cramer, 1998; Bugmann & Solomon, 2000). The effects of the climatic variables and light availability are combined to derive realized tree growth rates (Bugmann & Solomon, 2000). Tree establishment rates are determined as a function of species-specific responses to winter temperature, light availability on the forest floor and growing degree-days (Bugmann, 1996). A comprehensive description of the version of the model used in this study (ForClim v.3.0) can be found in Rasche *et al.* (2011, 2012).

We applied ForClim to the Pacific Northwest region of North America (PNW) to simulate the distribution of 18 tree species using species-specific parameters adopted from Bugmann & Solomon (2000). These parameters were either left unchanged or were recalculated to fit the definitions in the current version of the model (for details see Appendix S1 in Supporting Information). For the purpose of this study, we focused on re-estimating those species-specific parameters that denote climatic tolerances, i.e. the minimum degree-day requirement (kDDMin), minimum and maximum winter temperature tolerance (kWiTN and kWiTX, respectively) and drought tolerance

(kDrTol), hereafter referred to collectively as 'bioclimatic parameters'.

Calculation of bioclimatic tolerances from species range information

First, we extracted climatic information from the entire distribution range of the target species. We used digital representations of the species range in North America, as described by Little (1971). We randomly distributed 1000 points over each species range map, and then used those points to extract the corresponding climatic data from a gridded database (WorldClim, v.1.4, release 3; Hijmans *et al.*, 2005). WorldClim provides climatic information at high resolution (2.5 arcmin in this study) and captures the climatic complexity that is induced by topography. We used monthly precipitation and temperature data from WorldClim at each point to calculate the bioclimatic variables following the ForClim equations (Bugmann, 1996; Bugmann & Cramer, 1998) and emulating the stochastic calculations performed by the model. The soil-related parameters required for calculating the drought index were taken from Bugmann & Cramer (1998) and Bugmann & Solomon (2000).

The bioclimatic variables extracted from each point over the entire species range were used to estimate their probability density function (black line in Fig. S1), which was approximated by a polynomial spline with the unknown coefficients estimated by maximum likelihood (Kooperberg & Stone, 1992). We used these probability density functions to evaluate the probability of the original bioclimatic parameters from Bugmann & Solomon (2000) (orange bar in Fig. S1; hereafter *local* parameters). This procedure was repeated for each species and each bioclimatic variable.

Simulation of tree species distribution at the regional scale

Our regional study area covers the distributional ranges of the 18 tree species parameterized in the PNW by Bugmann & Solomon (2000). It was delimited by merging the distribution maps from all 18 species, and then constrained to the western slope of the Cascade Mountains. This latter condition was imposed to avoid several dominant tree species not yet parameterized in the model. We divided the study area into 328 strata of equal area, and selected a random location in each stratum. This stratified random sampling of locations ensured that the selected locations covered the entire study area. We used ForClim to build a presence–absence *prediction map* for each species based on the ForClim simulations of steady-state forest composition. For each of the 328 strata, 200 patches were simulated (corresponding to *c.* 16 ha) for 1500 years of succession, allowing unlimited seed availability for all species. Climatic parameters for running the model at each location were extracted from the WorldClim monthly temperature and precipitation data. A species was determined to be 'present' at a certain location if the sum of the basal area of trees > 12.7 cm

diameter at breast height was > 1 m^2 ha^{-1}. This threshold was found to produce the most accurate results (Fig. S2) and it is a commonly used tree size threshold for forest inventories (McRoberts *et al.*, 2005).

For each species we also built a *reference map*, which used the same 328 strata as above but species presence/absence was determined from Little (1971). The reference and prediction maps were then compared to assess the accuracy of presence–absence simulations from ForClim using Cohen's Kappa statistic (Goodman & Kruskal, 1954; Cohen, 1960). Kappa values range from −1 to +1, with values > 0.4 indicating a fair degree of agreement. Values < 0.2 indicate performance no better than random (Monserud & Leemans, 1992). We complemented Kappa with the following statistics (Allouche *et al.*, 2006): the area under the receiver operating characteristic curve (AUC) and the proportion of observed presences that are predicted as such (i.e. sensitivity, accounting for omission errors).

We used the range of probabilities described above to create 100 parameter sets using a Latin Hypercube design (Stein, 1987). The 100 sets covered the range of probabilities for each bioclimatic variable, but the actual parameter value used was extracted from the probability density function for each species. For example, a minimum degree-day probability of 0.1 would mean 609 degree-days for *Abies amabilis*, but 931 degree-days for *Acer macrophyllum*. Each species was simulated using all 100 probability sets, and the results were evaluated to determine which parameter values produced the most accurate presence–absence simulations. To select the bioclimatic parameters that maximized simulation accuracy for each species distribution, we ran simulations for the 328 strata using the 100 parameter sets (i.e. 32,800 simulations of steady-state forest composition). When comparing the reference and prediction maps for each species we selected the parameter set with the highest Kappa (hereafter the 'optimized' parameter set; blue bar in Fig. S1).

In addition, we ran simulations for the 328 strata using the bioclimatic parameters as estimated by Bugmann & Solomon (2000) (the 'local' parameter set, see also Appendix S1). The local parameter set was obtained in 1995 (Bugmann & Solomon 1995) by visually overlaying bioclimatic maps directly onto Little's species distribution maps (Little 1971) and modified to coincide with species bioclimatic requirements as described by Franklin & Dyrness (1988).

We further checked whether species traits indirectly influence the ability of the model to distinguish climatically suitable and unsuitable habitats (Guisan *et al.*, 2007). For this analysis, we correlated the Kappa statistics from the optimized parameters against species-specific parameter values denoting functional traits including maximum age, maximum height, growth rate and shade tolerance.

Finally, we compared ForClim predictions with species distribution maps obtained from a correlative species distribution model (SDM). We ran a SDM for each tree species (see additional methods in Appendix S1) and tested the prediction accuracy for the 328 strata. The accuracy of the SDM for presence–absence simulations was assessed following the same methods used for ForClim simulations (see also above).

Simulation of tree species distributions at the local scale

To test if regionally optimized bioclimatic parameters are valid at the local scale, we chose a longitudinal transect along which ForClim had already been comprehensively tested and had been found to produce accurate projections of tree species composition and biomass (Bugmann & Solomon, 2000). This 242-km transect is located at 44.13° N, extending from the Pacific Coast into the desert shrubland of the Great Basin in the interior of Oregon. Temperature and moisture vary from wet and warm at the coast to wet and cold in the Cascade Mountains, and into the hot dry conditions of the eastern Oregon desert shrubland. We simulated the same sites along the Bugmann & Solomon (2000) transect, using climatic data extracted from WorldClim and other site characteristics obtained from Bugmann & Solomon (2000) (i.e. soil water holding capacity and available nitrogen). At each site we simulated steady-state forest composition following Bugmann & Solomon (2000) using the local and optimized parameter sets. Simulation results were compared qualitatively and using the Jaccard index of similarity at each site. All statistical analyses were conducted in R (R Core Team, 2013).

RESULTS

Regional species distributions

Optimizing bioclimatic parameters consistently improved the accuracy of the presence–absence simulation of the DVM at the regional scale for all tree species compared with simulations that used the local parameters (Table 1; Kappa > 0.4 for 10 out of 18 tree species). The SDM predictions were more accurate than DVM predictions for all but three species (Table 1). However, DVM predictions with the optimized parameters approached the accuracy and sensitivity of SDM (Table 1, Figs 1 & 2, Appendices S2 & S3). Both SDM and DVM simulations produced commission errors towards the northern limits of species distributions (Table 2). In general, DVM predictions had lower commission errors than SDM predictions, but higher omission errors (Table 2). From the traits analysed, only the shade tolerance parameter was significantly correlated with regional presence–absence accuracy of the DVM (Spearman's rank correlation, $\rho < -0.6$, $P < 0.001$), indicating that predictions are less accurate for shade-intolerant species.

For some tree species, ForClim produced a highly accurate regional presence–absence map when using the optimized parameter set (Kappa > 0.7, e.g. *Picea sitchensis*; Fig. 1a, see also

Table 1 Summary of accuracy statistics for presence–absence predictions of tree species in the Pacific Northwest of North America. Three accuracy statistics are included: (1) the Kappa statistic, (2) AUC (the area under the curve of the receiver operating characteristic curve) and (3) sensitivity: the proportion of correctly predicted presences, calculated after Allouche *et al.* (2006). Species are ordered by DVM-optimized Kappa values. The highest value for each accuracy statistic, for each species is indicated in bold.

Species name	Kappa			AUC			Sensitivity		
	DVM-optimized	DVM-local	SDM	DVM-optimized	DVM-local	SDM	DVM-optimized	DVM-local	SDM
Picea sitchensis	**0.78**	0.1	0.77	0.88	0.54	**0.90**	0.82	0.09	**0.92**
Tsuga heterophylla	**0.78**	0.5	0.65	**0.89**	0.73	0.83	**0.87**	0.47	0.83
Thuja plicata	**0.71**	0.56	0.66	**0.84**	0.74	**0.84**	0.73	0.51	**0.79**
Arbutus menziesii	0.65	0.57	**0.76**	0.78	0.73	**0.91**	0.59	0.48	**0.87**
Abies grandis	0.61	0.38	**0.79**	0.79	0.81	**0.89**	0.62	0.86	**0.81**
Chamaecyparis nootkatensis	0.6	0.58	**0.83**	0.78	0.78	**0.93**	0.64	0.63	**0.92**
Acer macrophyllum	0.54	0.23	**0.64**	0.74	0.59	**0.85**	0.53	0.19	**0.91**
Abies lasiocarpa	0.5	−0.01	**0.67**	0.77	0.5	**0.84**	0.76	0.24	**0.80**
Tsuga mertensiana	0.49	0.43	**0.76**	0.78	0.74	**0.91**	0.79	0.72	**0.94**
Alnus rubra	0.42	0.08	**0.78**	0.7	0.53	**0.89**	0.47	0.07	**0.87**
Abies amabilis	0.37	0.31	**0.59**	0.69	0.72	**0.86**	0.5	0.73	**0.87**
Pseudotsuga menziesii	0.37	0.29	**0.46**	0.69	0.64	**0.74**	0.62	0.5	**0.77**
Pinus ponderosa	0.36	0.17	**0.37**	**0.75**	0.6	0.73	**0.81**	0.45	0.70
Picea engelmannii	0.35	−0.07	**0.36**	**0.76**	0.55	0.74	**0.77**	0.21	0.67
Pinus monticola	0.29	0.16	**0.30**	0.66	0.6	**0.70**	0.48	0.43	**0.61**
Quercus garryana	0.14	0.06	**0.64**	0.54	0.52	**0.83**	0.09	0.04	**0.70**
Abies procera	0	−0.03	**0.58**	0.5	0.55	**0.75**	0.2	0.1	**0.50**
Pinus contorta	0	0	**0.55**	0.5	0.5	**0.78**	0.03	0.01	**0.75**

DVM-optimized, simulations of the dynamic vegetation model (ForClim) using the parameter set for each species that maximized the accuracy of presence–absence simulations at the regional scale; DVM-local, simulations run with the parameter set used in Bugmann & Solomon (2000) for simulating a transect at 44.13° N latitude; SDM, results of a correlative species distribution model.

Optimised parameters   ## Local parameters
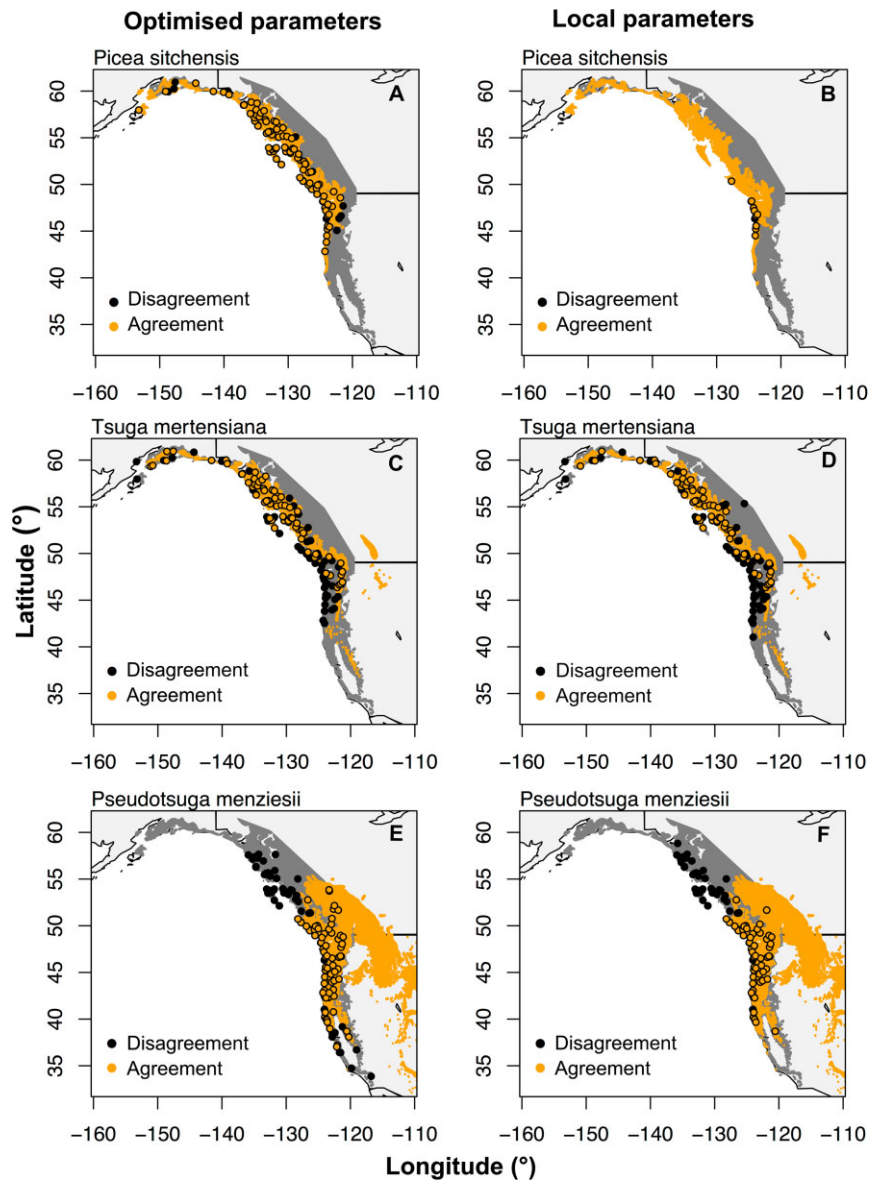
Figure 1 Predictive distribution map for three tree species in the Pacific Northwest of North America using a dynamic forest gap model (ForClim). The orange (online version only)/light grey area is the species range (reference maps from Little, 1971). Each circle indicates a location where the model predicted the species to occur. Orange/grey circles indicate agreement, where ForClim predicted the species to occur and the location was inside the species range. Black circles indicate disagreement, where ForClim predicted the species to occur but the location was outside the species range. Simulations were run using the optimized parameter sets obtained in this study and the local parameter set based on Bugmann & Solomon (2000). Simulations covered a delimited study area (indicated in dark grey) represented by 328 stratified random locations, where the model was run to predict steady-state forest composition.
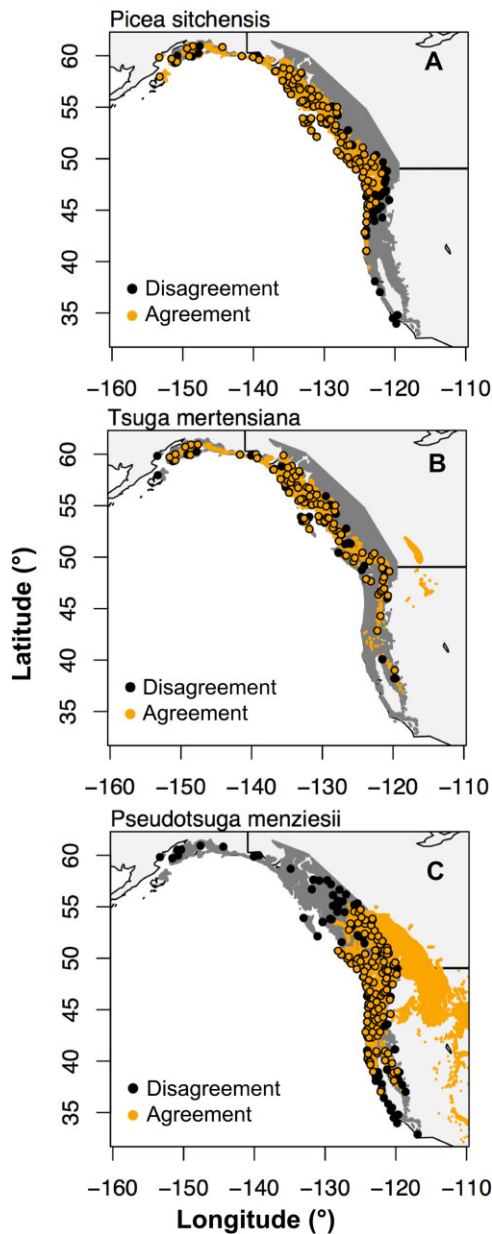
Appendix S2). In contrast, the local parameter set predicted the same species to be constrained to the area where the model had originally been calibrated (Fig. 1b). For other species, regional predictive maps were similar when using either the optimized or the local parameters (e.g. *Tsuga mertensiana*, Fig. 1c, d). The case of *Pseudotsuga menziesii* represents a third response type, where the optimized parameters improved the simulation results but resulted in an over-prediction beyond the potential limits of the species; in this case the optimization increased sensitivity (Table 1, Fig. 1e, f). Finally, ForClim was not able to predict the more southerly part of the regional distribution for some species (e.g. *T. mertensiana*, Fig. 1c; see also *Alnus rubra* in Appendix S2).

A sensitivity analysis across all 18 species revealed that there was no discernible pattern of improvement in regional presence–absence predictions to similar changes in bioclimatic parameters (Fig. 3). Instead, the variation in the accuracy in

response to changes in bioclimatic parameters was highly species specific. For example, *Picea sitchensis* was most sensitive to variation in the minimum degree-day requirement (kDDMin in Fig. 3a), whereas other species were sensitive to variation in drought tolerance (kDrTol in Fig. 3b, d). In general, however, changes in all bioclimatic parameters of a species were needed to obtain the most accurate prediction of regional presence–absence (Fig. 3). These changes restricted the range of probabilities to > 0.8 for maximum winter temperature and drought tolerance (Fig. 4a, b), and < 0.2 for minimum degree-day requirement (Fig. 4c); and we obtained a broad range of probabilities for minimum winter temperature (Fig. 4d).

Local species distributions

At the local scale, simulations using the local parameter set conformed to the forest composition zones described by

Figure 2 Predictive distribution map for three tree species in the Pacific Northwest of North America using a correlative species distribution model (SDM). Details are as in Fig. 1. Methods for running the SDM are described in Appendix S1.

Table 2 Summary of presence–absence simulation errors for 18 tree species in the Pacific Northwest of North America. Commission errors indicate that the model predicted a presence when it should be absent, and omission errors indicate the model predicted an absence when it should be present. Results were calculated using all of the species data (a total of 5904 trials, number of species × number of sites) and are expressed as a percentage of this total. As most of the errors were at the edge of the species ranges, we focused on these areas. Upper and lower elevation sites refer to sites located above the 75th and below the 25th quantiles of the elevational distribution for each species. Northern and southern sites were defined as those sites located north of the 75th and south of the 25th quantiles of the latitudinal distribution for each species. 'Overall total' refers to the total sum of errors across the entire distribution for all species. The names of model simulations as in Table 1.

|	Commission errors		Omission errors	
	DVM-optimized	SDM	DVM-optimized	SDM
Upper elevation	0.7	3.6	4.6	1.3
Lower elevation	1.0	0.6	0.6	0.2
Northern	5.1	8.4	3.6	0.8
Southern	2.6	2.6	3.4	1.6
Overall total	9.4	12.9	10.8	3.3

simulated open forests with biomass < 7 t ha^{-1} composed of *Abies lasiocarpa* and *T. mertensiana*, which corresponds well to forest descriptions near upper tree line (Fig. 5a, Franklin & Dyrness, 1988). Similar to Bugmann & Solomon (2000), our simulations showed forests dominated by *Abies grandis* and *P. menziesii*, followed by *Pinus ponderosa* from 121.50° W, and steppe from 121° W in the lower eastern Cascades (Fig. 5a).

When we used the regionally optimized parameter set to simulate the same transect, the results were quite different (Fig. 5b). Forest composition was characterized by the almost ubiquitous and erroneous dominance of *Picea sitchensis* and *T. heterophylla* along the transect, the disappearance of the *A. amabilis* zone and a strong over-prediction of biomass at the highest elevations (Fig. 5b). Compositional similarity between simulations using the optimized and local parameters was always < 0.5, falling to values of < 0.2 between longitudes 121.92 and 121.62° W (Fig. 5c).

DISCUSSION

Overall, our results demonstrate the potential of using forest gap models to predict regional species distributions. They also illustrate a potential trade-off between predicting geographical species distributions (generality) and capturing local forest composition accurately. At the regional scale, simulations of tree species distributions had fair to very good accuracy compared with the reference maps (Table 1), suggesting that ForClim includes the key processes required to capture regional distributions for most tree species. However, not all species were well represented, indicating that there may be ecological processes that are missing in the model.

Franklin & Dyrness (1988; Fig. 5a). *Picea sitchensis* forests from longitudes 124–123.50° W, replaced by forests dominated by *Pseudotsuga menziesii*, *Tsuga heterophylla* and *Tsuga plicata* from longitudes 123.50–122.25° W; increased dominance of *Abies amabilis* from 122.12° W, and forests dominated by *Abies amabilis* and *T. mertensiana* between longitudes 121.92 and 121.83° W as well as from 121.71 to 121.62° W (Fig. 5a). As a matter of fact, simulations with ForClim 3.0 were superior to results from Bugmann & Solomon (2000), particularly for forest composition at the highest elevations (i.e. longitudes 121.79–121.75° W; Fig. 5a). Using the local parameter set, ForClim

a) P. sitchensis
b) P. ponderosa
c) P. menziesii
d) T. plicata

Figure 3 Sensitivity analysis for the bioclimatic parameters used to maximize presence–absence simulation accuracy. The bioclimatic parameters are kDDMin (minimum degree-day requirement), kWiTN (minimum winter temperature tolerance), kWiTX, (maximum winter temperature tolerance), and kDrTol (drought tolerance). Black dots represent the 100 samples chosen from the range of probabilities for each bioclimatic parameter (see also Fig. 4). The squares are the species-specific bioclimatic parameters taken from Bugmann & Solomon (2000) and triangles are the species-specific optimized parameters. The model was run with all bioclimatic parameter sets and evaluated for accuracy, using the Kappa statistic. Horizontal lines indicates values of the Kappa statistics from which accuracy can be regarded as fair (Kappa > 0.4) to good (Kappa > 0.7).
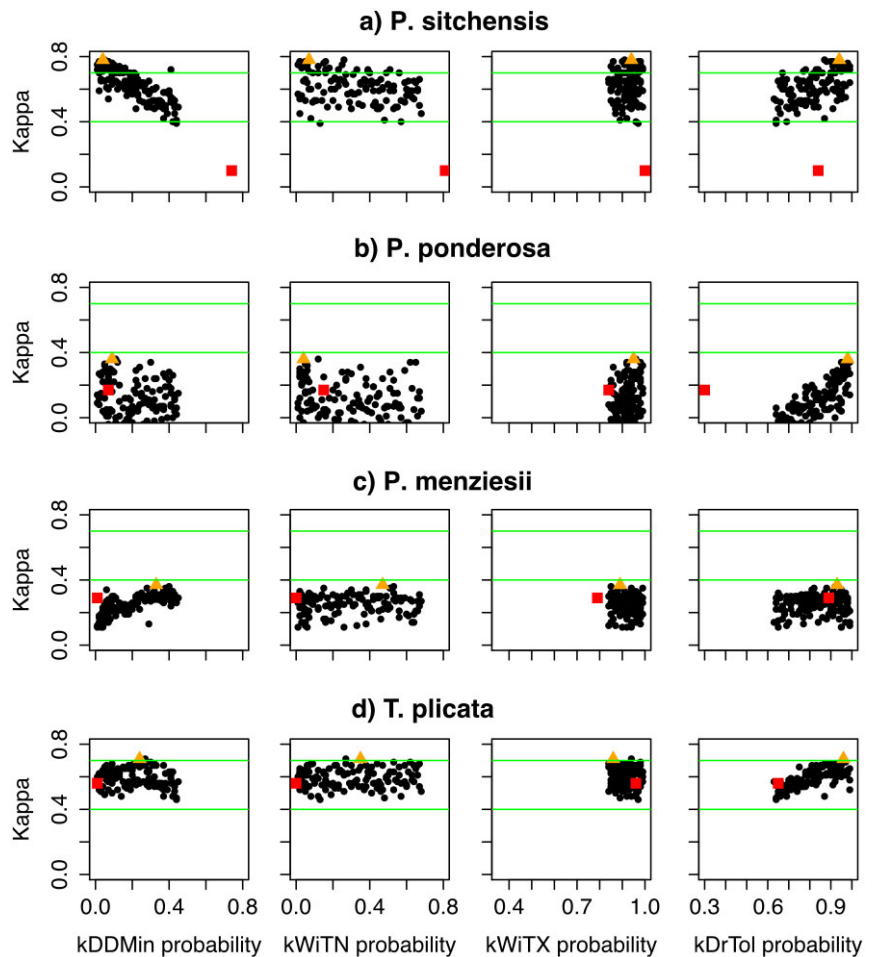
Figure 4 Density plots of bioclimatic parameter probabilities obtained for 18 tree species of the Pacific Northwest of North America. Probabilities were determined based on bioclimatic information taken from the entire tree species range (see Methods). 'Optimized' refers to the parameters that maximized the accuracy of presence–absence simulations at a regional scale. 'Local' parameters are those determined by Bugmann & Solomon (2000) along a transect at 44.13° N latitude. Optimized parameter values are shown on the x-axis as tick marks. Note the different scales on the y-axis.
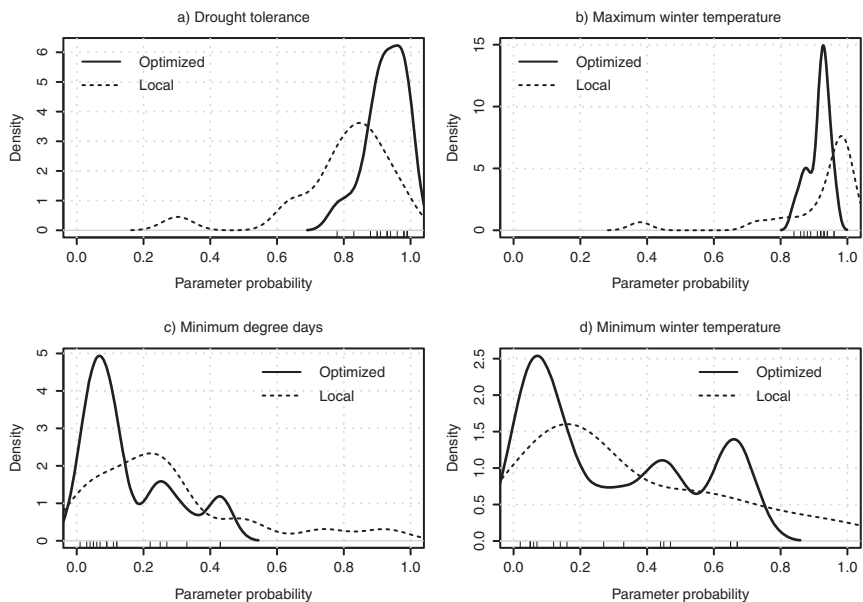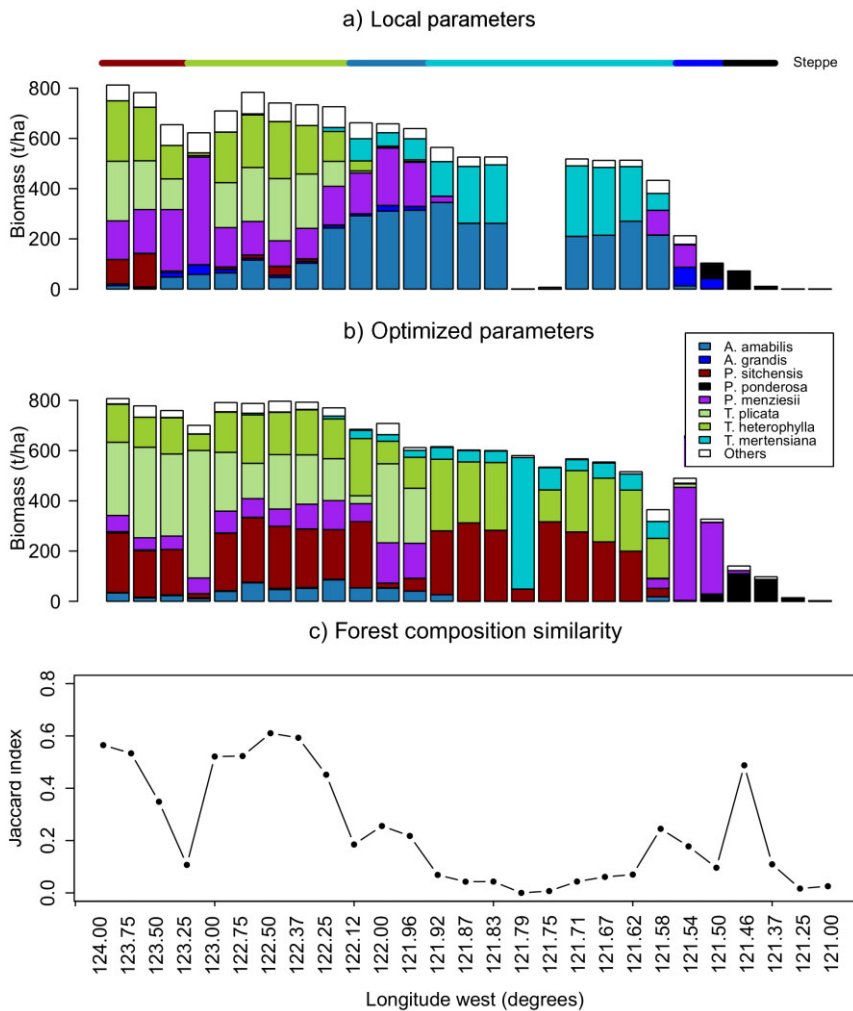
Figure 5 Steady-state forest composition simulated by ForClim along an environmental gradient at latitude 44.13° N in the Pacific Northwest of North America. The transect starts at the Pacific coast (left) and extends to the steppe in Oregon (right). The horizontal bar at the top illustrates the forest composition zones described by Franklin & Dyrness (1988) and follow the shades indicated in the legend. (a) Simulations run using the local parameter set (Bugmann & Solomon, 2000). (b) Simulations run using the optimized parameter set (this study). (c) Jaccard index of compositional similarity between simulated sites.

Below, we discuss the importance of: (1) parameterization schemes, (2) potentially missing processes in DVMs, (3) species-specific ecological characteristics that are likely to matter in this context, and (4) the strengths and limitations of using DVMs to simulate patterns and processes across spatial scales, i.e. from local to regional scales.

Bioclimatic parameterization

Our results emphasize the importance of testing the generality of parameterizations prior to using DVMs to predict species distributions. From our analysis, we found that the probabilities to determine the bioclimatic limits that best represent the regional species distribution commonly differed from the probabilities of the local parameters (Figs 1, 3, & 4). The local parameters in our study had originally been obtained by determining the climatic limits of tree species at their elevational range boundaries within Oregon and adjacent areas (cf. Bugmann & Solomon, 2000). Our results suggest that this calibration method does not capture the climatic information needed to determine range boundaries of tree species at the regional scale. In addition, the low accuracy of simulated species distributions when using the local parameters (Table 1) suggests that parameters calibrated in a small area may not be useful in other regions; this conforms to caveats regarding the application of correlative species distribution models (cf. Araújo & Guisan, 2006; Randin *et al.*, 2006).

The regionally optimized parameters were able to distinguish species that differed only slightly in their extreme climatic tolerances. For example, some tree species had the same drought tolerance according to the local parameter set (0.2 for *Chamaecyparis nootkatensis*, *T. heterophylla* and *T. plicata*), whereas the optimization identified markedly different drought tolerance values for these species (ranging from 0.21 to 0.29). Similarly, our method revealed minimum degree-day requirements that were offset by 200 degree-days for species with the same requirement in the local parameter set. Parameter changes such as these were essential for improving the accuracy of simulated tree species distributions (Table 1).

On the one hand these results demonstrate that locally and qualitatively calibrated DVMs are likely to require an adjustment of their parameters if they are applied to predict regional species distributions. On the other hand, the disparity of the parameter sets to simulate forest composition at local scales

(Fig. 5) suggests that regionally optimized parameters may not be appropriate for quantitatively assessing tree species abundance (e.g. biomass) at this scale.

Ecological processes potentially missing in the model

Our parameterization scheme relied on the assumption that a species' distribution is in equilibrium with the current climate (Guisan & Zimmermann, 2000; Svenning & Sandel, 2013) and that a species' presence/absence reflects the interaction between bioclimatic limits, competition and demography. These assumptions appear to be correct for eight of the studied species, i.e. those with the highest level of accuracy in the simulation of their distribution (Kappa > 0.5; Table 1).

However, many species were predicted to occur beyond their northern distribution limit (cf. Fig. 1e, Appendix S2), a commission error common to both ForClim and the SDM predictions (cf. Fig. 2c, Appendix S3). This suggests that the current northern limits of these species may not be in thermal equilibrium, or alternatively that current species distributions are a reflection of post-glacial migration lags (as neither the SDM nor ForClim include dispersal limitations). Similar to some European tree species (Svenning & Skov, 2004; Dullinger *et al.*, 2012; Randin *et al.*, 2013), pollen records from the PNW suggest that some conifers are still expanding northwards (Xie & Ying, 1994; Elias, 2013), and pollen records indicate a larger range prior to the Last Glacial Maximum (Hansen, 1941). Environmental factors (e.g. soil moisture) or biotic interactions (e.g. competition) could help explain the slow spread of PNW conifers towards their current northern limits (Elias, 2013). Most DVMs are based on the assumption of unlimited seed dispersal (cf. Snell *et al.*, 2014). Clearly, seed dispersal processes need to be included if one intends to use DVMs for predicting future tree range shifts.

Establishment in new areas is another crucial process that may be limiting the rate of range expansion. Although ForClim simulates tree establishment, it does so in a highly simplified manner, not capturing the microsite conditions upon which some tree species rely. For example, *Picea engelmannii* and *A. lasiocarpa* seeds have increased survival at the tree line ecotone due to favourable climatic and soil microsites (Germino *et al.*, 2002). *Pinus contorta* relies on fine-scale (< 500 m^2) habitat heterogeneity throughout its range, as it is specifically adapted to poor soils and areas where summer frost events preclude the establishment of many other species (McKenzie *et al.*, 2003). Other processes such as disturbance, land-use changes, human influence and other biotic interactions can further cause species distributions to be in disequilibrium with climate (Guisan & Zimmermann, 2000; Svenning & Sandel, 2013). The joint effects of these processes, dispersal limitation and establishment success are likely to influence the realized distribution of a species and to modify species pools at the regional scale.

Furthermore, tree species with a wide distribution range may have intrinsic variability in their climatic tolerance from one region to another because populations are genetically differentiated or narrowly adapted to their local environments (Morgenstern, 1996). For example, *Picea sitchensis* is distributed over 22° of latitude in the PNW, and has genetic clines that confer adaptation to local climate and enable competitive annual growth (Mimura & Aitken, 2007; Holliday *et al.*, 2010). Such trait variability is not captured by our optimized parameter scheme, which is based on the assumption of uniform parameter values across a species' range. Genetic variability can lead to potentially wider bioclimatic limits for a species. For example, the strong overprediction for *P. sitchensis* at the local scale using the regionally optimized parameters (Fig. 5) does not conform to the narrow dependence of this species on fog-modulated coastal environments at this latitude in the PNW. Thus, there may be a need to represent species as multiple subspecies that differ in their parameterization, if the model is applied to predict species distributions across spatial scales.

Ecological characteristics of tree species

Our results suggest that certain species traits, as well as experimental design, influence the ability of a DVM to accurately predict regional species distributions. Specifically, the relationship with shade tolerance indicates that the model is well suited to simulate late-successional tree species distributions in undisturbed forests. The three species with the highest accuracy (*P. sitchensis*, *T. heterophylla* and *T. plicata*; Table 1) are all late-successional, shade-tolerant trees that persist for a long time in undisturbed landscapes (Franklin & Hemstrom, 1981). In contrast, we generally found lower levels of accuracy for pioneer, shade-intolerant species (e.g. *Pinus ponderosa*, *P. contorta* and also *Pseudotsuga menziesii*, cf. Table 1). An absence of external disturbances in our simulations may explain the underprediction of pioneer, fire-tolerant species, such as *Pinus ponderosa* and *Quercus garryana*, in areas with historically high fire frequencies (Franklin & Hemstrom, 1981). Although fires and other stand-replacing disturbances can be included in ForClim via predictive parameters (Busing *et al.*, 2007; Rasche *et al.*, 2011), they were not included in the current study due to the difficulty of defining the heterogeneity of disturbance regimes across the study area. An obvious example of a species that would have benefited from the inclusion of fire is *P. contorta* (Franklin & Hemstrom, 1981). The SDM was able to capture the distribution of *P. contorta* without including fire; in ForClim, however, this highly shade-intolerant species was out-competed by later-successional species. In this case, the bioclimatic envelope approach (SDM) produced a more accurate distribution, although for the wrong reasons. Ignoring the interaction with fire would most likely cause severe errors under future climate scenarios.

In contrast, some late-successional species were overpredicted by ForClim towards the southern edge of their ranges (e.g. *T. mertensiana*), possibly due to increased competition by southern species not included in our set and thus a rather weak effect of climate limitations (Ettinger *et al.*, 2011). Finally, subdominant tree species (e.g. *A. amabilis*) tend to occur in a patchy manner within their distribution range, making it difficult to capture their actual distribution with a DVM. Overall, to accurately reflect species ranges the inclusion of additional processes

such as disturbances in combination with improved bioclimatic parameters may be critical for reducing biases in DVM predictions (cf. Brotons *et al.*, 2004; Snell *et al.*, 2014).

Pattern and processes across spatial scales

Predicting regional species distributions requires models that capture regional patterns while maintaining key interactions and ecological processes that occur at small scales. Although forest gap models were not specifically developed for predicting regional species distributions, we showed that they have a high potential in this regard (Table 1), provided that one is willing to sacrifice local-scale detail and current limitations due to parameterization schemes are addressed. Importantly, predictions of regional species distributions using a DVM consider the impact of competition, demography and the influence of climate and environmental factors at the level of the individual tree; this is a distinct advantage over correlative SDMs. Although in our example the DVM did not perform better than the SDM for presence–absence predictions of the current distribution, DVMs have distinct advantages over SDMs when it comes to predicting species abundances or range shifts in response to climate change (provided that dispersal and migration processes are integrated appropriately with the other demographic processes in DVMs).

Our simulation experiments further illustrate possible sources of uncertainty (cf. Ford *et al.*, 2013) in the predictions of species distributions across spatial scales using DVMs. Since it is unlikely that the *true* climatic tolerances of a particular species differ across spatial scales, our results may also reflect uncertainty in the climate data and/or species distribution maps, as discussed below.

Local-scale simulations of forest dynamics typically use climate data from individual weather stations (i.e. instrumental data); we had to use gridded (i.e. modelled) climatic data at both scales of analysis. The gridding procedure may conceal important local-scale climatic features that may be important for species distributions towards the range edges, where many species occur in favourable microsites only, and not under the 'generally prevailing' climate. Additionally, a finer resolution and higher accuracy of inferred soil properties (i.e. water holding capacity and nitrogen availability, $< 1\,km^2$) could improve predictions of regional species distribution in both the DVM and the SDM, especially at sites where precipitation is limiting.

Model uncertainty may also be due to the relatively coarse level of information in the species distribution data. On the one hand the reference distribution maps indicate presence only (not abundance). This makes it difficult to extract relationships between climatic factors and demographic processes, as all we have are the absolute limits to their distribution. On the other hand the Little (1971) maps are spatially contiguous, which is a clear advantage for regional-scale analyses, but the quality of the underlying data (or interpretations) is difficult to appreciate.

CONCLUSIONS

The ability of the dynamic vegetation model ForClim to predict regional species distributions corroborates that the DVM framework is an appropriate, although perhaps incomplete, representation of the processes involved in determining species ranges. The application of forest gap models for predicting species distributions requires the recognition of ecological processes that may cause disequilibrium with climate, such as dispersal limitations, migration lags and disturbances.

Despite of these caveats, it is important to appreciate that dynamic forest models include the effects of individual-based inter- and intraspecific competition as modified by climatic variability, and they further capture these fine-scale dynamics in a simple framework that is efficient to run at regional scales and relatively straightforward to analyse and interpret. These are clear advantages over purely correlative species distribution models. Future applications of gap models to understand and predict species distributions should examine how the strengths of biotic interactions are influenced by climate and disturbance, together with the spatially explicit modelling of key ecological processes such as dispersal. Including additional processes may be relevant when predicting range shifts (e.g. as a result of changing fire regimes due to climate change), but less important for predicting current species distributions. To face the challenges of predicting range dynamics across multiple spatial scales, future efforts should focus on assessing the relevance and the spatial scale at which ecological process operate. This is crucial if we intend to use DVMs to predict how species distributions and vegetation dynamics will be affected by future climate change.

ACKNOWLEDGEMENTS

REFERENCES

Allouche, O., Tsoar, A. & Kadmon, R. (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

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

Brotons, L., Thuiller, W., Araújo, M.B. & Hirzel, A.H. (2004) Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*, **27**, 437–448.

Bugmann, H. (1996) A simplified forest model to study species composition along climate gradients. *Ecology*, **77**, 2055–2074.

Bugmann, H. (2001) A review of forest gap models. *Climatic Change*, **51**, 259–305.

Bugmann, H. & Cramer, W. (1998) Improving the behaviour of forest gap models along drought gradients. *Forest Ecology and Management*, **103**, 247–263.

Bugmann, H.K.M. & Solomon, A.M. (1995) The use of a European Forest model in North America: a study of ecosystem response to climate gradients. *Journal of Biogeography*, **22**, 477–484.

Bugmann, H.K.M. & Solomon, A.M. (2000) Explaining forest composition and biomass across multiple biogeographical regions. *Ecological Applications*, **10**, 95–114.

Bugmann, H.K.M., Wullschleger, S.D., Price, D.T., Ogle, K., Clark, D.F. & Solomon, A.M. (2001) Comparing the performance of forest gap models in North America. *Climatic Change*, **51**, 349–388.

Busing, R.T., Solomon, A.M., McKane, R.B. & Burdick, C.A. (2007) Forest dynamics in Oregon landscapes: evaluation and application of an individual-based model. *Ecological Applications*, **17**, 1967–1981.

Chave, J. (2013) The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters*, **16**, 4–16.

Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**, 37–46.

Dormann, C.F., Schymanski, S.J., Cabral, J., Chuine, I., Graham, C., Hartig, F., Kearney, M., Morin, X., Römermann, C., Schröder, B. & Singer, A. (2012) Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.

Dullinger, S., Willner, W., Plutzar, C., Englisch, T., Schratt-Ehrendorfer, L., Moser, D., Ertl, S., Essl, F. & Niklfeld, H. (2012) Post-glacial migration lag restricts range filling of plants in the European Alps. *Global Ecology and Biogeography*, **21**, 829–840.

Elias, S.A. (2013) The problem of conifer species migration lag in the Pacific Northwest region since the last glaciation. *Quaternary Science Reviews*, **77**, 55–69.

Ettinger, A.K., Ford, K.R. & HilleRisLambers, J. (2011) Climate determines upper, but not lower, altitudinal range limits of Pacific Northwest conifers. *Ecology*, **92**, 1323–1331.

Ford, K.R., Ettinger, A.K., Lundquist, J.D., Raleigh, M.S. & HilleRisLambers, J. (2013) Spatial heterogeneity in ecologically important climate variables at coarse and fine scales in a high-snow mountain landscape. *PLoS ONE*, **8**, e65008.

Franklin, J.F. & Dyrness, C.T. (1988) *Natural vegetation of Oregon and Washington*. USDA Forest Service, Pacific Northwest Forest and Range Experiment Station, Portland, OR.

Franklin, J.F. & Hemstrom, M.A. (1981) Aspects of succession in the coniferous forests of the Pacific Northwest. *Forest succession* (ed. by D.C. West, H.H. Shugart and D.B. Botkin), pp. 212–229. Springer, New York.

García-Valdés, R., Zavala, M.A., Araújo, M.B. & Purves, D.W. (2013) Chasing a moving target: projecting climate change-induced shifts in non-equilibrial tree species distributions. *Journal of Ecology*, **101**, 441–453.

Gaston, K.J. (2003) *The structure and dynamics of geographic ranges*. Oxford University Press, Oxford.

Germino, M.J., Smith, W.K. & Resor, A.C. (2002) Conifer seedling distribution and survival in an alpine-treeline ecotone. *Plant Ecology*, **162**, 157–168.

Goodman, L.A. & Kruskal, W.H. (1954) Measures of association for cross classifications. *Journal of the American Statistical Association*, **49**, 732–764.

Grimm, V. & Railsback, S.F. (2005) *Individual-based modeling and ecology*. Princeton University Press, Princeton, NJ.

Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.

Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.

Guisan, A., Zimmermann, N.E., Elith, J., Graham, C.H., Phillips, S. & Peterson, A.T. (2007) What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs*, **77**, 615–630.

Hansen, H.P. (1941) Paleoecology of a bog in the spruce–hemlock climax of the Olympic Peninsula. *American Midland Naturalist*, **25**, 290–297.

Hartig, F., Dyke, J., Hickler, T., Higgins, S.I., O'Hara, R.B., Scheiter, S. & Huth, A. (2012) Connecting dynamic vegetation models to data – an inverse perspective. *Journal of Biogeography*, **39**, 2240–2252.

Hijmans, R., Cameron, S., Parra, J., Jones, P. & Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Holliday, J.A., Ritland, K. & Aitken, S.N. (2010) Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist*, **188**, 501–514.

Kooperberg, C. & Stone, C.J. (1992) Logspline density estimation for censored data. *Journal of Computational and Graphical Statistics*, **1**, 301–328.

Levin, S.A. (1992) The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. *Ecology*, **73**, 1943–1967.

Lischke, H., Zimmermann, N.E., Bolliger, J., Rickebusch, S. & Löffler, T.J. (2006) TreeMig: a forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. *Ecological Modelling*, **199**, 409–420.

Little, E.L. (1971) *Atlas of United States trees. Volume 1, conifers and important hardwoods*. Miscellaneous Publication 1146. US Department of Agriculture, Washington, DC.

Loehle, C. & LeBlanc, D. (1996) Model-based assessments of climate change effects on forests: a critical review. *Ecological Modelling*, **90**, 1–31.

MacArthur, R.H. (1972) *Geographical ecology: patterns in the distribution of species*. Princeton University Press, Princeton, NJ.

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**, 1093–1108.

McRoberts, R.E., Bechtold, W.A., Patterson, P.L., Scott, C.T. & Reams, G.A. (2005) The enhanced forest inventory and analysis program of the USDA Forest Service: historical perspective and announcement of statistical documentation. *Journal of Forestry*, **103**, 304–308.

Mimura, M. & Aitken, S.N. (2007) Adaptive gradients and isolation-by-distance with postglacial migration in *Picea sitchensis*. *Heredity*, **99**, 224–232.

Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, **62**, 275–293.

Morgenstern, E.K. (1996) *Geographic variation in forest trees: genetic basis and application of knowledge in silviculture*. UBC Press, Vancouver.

R Core Team (2013) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M. & Guisan, A. (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography*, **33**, 1689–1703.

Randin, C.F., Paulsen, J., Vitasse, Y., Kollas, C., Wohlgemuth, T., Zimmermann, N.E. & Körner, C. (2013) Do the elevational limits of deciduous tree species match their thermal latitudinal limits?: elevational and latitudinal limits of deciduous trees. *Global Ecology and Biogeography*, **22**, 913–923.

Rasche, L., Fahse, L., Zingg, A. & Bugmann, H. (2011) Getting a virtual forester fit for the challenge of climatic change. *Journal of Applied Ecology*, **48**, 1174–1186.

Rasche, L., Fahse, L., Zingg, A. & Bugmann, H. (2012) Enhancing gap model accuracy by modeling dynamic height growth and dynamic maximum tree height. *Ecological Modelling*, **232**, 133–143.

Schenk, H.J. (1996) Modeling the effects of temperature on growth and persistence of tree species: a critical review of tree population models. *Ecological Modelling*, **92**, 1–32.

Snell, R.S., Huth, A., Nabel, J.E.M.S., Bocedi, G., Travis, J.M.J., Gravel, D., Bugmann, H., Gutiérrez, A.G., Hickler, T., Higgins, S.I., Reineking, B., Scherstjanoi, M., Zurbriggen, N. & Lischke, H. (2014) Using dynamic vegetation models to simulate plant range shifts. *Ecography*, **37**, 1184–1197.

Stein, M. (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, **29**, 143–151.

Svenning, J.-C. & Sandel, B. (2013) Disequilibrium vegetation dynamics under future climate change. *American Journal of Botany*, **100**, 1266–1286.

Svenning, J.-C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.

Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.

Xie, C.Y. & Ying, C.C. (1994) Adaptedness of noble fir (*Abies procera* Rehd.) beyond its northern limit. *Forest Science*, **40**, 412–428.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Additional information about the modelling methodology.

Appendix S2 Predictive distribution maps using a dynamic forest gap model for the remaining 15 tree species.

Appendix S3 Predictive distribution maps using a correlative species distribution model for the remaining 15 tree species.

BIOSKETCHES

Alvaro G. Gutiérrez is a forest ecologist whose main research interests are in dynamics, impacts of global change and biogeography of temperate rain forests.

Rebecca Snell is a plant ecologist interested in understanding landscape to regional-level distribution patterns, and the processes that influence species composition, structure and dynamics.

Harald Bugmann is the leader of the Forest Ecology Group at ETH Zurich. His research focuses on the impacts of global change, and particularly climate change, on long-term forest dynamics and adaptive forest management.

Authors contributions: A.G.G. and H.B. conceived the ideas; A.G.G. performed the model runs, analysed the data and drafted the paper; R.S. conceived sensitivity analysis and interpreted results; all authors strongly contributed to writing the paper.

Editor: John-Arvid Grytnes