



Challenges in modelling the abundance of 105 tree species in eastern North America using climate, edaphic, and topographic variables

Dominic Chambers^a, Catherine Périé^b, Nicolas Casajus^c, Sylvie de Blois^{a,*}

^a McGill School of Environment and Department of Plant Science, Macdonald Campus of McGill University, 2111 Lakeshore, Sainte-Anne-de-Bellevue, Canada H9X 3V9

^b Direction de la recherche forestière, Ministère des Ressources naturelles et de la Faune, 2700 rue Einstein, Québec City, Canada G1P 3W8

^c Conservation of Northern Ecosystems and Centre of Nordic Studies, Quebec University at Rimouski, 300 allée des Ursulines, Rimouski, Canada G5L 3A1

ARTICLE INFO

Article history:

Received 26 July 2012

Received in revised form 26 September 2012

Accepted 27 October 2012

Keywords:

Climate change

Eastern North America

Predictive performance

Random Forest

Species abundance model

Temperate and boreal forests

ABSTRACT

Improving predictions of the location of suitable environmental conditions for species using species distribution models (SDM) is at the core of biodiversity/climate change research, but modelling species abundance, rather than distribution, is proving particularly challenging. Using data from more than 200,000 forest plots in eastern North America and Random Forest, we evaluated the performance of species abundance models (SAM) in predicting the relative abundance (measured as importance value) of each of 105 tree species in relation to climate, edaphic, and topographic variables. We calculated the coefficient of determination (R^2_{SAM}) between observed and predicted abundances as a measure of model performance for each species. We also performed multiple linear regressions to explain variation of R^2_{SAM} among species using five biogeographical or spatial attributes of species as explanatory variables. Predictive performances of SAM (R^2_{SAM}) were generally low, ranging from 0.016 to 0.815 (mean = 0.258). Black spruce (*Picea mariana*) had the best predictive model and Florida maple (*Acer barbatum*) and American chestnut (*Castanea dentata*) the worst. Thirty-seven of the 41 best performing species ($R^2_{SAM} \geq 0.3$) had climate ranked as the best and/or second best predictor. Species with the best performance tended to be those that could reach dominance, showed aggregated distribution of abundance, and/or had high latitudinal limits in the study area. Climate change is likely to affect patterns of dominance in communities by altering patterns of co-occurrences, but for many species that constitute the bulk of tree diversity, predictions based solely on the current distribution of relative abundances may not be reliable enough to inform conservation or management decisions. Predicting tree abundance in a warming climate using SAM remains a challenge, but it is only by reporting performances across a range of climate and statistical models, regions and species, as well as by highlighting model limitations and strengths, that we will improve the reliability of predictions and in turn better inform forest conservation and management.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The need to forecast potential changes in species distribution in response to climate warming and other environmental changes has stimulated the development of modelling approaches that use current geographic distribution to estimate suitable environmental conditions for a given species (Araújo et al., 2004; McKenney et al., 2007, 2011; Thuiller, 2003). Based on occurrence data, species distribution models (SDM) have generally reported moderate to high accuracy (Area Under the ROC Curve ≥ 0.60) in reproducing the current distribution of trees (McKenney et al., 2007), birds (Brotons et al., 2004; Hu et al., 2010), or amphibians (Lawler et al., 2010). This has increased confidence that SDM can be useful

for mapping biodiversity, assessing climate risk, or informing resource management (Millar et al., 2007). Improving confidence in predictions however, requires an understanding of the sources of uncertainty in SDM (Buisson et al., 2010). One of these sources is the modelling approach used; this has led to the development of consensus approaches where predictions are combined across a range of statistical models (Araújo and New, 2007). When different species are modelled using the same statistical approach, differences in accuracy among predictions for the same set of predictors have been attributed to the spatial, geographical, and/or biological attributes of the species themselves (Hanspach et al., 2011; Luoto et al., 2005; Marmion et al., 2009; Syphard and Franklin, 2010). The fact that SDM have been often criticised for not implicitly taking into account processes such as dispersal has not limited their development as an active area of research in ecology, conservation, and management.

* Corresponding author.

E-mail address: sylvie.deblois@mcgill.ca (S. de Blois).

Improving predictions of the potential location of species is important, but modelling species abundance, rather than occurrence, may prove particularly challenging (Anadón et al., 2010; Iversen et al., 2008; Pearce and Ferrier, 2001). Yet knowing about the relative abundance of a species in a community or landscape is crucial as abundance patterns determine a species' influence on other species or ecosystem processes, its conservation status, or its use as a resource. Information on the presence of a species at a site may mask patterns of low abundance or, on the contrary, patterns of very high abundance in populations of core importance to the species. In spite of recognising the importance of monitoring changes in abundance patterns for biodiversity research (McMahon et al., 2011), there are still relatively few studies addressing patterns of species abundances over broad geographical extent in relation to climate and other environmental gradients, and the available empirical observations show that the spatial distribution of abundances is more complex than previously acknowledged (Sagarin et al., 2006). Climatic suitability, as determined by SDM, positively correlates with abundance for some species (VanDerWal et al., 2009), suggesting that processes like dispersal or physiological constraints on establishment that determine a species' occurrence at a site also determine to some extent its abundance. Species abundance models (hereafter called SAM), however, have lagged behind SDM, partly because of the scarcity of good abundance data across large spatial extent to allow testing environmental drivers of species' abundance for a range of species.

When SAM have been used with climate data, they have produced mixed results. SAM have been apparently successful in modelling bird abundance in Africa in relation to climatic variables (Huntley et al., 2011), but less so for trees in North America (Canham and Thomas, 2010; Iversen et al., 2008); their performance may vary with taxa or regions. For trees, failure to incorporate edaphic variables along with climatic variables or failure to capture the northern limit of species (Canham and Thomas, 2010), or the lack of confidence in models built with only a portion of the species' range (Iversen et al., 2008), have all been invoked to explain poor performance of abundance models. It remains unclear whether predictive performance would be improved by extending datasets to include latitudinal limits or more environmental predictors. More importantly, it is unclear whether or how tree abundance patterns vary with climate at broad spatial scale (Canham and Thomas, 2010; Stegen et al., 2011). For animal species, examination of the relationship between climatic suitability predicted from presence/absence data and abundance showed that sites with low suitability consistently had low abundance, whereas suitability predicted the upper limit of abundance better than mean abundance (VanDerWal et al., 2009). Community processes such as competition can reduce species abundance at a site, but ecological theory also predicts that environmental conditions will have an impact on competitive outcomes and therefore on species abundance. All this suggests that species that can reach dominance in part of their range would be better modelled by SAM. There have been generally few attempts at explaining modelling variations in the predictive performance of SAM to identify the attributes of species that perform best or, conversely, of those that perform poorly.

To understand the relationship between tree abundance patterns and environmental predictors in a context of environmental changes, we tested how well we could predict the relative abundance of 105 tree species using abundance data from more than 200,000 forest plots across eastern North America in relation to climatic, edaphic, and topographic variables. We integrated ecological information across eastern Canada and the USA, thereby capturing the northern limit of tree species. We used the statistical modelling approach, Random Forest, that consistently performed best in exploratory analyses using a range of statistical models.

Finally, we went a step further by explaining variations in SAM performances using biogeographical and spatial attributes of species as explanatory variables to facilitate generalizations across species and regions.

2. Materials and methods

2.1. Study area

This study is part of a larger research project known as CC-Bio for "Effect of Climate Change on Quebec Biodiversity" (Berteaux et al., 2010). The focus of CC-Bio is on the impacts of climate change on the biodiversity of the province of Quebec, Canada. The extent of the study area, however, was defined to take into account the expected shift north of climatic envelope ($\sim +5^\circ\text{C}$ isotherm) as well as available data on species, climate, and edaphic conditions. The study area ranges in latitude from $30^\circ30'\text{N}$ to $53^\circ00'\text{N}$ and in longitude from $93^\circ00'\text{W}$ to $60^\circ30'\text{W}$ covering an area of more than 2,567,000 km^2 (Fig. 1).

Because of its size, the study area is characterized by strong climatic, geological, and topographical variability. It contains several climatic zones ranging from temperate in most of its range, to arid in the southwest of the Great Basin, to a maritime climate along the Atlantic coast. Annual mean temperature increases gradually from -5°C in the north to 20°C in the south, whereas annual total precipitation ranges from 670 to 2000 mm (USDA Forest Service, 2010).

Physiographic regions include the Canadian Shield (Laurentian Upland, North), the Great Lake plains, the Appalachian Highlands (North East), the Atlantic Plains (East), and the Interior Plains (West). The elevation ranges from sea level to 1250 m (mean = 300 m, SD = 165 m) (Canadian Council on Geomatics, 2003; USGS, 2010).

The study area was covered by a grid that contains 6418 cells of 400 km^2 ($20\text{ km} \times 20\text{ km}$) each. Each cell was considered as a sampling unit and contained information on tree abundance, climatic (temperature and precipitation), topographic (elevation) and edaphic variables (drainage and surface deposit). The grid in Canada was a convenient extension of the grid used by Prasad et al. (2006) in eastern USA, making it possible to merge datasets from different jurisdictions (see details below). Although climatic data can be obtained at finer or coarser scales, the cell size that we used was considered adequate to capture broad climatic variations over such a large study area.

2.2. Data and analysis

There were two main steps in the analysis (Fig. 2):

- Step (1) model tree abundance, measured as importance value in a cell (the response variable), for each species ($n = 105$) using climatic, edaphic, and topographic data as explanatory variables. Calculate a measure of predictive performance (R_{SAM}^2) for each modelled species.
- Step (2) explain the variation among species in the predictive performance of SAM by conducting a Multiple Linear Regression (MLR) using R_{SAM}^2 from step 1 as the response variable and five biogeographical and spatial attributes describing each species as explanatory variables.

2.2.1. Step 1: model tree abundance for each species

2.2.1.1. Data collection.

2.2.1.1.1. Tree importance value. The importance value (IV) was used as a measure of the relative density and dominance of each tree species ($n = 105$). In the American portion of the study area, IV for a given species was obtained online from the USDA Forest



Fig. 1. The CC-Bio study area covering more than 2,500,000 km².

Service tree Atlas website (Prasad and Iverson, 2003). The IVs were originally compiled from more than 100,000 forest inventory and analysis plots assessed by the USDA Forest Service (Iverson and Prasad, 1998). In the Canadian portion of the study area, IVs were calculated from an extensive dataset of >95,000 forest plots of the 3rd decennial inventory of the Ministère des Ressources naturelles et de la Faune (MRNF), Québec, using the same approach as in the USDA dataset and in other forest studies (Attorre et al., 2008; Murphy et al., 2010). IVs for each species in each plot were calculated according to the following formula (Maximum IV value = 100):

$$IV(i) = 100 \cdot \left(\frac{\frac{BA(i)}{BA(all\ species)} + \frac{NS(i)}{NS(all\ species)}}{2} \right) \quad (1)$$

where i is a particular species in a plot, BA is basal area, and NS is number of stems. IVs calculated at the plot level were then averaged over the 400 km² cell for each species. When cells overlapped both countries, a weighted average IV was calculated based on the percentage of occupancy in each jurisdiction. For $0 < IV < 1$, the IV was rounded to 1. The minimum number of cells occupied by a

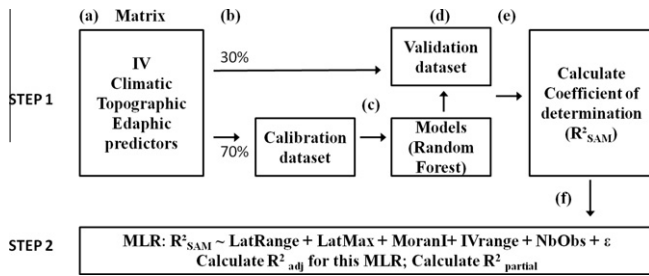


Fig. 2. The two-step analytical approach used to build and evaluate species abundance (importance value; IV) models for 105 species. Step 1: (a) for each species, we start with a matrix of species IV with climatic, topographic, and edaphic predictors (for detailed description of predictors, see Table 1) in each cell. (b) The matrix is randomly split into a calibration set (70% of the original matrix) and an evaluation set (30%). This process is repeated 60 times to reduce potential biases due to random selection. (c) Random Forest is used each time as the statistical algorithm for model calibration. (d) The performance of Random Forest is then tested against each of the 60 validation datasets (e) so 60 coefficients of determination (R^2_{SAM}) are calculated for each species and averaged. STEP 2: (f) To explain the variation of these R^2_{SAM} among species, multiple linear regression (MLR) is performed using as predictors the biogeographical and spatial attributes of a given species. The percentage of the variation explained by the full model (R^2_{adj}) is calculated as well as ($R^2_{partial}$) to determine the contribution of each biogeographical and spatial attributes to the MLR. (IVRange = importance value range, LatMax = latitudinal maximum, LatRange = latitudinal range, NbObs = Number of observation, Moran I = spatial aggregation).

Table 1
Predictor variables used in species abundance models (SAM) in step 1.

| Climate | Surface deposit (%) | Drainage (%) | Topographic |
|---|---|--------------------|-----------------------|
| Annual mean temperature (°C) | Eolian deposit | Humid water regime | Average elevation (m) |
| Annual total precipitation (mm) | Fluvio-glacial deposit | Mesic water regime | |
| Ratio of summer precipitation over annual precipitation | Littoral, marine, or lacustrine deposit | Xeric water regime | |
| | Glacial deposit | | |
| | Organic deposit | | |
| | Rocky substrate | | |
| | Slope or altered deposit | | |

species was 59 and the maximum was 3442 cells. Only cells with IV > 0 (i.e., where the species was present) were considered for analysis.

2.2.1.1.2. Climatic variables. For climate data, we used monthly long term averages for the period 1961–1990, a period also coinciding with forest surveys. Raw climatic surfaces for maximum, minimum and average temperature, as well as total precipitation, were downloaded from the US Forest Service, Rocky Mountain Research station website (USDA Forest Service, 2010). These data covered the whole study area and were obtained using Anusplin thin-plate splines to interpolate weather station data available from the USA and Canada (Rehfeldt, 2006). They were downloaded with a resolution of 0.0083 decimal degrees (~1 km), and were subsequently averaged for each 400 km² grid cell. For SAM, we selected, using clustering analysis to limit redundancy among explanatory variables, three climatic variables out of 20 variables derived from temperature and precipitation (package Hclust; R Development Core Team, 2009). They were annual total precipitation (mm), annual mean temperature (°C), and the ratio of summer precipitation (monthly precipitations for June to September) over annual total precipitation (Table 1).

2.2.1.1.3. Edaphic variables. Edaphic variables are used to capture suitable combinations of climate and edaphic conditions potentially influencing species abundance. The selected edaphic variables had to be available over the whole study area and at the scale of our grid (400 km²). We included in our models surface deposit (7 classes) and soil drainage conditions (3 classes; Table 1). Surface deposit, a stable characteristic of the landscape, affects soil development and properties and the distribution of forest types, whereas drainage relates to soil texture and topography and is known to limit plant distribution (Burns and Honkala, 1990). For the USA, data were obtained from the Soil Survey Geographic (SSURGO) Database (version 2.1, scale 1:24,000; USGS, 2010). Data are collected from field surveys and photo-interpretation and assigned to map units (polygons) for each state. For Canada, data came from the Ministère des Ressources naturelles et de la Faune, Québec (Forest cover 3rd decennial, scale 1:20,000). Data are obtained from the photo-interpretation of forest polygons. The detailed classification of surface deposits used in both jurisdictions was simplified in consultation with a geomorphologist to retain seven main classes according to the origin of the deposit and the size of particles (Table 1). For drainage, seven similar classes were used by both jurisdictions; they were grouped into three main categories to use in our model (Table 1). For each grid cell, we calculated the percentage of land occupied by each edaphic variable and used these values in our model. This was the best compromise to account for possible variation in edaphic conditions within a cell given that the precise location of USDA plots in the cell was unknown in this study and that tree abundance was averaged within a cell.

2.2.1.1.4. Topographic variable. Elevation was taken from the database used by Iverson et al. (2008) for the USA, whereas for Canada it was obtained from the Canadian Digital Elevation Data (Geobase, 2003) at a resolution of ca. 30 m and averaged to match our grid.

2.2.1.2. Statistical analyses. We related species abundance (importance value) to the environmental predictors using the Random Forest algorithm (Breiman, 2001). In our study, as in others (Kampichler et al., 2010; Prasad et al., 2006), Random Forest was consistently the best performing model in exploratory analyses using a range of algorithms, including generalised additive models, generalised linear models, regression tree analysis, artificial neural network, and generalised boosted models. Random Forest selects a random subset of cells and an independent random number of predictors to build the most complex classification tree (no pruning step) from these predictors. This step is repeated as many times as set by the user (1000 trees in our case) to identify complex relationships among data. The model was calibrated using species abundance, i.e., using only cells with observed IV > 0 as we were interested in modelling variation in species abundance, not presence/absence. Preliminary analysis showed that SAM were efficient at predicting zeros, therefore boosting performance for some species. We randomly picked 70% of the data to calibrate the model and the remaining 30% to assess the predictive performance of Random Forest using the coefficient of determination (R^2_{SAM}), which corresponds to the square of the Pearson correlation coefficient between the observed and the predicted IV values of the validation dataset (Hanspach et al., 2011; Kissling and Gudrum, 2008; Whitton et al., 2012). R^2_{SAM} is only one dimension of predictive performance. Iverson et al. (2008), for instance, have used more complex measures to assess reliability of Random Forest models, but we found their measures (model scores) to be highly correlated with the R^2 they reported for Random Forest. The data splitting step was performed 60 times for each species to reduce potential biases due to data selection. So, for one species, 60 SAM were built

and 60 R_{SAM}^2 were obtained (Fig. 2). $\overline{R_{SAM_i}^2}$, the average of the 60 $R_{SAM_i}^2$, was calculated for each species i and used in subsequent analysis. Modelling was performed using R (Random Forest package; R Development Core Team, 2009).

2.2.2. Step 2: explain variation in predictive performance of SAM

2.2.2.1. Data describing the biogeographical and spatial structure of each species distribution. For each species, we calculated the IV range (IVRange) as the difference between the average of the 10% highest IV values and the average of the 10% lowest IV values. Latitudinal range (LatRange) was calculated the same way by averaging the latitudes for the 10% most northerly and the most southerly cells where the species was present and calculating the range in decimal degrees. The maximum latitude (LatMax) was calculated as the average of the 10% most northerly cells. For each species, the number of observations (NbObs) is the number of cells occupied (IV > 0).

The Moran index (Moran I) calculates the degree of similarity among IV based on their spatial distribution, with Moran I expected to decrease as distance between cells increases (Legendre, 1993). Moran I ranges from -1 (dispersed distribution) to $+1$ (clustered distribution). Although we initially calculated Moran I for all distances, we used the value based only on the 8 nearest neighbour cells where there was most variation among species. Moran I in this study is therefore a measure of spatial aggregation of IV at the landscape scale.

2.2.2.2. Statistical analyses. A Multiple Regression Model (MLR) Eq. (2) was used to explain the variability in $\overline{R_{SAM_i}^2}$ among species using, as predictors, the five variables (IVRange, LatRange, LatMax, NbObs, and Moran I) describing the biogeographical and spatial structure of each species distribution.

$$\overline{R_{SAM_i}^2} = a + b \cdot \text{IVRange}_i + c \cdot \text{LatRange}_i + d \cdot \text{LatMax}_i + e \cdot \text{NbObs}_i + f \cdot \text{MoranI}_i + \varepsilon_i \quad (2)$$

where a, b, \dots, f are unknown parameters, ε_i is the residual random effect ($\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$).

Preliminary analyses also included (1) IV mean, which was not included in the final model as it was highly correlated with IV range, (2) longitudinal range, which was also not retained since it was correlated with maximum latitude and because longitudinal ranges were truncated in our dataset in the northern portion of the study area (Fig. 1), and (3) temperature range, which was highly correlated with latitudinal range. We transformed the variables to obtain a normal distribution (square root for $R_{SAM_i}^2$, LatRange,

LatMax, and Moran I; log transformation for IVRange and NbObs). We used the adjusted squared correlation coefficient (R_{adj}^2) to assess the percentage of variation explained by a full MLR that included all predictors, whereas partial R^2 in the MLR was used to assess the relative contribution of each predictor variable in step-wise selection.

Note for clarity that the latter measures of R^2 (R_{adj}^2 and $R_{partial}^2$) refer to the MLR and should not be confused with the average coefficient of determination calculated for each species in the first step ($\overline{R_{SAM_i}^2}$) as a measure of predictive performance and referred to from now on as $R_{SAM_i}^2$. The MLR was performed with R (R Development Core Team, 2009).

3. Results

3.1. Modelling tree abundance – SAM

The coefficients of determination ($R_{SAM_i}^2, n = 105$) as a measure of predictive performance ranged from 0.016 to 0.815, with a mean of 0.258 (SD = 0.169). Most performances over all model iterations were in the lowest portion of the $R_{SAM_i}^2$ distribution (Fig. 3). Based on the average for 60 iterations, the SAM for black spruce (*P. mariana*, $R_{SAM_i}^2 = 0.815$) had the best predictive performance and the ones for Florida maple (*A. barbatum*, $R_{SAM_i}^2 = 0.016$) and American chestnut (*C. dentata*) the worst (Table 2). Only 41 SAM out of 105 (39%) had $R_{SAM_i}^2 \geq 0.3$, (corresponding to Pearson correlation coefficients between the observed and the predicted IV values ≥ 0.55 ; Table 2). For 33 of these species, a climatic variable was the first predictor according to Random Forest, whereas 8 of the best performing SAM had a variable other than temperature or precipitation as the first predictor (Table S1). These included SAM for *Carya ovata*, *Picea glauca*, *Picea rubens*, *Thuja occidentalis*, *Pinus resinosa*, *Pinus rigida*, *Quercus macrocarpa*, and *Ulmus alata*. The latter four species had non-climatic variables as their first two best predictors (Table S1).

3.2. Explaining the variation in the predictive performance of SAM

The MLR explained 78.6% ($R_{adj}^2 = 0.786$) of the variation in $R_{SAM_i}^2$ among species. Three out of 5 attributes were highly significant and positively related to predictive performance (Table 3). The log of IV Range explained most of the variation ($R_{partial}^2 = 0.506$, $P < 0.001$), followed by the square root of the Moran I ($R_{partial}^2 = 0.180$, $P < 0.001$), and the square root of the latitudinal maximum

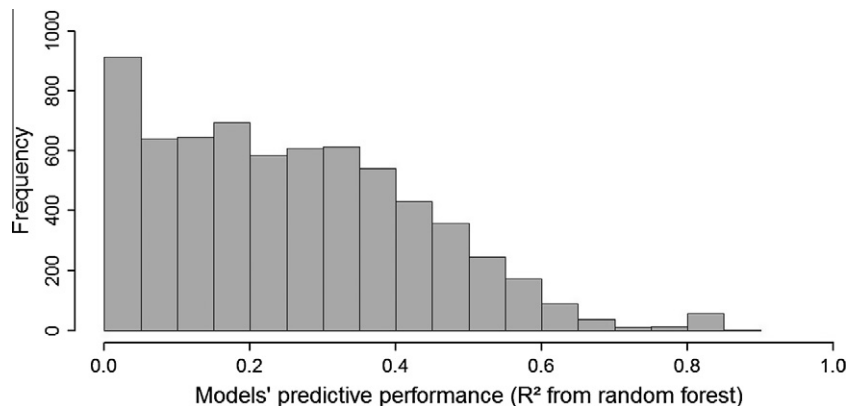


Fig. 3. The frequency distribution of the $R_{SAM_i}^2$ from Random Forest assessing the predictive performance of species abundance models (SAM) (60 iterations for each of 105 tree species = 6300 $R_{SAM_i}^2$).

Table 2

Tree species and R^2_{SAM} averaged over 60 iterations. Species are ordered alphabetically (Latin name). The species ranked “1” has the highest R^2_{SAM} (*Picea mariana*) and “104” the lowest (*Acer barbatum* and *Castanea dentata*).

| English name | Latin | R^2_{SAM} | Rank |
|---------------------|--------------------------------|-------------|------|
| Balsam fir | <i>Abies balsamea</i> | 0.596 | 4 |
| Florida maple | <i>Acer barbatum</i> | 0.016 | 104 |
| Manitoba maple | <i>Acer negundo</i> | 0.192 | 62 |
| Black maple | <i>Acer nigrum</i> | 0.162 | 71 |
| Striped maple | <i>Acer pensylvanicum</i> | 0.298 | 43 |
| Red maple | <i>Acer rubrum</i> | 0.494 | 11 |
| Silver maple | <i>Acer saccharinum</i> | 0.237 | 54 |
| Sugar maple | <i>Acer saccharum</i> | 0.363 | 28 |
| Mountain maple | <i>Acer spicatum</i> | 0.595 | 5 |
| Ohio buckeye | <i>Aesculus glabra</i> | 0.120 | 81 |
| Yellow buckeye | <i>Aesculus octandra</i> | 0.025 | 100 |
| European alder | <i>Alnus glutinosa</i> | 0.076 | 88 |
| Pawpaw | <i>Asimina triloba</i> | 0.026 | 99 |
| Yellow birch | <i>Betula alleghaniensis</i> | 0.449 | 16 |
| Sweet birch | <i>Betula lenta</i> | 0.391 | 22 |
| River birch | <i>Betula nigra</i> | 0.061 | 91 |
| White birch | <i>Betula papyrifera</i> | 0.563 | 6 |
| Gray birch | <i>Betula populifolia</i> | 0.144 | 74 |
| American hornbeam | <i>Carpinus caroliniana</i> | 0.051 | 95 |
| Bitternut hickory | <i>Carya cordiformis</i> | 0.180 | 65 |
| Pignut hickory | <i>Carya glabra</i> | 0.186 | 64 |
| Shellbark hickory | <i>Carya laciniosa</i> | 0.083 | 86 |
| Shagbark hickory | <i>Carya ovata</i> | 0.330 | 37 |
| Mockernut hickory | <i>Carya tomentosa</i> | 0.207 | 57 |
| American chestnut | <i>Castanea dentata</i> | 0.016 | 104 |
| Sugarberry | <i>Celtis laevigata</i> | 0.046 | 96 |
| Hackberry | <i>Celtis occidentalis</i> | 0.161 | 72 |
| Eastern redbud | <i>Cercis canadensis</i> | 0.085 | 85 |
| Flowering dogwood | <i>Cornus florida</i> | 0.263 | 48 |
| Common persimmon | <i>Diospyros virginiana</i> | 0.062 | 90 |
| American beech | <i>Fagus grandifolia</i> | 0.391 | 22 |
| White ash | <i>Fraxinus americana</i> | 0.430 | 17 |
| Black ash | <i>Fraxinus nigra</i> | 0.368 | 26 |
| Green ash | <i>Fraxinus pennsylvanica</i> | 0.277 | 46 |
| Blue ash | <i>Fraxinus quadrangulata</i> | 0.044 | 97 |
| Honeylocust | <i>Gleditsia triacanthos</i> | 0.175 | 68 |
| Loblolly bay | <i>Gordonia lasianthus</i> | 0.252 | 50 |
| American holly | <i>Ilex opaca</i> | 0.459 | 14 |
| Butternut | <i>Juglans cinerea</i> | 0.127 | 79 |
| Black walnut | <i>Juglans nigra</i> | 0.187 | 63 |
| Eastern redcedar | <i>Juniperus virginiana</i> | 0.256 | 49 |
| Tamarack | <i>Larix laricina</i> | 0.205 | 58 |
| Sweetgum | <i>Liquidambar styraciflua</i> | 0.524 | 7 |
| Osage orange | <i>Maclura pomifera</i> | 0.135 | 76 |
| Cucumber tree | <i>Magnolia acuminata</i> | 0.029 | 98 |
| Southern magnolia | <i>Magnolia grandiflora</i> | 0.022 | 102 |
| Sweetbay | <i>Magnolia virginiana</i> | 0.274 | 47 |
| Red mulberry | <i>Morus rubra</i> | 0.198 | 59 |
| Water tupelo | <i>Nyssa aquatica</i> | 0.052 | 93 |
| Blackgum | <i>Nyssa sylvatica</i> | 0.252 | 50 |
| Eastern hophornbeam | <i>Ostrya virginiana</i> | 0.252 | 50 |
| Sourwood | <i>Oxydendrum arboreum</i> | 0.381 | 25 |
| Redbay | <i>Persea borbonia</i> | 0.347 | 31 |
| White spruce | <i>Picea glauca</i> | 0.414 | 19 |
| Black spruce | <i>Picea mariana</i> | 0.815 | 1 |
| Red spruce | <i>Picea rubens</i> | 0.494 | 11 |
| Jack pine | <i>Pinus banksiana</i> | 0.504 | 9 |
| Shortleaf pine | <i>Pinus echinata</i> | 0.291 | 44 |
| Slash pine | <i>Pinus elliotii</i> | 0.612 | 3 |
| Spruce pine | <i>Pinus glabra</i> | 0.022 | 102 |
| Longleaf pine | <i>Pinus palustris</i> | 0.347 | 31 |
| Table Mountain pine | <i>Pinus pungens</i> | 0.055 | 92 |
| Red pine | <i>Pinus resinosa</i> | 0.366 | 27 |
| Pitch pine | <i>Pinus rigida</i> | 0.635 | 2 |
| Pond pine | <i>Pinus serotina</i> | 0.509 | 8 |
| Eastern pine | <i>Pinus strobus</i> | 0.415 | 18 |
| Loblolly pine | <i>Pinus taeda</i> | 0.503 | 10 |
| Virginia pine | <i>Pinus virginiana</i> | 0.282 | 45 |
| Sycamore | <i>Platanus occidentalis</i> | 0.113 | 82 |
| Balsam poplar | <i>Populus balsamifera</i> | 0.139 | 75 |
| Common cottonwood | <i>Populus deltoides</i> | 0.179 | 67 |
| Bigtooth aspen | <i>Populus grandidentata</i> | 0.314 | 39 |

Table 2 (continued)

| English name | Latin | R^2_{SAM} | Rank |
|-----------------------|--------------------------------|-------------|------|
| Quaking aspen | <i>Populus tremuloides</i> | 0.482 | 13 |
| Pin cherry | <i>Prunus pensylvanica</i> | 0.089 | 84 |
| Black cherry | <i>Prunus serotina</i> | 0.454 | 15 |
| Chokecherry | <i>Prunus virginiana</i> | 0.160 | 73 |
| White oak | <i>Quercus alba</i> | 0.299 | 42 |
| Swamp white oak | <i>Quercus bicolor</i> | 0.128 | 78 |
| Scarlet oak | <i>Quercus coccinea</i> | 0.180 | 65 |
| Northern pin oak | <i>Quercus ellipsoidalis</i> | 0.194 | 60 |
| Southern red oak | <i>Quercus falcata falcata</i> | 0.164 | 70 |
| Turkey oak | <i>Quercus laevis</i> | 0.391 | 22 |
| Laurel oak | <i>Quercus laurifolia</i> | 0.331 | 36 |
| Bur oak | <i>Quercus macrocarpa</i> | 0.300 | 41 |
| Water oak | <i>Quercus nigra</i> | 0.355 | 30 |
| Pin oak | <i>Quercus palustris</i> | 0.222 | 55 |
| Willow oak | <i>Quercus phellos</i> | 0.071 | 89 |
| Chestnut oak | <i>Quercus prinus</i> | 0.392 | 21 |
| Northern red oak | <i>Quercus rubra</i> | 0.343 | 33 |
| Shumard oak | <i>Quercus shumardii</i> | 0.024 | 101 |
| Post oak | <i>Quercus stellata</i> | 0.173 | 69 |
| Black oak | <i>Quercus velutina</i> | 0.327 | 38 |
| Live oak | <i>Quercus virginiana</i> | 0.091 | 83 |
| Black locust | <i>Robinia pseudoacacia</i> | 0.132 | 77 |
| Black willow | <i>Salix nigra</i> | 0.052 | 93 |
| Sassafras | <i>Sassafras albidum</i> | 0.240 | 53 |
| American mountain ash | <i>Sorbus americana</i> | 0.124 | 80 |
| Baldcypress | <i>Taxodium distichum</i> | 0.082 | 87 |
| Northern white cedar | <i>Thuja occidentalis</i> | 0.360 | 29 |
| American basswood | <i>Tilia americana</i> | 0.335 | 35 |
| Eastern hemlock | <i>Tsuga canadensis</i> | 0.311 | 40 |
| Winged elm | <i>Ulmus alata</i> | 0.338 | 34 |
| American elm | <i>Ulmus americana</i> | 0.404 | 20 |
| Slippery elm | <i>Ulmus rubra</i> | 0.193 | 61 |
| Rock elm | <i>Ulmus thomasi</i> | 0.217 | 56 |

(northern limit) of species distribution in the study area ($R^2_{\text{partial}} = 0.106$, $P < 0.001$).

Examining the distribution of IVs across mean annual temperature for each of the 105 species (Fig. S1) shows that our dataset likely captured the northern temperature limit in our study area for all species except the northernmost ones (*P. mariana* and *Pinus banksiana*). It shows also that several species with poor SAM performance did not reach high abundance or their abundance did not vary much with climate. Finally, we found a significant negative relationship between species richness (number of species in a grid cell) and latitude in our dataset (Pearson $r = -0.54$; $P < 0.001$) and a significant positive one between species richness and temperature (Pearson $r = 0.53$; $P < 0.001$). Species reaching higher latitude in our dataset also tended to have a broader latitudinal range (Pearson r between maximum latitude and latitudinal range in our dataset = 0.56, $P < 0.001$).

4. Discussion

The accurate modelling of tree species abundance is important for the development of reliable predictions of the impact of global changes on forest, yet predictive performances remained poor or moderate for the majority of species modelled. Conversely, SDMs based on presence/absence for the same dataset showed very high accuracy (mean Area Under the ROC Curve: 0.905 ± 0.056 ; Périé et al., unpublished data). Although direct comparison of performance with SDM is not appropriate given the different models used, this reveals a greater challenge in forecasting tree abundance compared to presence/absence at spatial scales relevant to global changes. The way we accounted for edaphic conditions in a cell could have contributed to obscure their spatial relationship to tree abundance but models constructed for a subset of sites for which we had a precise match between tree abundance and edaphic

Table 3

The values of the biogeographical and spatial attributes considered for the multiple linear regression (MLR) in step 2. R^2_{SAM} = Coefficient of determination, IVRange = Importance value range, LatMax = Latitudinal maximum, LatRange = Latitudinal range, NbObs = Number of observation, Moran I = Spatial aggregation.

| Latin | R^2_{SAM} | IVRange ^a | LatRange | LatMax ^a | NbObs | Moran I ^a |
|--------------------------------|-------------|----------------------|----------|---------------------|-------|----------------------|
| <i>Abies balsamea</i> | 0.596 | 49.977 | 7.538 | 51.367 | 2139 | 0.82 |
| <i>Acer barbatum</i> | 0.016 | 2.381 | 4.886 | 36.283 | 210 | 0.221 |
| <i>Acer negundo</i> | 0.192 | 20.608 | 11.054 | 44.555 | 970 | 0.193 |
| <i>Acer nigrum</i> | 0.162 | 8.8 | 7.879 | 44.53 | 96 | 0.035 |
| <i>Acer pensylvanicum</i> | 0.298 | 6.408 | 11.044 | 48.244 | 1195 | 0.571 |
| <i>Acer rubrum</i> | 0.494 | 25.51 | 15.886 | 47.965 | 4775 | 0.63 |
| <i>Acer saccharum</i> | 0.363 | 27.675 | 12.032 | 47.681 | 3174 | 0.559 |
| <i>Acer saccharinum</i> | 0.237 | 24.825 | 9.817 | 46.021 | 627 | 0.176 |
| <i>Acer spicatum</i> | 0.595 | 14.155 | 7.969 | 49.678 | 1156 | 0.752 |
| <i>Aesculus glabra</i> | 0.12 | 8.429 | 5.296 | 41.719 | 206 | 0.138 |
| <i>Aesculus flava</i> | 0.025 | 4.133 | 5.685 | 39.983 | 297 | 0.272 |
| <i>Alnus glutinosa</i> | 0.076 | 4.363 | 9.415 | 44.36 | 906 | 0.389 |
| <i>Asimina triloba</i> | 0.026 | 6.875 | 3.632 | 40.546 | 243 | 0.126 |
| <i>Betula alleghaniensis</i> | 0.449 | 10.404 | 9.988 | 48.442 | 1832 | 0.654 |
| <i>Betula lenta</i> | 0.391 | 10.353 | 10.213 | 43.353 | 1158 | 0.594 |
| <i>Betula nigra</i> | 0.061 | 6.44 | 11.434 | 43.607 | 251 | 0.164 |
| <i>Betula papyrifera</i> | 0.563 | 20.059 | 8.443 | 51.034 | 2386 | 0.729 |
| <i>Betula populifolia</i> | 0.144 | 7.608 | 6.243 | 47.218 | 512 | 0.425 |
| <i>Carpinus caroliniana</i> | 0.051 | 5.222 | 13.059 | 45.079 | 1981 | 0.235 |
| <i>Carya cordiformis</i> | 0.18 | 8.159 | 9.71 | 45.304 | 824 | 0.184 |
| <i>Castanea dentata</i> | 0.016 | 1.75 | 7.788 | 42.793 | 243 | 0.245 |
| <i>Carya glabra</i> | 0.186 | 7.062 | 9.838 | 42.226 | 2409 | 0.448 |
| <i>Carya laciniosa</i> | 0.083 | 5.636 | 7.24 | 42.505 | 109 | 0.045 |
| <i>Carya ovata</i> | 0.33 | 14.855 | 8.257 | 43.734 | 1171 | 0.277 |
| <i>Carya alba</i> | 0.207 | 6.556 | 9.743 | 42.059 | 2230 | 0.464 |
| <i>Cercis canadensis</i> | 0.085 | 6.396 | 6.993 | 40.077 | 908 | 0.17 |
| <i>Celtis laevigata</i> | 0.046 | 12.526 | 6.89 | 38.364 | 193 | 0.261 |
| <i>Celtis occidentalis</i> | 0.161 | 19.04 | 9.335 | 42.828 | 748 | 0.241 |
| <i>Cornus florida</i> | 0.263 | 11.516 | 10.054 | 41.899 | 2435 | 0.545 |
| <i>Diospyros virginiana</i> | 0.062 | 4.763 | 7.701 | 39.082 | 1140 | 0.255 |
| <i>Fagus grandifolia</i> | 0.391 | 17.725 | 13.453 | 46.629 | 2805 | 0.533 |
| <i>Fraxinus americana</i> | 0.43 | 20.154 | 11.227 | 46.238 | 2918 | 0.536 |
| <i>Fraxinus nigra</i> | 0.368 | 11.412 | 6.994 | 48.229 | 1186 | 0.308 |
| <i>Fraxinus pennsylvanica</i> | 0.277 | 16.868 | 13.323 | 46.065 | 1362 | 0.251 |
| <i>Fraxinus quadrangulata</i> | 0.044 | 5.75 | 6.023 | 41.42 | 84 | 0.255 |
| <i>Gleditsia triacanthos</i> | 0.175 | 18.333 | 8.241 | 41.93 | 448 | 0.168 |
| <i>Gordonia lasianthus</i> | 0.252 | 9.5 | 4.801 | 35.628 | 139 | 0.305 |
| <i>Ilex opaca</i> | 0.459 | 8.234 | 7.232 | 38.586 | 766 | 0.647 |
| <i>Juglans cinerea</i> | 0.127 | 4.093 | 9.586 | 45.956 | 430 | 0.129 |
| <i>Juglans nigra</i> | 0.187 | 11.352 | 8.25 | 42.889 | 1417 | 0.204 |
| <i>Juniperus virginiana</i> | 0.256 | 17.055 | 10.454 | 42.672 | 1277 | 0.514 |
| <i>Larix laricina</i> | 0.205 | 6.532 | 8.055 | 51.605 | 1409 | 0.216 |
| <i>Liquidambar styraciflua</i> | 0.524 | 22.518 | 7.655 | 38.956 | 1698 | 0.759 |
| <i>Magnolia acuminata</i> | 0.029 | 1.913 | 8.143 | 41.968 | 460 | 0.472 |
| <i>Magnolia grandiflora</i> | 0.022 | 3.864 | 6.146 | 37.045 | 218 | 0.234 |
| <i>Maclura pomifera</i> | 0.135 | 33.5 | 7.088 | 41.488 | 262 | 0.182 |
| <i>Magnolia virginiana</i> | 0.274 | 7.574 | 7.014 | 37.888 | 613 | 0.559 |
| <i>Morus rubra</i> | 0.198 | 11.614 | 9.573 | 41.741 | 570 | 0.054 |
| <i>Nyssa aquatica</i> | 0.052 | 8.115 | 6.593 | 37.548 | 256 | 0.28 |
| <i>Nyssa sylvatica</i> | 0.252 | 8.409 | 9.835 | 41.417 | 2354 | 0.52 |
| <i>Ostrya virginiana</i> | 0.252 | 8.487 | 13.364 | 46.57 | 2298 | 0.226 |
| <i>Oxydendrum arboreum</i> | 0.381 | 9.103 | 6.691 | 38.855 | 1255 | 0.68 |
| <i>Persea borbonia</i> | 0.347 | 6.914 | 5.294 | 36.117 | 349 | 0.547 |
| <i>Pinus banksiana</i> | 0.504 | 49.095 | 8.478 | 51.74 | 1156 | 0.528 |
| <i>Pinus echinata</i> | 0.291 | 11.152 | 7.021 | 38.584 | 1121 | 0.591 |
| <i>Pinus elliotii</i> | 0.612 | 46.244 | 3.991 | 34.74 | 414 | 0.801 |
| <i>Pinus glabra</i> | 0.022 | 2.75 | 1.948 | 32.825 | 116 | 0.468 |
| <i>Picea glauca</i> | 0.414 | 11.729 | 6.985 | 50.875 | 1660 | 0.49 |
| <i>Picea mariana</i> | 0.815 | 88.879 | 7.131 | 51.914 | 1980 | 0.9 |
| <i>Pinus palustris</i> | 0.347 | 19.86 | 4.275 | 35.097 | 499 | 0.509 |
| <i>Pinus pungens</i> | 0.055 | 4.3 | 5.03 | 40.135 | 95 | 0.275 |
| <i>Pinus resinosa</i> | 0.366 | 15.889 | 7.55 | 47.792 | 812 | 0.343 |
| <i>Pinus rigida</i> | 0.635 | 18.978 | 8.246 | 43.439 | 452 | 0.636 |
| <i>Picea rubens</i> | 0.494 | 15.875 | 7.393 | 48.639 | 879 | 0.66 |
| <i>Pinus serotina</i> | 0.509 | 19.704 | 5.53 | 36.595 | 274 | 0.506 |
| <i>Pinus strobus</i> | 0.415 | 17.607 | 12.013 | 48.091 | 2059 | 0.542 |
| <i>Pinus taeda</i> | 0.503 | 40.898 | 6.686 | 37.832 | 1372 | 0.83 |
| <i>Pinus virginiana</i> | 0.282 | 18.586 | 6.283 | 40.006 | 989 | 0.617 |
| <i>Platanus occidentalis</i> | 0.113 | 7.449 | 8.861 | 41.483 | 1175 | 0.196 |
| <i>Populus balsamifera</i> | 0.139 | 6.224 | 7.28 | 49.226 | 670 | 0.22 |
| <i>Populus deltoides</i> | 0.179 | 11.733 | 11.692 | 45.059 | 599 | 0.141 |
| <i>Populus grandidentata</i> | 0.314 | 8.46 | 8.362 | 47.447 | 1614 | 0.484 |
| <i>Populus tremuloides</i> | 0.482 | 20.457 | 8.885 | 50.356 | 2432 | 0.632 |

Table 3 (continued)

| Latin | R^2_{SAM} | IVRange ^a | LatRange | LatMax ^a | NbObs | Moran I ^a |
|------------------------------|-------------|----------------------|--------------|---------------------|---------|----------------------|
| <i>Prunus pensylvanica</i> | 0.089 | 3.806 | 9.816 | 49.648 | 1435 | 0.346 |
| <i>Prunus serotina</i> | 0.454 | 18.799 | 13.59 | 45.866 | 3940 | 0.507 |
| <i>Prunus virginiana</i> | 0.16 | 4.921 | 9.932 | 48.593 | 762 | 0.197 |
| <i>Quercus alba</i> | 0.299 | 16.217 | 12.079 | 44.468 | 3324 | 0.402 |
| <i>Quercus bicolor</i> | 0.128 | 9.194 | 10.747 | 45.095 | 357 | 0.103 |
| <i>Quercus coccinea</i> | 0.18 | 7.047 | 8.512 | 41.762 | 1704 | 0.372 |
| <i>Quercus ellipsoidalis</i> | 0.194 | 15.75 | 7.267 | 46.143 | 277 | 0.408 |
| <i>Quercus falcata</i> | 0.164 | 6.913 | 7.028 | 38.399 | 1503 | 0.517 |
| <i>Quercus laevis</i> | 0.391 | 13.714 | 4.463 | 35.31 | 207 | 0.445 |
| <i>Quercus laurifolia</i> | 0.331 | 11.5 | 5.108 | 36.025 | 682 | 0.595 |
| <i>Quercus macrocarpa</i> | 0.3 | 22.304 | 7.655 | 45.942 | 459 | 0.256 |
| <i>Quercus nigra</i> | 0.355 | 14.66 | 5.684 | 36.728 | 1060 | 0.688 |
| <i>Quercus palustris</i> | 0.222 | 14.639 | 6.033 | 42.554 | 357 | 0.164 |
| <i>Quercus phellos</i> | 0.071 | 5.194 | 6.81 | 38.334 | 717 | 0.308 |
| <i>Quercus prinus</i> | 0.392 | 16.348 | 8.603 | 42.118 | 1407 | 0.648 |
| <i>Quercus rubra</i> | 0.343 | 13.03 | 12.318 | 46.157 | 3359 | 0.435 |
| <i>Quercus shumardii</i> | 0.024 | 2.118 | 6.627 | 38.222 | 170 | 0.085 |
| <i>Quercus stellata</i> | 0.173 | 6.458 | 7.253 | 38.71 | 1425 | 0.4 |
| <i>Quercus velutina</i> | 0.327 | 12.262 | 10.631 | 43.917 | 2557 | 0.409 |
| <i>Quercus virginiana</i> | 0.091 | 9.333 | 3.16 | 33.816 | 151 | 0.404 |
| <i>Robinia pseudoacacia</i> | 0.132 | 11.723 | 7.659 | 42.539 | 1125 | 0.21 |
| <i>Sassafras albidum</i> | 0.24 | 10.353 | 9.517 | 42.322 | 1902 | 0.399 |
| <i>Salix nigra</i> | 0.052 | 14.489 | 7.823 | 45.207 | 471 | 0.076 |
| <i>Sorbus americana</i> | 0.124 | 2.288 | 6.096 | 50.454 | 1036 | 0.605 |
| <i>Taxodium distichum</i> | 0.082 | 5.438 | 6.13 | 37.079 | 316 | 0.267 |
| <i>Thuja occidentalis</i> | 0.36 | 17.438 | 4.946 | 48.721 | 1122 | 0.546 |
| <i>Tilia americana</i> | 0.335 | 9.25 | 10.147 | 46.487 | 1725 | 0.357 |
| <i>Tsuga canadensis</i> | 0.311 | 14.433 | 10.866 | 46.762 | 1566 | 0.55 |
| <i>Ulmus alata</i> | 0.338 | 13 | 6.348 | 38.199 | 521 | 0.465 |
| <i>Ulmus americana</i> | 0.404 | 20.193 | 11.777 | 46.336 | 2429 | 0.415 |
| <i>Ulmus rubra</i> | 0.193 | 10.404 | 11.197 | 45.113 | 1358 | 0.279 |
| <i>Ulmus thomasii</i> | 0.217 | 7 | 9.654 | 46.182 | 200 | 0.118 |
| Mean | 0.258 | 13.596 | 8.278 | 43.191 | 1153 | 0.398 |
| Standard deviation | 0.169 | 11.83 | 2.555 | 4.646 | 936 | 0.204 |
| Range | 0.016–0.815 | 1.750–88.879 | 1.948–15.866 | 32.825–51.914 | 84–4775 | 0.035–0.900 |
| Data trans. | Square root | Log | Square root | Square root | Log | Square root |

^a Indicates significant ($P < 0.001$) predictor in the MLR.

conditions did not lead to better performance (C. Périé, unpublished data).

The modest performance of many species raises the issue of climate control over tree species' relative abundance at the site or landscape level. Using Gaussian functions, Canham and Thomas (2010) examined the percentage of sites occupied in a given climate (i.e., species frequency, which is effectively a measure of occurrence) as well as the relative abundance (an estimate of relative biomass) of trees as a function of mean average temperature and annual precipitation for the 24 most common tree species in Northeastern USA, 23 of which are also included in our study. They found a strong relationship between climate and frequency of occurrence, but not with relative abundance. They concluded that climate is generally a poor predictor of relative tree abundance, which is supported by our data, while acknowledging that the inclusion of edaphic variables and of the northern limit of species to capture temperature extremes could have improved their model performance. The poor predictive performance may also be related to the fact that their abundance measure, like ours, is a relative one and that for some species it varied little across the range and a temperature gradient. Relative measures of abundance integrate species interactions (competition, facilitation) at the community or landscape level as well as any other disturbances acting selectively on species. For species to reach high importance value at a site, they need to be released from competition in low diversity areas, for instance at high latitude in the boreal forest where climate constrains competitors (e.g., *P. mariana*, Fig. S2). Nevertheless, the 23 species that were considered in both studies were among those with the highest model performances in our study ($R^2_{SAM} \geq 0.3$). We used a different abundance metric, a non-linear

model, included edaphic variables along with climatic variables, and covered a larger portion of the range that included the northern limit for most species, thereby testing some of the hypotheses raised by Canham and Thomas (2010) to explain the limitations of their model. Predictive performance remained nevertheless relatively modest for several species.

If tree abundance is poorly predicted by climate, projecting potential future abundance in the context of climate change will remain a challenge. Tree abundance in the eastern US was modelled with Random Forest by Iverson et al. (2008) in relation to a comprehensive set of 38 climatic and edaphic variables and projected under a future climate. We shared the same data on tree abundance in the US portion of the range for 105 out of the 134 species they modelled. Again direct comparison of predictive performances is not appropriate given the different predictors used, but looking at R^2_{SAM} from Random Forest can give some indications about the patterns in each study. The mean R^2_{SAM} from Random Forest in both studies was low, their mean R^2_{SAM} being 0.28 (SD = 0.22; see Appendix A in Iverson et al., 2008). This is true also when comparing only the 105 species modelled by both studies (mean R^2_{SAM} Iverson et al., 2008 for these 105 species = 0.3, see Appendix A in Iverson et al. 2008; mean R^2_{SAM} this study = 0.26, Table 3), suggesting that adding more predictors to their models resulted in a marginal increase in model fit compared to ours. The authors also observed that species with “most data” tended to have more reliable models. These species may also be the ones likely to reach high IV. It was suggested that poor model reliability for some species could have resulted from the failure to capture some portions of the range or by the fact that some species were small-ranged. We identified only seven species for which extending the range

into Canada could have resulted in better predictions, but in most cases it went from poor (i.e., $R^2_{\text{SAM}} < 0.3$) to moderate ($R^2_{\text{SAM}} \geq 0.3$ but < 0.5) performance. Moreover, the extent of the latitudinal range was not significant in our study and this may be partly explained by the fact that even species with small range can vary in abundance along local edaphic gradients, resulting in good model performance (e.g., *Pinus elliottii*, $R^2_{\text{SAM}} = 0.61$). The species that performed best in our study (i.e., $R^2_{\text{SAM}} \geq 0.3$; $n = 41$) were also all categorized among the most or second most reliable models in Iversen et al. (2008). The authors emphasised their lack of confidence in future projections for species with poor model reliability. In Europe, a study on predicting the effect of climate change on the distribution and abundance of 16 of the most abundant tree species reported a mean Pearson's correlation coefficient of 0.698 (Attorre et al., 2008). Once adjusted to the coefficient of determination R^2 , the mean is 0.497 (SD = 0.145).

Our models attempted to identify, with mixed results, suitable combinations of climate and edaphic conditions potentially influencing species' relative abundance in a grid cell. Nevertheless, the interpretation of SAM, as opposed to SDM, remains problematic when projecting under a future climate. Researchers using SDM for projections in a future climate space emphasize that it is not the presence or absence of species that is being predicted, but rather the potential location of suitable climatic conditions. It is not entirely clear, however, how to interpret abundance projections, especially for relative measurements that depend on co-occurrence patterns. One approach to refine SAM for climate change research could be to infer future relative importance patterns from reliable SDM predicting the co-occurrence of species, coupled with more mechanistic knowledge of how abundance is expected to vary with species interactions in communities and landscapes. This could be based on species traits and a better knowledge of the individual response of species to climate (Clark et al., 2011; Gunderson et al., 2012; Meier et al., 2011). Some species, for instance, never reach dominance across their range, but it is still largely unclear how species will reassemble into communities under novel climate conditions.

In spite of low average model performance, some species performed better than others in our study. IV range, which was positively correlated with IV mean, was a good predictor of model performance. The species with the largest range of IV tended to be dominant species that form pure stands (IV = 100%) in part of their range (e.g., *P. mariana*, *P. banksiana*, *Abies balsamea*, *Pinus taeda*). For some of the cold tolerant species, dominance can increase with latitude as tree species diversity drops with cold temperatures (Gaston, 2000; Purves, 2009; Willig et al., 2003); it will, on the other hand, decrease with competition in the warmer portion of their range. This pattern may be amplified by the fact that range size also increases with latitude (Morin and Lechowicz, 2011), so a larger temperature gradient is captured by the distribution of northern species. Latitudinal extent, however, did not explain model performance in our study.

Many species (e.g., *Castanea dentata*, *Magnolia acuminata*, *Prunus pensylvanica*) rarely reach dominance in their community and therefore show little variation in their relative abundance along a climate gradient, resulting in poor model performance (Murphy et al., 2006). Relative species abundance curves commonly reveal the dominance of a few species, with most others being in the tail of the abundance distribution (Magurran, 2004; McGill et al., 2007). Whereas the dominant and widespread species are important in structuring communities, biodiversity lies in all these other species. It is therefore important to understand how they will behave in a context of rapid environmental changes, but abundance models with the predictors used in this study appear the least informative for these species.

The contribution of the Moran index to the MLR indicates that SAM performed better for species with aggregated IV at the landscape scale. Moran I also tended to increase with IV range, suggesting that dispersal process for some of the dominant species could have led to clumped distribution over a range of climate and edaphic conditions. Adjacent cells are expected to share similar climatic and possibly edaphic conditions and the structure of ecological data sets is rarely free of spatial dependency effect, which also raises statistical issues (Legendre, 1993). However, spatial aggregation also provides useful information on underlying structuring process (Legendre and Legendre, 1998; Wulder and Boots, 1998). The sub-sampling strategy that we used with a classification algorithm could have also contributed to attenuate inflation effect related to spatial correlation in our models (Segurado et al., 2006).

Overall, the SAM tested were only partially successful in reproducing patterns of relative abundance in relation to the climatic, edaphic, and topographic predictors used in this study for trees of eastern North-America. Species with the best performance tended to be those that could reach dominance, showed aggregated distribution of abundance, and/or had high latitudinal limits in the study area. Climate change is very likely to affect patterns of dominance in communities by altering patterns of co-occurrences, but for many species that constitute the bulk of tree diversity, predictions based solely on the current distribution of relative abundances may not be reliable enough to inform conservation or management decisions. Predicting tree abundance in a warming climate using SAM remains a challenge, but we argue that it is only by reporting performances across a range of climate and statistical models, regions and species, as well as by highlighting model limitations and strengths, that we will improve the reliability of predictions and in turn better inform forest conservation and management.

Acknowledgements

This research was made possible thanks to funding to S. de Blois from the Natural Sciences and Engineering Research Council of Canada (NSERC) and from the Ministère des Ressources naturelles et de la Faune du Québec (MRNF). We acknowledge all members of the CC-Bio project for support. We thank Nicholas Crookston from the Moscow Forestry Sciences Laboratory (USDA Forest Service) for providing climate data and the consortium OURANOS on regional climatology and adaptation to climate change, particularly Travis Logan, for helping process them. We acknowledge the USDA Forest Service tree Atlas website for IV data in the U.S. portion of the range and the MRNF for IV in the Canadian portion. We thank André Robitaille (geomorphology), Marie-Claude Lambert (statistics), and Jean Noël (GIS) from MRNF for technical help, Brian McGill (University of Maine) and Volker Bahn (Wright University) for comments during the initial phase of the research, and James Fyles (McGill) for helpful discussion. We finally thank Christian Messier (UQAM) and Jacqueline Bede (McGill) for providing comments on an earlier draft of the manuscript and two anonymous reviewers for helping improve it.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.foreco.2012.10.046>.

References

- Anadón, J.D., Gimenez, A., Ballestar, R., 2010. Linking local ecological knowledge and habitat modelling to predict absolute species abundance on large scales. *Biodivers. Conserv.* 19, 1443–1454.

- Araújo, M.B., Cabeza, M., Thuiller, W., Hannah, L., Williams, P.H., 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biol.* 10, 1618–1626.
- Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22, 42–47.
- Attorre, F., Francesconi, F., Scarnati, L., De Sanctis, M., Alfò, M., Bruno, F., 2008. Predicting the effect of climate change on tree species abundance and distribution at a regional scale. *iForest – Biogeosci. Forestry* 1, 132–139.
- Berteaux, D., de Blois, S., Angers, J.F., Bonin, J., Casajus, N., Darveau, M., Fournier, F., Humphries, M.M., McGill, B.J., Larivée, J., Logan, T., Nantel, P., Périé, C., Poisson, F., Rodrigue, D., Rouleau, S., Siron, R., Thuiller, W., Vescovi, L., 2010. The CC-Bio project: studying the effects of climate change on Québec biodiversity. *Diversity* 2, 1181–1204.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brotans, 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.
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., Grenouillet, G., 2010. Uncertainty in ensemble forecasting of species distribution. *Global Change Biol.* 16, 1145–1157.
- Burns, R., Honkala, B.H. (tech. coords.), 1990. *Silvics of North America: 1. Conifers; 2. Hardwoods*. Agriculture Handbook 654, U.S. Department of Agriculture, Forest Service, Washington, D.C. vol. 2, pp. 877.
- Canadian Council on Geomatics., 2003. Canadian geographical names. Government of Canada, Natural Resources Canada, Earth sciences sector, Geomatics Canada, center for topographic information, geographical names section, Ottawa, ON, <<http://www.geobase.ca/geobase/en/data/cded/index.html>>.
- Canham, C.D., Thomas, R.Q., 2010. Frequency, not relative abundance, of temperate tree species varies along climate gradients in eastern North America. *Ecology* 91, 3433–3440.
- Clark, J.S., Bell, D.M., Hersh, M.H., Nichols, L., 2011. Climate change vulnerability of forest biodiversity: climate and competition tracking of demographic rates. *Global Change Biol.* 17, 1834–1849.
- Gaston, K.J., 2000. Global patterns in biodiversity. *Nature* 405, 220–227.
- Gunderson, C.A., Edwards, N.T., Walker, A.V., O'Hara, K.H., Campion, M., Hanson, P.J., 2012. Forest phenology and a warmer climate – growing season extension in relation to climatic provenance. *Global Change Biol.* 18, 2008–2025.
- Hanspach, J., Kühn, I., Schweiger, O., Pompe, S., Klotz, S., 2011. Geographical patterns in prediction errors of species distribution models. *Global Ecol. Biogeogr.* 20, 779–788.
- Hu, J., Hu, H., Jiang, Z.G., 2010. The impacts of climate change on the wintering distribution of an endangered migratory bird. *Oecologia* 164, 555–565.
- Huntley, B., Altwegg, R., Barnard, P., Collingham, Y.C., Hole, D.G., 2011. Modelling relationships between species spatial abundance patterns and climate. *Global Ecol. Biogeogr.* 20, 744–754.
- Iverson, L.R., Prasad, A.M., 1998. Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecol. Monogr.* 68, 465–485.
- Iverson, L.R., Prasad, A.M., Matthews, S.N., Peters, M., 2008. Estimating potential habitat for 134 eastern US tree species under six climate scenarios. *Forest Ecol. Manage.* 254, 390–406.
- Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: a comparison of five machine-learning methods. *Ecol. Infor.* 5, 441–450.
- Kissling, D.W., Gudrum, C., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecol. Biogeogr.* 20, 779–788.
- Lawler, J.J., Shafer, S.L., Blaustein, A.R., 2010. Projected climate impacts for the amphibians of the western hemisphere. *Conserv. Biol.* 24, 38–50.
- Legendre, P., 1993. Spatial autocorrelation – trouble or new paradigm. *Ecology* 74, 1659–1673.
- Legendre, P., Legendre, L., 1998. *Numerical Ecology*. Elsevier, Amsterdam, New York, NY.
- Luoto, M., Pöyry, J., Heikkinen, R.K., Saarinen, K., 2005. Uncertainty of bioclimate envelope models based on the geographical distribution of species. *Global Ecol. Biogeogr.* 14, 575–584.
- Magurran, A.E., 2004. *Measuring Biological Diversity*. Blackwell Scientific, Oxford.
- Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecol. Model.* 220, 3512–3520.
- McKenney, D.W., Pedlar, J.H., Lawrence, K., Campbell, K., Hutchinson, M.F., 2007. Potential impacts of climate change on the distribution of North American trees. *Bioscience* 57, 939–948.
- McKenney, D.W., Pedlar, J.H., Rood, R.B., Price, D., 2011. Revisiting projected shifts in the climate envelopes of North American trees using updated general circulation models. *Global Change Biol.* 17, 2720–2730.
- McGill, B.J., Etienne, R.S., Gray, J.S., Alonso, D., Anderson, M.J., Benecha, H.K., Dornelas, M., Enquist, B.J., Green, J.L., He, F., Hurlbert, A.H., Magurran, A.E., Marquet, P.A., Maurer, B.A., Ostling, A., Soykan, C.U., Ugland, K.I., White, E.P., 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* 10, 995–1015.
- McMahon, S.M., Harrison, S.P., Armbruster, W.S., Bartlein, P.J., Beale, C.M., Edwards, M.E., Kattge, J., Midgley, G., Morin, X., Prentice, I.C., 2011. Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. *Trends Ecol. Evol.* 26, 249–259.
- Meier, E.S., Edwards, T.C., Kienast, F., Dobbstein, M., Zimmermann, N.E., 2011. Co-occurrence patterns of trees along macro-climatic gradients and their potential influence on the present and future distribution of *Fagus sylvatica* L. *J. Biogeogr.* 38, 371–382.
- Millar, C.I., Stephenson, N.L., Stephens, S.L., 2007. Climate change and forests of the future: managing in the face of uncertainty. *Ecol. Appl.* 17, 2145–2151.
- Morin, X., Lechowicz, M., 2011. Geographical and ecological patterns of range size in North American trees. *Ecography* 34, 738–750.
- Murphy, H.T., VanDerWal, J., Lovett-Doust, J., 2006. Distribution of abundance across the range in eastern North American trees. *Global Ecol. Biogeogr.* 15, 63–71.
- Murphy, H.T., VanDerWal, J., Lovett-Doust, J., 2010. Signatures of range expansion and erosion in eastern North American trees. *Ecol. Lett.* 13, 1233–1244.
- Pearce, J., Ferrier, S., 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biol. Conserv.* 98, 33–43.
- Prasad, A.M., Iverson, L.R., 2003. Little's Range and FIA Importance Value Database for 135 eastern US Tree Species. In: Northeastern Research Station, editor. USDA Forest Service, Delaware, OH, <<http://www.fs.fed.us/ne/delaware/4153/global/littlefia/index.html>>.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- Purves, D.W., 2009. The demography of range boundaries versus range cores in eastern US tree species. *Proc. Royal Soc. Ser. B: Biol. Sci.* 276, 1477–1484.
- R Development Core Team, 2009. *A Language and Environment for Statistical Computing*. R. Foundation for Statistical Computing, Vienna, Austria. <<http://cran.r-project.org>>.
- Rehfeldt, G.E., 2006. *A Spline Model of Climate for the Western United States*. General technical report RMRS-GTR-165, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO.
- Sagarin, R.D., Gaines, S.D., Gaylord, B., 2006. Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends Ecol. Evol.* 21, 524–530.
- Segurado, P., Araújo, M.B., Kunin, W.E., 2006. Consequences of spatial autocorrelation for niche-based models. *J. Appl. Ecol.* 43, 433–444.
- Stegen, J.C., Swenson, N.G., Enquist, B.J., White, E.P., Phillips, O.L., Jørgensen, P.M., Weiser, M.D., Monteagudo, A., Vargas, P.N., 2011. Variation in above-ground forest biomass across broad climatic gradients. *Global Ecol. Biogeogr.* 20, 744–754.
- Syphard, A.D., Franklin, J., 2010. Species traits affect the performance of species distribution models for plants in Southern California. *J. Veg. Sci.* 21, 177–189.
- Thuiller, W., 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biol.* 9, 1353–1362.
- USDA Forest Service, 2010. Moscow Forestry Sciences Laboratory. Moscow Forestry Sciences Laboratory, Moscow, ID, <<http://forest.moscowfsi.wsu.edu/>>.
- USGS, 2010. USGS Geographic Data Download. United States Geological Survey, Reston, VA, <<http://edc2.usgs.gov/geodata/index.php>> (data available in September 2012 at <http://datagateway.nrcs.usda.gov/>).
- VanDerWal, J., Shoo, L.P., Johnson, C.N., Williams, S.E., 2009. Abundance and the environmental niche: environmental suitability estimated from niche models predicts the upper limit of local abundance. *Am. Nat.* 174, 282–291.
- Whitton, F.J.S., Purvis, A., David, C., Orme, L., Olalla-Tárraga, M.A., 2012. Understanding global patterns in amphibian geographic range size: does Rapoport rule? *Global Ecol. Biogeogr.* 21, 179–190.
- Willig, M.R., Kaufman, D.M., Stevens, R.D., 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. *Ann. Rev. Ecol. Syst.* 34, 273–309.
- Wulder, M., Boots, B., 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic. *Int. J. Remot. Sens.* 19, 2223–2231.