Relating modelled habitat suitability for *Abies balsamea* to on-the-ground species structural characteristics in naturally growing forests

Yaquan Chang[a,b], Charles P.-A. Bourque[a,c,*]

[a] *Faculty of Forestry and Environmental Management, University of New Brunswick, 28 Dineen Drive, Fredericton, New Brunswick E3B 5A3, Canada*
[b] *Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, Universitetsparken 15, Copenhagen 2100, Denmark*
[c] *School of Soil and Water Conservation, Beijing Forestry University, Haidan District, 35 East Qinghua Road, Beijing 100083, China*

ARTICLE INFO

ABSTRACT

Modelling tree habitat suitability (HS) is a common practice to assess tree species distribution across a broad range of spatial scales. However, it is seldom used to test the extent that modelled HS-scores (probability of species occurrence) can represent on-the-ground measures of species-structural characteristics. In this study, we compare four parametric and non-parametric models generated with the R-package, *sdm*, to assess the potential for these models to estimate tree species distribution of balsam fir [*Abies balsamea* (L.) Mill.] in naturally-growing forests across an extensive landscape. Central to this development are inventory plot data of species presence-absence and four abiotic factors linked to plant growth and distribution. The study's abiotic factors include: (1) photosynthetically active radiation; (2) growing degree-days; (3) relative extractable soil water content; and (4) near-surface wind speed, all expressed spatially at 30-m resolution. To gauge whether modelled HS can explain structural characteristics in balsam fir-dominated stands, field-based estimates of site index (*SI*) and cumulative aboveground biomass (*ABG*) were compared against independently-derived HS-scores. The results showed that: (i) random forest was the most successful at representing species distribution of balsam fir among the four methods considered; (ii) overall growing conditions for balsam fir was observed to be most favourable on north-facing slopes, particularly in the northwest part of the target landscape, where near-surface air temperatures are cooler, soils are moderately wetter, and wind speeds are lower; (iii) tree-based calculations of *SI* were partially characterised by patterns in modelled HS-scores, due to scale differences (i.e., from individual tree to 30-m grid cells) and an inadequate number of sample trees; and (iv) patterns of cumulative *AGB* were more accurately represented by species HS. Modelled HS-scores, as potential indicators of tree species habitat preference, *AGB*, and species distribution, can offer key ecological information essential to inform forest management and conservation planning at the landscape level.

1. Introduction

Forests, as important dynamic systems, provide essential ecological services to humans (e.g., through the provision of food and sequestration of carbon; Nadrowski et al., 2010). Forest management impacts ecosystem processes by affecting species diversity and forest-structural characteristics (Tilman et al., 1997). Understanding tree-growth processes are central to making informed forest-management decisions. Some factors controlling tree growth and forest-landscape development processes, include incident solar radiation (Austin, 2002), growing degree-days (Zimmermann and Kienast, 1999), plant extractable soil water content (Austin, 2007), near-surface wind speed (Bourque and Bayat, 2015), soil fertility, among other factors.

With the development of new statistical techniques, Geographic Information Systems (GIS), and associated spatial procedures, species distribution models (SDM's) have become common in ecological and resource-management studies (Guisan and Zimmermann, 2000). Species distribution models correlate species occurrence with site environmental variables in order to predict tree habitat suitability (HS). These models are powerful tools in predicting species' occurrence in both space and time (Elith and Leathwick, 2009) and have a role, for instance, in facilitating (i) the prediction of potential tree species and animal habitat shifts projected with future climate change (Bourque et al., 2000), (ii) classification of forestland types for forest-management objectives (Baah-Acheamfour et al., 2013), and (iii) searching efforts, when planning field surveys (Fois et al., 2015). Although the method is crucial to the field of ecology and resource management, the method has seldom been tested to determine the extent HS-scores

---
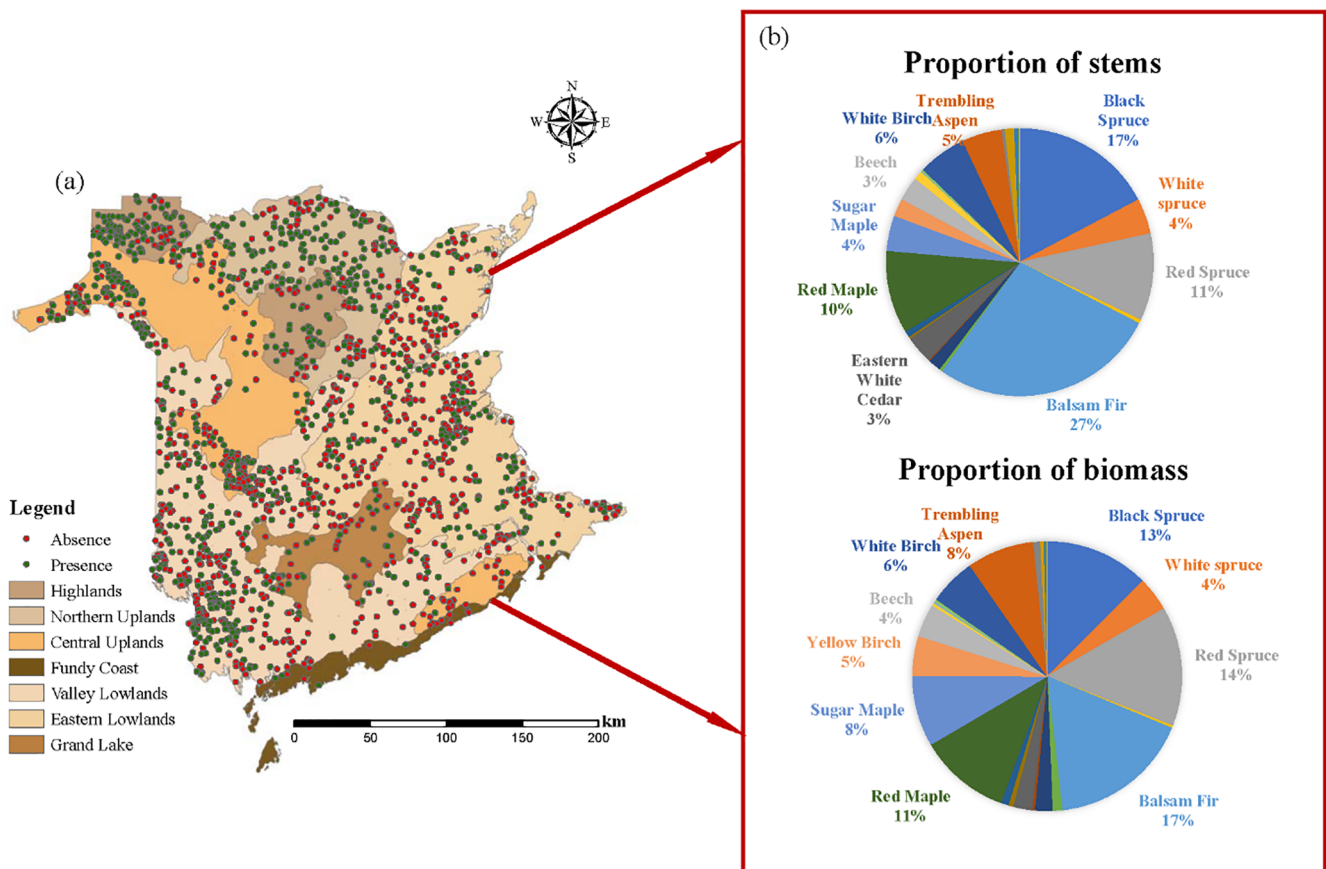
Fig. 1. Ecoregions of the Province of New Brunswick, Canada (NB; a) and within-plot proportion of aboveground biomass and tree stems according to species (b); note that stem density in NB is roughly 1,841 stems ha^{-1}. The green dots in panel (a) represent permanent sample plots (PSP's) with balsam fir (bF) contributing to $\geq 10\%$ of total species basal area (BA) and are identified as plots with "bF being present", whereas those with $< 10\%$ total species BA (red dots) are identified as "bF being absent".

(probability of species occurrence) can be used as indicators of on-the-ground forest conditions (Fois et al., 2018). Being able to relate modelled values of species distribution and HS to on-the-ground conditions can convey important ecological information to forest management and conservation planning.

Understanding the relationship between plant species and their environment is fundamental to plant species distribution modelling (Guisan and Zimmermann, 2000). Associated analytical methods use data from biophysical surfaces (Bourque et al., 2000; Zimmermann et al., 2007), which enhances our modelling capabilities by taking into account many of the environmental conditions prevailing at a site (Elith and Leathwick, 2009). There are various methods to parameterise SDM's. Convex hulls are often used to model sparse data (Burgman and Fox, 2003), and expert opinion can at times be used to enhance actual species records (Elith and Leathwick, 2009).

Bioclimatic-envelop models, as the earliest form of SDM's, define the limit of environmental factors in multi-dimensional space (Box, 1981). Statistical models, including those constructed by nonlinear regression, are regularly used in ecology because of species nonlinear interactions to site conditions (Elith and Leathwick, 2009; Latimer et al., 2006). Forest gap models often define total species response by multiplying individual tree species response to specific site conditions (Acevedo et al., 1995; Bourque et al., 2000). Some researchers have demonstrated that SDM's constructed on presence-only data can be

closely associated with population characteristics, such as population density (Tôrres et al., 2012), abundance of individuals (de Moraes Weber and Viveiros Grelle, 2012; VanDerWal et al., 2009), and extent of coverage (Fois et al., 2018).

Although some studies have found relationships between SDM-results and species demography (de Moraes Weber and Viveiros Grelle, 2012; VanDerWal et al., 2009), many other studies of this kind have been less successful (Filz et al., 2013; Nielsen et al., 2005; Pearce and Ferrier, 2001; Tellería et al., 2012), especially for plant species (Weber et al., 2017). Clearly, there is a need for further research in this area (Weber et al., 2017). The spatial scales of SDM's vary from continent to grid-based scales of a few tens of metres for different analytical objectives (Elith and Leathwick, 2009). Conservation planning and forest management tend to use finer resolutions to generate important analytical detail (Bourque and Bayat, 2015). This effort has been greatly enhanced in recent years with the introduction of LiDAR (Light Detection and Ranging) technologies and data in the generation of high-resolution digital elevation models (DEM's), fundamental to the development of highly detailed surfaces of multi-factor environments.

The main objective of this study was to develop a series of SDM's with the R-package, *sdm* (Naimi and Araújo, 2016), and to examine the extent modelled HS-scores can be used as an indicator of actual on-the-ground structural characteristics in balsam fir [*Abies balsamea* (L.) Mill]-dominated stands. Construction of SDM's will be based on: (i)

balsam fir presence-absence data gleaned from a network of forest inventory plots (Porter et al., 2001); and (ii) digital surfaces of incident photosynthetically active radiation (PAR; Bourque and Gullison, 1998), growing degree-days (GDD; Bourque et al., 2000; Hassan, 2008), plant extractable soil water content (SWC; Bourque et al., 2000), and near-surface wind speed (WS; Bourque and Bayat, 2015). Given the absence of scale-appropriate information of soil fertility for the study area, we currently ignore its influence on HS.

2. Methods

2.1. Study area

The Province of New Brunswick (158,186 km^2), which serves as our study area, is located in eastern Canada adjacent to the Atlantic Ocean. Eighty three percent of the province is estimated to be forested. The province is divided into seven ecoregions, which are defined by climate differences (generally, defined in terms of air temperature and precipitation), elevation, landforms, and marine influences (Zelazny et al., 2007). New Brunswick (NB) has a humid continental climate influenced by the cool Bay of Fundy to the south and the warmer Northumberland Strait to the east of the province (Fig. 1a). The elevation of NB ranges from 0 to 834 m above mean sea level (amsl), with mean temperature predictably decreasing with elevation. Based on long-term weather records from several airport weather stations in NB, the mean annual temperature and total precipitation range from 2.0 to 6.3 °C and 997–1510 mm, respectively (Environment and Climate Change Canada, 2018). Tree species distribution mostly follows the ecoregions of NB (Fig. 1a). The province is resident to both broad- and needle-leaf forest species common to the Acadian Forest, including 55% spruce-fir, 14% other softwood species, and 32% temperate hardwood species (Chen et al., 2017).

Balsam fir is one of two dominant species in NB (Erdle and Ward, 2008), making up the largest proportion of public forests at 27%, followed by black spruce (*Picea mariana*, 17%), red spruce (*Picea rubens*, 11%), and red maple (*Acer rubrum*, 10%; Fig. 1b). Likewise, above-ground biomass of balsam fir accounts for 17% of the total tree aboveground biomass, which contributes to the single largest biomass component in NB forests, followed by red spruce (14%), black spruce (13%), red maple (11%), trembling aspen (*Populus tremuloides*, 8%), and sugar maple (*Acer saccharum*, 8%; Fig. 1b).

2.2. Species distribution modelling

The R-package, *sdm*, uses individual and community-based approaches, not only to generate ensembles and evaluate different models within the same framework (Borregaard and Hart, 2016), but also to develop new species-distribution modelling approaches by users (Naimi and Araújo, 2016). Typically, the numerical, non-dimensional output of SDM's (i.e., HS) ranges from 0.0 to 1.0, where 0.0 represents poor site quality and potential absence of the species and 1.0, optimal site quality and the potential presence of the species (Hirzel et al., 2006). The work proceeds incrementally along three main lines of action, namely:

1. **Data management and pre-processing**. Input data to *sdm* includes geographical occurrence (i.e., presence-absence) of balsam fir (Fig. 1b), a digital elevation model (DEM) of NB based on the NASA Shuttle Radar Topography Mission (SRTM3) at one-arc second resolution (https://earthexplorer.usgs.gov/), and four environmental variables rasterised at 30-m resolution (Fig. A1);

2. **Model development.** We chose three fundamental models engaging both statistical and machine-learning procedures, i.e., generalized linear and additive models, centred on statistical analyses (GLM and GAM; McCullagh and Nelder, 1989; Hastie and Tibshirani, 1987), and classification and regression trees, centred on machine learning (CART; Breiman, 2001). As a natural extension to CART, we also use random forest (RF; Breiman, 2001) for its exceptional performance in modelling species distribution (Gama et al., 2017); and

3. **Model performance assessment and analysis**. HS-scores from the SDM that best approximates the actual province-wide distribution of balsam fir are subsequently used in a comparison against tree- and plot-level measures of site quality, i.e., site index (*SI*) and species aboveground biomass (*ABG*).

2.2.1. Data management and pre-processing

2.2.1.1. Presence-absence data.
The presence-absence of balsam fir were based on tree data recorded for 1800+ forest permanent sample plots (PSPs). The provincial network of PSP's was established by the Canadian Forest Service and NB Department of Energy and Resource Development in 1987 (Porter et al., 2001). Those PSP's with balsam fir (bF) contributing to ≥10% of total species basal area (BA) were designated as "bF being present", whereas those with < 10% BA were identified as "bF being absent". Altogether, plots with and without balsam fir totalled 982 and 883 plots, respectively (Fig. 1a).

2.2.1.2. Biologically-relevant abiotic surfaces.
Species distribution and growth patterns vary according to incident PAR. To calculate the effects of topography on available PAR (Fig. A1a), DEM-derived terrain features of slope, aspect, horizon angle, sky-view factor, and terrain configuration were considered (Bourque and Gullison, 1998). A mean mid-afternoon atmospheric transmissivity of 0.7 (non-dimensional) was used to correct for the diffused PAR that rarely reaches the earth's surface (Bourque and Gullison, 1998). Photosynthetically active radiation (~45% of incident solar radiation), for cloud-free conditions, was estimated with the **Lan**dscape **D**istribution of **S**oil moisture, **E**nergy, and **T**emperature model (LanDSET; Bourque and Gullison, 1998; Bourque et al., 2000) as an integration of hourly estimates of PAR over the entire growing season (Hassan et al., 2006).

Temperature is one of the primary determinants of plant metabolic processes (Gillooly et al., 2001). Growing degree-days (GDD) represent the level of warmth plants require for growth over the growing season. The GDD map at 28.5-m resolution (Fig. A1b) was developed from a compilation of Landsat-7 ETM + and Moderate Resolution Imagining Spectroradiometer (MODIS) sensor images (Hassan et al., 2007a) and an application of the standard equation of growing degree-day, i.e.,

$$\text{GDD} = \sum_{i=1}^{n} (T_{mean} - T_{base}), \text{ when } (T_{mean} - T_{base}) > 0, \tag{1}$$

where T_{mean} is the average daytime temperature, T_{base} is a base temperature of 5.6 °C below which balsam fir ceases to grow, and *i* goes from 1 to *n*, where 1 and *n* represent the beginning and last day of the growing season (Hassan et al., 2007a). The final GDD map was subsequently standardised for a normal (30-year) period using long-term temperature data from selected weather stations within and adjacent to NB (Hassan et al., 2007b).

Tree species have different soil water requirements (Oliver and Larson, 1996). The annual mean SWC (Fig. A1c) was calculated with LanDSET. A DEM-level cell-by-cell water budget calculation of SWC was performed by taking into account annual total precipitation, lateral flow from upslope to downslope positions, infiltration, deep percolation, evapotranspiration, surface runoff, and changes in soil water

storage (all in mm day^{-1}). Other model inputs include cell-size specification (30-m) and all-wave radiative fluxes (in MJ m^{-2}; Bourque et al., 2000) for a Priestley-Taylor-based calculation of evapotranspiration. Relative extractable soil water content (SWC) is represented by a non-dimensional value that ranges between 0.0 and 1.0, with SWC approaching 1.0 for wetter sites at field capacity and 0.0, for drier sites at the permanent wilting point.

Wind speed (Fig. A1d) affects plant growth due to both mechanical (e.g., flagging) and physiological effects (Retuerto and Woodward, 1992). Wind-tunnel studies of tree growth have shown relative growth-rate increases with increasing wind speed, but these growth rates soon decline as wind speeds continue to increase beyond an optimal level (~1.5–2 m s^{-1} for balsam fir), especially in young trees (Wadsworth, 1959). A three-dimensional wind field for a mountainous region in central NB is determined by means of a small-scale numerical solution of the Reynolds-averaged Navier-Stokes equations (at 30-m spatial resolution), incorporating the effects of both atmospheric turbulence and thermal processes (Lopes, 2003). The calculation uses input from the DEM and initialising near-surface atmospheric conditions (i.e., air temperature, wind speed and direction) from an upwind weather station located at 47° 09′ 27″ North latitude and 67° 50′ 5″ West longitude (Saint Leonard Airport, NB). Wind speed at 500 m amsl is specified as corresponding to a mean value of 1.3 × near-surface wind speed (Franklin et al., 2003). Temperature stratification of the atmosphere is assumed neutral. Small-scale estimates of wind are subsequently upscaled to the entire province with nonlinear regression and several DEM-based, SAGA GIS-derived indices of terrain (Conrad et al., 2015), including those for wind effect and terrain ruggedness, point maximal curvature, and relative elevation as independent variables in the regression.

2.2.2. Model development

Each modelling procedure, because of their underlying assumptions and internal sensitivities, can generate slightly different empirical relationships, with potentially different goodness-of-fit. Although ensemble models (combining two or more algorithms in solving a problem) usually provide better results (Araújo and New, 2007), we do not invoke such an ensemble here as we are more interested in determining the utility of the different procedures in approximating HS as a function of the different growth-related factors.

2.2.2.1. Statistical methods. Statistical models have best-practice applications if the aim of the model is to capture the interaction between species distribution factors (Vasconcelos et al., 2017). Habitat suitability of tree species is typically described mathematically by means of general linear, logistic, or logarithmic functions in developing emprical relationships between response and explanatory variables, with 1.0 and 0.0 corresponding to species presence-absence (Carrete et al., 2007; Maunder and Punt, 2004). General linear models (GLM's) are parametric and can involve one of several distributions (e.g., binomial, Poisson, multinomial, etc.), whereas general additive models (GAM's) are non-parametric, commonly viewed as extensions of GLM's (Guisan et al., 2002). In general, the degree of fit with both procedures depend on the inherent complexity of the target dataset used in establishing the relationship. GAM's are generally more flexible than GLM's in addressing nonlinear interactions in the data (Guisan et al., 2002; Yu et al., 2013). Here, model development with GLM's is done through the R *stats* package called by *sdm*, whereas GAM's involve calls to *mgcv* and *gam* (Naimi and Araújo, 2016).

2.2.2.2. Machine-learning methods. Classification and regression tree (CART, based on the R-package *tree*) is a non-parametric approach based on binary-recursive partitioning with response to a specified formula (Leo et al., 1984) and can be used to build predictive models directly from observations (Evans et al., 2011; Guisan and Zimmermann, 2000). Random forest (RF, based on the R-package *randomForest*) is also non-parametric, which uses key elements of CART along with various bagging algorithms (Breiman, 2001). This approach uses a large number of trees (1000 trees, for instance) to find an optimal solution. This approach can be implemented both in classification and in function development (Breiman, 2001).

2.2.3. Model assessment

To assess the performance of distribution models developed for balsam fir, we performed 50 independent model runs based on the selection of 75% of the original input data to build the model and the remaining 25%, to test the model at the end of each bootstrapping iteration.

Model performance was assessed according to the area under the receiver operating characteristic (ROC) curve (or AUC) and true skill statistics (TSS). The AUC is a signal threshold-independent approach (Manel et al., 2001), unbiased by prevalence and interpreted as the ratio of the average sensitivity (true positive rate) to a fixed range of *1-specificity* (false positive rate), requiring the whole probability surface be calculated, providing an objective measurement of discrimination in SDM's. Models, whose AUC > 0.8, are viewed as significantly robust (Pellissier et al., 2013; Swets, 1988). Although kappa statistics are widely used to assess model performance in this type of work, kappa analysis always responds to variation in a unimodal fashion, showing an inherent dependence on prevalence, leading to biases in estimates of accuracy (Allouche et al., 2006). While TSS keeps balance between these two different types of errors (i.e., false positives and false negatives) with high correlation of sensitivity and specificity (Breiner et al., 2015), it depends on the threshold of prevalence. Choosing prevalence as a threshold can preserve true occurrence, while the results might be slightly lower for individual plots (across kappa statistics). Therefore, the number of false positives balances the number of false negatives (Allouche et al., 2006; Freeman and Moisen, 2008). We set the TSS-threshold as prevalence (ratio of plots with balsam fir present to the total number of plots; 0.527, in this case) during training (Rubidge et al., 2011). To assist with the interpretation of SDM-results later on, HS-scores were binned into five site-quality categories in order to address the range of site quality present and to ensure that the number of observations within individual bins were approximately equal. The five site-quality categories defined, include poor (with HS-scores < 0.5), low (0.5 – 0.6), moderate (0.6 – 0.7), good (0.7 – 0.8), and excellent (> 0.8).

2.3. Relating modelled HS to actual forest conditions

Site index *(SI)* and cumulative aboveground biomass of balsam fir were used as indicators of site quality in exploring the relationships between modelled HS and on-the-ground measurements of forest site quality. Increment cores were used to determine tree age from a minimum of two trees neighbouring some of the PSP's, according to species and age-class requirements. Diameter at breast height (DBH; in cm) was measured for all commercial tree species, whereas tree heights (H; in m) were collected from trees with a DBH > 9 cm (Dunlap, 1987). In this study, we used DBH of all living trees to estimate aboveground biomass (based on species-specific biomass equations given in Lambert et al., 2005), and a subset of these trees with a field-record of age at breast height and H to estimate *SI* (Ker and Bowling, 1991).

2.3.1. Site index

Variation in *SI* leads to differences in polymorphic tree H growth (Carmean, 1972). *SI*-equations are widely used throughout the world, especially in North America (Aertsen et al., 2010; Doolittle, 1958; Ker and Bowling, 1991; Watt et al., 2015). Calculating *SI* follows a simple approach by taking into account tree H and age at breast height (Carmean and Lenthall, 1989). The *SI*-equation, i.e.,

$$H = 1.3 + (SI - 1.3)[1 - exp(-b_2 50)]^{-b_3 SI^{b_4}} \bullet [1 - exp(-b_2 A)]^{b_3 SI^{b_4}}$$

(2)

is based on a generalized Chapman-Richards' equation (Clutter et al., 1983; Richards, 1959) that calculates tree height or *SI* at the index age of 50 years (Ker and Bowling, 1991); **H** in eqn. [2] is the total tree height (m), *A* is the recorded tree age at breast height in years, and are species-specific regression coefficients (Ker and Bowling, 1991). Site index in eqn. [2] is commonly solved numerically using, e.g., the Newton-Raphson or bisection-secant rule, given actual tree H and age at breast height.

2.3.2. Aboveground biomass

Biomass is one of the important factors in gauging tree competition (Bonser and Reader, 1995; Reader et al., 1994) and site productivity. Allometric equations are commonly used to estimate biomass to avoid using destructive harvesting methods (Brown, 1997). However, equations can vary from area to area, mainly in their coefficients (Chave et al., 2014). Aboveground biomass can be calculated by tree biomass equations based on DBH alone or in combination with H (Lambert et al., 2005). In this study, aboveground biomass was calculated by summing individual calculations of aboveground biomass for foliage, branches, stemwood, and bark, using DBH as the only independent variable (predictor). DBH-based equations of aboveground biomass follow those developed by Lambert et al. (2005), i.e.,

$$y_{wood} = \beta_{wood1} D^{\beta_{wood2}}$$

$$y_{bark} = \beta_{bark1} D^{\beta_{bark2}}$$

$$y_{foliage} = \beta_{foliage1} D^{\beta_{foliage2}}$$

(3)

$$y_{branches} = \beta_{branches1} D^{\beta_{branches2}}$$

$$y_{AGB} = y_{wood} + y_{bark} + y_{foliage} + y_{branches},$$

where y_i is the dry biomass of the aboveground (AGB) components (j = foliage, branches, stemwood, or bark) of living trees (kg), *D* is the DBH (cm), and β_{jk}'s are coefficients (k = 1) or exponents (k = 2; Lambert et al., 2005) in the equation set; for their species-specific values, refer to Table A1.

Table 1

Mean model performance statistics in test data, including the area under the curve (AUC, based on ROC-curves in Fig. A2), true skill statistic (TSS), deviance, and a statement of model performance (after Moradi et al., 2019).

Method	AUC	TSS	Deviance	Model Performance
GLM	0.601 ± 0.023	0.148 ± 0.042	1.357 ± 0.017	Poor
GAM	0.625 ± 0.023	0.175 ± 0.044	1.342 ± 0.029	Poor
CART	0.604 ± 0.016	0.177 ± 0.040	1.364 ± 0.027	Poor
RF	0.862 ± 0.014	0.539 ± 0.037	0.957 ± 0.028	Good

3. Results

3.1. Comparison of processing methods

Model-performance evaluation was based on the AUC and true skill statistics (TSS; Fig. A2, Table 1). In general, all measures of performance revealed that RF provided the best overall performance among the four methods considered, yielding the highest overall AUC (0.862 ± 0.014; Swets, 1988), TSS (0.539 ± 0.037), and lowest deviance (Table 1). Random forest as an extension to CART, processes an optimal solution while avoiding main drawbacks associated with CART (e.g., overfitting; Evans et al., 2011), by generating a set of weak-learners based on bootstrapping. All methods considered produced non-random patterns (AUC > 0.5), with RF offering a slightly stronger predictive response. All subsequent analyses were based exclusively on HS-scores generated with RF.

3.2. Modelling results

According to the overall importance of variables, GDD is capable of explaining a significant portion of the spatial variation in balsam fir presence-absence, with COR (i.e., '*1.0-correlation*'; Naimi and Araújo, 2016) and AUC equalling 0.422 ± 0.053 and 0.340 ± 0.033, respectively, in the test data alone (Table 2). Median difference in GDD between plots with and without balsam fir (Fig. 2) is statistically significant (median GDD of 1510.1 *vs.* 1566.3 degree-days, with $p < 0.01$; see Fig. 2b), although small (~56 degree-days). Wind speed is the second most important variable (i.e., COR = 0.332 ± 0.041 and AUC = 0.280 ± 0.030; Table 2), affecting both growing characteristics and structural stability of associated trees. The small median difference between plots representing presence-absence of balsam fir is also statistically significant ($p < 0.01$, Fig. 2d). Photosynthetically active radiation, as the third most important variable, affects less than a third of the spatial variation in balsam fir presence-absence data, with statistically-significant difference between plots of presence and absence ($p < 0.05$, Fig. 2a; Table 2). Relative extractable soil water content has the least impact on balsam fir HS, presenting no statistically-significant difference in the medians (Table 2, $p = 0.06$, Fig. 2c).

The best habitat for balsam fir (highest HS-scores) was determined to be in the highlands and northern uplands of NB and along with a narrow band adjacent to the Bay of Fundy (Fig.'s 1b and 3a). Based on long-term weather records from the Saint Leonard Airport weather station, adjacent to the highlands of the province, this area normally experiences annual temperatures and precipitation of approximately 3.5 °C and 1,104 mm, respectively. On closer inspection of the growing area (Location A, Fig. 3b), balsam fir trees are shown to benefit (high HS-scores) along the north-facing slopes of prominent valleys, close to zones of moderate SWC, adjacent to rivers and streams, and in areas where PAR and WS tend to be low (Fig. 3b). In the Grand Lake ecoregion (Fig. 1a; Location B, Fig.'s 3a and 3b), although the warmest part of the province, balsam fir will grow next to Grand Lake and surrounding areas (Fig. 3b). Likewise, balsam fir trees are projected to occupy areas adjacent to other large lakes in the southwest of the province (Location C, Fig. 3b). In contrast, HS for balsam fir is shown to drop dramatically (Fig. 3a) in the eastern lowlands of NB (Fig. 1a), where climate is warmer and soils are frequently wetter; mean annual air temperature and total precipitation in the region = 6.1 °C and 1,124 mm, respectively (Environment and Climate Change Canada, 2018).
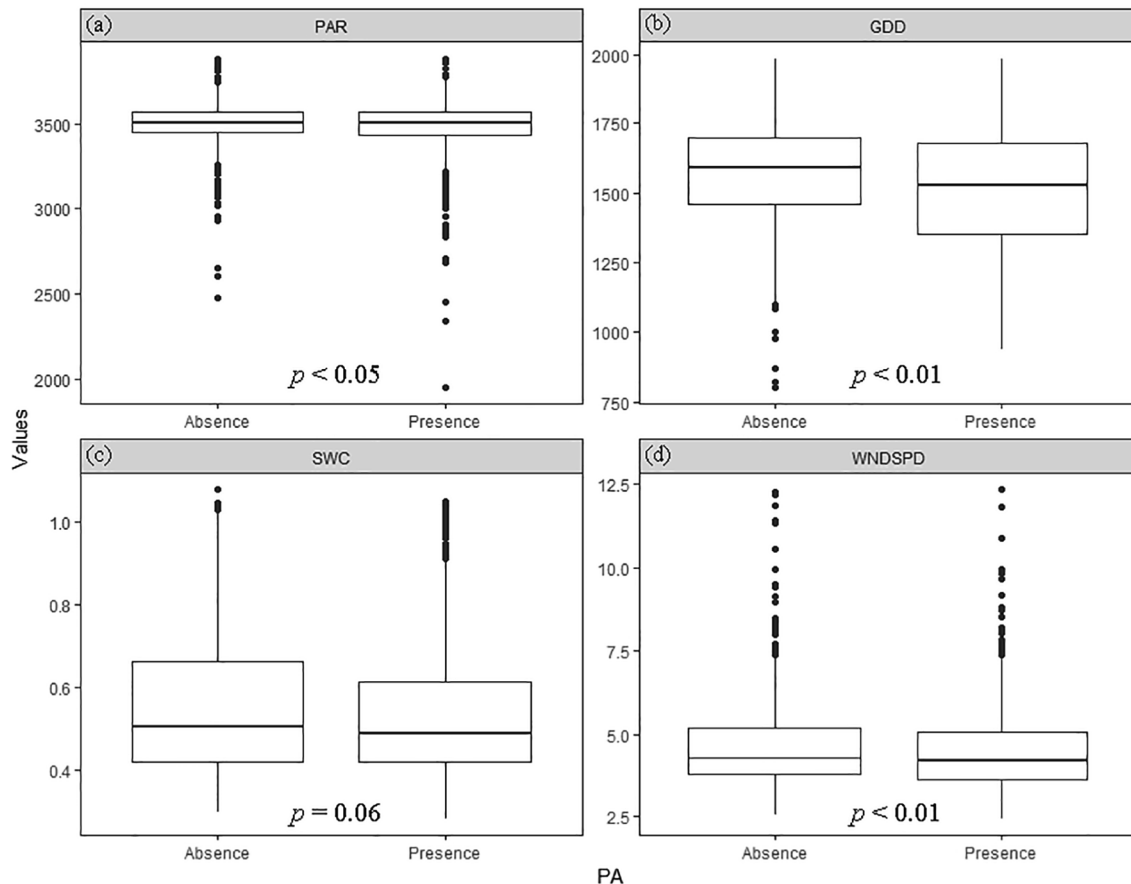
Table 2

Variable importance based on COR (i.e., '1-correlation') and AUC. Also, provided are general trends in habitat suitability (HS) as a function of the various environmental variables (column 5). Values of relative importance were generated during model training. Relative importance increases as COR and AUC increase (after Naimi and Araújo, 2016).

| Variable	Order of Relative Importance	COR	AUC	Summary
GDD	1	0.422 ± 0.053	0.340 ± 0.033	HS for balsam fir is satisfactory in low GDD areas
WS	2	0.332 ± 0.041	0.280 ± 0.030	HS for balsam fir is satisfactory in low WS areas
PAR	3	0.306 ± 0.042	0.258 ± 0.025	HS for balsam fir is satisfactory in low PAR areas
SWC	4	0.299 ± 0.032	0.253 ± 0.028	HS for balsam fir is satisfactory close to rivers and large streams and in moderately wet areas, particularly in northwestern NB (Amos-Binks et al., 2010)

3.3. Relating HS-scores to SI

Analysis of variance (ANOVA) revealed a statistically significant increase in SI with increasing HS-scores (site quality, $p < 0.001$, Fig. 4 and Table 3). Although plots with poor HS-scores contain some of the lowest SI's (median $SI = 9.49$ m), related error bars indicate considerable variation in SI (mean $SI = 10.81 \pm 4.69$ m, n = 5; Fig. 4).

Consequently, the mean difference between plots of poor and excellent HS-scores, although statistically significant, is not particularly strong, when compared with a Tukey test ($p = 0.022$, Table 3). Plots with low HS-scores, unexpectedly incorporate the third largest SI, with minor variation among values (median $SI = 12.98$ m, mean $SI = 12.67 \pm 1.05$ m, n = 19), leading to no statistically-significant difference with plots of low-to-good site quality (Fig. 4; Table 3). Plots with moderate

Fig. 2. Boxplots of environmental variables associated with balsam fir in NB. Environmental variables include (a) photosynthetically active radiation (PAR; MJ m^{-2}), (b) growing degree-days (GDD; degrees), (c) relative plant extractable soil water content (SWC; non-dimensional and values > 1.0, coincide with SWC > field capacity), and (d) wind speed (WS; m s^{-1}). The thick horizontal lines in the boxes represent the median of plotted values. The bottom and top edges of the boxes denote the 25th and 75th percentiles, respectively. The ends of the whiskers correspond to the 10th and 90th percentile. Statistical tests are based on Mann-Whitney Rank Sum tests, given absence of normality in the data.
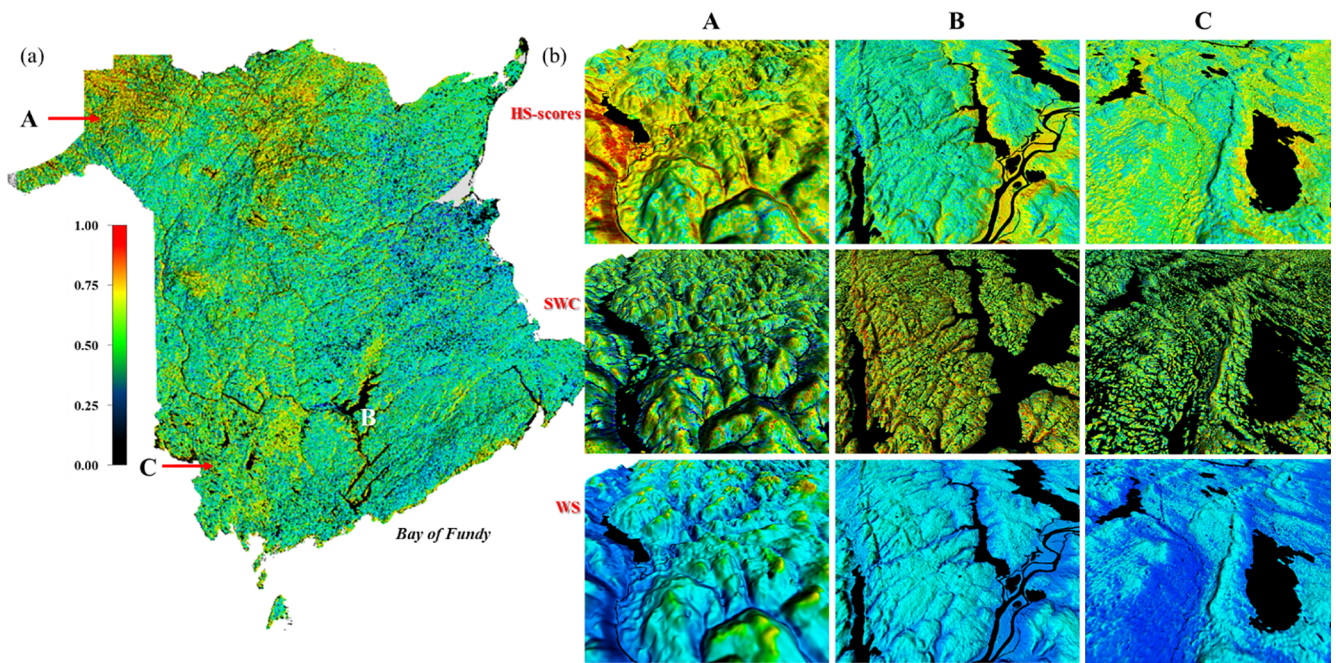
Fig. 3. Modelled habitat suitability (HS) for balsam fir for the entire province (a) and for selected areas in the province (b) labelled by letters A through C. Black patches in both sets of maps coincide with the presence of surface water. Close-ups of HS-scores for the selected areas (b) are presented together with corresponding maps of extractable soil water content (SWC) and near-surface wind speed (WS). Colours associated with the legend next to the map of the province (panel a) correspond to variation in HS-scores and probability of balsam fir being present. Reddish-brown to green, to blue colours in maps of SWC (b) coincide with progression in soil wetness from dry soils (SWC ~ 30% of field capacity) to wet soils at field capacity. Soils with SWC beyond field capacity are emphasised in black. Medium-blue colours in maps of WS identify low wind speeds, while cyan to yellowish colours especially in the high terrain of location A coincide with elevated wind speeds.
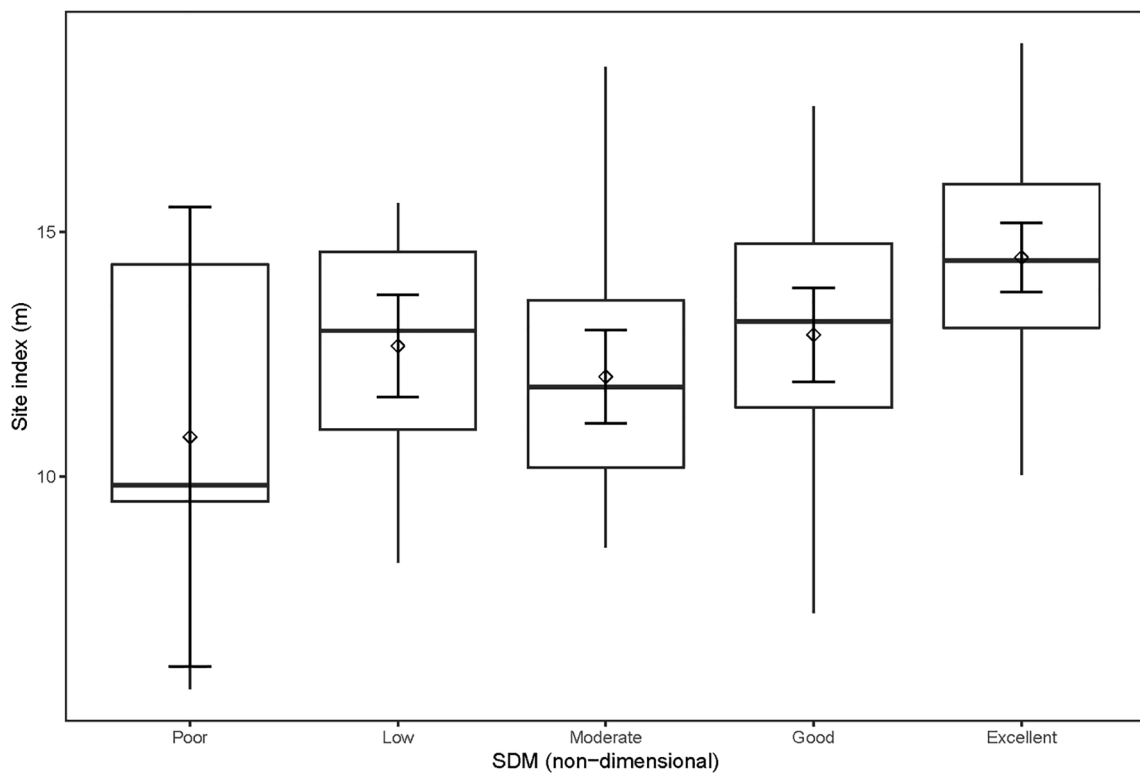
Fig. 4. Boxplots of *SI* (m) for variable ranges of *sdm*-generated HS-scores; i.e., poor (< 0.5), low (0.5–0.6), moderate (0.6–0.7), good (0.7–0.8), and excellent (> 0.8). The thick horizontal lines in the boxes represent the median of plotted values. The bottom and top edges of the boxes denote the 25th and 75th percentiles, respectively. The ends of the whiskers correspond to the 10th and 90th percentile. The open diamond symbols in the boxes represent the mean of plotted values; cap-tipped lines coincide with the lower and upper 95% confidence limits based on ANOVA.

Table 3
Probabilities (i.e., *p*) for multiple pairwise Tukey comparisons of *SI* and aboveground biomass (AGB) for individual site-quality categories from poor to excellent. Note that a column for "Excellent" site quality does not appear in the table, as this would coincide with a comparison of the site-quality category with itself.

Variable	Range of HS-scores	Poor	Low	Moderate	Good
SI	Poor	–	–	–	–
	Low	0.584	–	–	–
	Moderate	0.854	0.922	–	–
	Good	0.414	0.998	0.674	–
	Excellent	0.022*	0.08	0.002**	0.053
AGB	Poor	–	–	–	–
	Low	0.272	–	–	–
	Moderate	0.030*	0.060	–	–
	Good	0.003**	0.***	0.062	–
	Excellent	0***	0***	0***	0***

Note: ***, **, and * indicate statistical significance at *p*-values < 0.001, < 0.01, and < 0.05, respectively.

HS-scores contain the second lowest *SI* (median $SI = 11.83$ m, mean $SI = 12.04 \pm 0.95$ m, n = 26). These plots exhibit no significant difference with plots of excellent site quality (Fig. 4; Table 3). Plots of good site quality include the second highest *SI* (median $SI = 13.17$ m, mean $SI = 12.9 \pm 0.97$ m, n = 27).

3.4. Relating HS to aboveground biomass of balsam fir

Aboveground biomass (AGB) of balsam fir strongly differ among the different site-quality categories ($p < 0.05$; Table 3). Aboveground biomass in plots designated as having excellent site quality (i.e., high HS-score) is significantly greater than all other site-quality categories ($p = 0$); median AGB = 62.01 tonnes ha^{-1}, mean AGB = 64.04 $\pm$ 7.24 tonnes ha^{-1} (n = 104; Fig. 5 and Table 3). Plots with good site quality contain the second greatest amount of AGB, with a median = 33.32 and a mean = 40.22 $\pm$ 3.47 tonnes ha^{-1} (n = 288; Fig. 5), with significant difference with plots of low through poor site quality (p-value = 0.0 and 0.003, respectively). Although AGB in plots of moderate-to-poor site quality slightly decreases with HS-score, i.e., mean = 35.97 $\pm$ 2.88, 31.30 $\pm$ 3.40, and 26.60 $\pm$ 5.95 tonnes ha^{-1} (n = 346, 87, and 43, respectively; Fig. 5), only moderate-to-poor site quality shows statistical difference ($p = 0.030$; Table 3).

4. Discussion

Abundance of balsam fir is strongly influenced by management and disturbance history due to its high economic value in NB (Erdle and Ward, 2008). For current forest conditions, balsam fir is a competitive species that grows almost anywhere in the province (Fig. 3a), as in other Atlantic Provinces of Canada (Ritchie, 1996). The relationship between GDD and HS acknowledged here (Fig.'s A1b and 3a) is con-
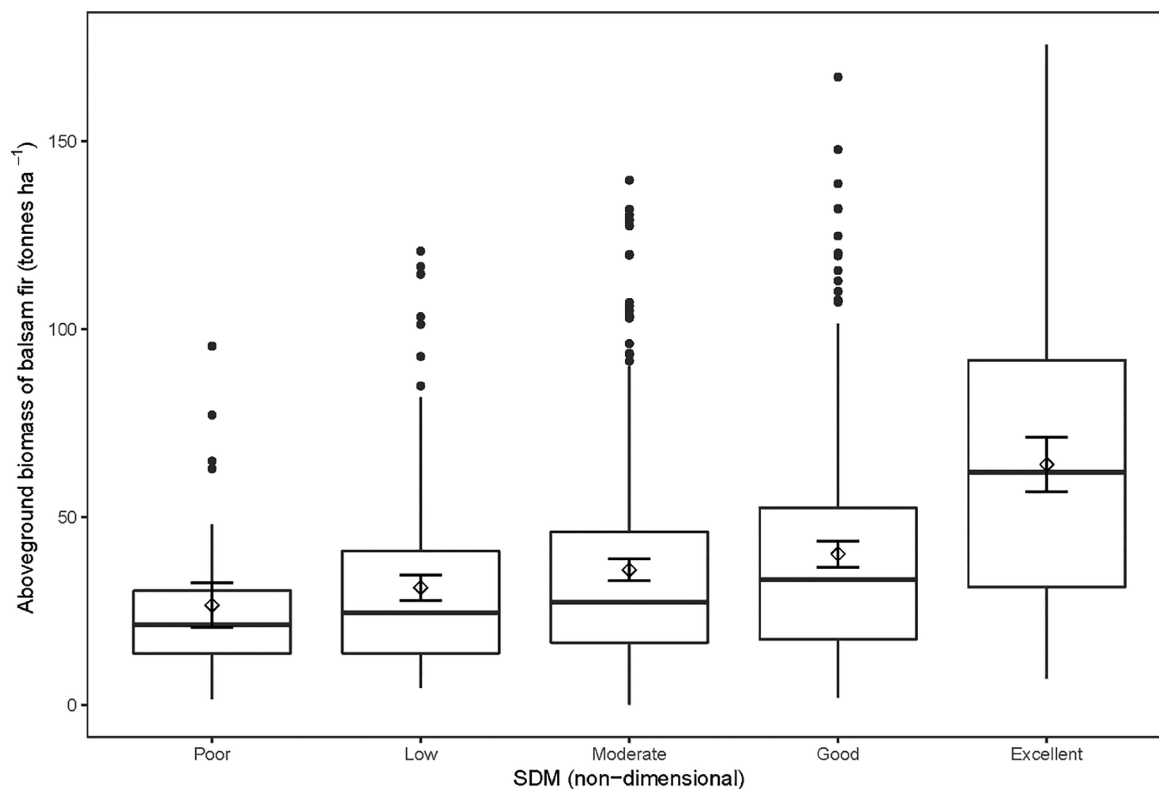
Fig. 5. Boxplots of aboveground biomass (tonnes ha^{-1}) for variable ranges of *sdm*-generated HS-scores; i.e., poor (< 0.5), low (0.5–0.6), moderate (0.6–0.7), good (0.7–0.8), and excellent (> 0.8). The thick lines in the boxes represent the median of plotted values. The bottom and top edges of the boxes coincide with the 25th and 75th percentiles, respectively. The ends of the whiskers correspond to the 10th and 90th percentile; closed circles above the whiskers represent data points beyond the 90th percentile. The open diamond symbols in the boxes represent the mean of plotted values; cap-tipped lines coincide with the lower and upper 95% confidence limits based on a Kruskal-Wallis one-way ANOVA.

sistent with other studies (Broders et al., 2006; Hennigar et al., 2008), suggesting that balsam fir trees tend to grow best in areas where cool summer temperatures persist (Fig. A1b; Table 2; Burns and Honkala, 1990).

Balsam fir on shallow, wet soils are exceptionally vulnerable to windthrow, particularly when winds are strong (Erdle and Ward, 2008). The possibility of frequent episodes of windthrow over the moderate-to-long term could potentially exclude the species from occurring in the windiest parts of the landscape (Fig. A1d). High WS along the Bay of Fundy and past harvesting practices, for instance, may have prevented balsam fir from establishing and growing in the southcentral-most part of the province, notwithstanding the occasional pockets of high HS-scores along the coast (Fig. 3a). Relative extractable soil water content is the least important variable (Table 2) most likely because of scale differences driving balsam fir distribution regionally. Extractable soil water content affects tree growth and development at the individual-tree scale, which remains mostly undetected, especially when the target resolution is 30 m. Furthermore, plant response is marginally impacted when actual extractable soil water content lies between the lower and upper limits of plant-available water, i.e., permanent wilting point and field capacity. Only at the extremes and beyond (i.e., extreme soil dryness and waterlogged conditions) does soil water content begin to have an appreciable impact on plant growth. Waterlogged conditions are a common feature of NB landscapes, with all its wetlands and lakes. Historically, droughts have happened in the province, but not frequently (Turkkan et al., 2011).

Although balsam fir habitat is fairly high-quality (high HS-scores) in northwestern and northcentral NB (Location A, Fig. 3b) due to low PAR and below average WS, balsam fir trees in the area tend to be younger, due to recurring disturbance associated with spruce budworm [*Choristoneura fumiferana* (Clem.)] infestation, episodic windstorms, commercial thinning, and other industrial operations in the region (Etheridge et al., 2006). Although the warmest region in the province, good-quality growing sites adjacent to Grand Lake and surrounding areas (Location B, Fig.'s 3a and 3b) are mostly there as a result of the lowering of growing-season temperatures brought on by the cooling effects of the Lake (Akalusi and Bourque, 2018). This pattern is replicated in many other areas adjacent to large lakes throughout the province, particularly in the southwest of the province (Location C, Fig. 3b). Unlike large deep lakes, small shallow and potentially warmer waterbodies have a limited role in controlling local climate.

In general, *SI* has a weak functional relationship with HS-scores (Fig. 4). We found some inventory plots with low HS-scores, possess *SI*'s greater than one would expect. Some studies have shown that there is sometimes greater temporal affinity between HS and species vital attributes (Pellissier et al., 2013), human density and urbanisation (Araújo, 2003), or management activities (Fois et al., 2018) that help to structure species distribution. Here, we use static, long-term occurrence data (from 1987 to 2014) in the development of the SDM, essentially to eliminate some of the temporal effects. However, their removal may not have been completely realised, introducing some level of inconsistency between data sources. This inconsistency may also, in part, be related to artefacts that may have been created because of the inherent resolution differences (i.e., from the single tree-level to 30-m grid cells) and the inadequate number of sample trees used in the determination of *SI*. Although intraspecific heights may not vary greatly across climatic gradients (Lines et al., 2012), our findings, albeit tenuous, suggest some functional relationship between tree height (and *SI*) and site factors

(Fig. 4). Some of these discrepancies may vanish with the application of DEM's of improved quality and resolution, e.g., LiDAR-data based products.

Aboveground biomass (AGB) is shown to increase with increasing HS-scores. This is particularly noticeable among the high-quality sites, illustrating an increasing trend from poor-to-excellent site quality as approximately exponential (Fig. 5; Table 3). As AGB is calculated from many trees and approximates a stand-level estimate, unlike *SI*, AGB provides a stronger relationship with site quality and consequently modelled HS-scores. Some studies have shown that relative abundance and AGB have a closer relationship with HS at small spatial scales (de Moraes Weber and Viveiros Grelle, 2012; VanDerWal et al., 2009).

Although correlation-based SDM's aim to use abiotic surfaces and statistical/machine-learning techniques to evaluate species occurrence, other characteristics of plant species, such as density (Fois et al., 2018), reproductive performance (Brambilla and Ficetola, 2012; Sergio et al., 2003), and inter- and intraspecific competition (Baah-Acheamfour et al., 2017), can be integrated into these non-dimensional values, re-defining the shared impact of abiotic variables in determining species habitat selection (Baah-Acheamfour et al., 2017). Our analysis confirms that variation in HS-scores can be used to address variation in *SI* and AGB, given current technologies and resolutions. Establishing a relationship between HS-scores and on-the-ground-forest conditions is particularly important to inform decision-making and planning in forestry and conservation of biodiversity (Hirzel et al., 2006).

5. Conclusion

RandomForest provided the best overall description of balsam fir HS spatially. Balsam fir is shown to grow best in the highlands and northern uplands of the province and in the vicinity of large deep lakes, because of the cooler summer temperatures and low seasonal wind speeds that prevail in these locations. Growing degree-day and wind speed were shown to provide a stronger control on HS regionally, than relative extractable SWC, mostly because of differences in scale dependencies. Modelled HS were partially supported by tree-based calculations of *SI*. Aboveground biomass provided a stronger association with modelled HS because AGB represents an areal integration of tree information at the plot level. Although calculations of HS were based entirely on abiotic-based inputs, the calculations provided a reasonable assessment of tree species distribution (AUC ~ 0.86), as expressed in regional and local differences in AGB of balsam fir and forest inventories.
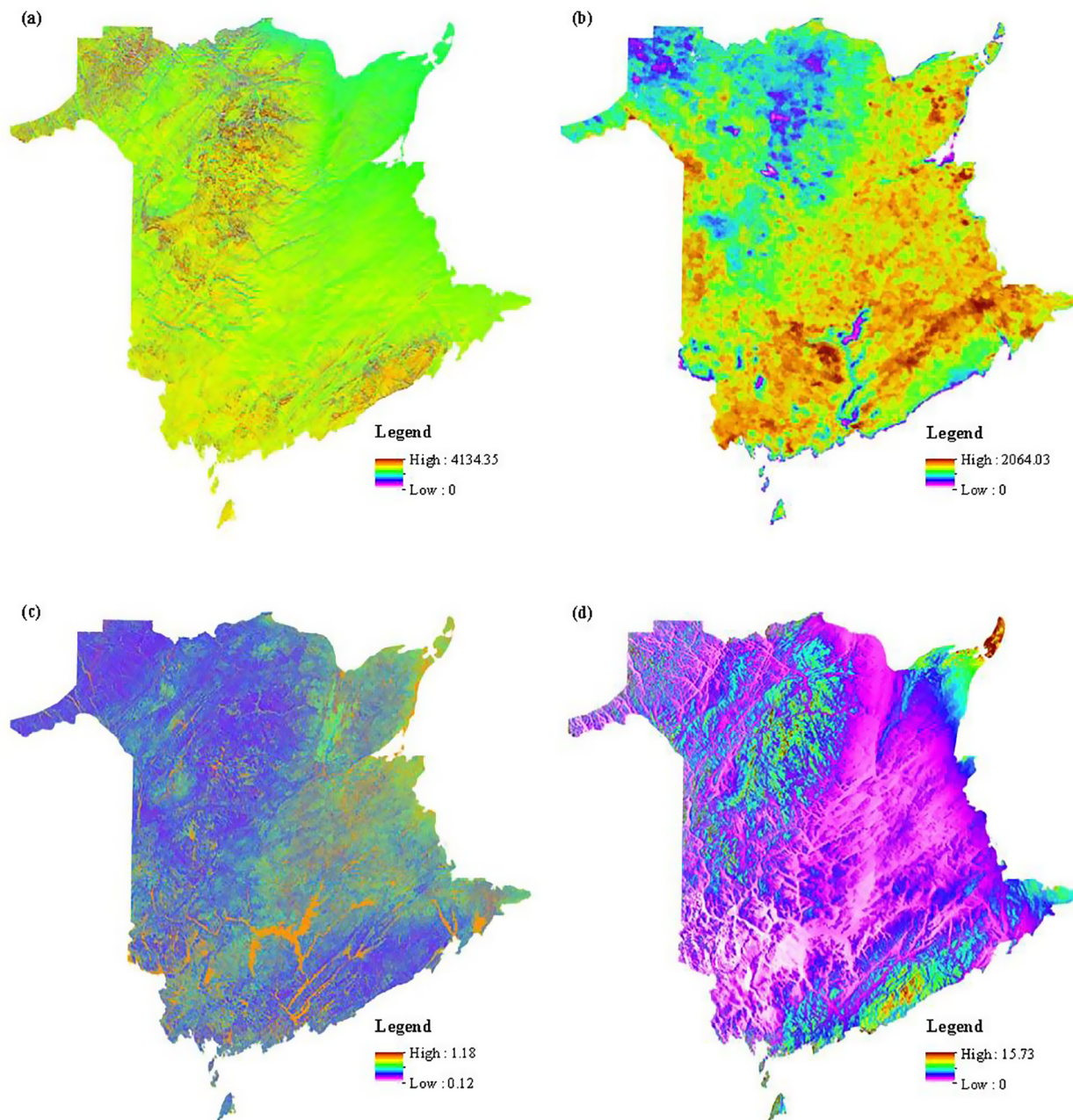
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix

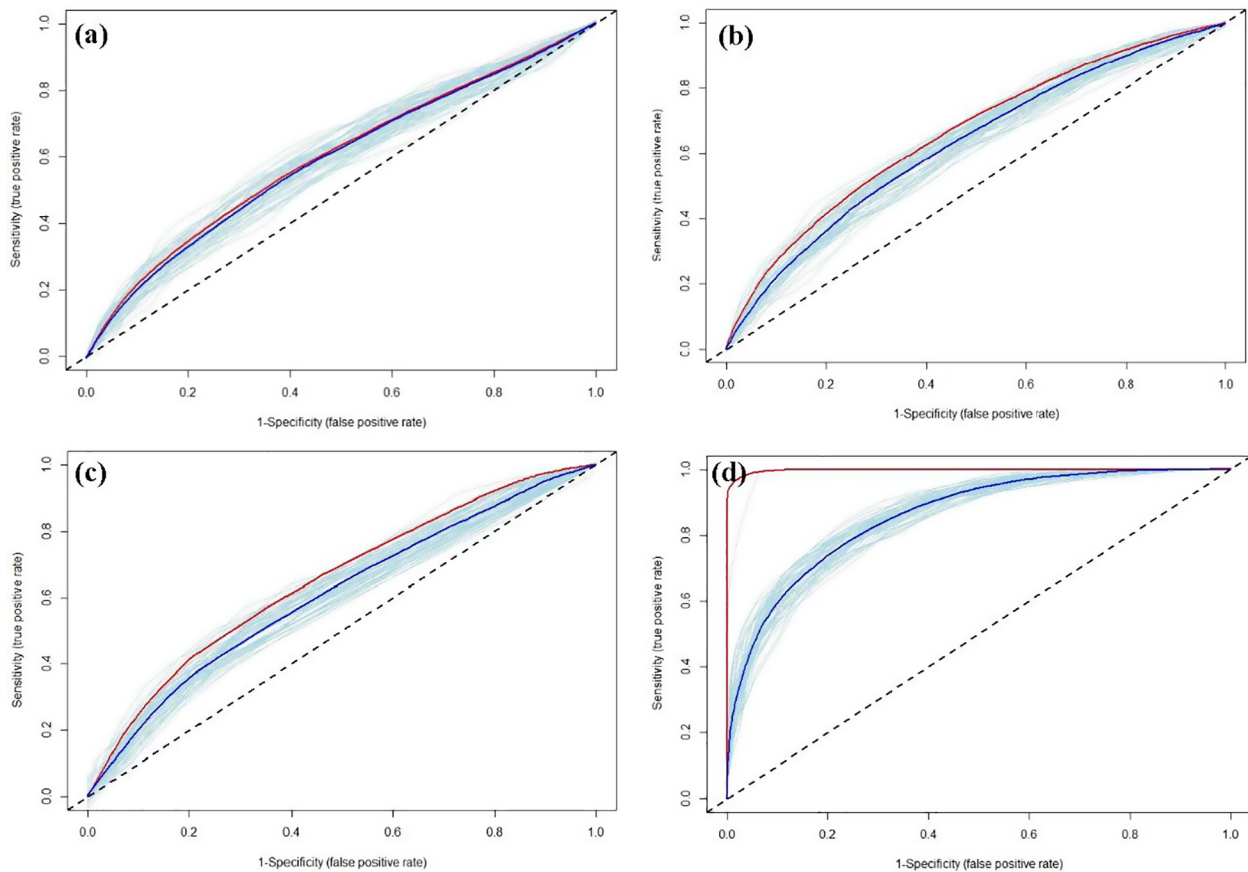See Figs. A1 and A2 and Table A1.

Fig. A1. Long-term growing-season distribution of abiotic variables, i.e., (a) photosynthetically active radiation (PAR; MJ m^{-2}), (b) growing degree-days (GDD; degrees), (c) relative plant extractable soil water content (SWC; non-dimensional and values > 1.0, coincide with SWC > field capacity), and (d) wind speed (WS; m s^{-1}).

Fig. A2. Receiver operating characteristic (ROC) curves in the application of Generalized Linear Model (GLM; a), Generalized Additive Model (GAM; b), Classification and Regression Tree (CART; c), and Random Forest classifiers (RF; d). The light blue curvilinear lines in the background correspond to ROC-curves generated by repeated application of the models by bootstrapping (n = 50 replicates). Red and dark blue lines represent mean ROC-curves for the training and validation data, respectively.

Table A1

Equation coefficients and exponents in the calculation of aboveground biomass for stemwood, bark, branches, and foliage for 33 different tree species (after Lambert et al., 2005), including balsam fir and other tree species addressed in Fig. 1b.

Species	β_{wood1}	β_{wood2}	β_{bark1}	β_{bark2}	$\beta_{branches1}$	$\beta_{branches2}$	$\beta_{foliage1}$	$\beta_{foliage2}$
Black spruce	0.0477	2.5147	0.0153	2.2429	0.0278	2.0839	0.1648	1.4143
White spruce	0.0229	2.5775	0.0116	2.3022	0.0283	2.0823	0.1601	1.467
Red spruce	0.0989	2.2814	0.0220	2.0908	0.0005	3.275	0.0066	2.4213
Cherry	0.3743	1.9406	0.0679	1.8377	0.0796	2.0103	0.084	1.2319
Balsam fir	0.0534	2.403	0.0115	2.3484	0.007	2.5406	0.084	1.6695
White pine	0.0997	2.2709	0.0192	2.2038	0.0056	2.6011	0.0284	1.9375
Jack pine	0.0804	2.4041	0.0184	2.0703	0.0079	2.4155	0.0389	1.729
Red pine	0.0564	2.4465	0.0188	2.0527	0.0033	2.7515	0.0212	2.069
Eastern white cedar	0.0654	2.2121	0.0114	2.1432	0.0322	1.9367	0.0499	1.7278
Eastern hemlock	0.0619	2.3821	0.0139	2.3282	0.0217	2.2653	0.0776	1.6995
Tamarack	0.0625	2.4475	0.0174	2.1109	0.0196	2.2652	0.0801	1.4875
Red maple	0.1014	2.3448	0.0291	2.0893	0.0175	2.4846	0.0515	1.5198
Sugar maple	0.1315	2.3129	0.0631	1.9241	0.033	2.3741	0.0393	1.693
Yellow birch	0.1932	2.1569	0.0192	2.2475	0.0305	2.4044	0.1119	1.3973
Beech	0.1478	2.2986	0.012	2.2388	0.037	2.368	0.0376	1.6164
Grey birch	0.072	2.3385	0.0168	2.2569	0.0088	2.5689	0.0099	1.8985
Red oak	0.1754	2.1616	0.0381	2.0991	0.0085	2.779	0.0373	1.674
White ash	0.1861	2.1665	0.0406	1.9946	0.0461	2.2291	0.1106	1.2277
White birch	0.0593	2.5026	0.0122	2.4053	0.0122	2.5532	0.0546	1.6221
Trembling aspen	0.0605	2.475	0.0168	2.3949	0.008	2.5214	0.0261	1.6304
Large-tooth aspen	0.0959	2.343	0.0308	2.224	0.0047	2.653	0.008	2.0149
Hardwood	0.0871	2.3702	0.0241	2.1969	0.0167	2.4807	0.039	1.6229
Ash	0.0941	2.3491	0.0323	2.0761	0.0448	1.9771	0.0538	1.2284
Balsam poplar	0.051	2.4529	0.0297	2.1131	0.012	2.4165	0.0276	1.6215
Elm	0.0402	2.5804	0.0073	2.4859	0.0401	2.1826	0.075	1.3436

Note that the **hardwood** category groups striped and mountain maple, witch hazel, willow, serviceberry, and alder together; while **ash** groups black and American mountain ash; **cherry** groups pin cherry, black cherry, and choke cherry together.

References

Acevedo, M.F., Urban, D., Ablan, M., 1995. Transition and gap models of forest dynamics. Ecol. Appl. 5 (4), 1040–1055.

Aertsen, W., Kint, V., Van Orshoven, J., Özkan, K., Muys, B., 2010. Comparison and ranking of different modelling techniques for prediction of site index in Mediterranean mountain forests. Ecol. Model. 221 (8), 1119–1130.

Akalusi, M.E., Bourque, C.P.A., 2018. Effect of climatic variation on the morphological characteristics of 37-year-old balsam fir provenances planted in a common garden in New Brunswick, Canada. Ecol. Evol. 8 (6), 3208–3218.

Allouche, O., Tsoar, A., Kadmon, R., 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. 43 (6), 1223–1232.

Amos-Binks, L.J., MacLean, D.A., Wilson, J.S., Wagner, R.G., 2010. Temporal changes in species composition of mixedwood stands in northwest New Brunswick: 1946–2008. Can. J. For. Res. 40 (1), 1–12.

Araújo, M.B., 2003. The coincidence of people and biodiversity in Europe. Glob. Ecol. Biogeogr. 12 (1), 5–12.

Araújo, M.B., New, M., 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. 22 (1), 42–47.

Austin, M., 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecol. Model. 157 (2–3), 101–118.

Austin, M., 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecol. Model. 200 (1–2), 1–19.

Baah-Acheamfour, M., Bourque, C.P.-A., Meng, F.-R., Swift, D.E., 2013. Classifying forestland from model-generated tree species habitat suitability in the Western Ecoregion of Nova Scotia, Canada. Can. J. For. Res. 43 (6), 517–527.

Baah-Acheamfour, M., Bourque, C.P.-A., Meng, F.R., Swift, D.E., 2017. Incorporating interspecific competition into species distribution mapping by upward-scaling of small-scale model projections to the landscape. PLoS One 12 (2), e0171487.

Bonser, S., Reader, R., 1995. Plant competition and herbivory in relation to vegetation biomass. Ecology 76 (7), 2176–2183.

Borregaard, M.K., Hart, E.M., 2016. Towards a more reproducible ecology. Ecography 39 (4), 349–353.

Bourque, C.P.-A., Bayat, M., 2015. Landscape variation in tree species richness in northern Iran forests. PLoS ONE 10 (4), e0121172.

Bourque, C.P.-A., Gullison, J.J., 1998. A technique to predict hourly potential solar radiation and temperature for a mostly unmonitored area in the Cape Breton Highlands. Can. J. Soil Sci. 78 (3), 409–420.

Bourque, C.P., Meng, F.-R., Gullison, J.J., Bridgland, J., 2000. Biophysical and potential vegetation growth surfaces for a small watershed in northern Cape Breton Island, Nova Scotia Canada. Can. J. For. Res. 30 (8), 1179–1195.

Box, E.O., 1981. Predicting physiognomic vegetation types with climate variables. Vegetatio 45 (2), 127–139.

Brambilla, M., Ficetola, G.F., 2012. Species distribution models as a tool to estimate reproductive parameters: A case study with a passerine bird species. J. Anim. Ecol. 81 (4), 781–787.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.

Breiner, F.T., Guisan, A., Bergamini, A., Nobis, M.P., 2015. Overcoming limitations of modelling rare species by using ensembles of small models. Methods Ecol. Evol. 6 (10), 1210–1218.

Broders, H.G., Forbes, G.J., Woodley, S., Thompson, I.D., 2006. Range extent and stand selection for roosting and foraging in forest-dwelling northern long-eared bats and little brown bats in the Greater Fundy Ecosystem, New Brunswick. J. Wildl. Manage. 70 (5), 1174–1184.

Brown, S., 1997. Estimating biomass and biomass change of tropical forests: A primer, 134. UN Food and Agriculture Organization.

Burgman, M.A., Fox, J.C., 2003. Bias in species range estimates from minimum convex polygons: Implications for conservation and options for improved planning, Animal Conservation Forum. Cambridge University Press, pp. 19–28.

Burns, R.M., Honkala, B.H., 1990. Silvics of north America, 2. United States Department of Agriculture Washington, DC.

Carmean, W.H., 1972. Site index curves for upland oaks in the Central States. For. Sci. 18 (2), 109–120.

Carmean, W.H., Lenthall, D.J., 1989. Height-growth and site-index curves for jack pine in north central Ontario. Can. J. For. Res. 19 (2), 215–224.

Carrete, M., et al., 2007. Habitat, human pressure, and social behavior: Partialling out factors affecting large-scale territory extinction in an endangered vulture. Biol. Conserv. 136 (1), 143–154.

Chave, J., et al., 2014. Improved allometric models to estimate the aboveground biomass of tropical trees. Glob. Change Biol. 20 (10), 3177–3190.

Chen, C., Weiskittel, A., Bataineh, M., MacLean, D.A., 2017. Evaluating the influence of varying levels of spruce budworm defoliation on annualized individual tree growth and mortality in Maine, USA and New Brunswick, Canada. For. Ecol. Manage. 396, 184–194.

Clutter, J.L., Fortson, J.C., Pienaar, L.V., Brister, G.H., Bailey, R.L., 1983. Timber management: A quantitative approach. John Wiley & Sons, Inc.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007.

de Moraes Weber, M., Viveiros Grelle, C., 2012. Does environmental suitability explain the relative abundance of the tailed tailless bat, Anoura caudifer. Natureza and Conservaçao 10, 221–227.

Doolittle, W.T., 1958. Site Index Comparisons for Several Forest Species in the Southern

Appalachians. Soil Sci. Soc. Am. J. 22 (5), 455–458.

Dunlap, J., 1987. Permanent sampling program annual report. Timber Management Branch, New Brunswick Department of Natural Resources and Energy, Fredericton, NB.

Elith, J., Leathwick, J.R., 2009. Species distribution models: Ecological explanation and prediction across space and time. Annu. Rev. Ecol. Evol. Syst. 40, 677–697.

Erdle, T., Ward, C., 2008. Management alternatives for New Brunswick's public forest. NB Department of Natural Resources, Fredericton, NB, Canada.

Etheridge, D.A., MacLean, D.A., Wagner, R.G., Wilson, J.S., 2006. Effects of intensive forest management on stand and landscape characteristics in northern New Brunswick, Canada (1945–2027). Landscape Ecol. 21 (4), 509–524.

Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A., 2011. Modeling species distribution and change using random forest, Predictive species and habitat modeling in landscape ecology. Springer, pp. 139–159.

Filz, K.J., Schmitt, T., Engler, J.O., 2013. How fine is fine-scale? Questioning the use of fine-scale bioclimatic data in species distribution models used for forecasting abundance patterns in butterflies. Eur. J. Entomol. 110 (2), 311–317.

Fois, M., Fenu, G., Lombrana, A.C., Cogoni, D., Bacchetta, G., 2015. A practical method to speed up the discovery of unknown populations using Species Distribution Models. J. Nat. Conserv. 24, 42–48.

Fois, M., Cuena-Lombraña, A., Fenu, G., Cogoni, D., Bacchetta, G., 2018. Does a correlation exist between environmental suitability models and plant population parameters? An experimental approach to measure the influence of disturbances and environmental changes. Ecol. Ind. 86, 1–8.

Franklin, J.L., Black, M.L., Valde, K., 2003. GPS dropwindsonde wind profiles in hurricanes and their operational implications. Weather Forecasting 18 (1), 32–44.

Freeman, E.A., Moisen, G.G., 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. 217 (1–2), 48–58.

Gama, M., Crespo, D., Dolbeth, M., Anastácio, P.M., 2017. Ensemble forecasting of Corbicula fluminea worldwide distribution: Projections of the impact of climate change. Aquat. Conserv. Mar. Freshwater Ecosyst. 27 (3), 675–684.

Gillooly, J.F., Brown, J.H., West, G.B., Savage, V.M., Charnov, E.L., 2001. Effects of size and temperature on metabolic rate. Science 293 (5538), 2248–2251.

Guisan, A., Edwards Jr, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: Setting the scene. Ecol. Model. 157 (2–3), 89–100.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2–3), 147–186.

Hassan, Q.K., Bourque, C.P.-A., Meng, F.-R., 2006. Estimation of daytime net ecosystem CO_2 exchange over balsam fir forests in eastern Canada: Combining averaged tower-based flux measurements with remotely sensed MODIS data. Can. J. Remote Sens. 32 (6), 405–416.

Hassan, Q.K., Bourque, C.P., Meng, F.-R., 2007a. Application of Landsat-7 ETM+ and MODIS products in mapping seasonal accumulation of growing degree days at an enhanced resolution. J. Appl. Remote Sens. 1 (1), 013539.

Hassan, Q.K., Bourque, C.P., Meng, F.-R., Richards, W., 2007b. Spatial mapping of growing degree days: An application of MODIS-based surface temperatures and enhanced vegetation index. J. Appl. Remote Sens. 1 (1), 013511.

Hassan, Q.K., 2008. A framework for modelling species-specific site quality index based on data generated from remote sensing imagery and a process-based model. Unpubl. PhD Dissertation, University of New Brunswick, Fredericton, New Brunswick, Canada, 164pp.

Hastie, T., Tibshirani, R., 1987. Generalized additive models: Some applications. J. Am. Stat. Assoc. 82 (398), 371–386.

Hennigar, C.R., MacLean, D.A., Quiring, D.T., Kershaw Jr, J.A., 2008. Differences in spruce budworm defoliation among balsam fir and white, red, and black spruce. For. Sci. 54 (2), 158–166.

Hirzel, A.H., Le Lay, G., Helfer, V., Randin, C., Guisan, A., 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Model. 199 (2), 142–152.

Ker, M., Bowling, C., 1991. Polymorphic site index equations for four New Brunswick softwood species. Can. J. For. Res. 21 (5), 728–732.

Lambert, M., Ung, C., Raulier, F., 2005. Canadian national tree aboveground biomass equations. Can. J. For. Res. 35 (8), 1996–2018.

Latimer, A.M., Wu, S., Gelfand, A.E., Silander, J.A., 2006. Building statistical models to analyze species distributions. Ecol. Appl. 16 (1), 33–50.

Leo, B., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and regression trees. Wadsworth International Group.

Lines, E.R., Zavala, M.A., Purves, D.W., Coomes, D.A., 2012. Predictable changes in aboveground allometry of trees along gradients of temperature, aridity and competition. Glob. Ecol. Biogeogr. 21 (10), 1017–1028.

Lopes, A., 2003. WindStation – A software for the simulation of atmospheric flows over complex topography. Environ. Modell. Software 18 (1), 81–96.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: The need to account for prevalence. J. Appl. Ecol. 38 (5), 921–931.

Maunder, M.N., Punt, A.E., 2004. Standardizing catch and effort data: A review of recent approaches. Fish. Res. 70 (2–3), 141–159.

McCullagh, P., Nelder, J.A., 1989. Generalized linear models, 37. CRC press.

Moradi, S., Ilanloo, S.S., Kafash, A., Yousefi, M., 2019. Identifying high-priority conservation areas for avian biodiversity using species distribution modeling. Ecol. Indic. 97, 159–164.

Nadrowski, K., Wirth, C., Scherer-Lorenzen, M., 2010. Is forest diversity driving ecosystem function and service? Curr. Opin. Environ. Sustain. 2 (1–2), 75–79.

Naimi, B., Araújo, M.B., 2016. *sdm*: A reproducible and extensible R platform for species

distribution modelling. Ecography 39 (4), 368–375.

Nielsen, S.E., Johnson, C.J., Heard, D.C., Boyce, M.S., 2005. Can models of presence-absence be used to scale abundance? Two case studies considering extremes in life history. Ecography 28 (2), 197–208.

Oliver, C.D. and Larson, B.C., 1996. Forest stand dynamics: Updated edition. John Wiley and sons.

Pearce, J., Ferrier, S., 2001. The practical value of modelling relative abundance of species for regional conservation planning: A case study. Biol. Conserv. 98 (1), 33–43.

Pellissier, L., et al., 2013. Suitability, success and sinks: How do predictions of nesting distributions relate to fitness parameters in high arctic waders? Divers. Distrib. 19 (12), 1496–1505.

Porter, K.B., MacLean, D.A., Beaton, K.P. and Upshall, J., 2001. New Brunswick permanent sample plot database (PSPDB v1. 0): User's guide and analysis, 209. Fredericton, NB: Atlantic Forestry Centre.

Reader, R., et al., 1994. Plant competition in relation to neighbor biomass: An intercontinental study with Poa pratensis. Ecology 75 (6), 1753–1760.

Retuerto, R., Woodward, F.I., 1992. Effects of windspeed on the growth and biomass allocation of white mustard Sinapis alba L. Oecologia 92 (1), 113–123.

Richards, F., 1959. A flexible growth function for empirical use. J. Exp. Bot. 10 (2), 290–301.

Ritchie, G.A., 1996. A Handbook of Maritime Trees. Canadian Forest Service, Maritime Services.

Rubidge, E.M., Monahan, W.B., Parra, J.L., Cameron, S.E., Brashares, J.S., 2011. The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. Glob. Change Biol. 17 (2), 696–708.

Sergio, F., Pedrini, P., Marchesi, L., 2003. Spatio-temporal shifts in gradients of habitat quality for an opportunistic avian predator. Ecography 26 (2), 243–255.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science 240 (4857), 1285–1293.

Tellería, J.L., Santos, T., Refoyo, P., Muñoz, J., 2012. Use of ring recoveries to predict habitat suitability in small passerines. Divers. Distrib. 18 (11), 1130–1138.

Tilman, D., et al., 1997. The influence of functional diversity and composition on ecosystem processes. Science 277 (5330), 1300–1302.

Tôrres, N.M., et al., 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. Divers. Distrib. 18 (6), 615–627.

Turkkan, N. et al., 2011. Floods and droughts under different climate change scenarios in New Brunswick. Report 0706-6457, DFO, Moncton, NB (Canada).

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 (2), 282–291.

Vasconcelos, T.S., Antonelli, C.P., Napoli, M.F., 2017. Mutualism influences species distribution predictions for a bromeliad-breeding anuran under climate change. Austral Ecol. 42 (7), 869–877.

Wadsworth, R., 1959. An optimum wind speed for plant growth. Ann. Bot. 23 (1), 195–199.

Watt, M.S., Dash, J.P., Bhandari, S., Watt, P., 2015. Comparing parametric and non-parametric methods of predicting Site Index for radiata pine using combinations of data derived from environmental surfaces, satellite imagery and airborne laser scanning. For. Ecol. Manage. 357, 1–9.

Weber, M.M., Stevens, R.D., Diniz-Filho, J.A.F., Grelle, C.E.V., 2017. Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. Ecography 40 (7), 817–828.

Yu, H., Jiao, Y., Carstensen, L.W., 2013. Performance comparison between spatial interpolation and GLM/GAM in estimating relative abundance indices through a simulation study. Fish. Res. 147, 186–195.

Zelazny, V.F. et al., 2007. Our landscape heritage: The story of ecological land classification in New Brunswick. New Brunswick Department of Natural Resources, Hugh John Fleming Forestry Centre, Fredericton, NB E3C 2G6, Canada.

Zimmermann, N., Edwards, T., Moisen, G.G., Frescino, T., Blackard, J., 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. J. Appl. Ecol. 44 (5), 1057–1067.

Zimmermann, N.E., Kienast, F., 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. J. Veg. Sci. 10 (4), 469–482.