Mapping Dominant Tree Species over Large Forested Areas Using Landsat Best-Available-Pixel Image Composites

Shanley D. Thompson, Trisalyn A. Nelson, Joanne C. White & Michael A. Wulder

Taylor & Francis
Taylor & Francis Group

∂ OPEN ACCESS

Mapping Dominant Tree Species over Large Forested Areas Using Landsat Best-Available-Pixel Image Composites

Shanley D. Thompson[1,*], Trisalyn A. Nelson[1], Joanne C. White[2], and Michael A. Wulder[2]

[1]Spatial Pattern Analysis and Research Lab, Department of Geography, University of Victoria, PO BOX 3060, Victoria, British Columbia V8W 3R4, Canada
[2]Canadian Forest Service (Pacific Forestry Centre), Natural Resources Canada, 506 Burnside Road West, Victoria, British Columbia V8Z 1M5, Canada

Abstract. Remotely sensed image composites that are pixel based rather than scene based are increasingly feasible to use over large areas and fine spatial resolutions. For large jurisdictions that utilize remotely sensed imagery for ecosystem mapping and monitoring, pixel-based composites enable a wider range of applications, at higher quality. The goal of this study was to model spatial distributions of 6 tree species over a large forested area of Saskatchewan, Canada (>39 million ha) at 30-m spatial resolution using a multiyear Best-Available-Pixel (BAP) Landsat composite. We tested the influence of the BAP composite on the resultant maps by comparing species composition and configuration for areas where imagery was from a single sensor, year, and day of year, to areas with variable composite characteristics. Model error rates ranged from 0.09% to 0.24%, area under the curve values approaching 1, and met ecological expectations. The BAP composite was found to have little effect on model outcomes, with composition and configuration values in nonreference areas being similar for all species but one, which had an unexpected configuration. Moreover, sensor, year, and day of year were similar for reference and nonreference blocks for all species. Results indicate that Landsat BAP image composites are useful for generating large-area maps of tree species distributions.

Résumé. Des images composites de télédétection qui sont basées sur des pixels, plutôt que sur des scènes, sont de plus en plus possibles sur des grandes superficies et à des résolutions spatiales fines. Pour les grandes régions administratives qui utilisent des images de télédétection pour la cartographie et la surveillance des écosystèmes, des composites à base de pixels permettent une plus large gamme d'applications de meilleures qualités. Le but de cette étude était de modéliser les distributions spatiales de 6 espèces d'arbres sur une grande superficie boisée de la Saskatchewan, Canada (>39 millions ha) à une résolution spatiale de 30 m en utilisant un composite pluriannuel du meilleur pixel disponible (Best-Available-Pixel; BAP) de Landsat. Nous avons testé l'influence du composite BAP sur les cartes résultantes en comparant la composition des espèces et la configuration pour les zones où l'imagerie provenait d'un seul capteur, d'une seule année et d'un seul jour de l'année, aux zones ayant des caractéristiques de composites variables. Les taux d'erreur des modèles variaient de 0,09 % à 0,24 %, les valeurs de surface sous la courbe étaient proches de 1, et ont répondu aux attentes écologiques. Le composite BAP s'est révélé avoir peu d'effet sur les résultats des modèles. Les valeurs de composition et de configuration dans les zones non-références étant similaires pour toutes les espèces, à part pour une qui avait une configuration inattendue. En outre, le capteur, l'année et le jour de l'année étaient semblables pour les blocs de référence et de non-référence pour toutes les espèces. Les résultats indiquent que les images composites BAP de Landsat sont utiles pour générer des cartes de la répartition des espèces d'arbres pour de grandes surfaces.

INTRODUCTION

Effective forest management requires knowledge of the spatial distribution of tree species composition and abundance. Species information is used to assess risks and impacts associated with a variety of natural or anthropogenic disturbances,

including fires, insect and invasive plant infestations, and resource extraction. Species composition and abundance, together with other metrics such as species richness, species endemism, and rarity are also important metrics of biodiversity that can be used to guide conservation planning (Fleishman et al. 2006) and ecosystem service assessment (Kremen 2005). In addition, climate change may alter the distribution of tree species in the future (Coops and Waring 2010; Hamann and Wang 2006; Pfeifer-Meister et al. 2013; Thuiller et al. 2005), and knowledge of the current distribution is the first step in attempting to understand, monitor, and as possible, manage those changes.

Despite the importance and various needs for tree species distribution data, the availability of these data is limited. Some

projects might benefit from the tree observations and vegetation plot data compiled and shared via online databases such as the Global Biodiversity Information Facility[1] and the Global Index of Vegetation-Plot Databases.[2] However, these data remain spatially incomplete and biased toward easily accessible or protected areas (García Márquez et al. 2012; Hortal et al. 2007). Expert range maps (e.g., Little 1971) provide a general indication of where species occur, but overestimate the true distribution of a species (Jetz et al. 2007; McPherson and Jetz 2007). Strategic-level forest inventories are typically undertaken in areas that have the capacity to support commercial timber production. Outside these areas, forest inventory data might be available, generally with less spatial and attributional detail, according to the forest monitoring needs in the area. In Canada, for example, forest inventories are common in the more intensively managed southern forests (Falkowski et al. 2009), whereas only the sample-based National Forest Inventory aims to systematically characterize forest resources outside of managed forest areas (Gillis et al. 2005; Wulder, Kurz, et al. 2004).

The paucity of detailed inventory data makes satellite remote sensing a necessary source of information from which species distributions can be mapped or modeled over large areas. High-spatial-resolution imagery can offer opportunities for mapping individual tree structure and composition (Wulder, Hall, et al. 2004), but the spatial image extents are limited (e.g., 10 by 10 km) requiring many images (with variable view angles and illumination conditions) to map a given area, leading to high cost. Imagery with larger extents are often of interest for mapping larger areas; however, these data come with pixel sizes that subsume many individual objects and landscape features, diminishing the variance between pixels, and limiting the capacity to map high levels of categorical detail, such as tree species composition. Issues such as atmospheric contamination (i.e., clouds, haze) further limit scene availability and confound mapping efforts. Several recent advances address some of these shortcomings and offer increased capacity to use remotely sensed data for predictive species mapping over large areas at higher spatial resolutions. Specifically, as of 2008, satellite imagery from the Landsat series of sensors, extending from 1972 to the present, are freely available to the public (Woodcock et al. 2008; Wulder et al. 2012). Free and open access to analysis-ready Landsat data has enabled considerable innovative capacity (Wulder and Coops 2014). Combined with improved computing power that facilitates large-area image compositing approaches (Roy et al. 2010) as well as the applications-focused, best-available-pixel (BAP) approaches of Griffiths et al. (2007) and White et al. (2014), spatially exhaustive coverage of large areas at a spatial resolution of 30 m in a systematic and transparent fashion is now possible. For instance, compositing approaches can be based on use of the best available observation for each pixel, with "best"

being defined according to a set of scores for characteristics such as year, day of year (DOY), distance to cloud and cloud shadow, and sensor (Griffiths et al. 2007). Regional composites of medium spatial resolution imagery can be expected to become increasingly common (Griffiths et al. 2014). Detailed, efficient (large-area) maps of species distributions are a likely product of these compositing approaches, as long as the models are robust to some composite-imposed spectral variability. The goal of this research is to generate spatially detailed (30 m) distribution maps for 6 tree species over a large forested area of Canada using a multiyear best-available-pixel (BAP) Landsat composite. A specific objective was to evaluate the impact of composite characteristics (i.e., DOY, year, and sensor) on model outcomes, through assessment of the spatial pattern (composition and configuration) of the predictions.

METHODS

Study Area

The study area is approximately 39 million ha, and comprises the 3 forested ecozones of Saskatchewan, Canada. From North to South, these are the Taiga Shield, the Boreal Shield, and the Boreal Plains (Ecological Stratification Working Group 1996; Figure 1). The Boreal Plains is adjacent to the Prairie ecozone further south, and consists of rolling uplands and plains with a mixture of deciduous and coniferous vegetation species (McLaughlan et al. 2010). The Boreal Shield and Taiga Shield are characterized by a harsh climate and poorer soils, a greater proportion of coniferous tree species, and a lower diversity of plant species (McLaughlan et al. 2010; Pastor et al. 1996). A variety of provincial forest inventory data exist in the managed forest area of the province (Figure 1), with varying spatial scales, levels of attribution, and temporal frequencies (Gillis et al. 2005; Saskatchewan Environment—Forest Service 2004; Saskatchewan Ministry of Environment 2009).

Tree Species Distribution Data

Tree species distribution data were acquired from Canada's National Forest Inventory (NFI). The NFI consists of a grid of permanent sample plots distributed across the country, the majority of which are 2 km x 2 km "photo plots" (i.e., derived from air photo interpretations), within which multiple polygons are delineated indicating species composition and relative abundance. To reduce uncertainty at polygon edges, we removed 30 m (one pixel) from the inner edge of each polygon to ensure agreement between the inventory data and our predictor data (Verbyla and Hammond 1995). To ensure adequate sample size and quality, we did not model species that occurred relatively infrequently or that occurred exclusively in polygons of heterogeneous composition (≤90% of one species). In addition, polygons that were observed to have burned or been harvested since they were inventoried were removed, as were polygons that were very small in size (less than one pixel). Thus, of the 11 tree

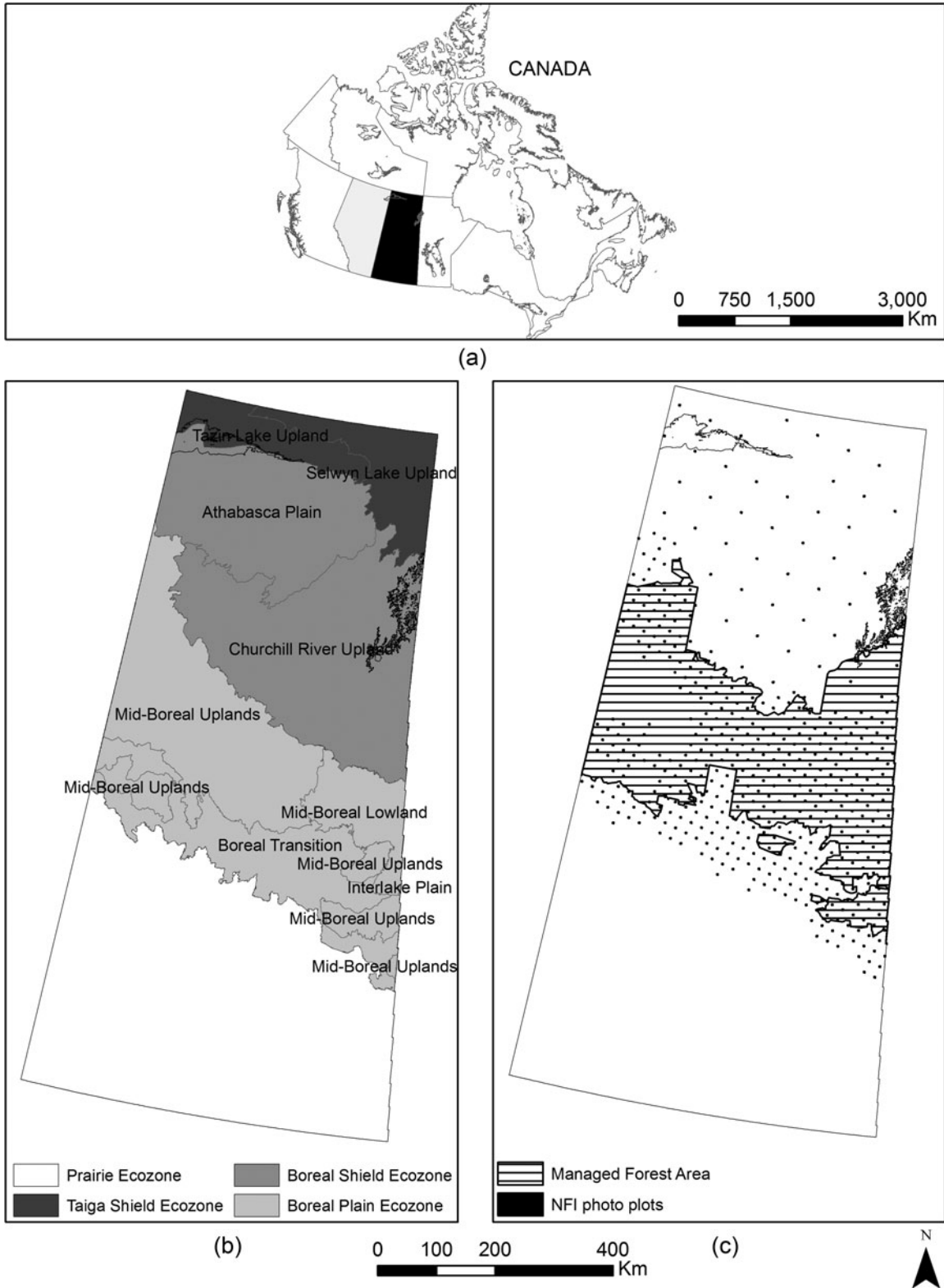FIG. 1. Study area. (a) The province of Saskatchewan (highlighted in black), in central Canada. (b) The forested ecozones and ecoregions of Saskatchewan. (c) The distribution of Canada's National Forest Inventory 2 km × 2 km photo plots across Saskatchewan, including inside and outside of the Managed Forest Area. These inventory data were the source of the training data used to model tree species distributions.

TABLE 1
Species modeled, area and number of National Forest Inventory Photo Plot (NFI PP) polygons dominated by each

English Name	Latin Name	Dominant (present at ≥90% abundance)[a]		Nondominant (present at ≤10% abundance)[a]
		Number of Polygons	Area (km^2)	Number of Polygons
White birch	*Betula papyrifera*	29	0.71	5404
Tamarack	*Larix laricina*	459	27.11	4974
White spruce	*Picea glauca*	49	0.73	5348
Black spruce	*Picea mariana*	2057	55.81	3376
Jack pine	*Pinus banksiana*	856	38.63	4577
Trembling aspen	*Populus tremuloides*	1983	88.48	3450

[a]Count follows exclusion of very small polygons (less than 900 m^2) and those that were burned or harvested in the time since the polygon was delineated and attributed.

species identified in the NFI photo-plot data, we modeled the 6 most common (Table 1) : black spruce (*Picea mariana*), trembling aspen (*Populus tremuloides*), jack pine (*Pinus banksiana*), white birch (*Betula papyrifera*), tamarack (*Larix laricina*), and white spruce (*Picea glauca*). Species that were present in insufficient numbers for modeling were *Abies balsamea, Acer negundo, Fraxinus pennsylvanica, Pinus contorta,* and *Populus balsamifera.* Our 6 target species can be found across all 3 forested ecozones in the province (McLaughlan et al. 2010).

Image Composite Data

We used a *multiyear* BAP surface reflectance composite as the source of spectral information in our distribution models. A detailed description of this and other compositing methods is provided in White et al. (2014). Briefly, candidate pixel observations were scored according to sensor (Landsat TM or ETM+), year, DOY, distance to clouds or cloud shadows, and haze, and the pixels with the highest score were used to populate the final image composite. Our target was Landsat 5 TM imagery from August 1, 2010; however, candidate pixels included all observations acquired ± 30 days of August 1, 2009, 2010, and 2011, from Landsat 5 TM or 7 ETM+, as required to provide complete, cloud-free coverage of the study area. In Table 2 we show the number of unique images considered and selected for the final BAP composite for Saskatchewan as a whole (encompassing our study area as well as the Prairie Ecoregion). After scoring, 5% of pixel observations in the final composite of our study area were acquired from 2009 imagery, 69% from 2010 imagery, and 2% from 2011 imagery. The remaining 24% of pixels had BAP observations for both 2009 and 2011, and a proxy value was generated by taking the average of the 2009 and 2011 observations (see White et al. 2014). Pixels with proxy values were excluded from the analyses because there were no logical corresponding DOY or sensor values with which to assess relationships. In total, 95% of pixel observations were from Landsat 5 TM, and 5% from Landsat 7 ETM+. Almost 30% of pixel observations came within 7 days of the target DOY

(August 1), with the remaining pixels acquired within 30 days of August 1. In the context of this study, which was designed to model tree species distributions, it is important to note that the majority of pixels (by area) in the composite came from imagery acquired in the 2010 target year and, furthermore, that tree species distributions tend to change slowly over longer time horizons. As such, the multiyear image composite used in this study is appropriate for modeling species distributions.

In order to reduce undesirable variability in the spectral reflectance values of the predictor variables used in our models (see Loveland et al. 1991), we calculated several spectral indices from the image composite data and used these to exclude nonvegetated areas and, to the extent possible, nonforested areas from our analyses (Figure 2). Specifically, the Normalized Difference Vegetation Index (NDVI), and the Tasseled Cap (Crist and Cicone 1984; Kauth and Thomas 1976) Greenness (TCG), Brightness (TCB), and Wetness (TCW) indices were used to remove water, bare ground, urban areas, and sparsely vegetated areas. As well, the normalized burn ratio (NBR; Key and Benson 2006), was used to identify and remove areas that had experienced fire in recent years, with the threshold NBR value chosen for this analysis (0.15) validated by data from the Canadian National Fire Database (Canadian Forest Service 2013). An unsupervised classification of the Landsat spectral bands for the remaining pixels served to remove additional areas subsequently identified as cultivated land.

Topographic Data

Topographic data were acquired from the freely available Canadian Digital Elevation Data.[3] These elevation data are derived from provincial and national topographic data sources, and are provided as a 1:50,000 digital elevation model (DEM). The DEM, which has a native spatial resolution of approximately 23 m (0.75 arc seconds), was resampled to match the 30 m

_____
[3]http://www.geobase.ca/geobase/en/data/cded/

TABLE 2
Number of Landsat images used in best-available-pixel composite for Saskatchewan, Canada

Year	Candidate*			Composite		
	TM	ETM+	Total	TM	ETM+	Total
2009	379	328	707	345	188	533
2010	362	367	729	342	284	626
2011	406	398	804	333	218	551

*Candidate images are ± 30 days of August 1 with less than 70% cloud cover.

spatial resolution of our image composite using bilinear resampling. From the 30 m DEM we calculated slope (in degrees), the Topographic Solar Radiation Index (TRASP; a transformed measure of aspect), and the Topographic Wetness Index (TWI; Table 3).

Species Distribution Modeling

Species distribution modeling was conducted using Random Forests (RF) in R 3.1 (Breiman 2001). We chose RF, a type of decision tree, because it can accommodate nonnormal responses and nonlinear relationships and automatically account for interactions among predictors (De'ath and Fabricus 2000; Elith et al. 2008; Hawkins 2012). Decision trees involve a sequence of binary splits at values of the predictor variables that result in the maximum differentiation of values of the response variable (in this case, species dominance or nondominance at a given location). In RF, many (500 to 2000) single trees are developed, each constructed from a different bootstrapped sample of the training data and a randomly selected subset of the predictor variables (Prasad et al. 2006). The predictions are averaged over all trees to generate an overall probability while minimizing the chances of over-fitting to the training data (Franklin 2009; Prasad et al. 2006). Ensemble tree methods have been found to perform well, relative to most other predictive methods across many regions and species, including plants (Elith et al. 2006; Guisan, Zimmerman, et al. 2007; Prasad et al. 2006).

The use of RF for species distribution modeling involved both fitting and prediction stages. First, each model was fit using the species observation data from Table 1 and the mean of each predictor variable. The predictor variables used in the modeling were selected from among the multiple aforementioned topographic and spectral variables after conducting a Spearman's rank correlation analysis and assessing variable utility through boxplots for each predictor variable across all species. Specifically, we selected the following 3 spectral and 2 topographic indices for use as inputs, all with correlations of less than ± 0.15: the Tasseled Cap TCG, TCB, and TCW, the TRASP, and the TWI (Table 3). We used 500 decision trees, with a random subset of 2 of the explanatory variables chosen for input for each of these individual trees. Because of the unbalanced number of observations of dominance and nondominance in the training data for a given species (Table 1), we used a *down-sampling*

approach, specifying that each model should use all samples from the least common class (dominance), and an equal number of samples from the more common class (nondominance; Chen et al. 2004). A separate RF model was generated for each species, and thus, the number of samples varied across models.

Each output RF model contained probabilities of dominance ranging from 0 to 1, which are classified by default into the binary classes dominance or nondominance, if probabilities are ≥0.5 or <0.5, respectively. However, often a threshold probability other than 0.5 is preferred (Nenzén and Araújo 2011), particularly for rare species (Freeman and Moisen 2008). To choose appropriate threshold probabilities for our models, we generated a receiver operating characteristic (ROC) plot for each species using the auc.roc.plot() function in the PresenceAbsence library in R (v3.1.2). An ROC plot shows how the rate of true positives (*y*-axis) versus false positives (*x*-axis) of a model varies for all threshold probabilities between 0 and 1. A perfect classification would pass through the upper-left corner of the plot (100% true positives and 0% false positives). A threshold value that achieves the minimal distance between this place of perfect classification and the curve is an appropriate value to transform continuous outputs to binary classifications (Liu et al. 2005).

Finally, each of the 6 models were re-run using the values of the predictor variables for locations (pixels) where species observations were absent and the predictions of dominance and nondominance output as continuous raster surfaces using the determined thresholds. A secondary goal was to create a forest composition map with all species combined. A composite species map was generated by evaluating the individual probabilities of dominance resulting from each of the 6 RF models were compared at each location (pixel) and that species with the highest overall probability was selected as the appropriate classification value for that location.

Model Evaluation

Model performance was assessed using the *out-of-bag* (OOB) error generated internally by the RF method, eliminating the need for a separate cross-validation (Breiman 2001). Specifically, the ability of the classifier to correctly predict observed values was assessed, and discrete class predictions were based
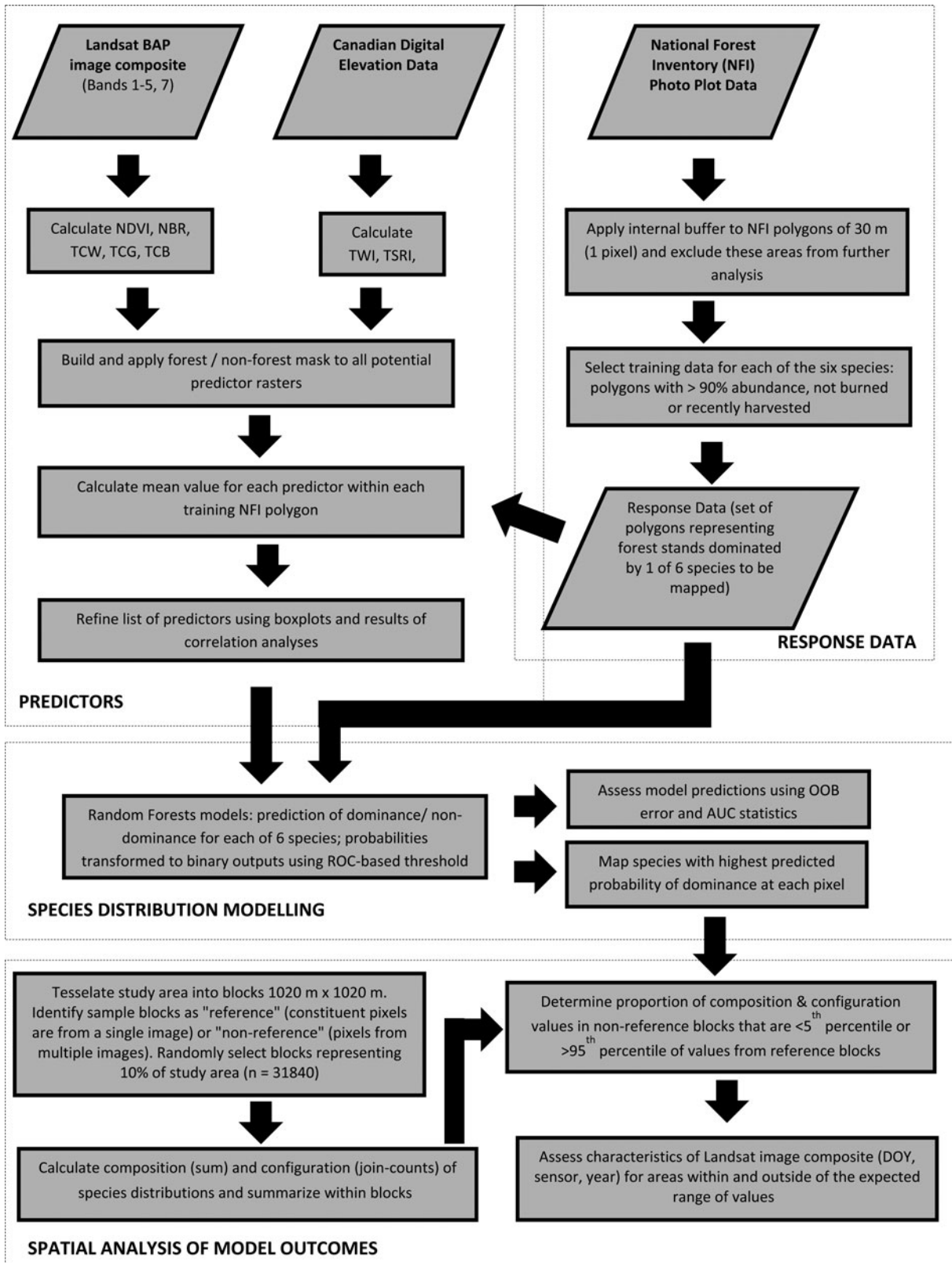
FIG. 2. Flowchart of the data and methods followed to model species distributions and asses the impact of BAP composites of the resultant maps.

TABLE 3
Predictor variables used in the Random Forests models of individual tree species distributions

Name	Description	Rationale
Brightness	Tasseled Cap components (Crist and Cicone 1984; Crist 1985)	TC brightness can differentiate between soil and vegetated surfaces (Crist et al. 1986) and can help differentiate among successional stages/stand age (Song et al. 2007).
Greenness		Greenness relates to biomass and vegetation vigor and is highly correlated to the Normalized Difference Vegetation Index (NDVI), which is useful for general land cover classification (Defries et al. 1995; Running et al. 1994)
Wetness		TC Wetness correlated to structural complexity (Hansen et al. 2001), perhaps particularly for successional stages/stand age (Wulder, Skakun, et al. 2004).
Topographic Wetness Index (TWI)	Model of potential surface moisture, based on topographic position (Beven and Kirkby 1979): *Ln(specific catchment area/tan(slope in radians))*	Soil moisture directly affects plant growth.
Topographic Solar Radiation Aspect Index (TRASP)	Values range from 0 to 1, with 0 indicating cool, NE slopes, and 1 indicating warm SW slopes (Roberts and Cooper 1989).	Solar radiation affects soil moisture and heat load, and directly affects plant growth.

on species-specific probability thresholds determined through the ROC analysis described. We also examined the area under the curve (AUC) associated with an ROC plot, as calculated by back-predicting on our observed data (essentially, but not exactly the same data used for training because of the subsampling and consensus approach used in RandomForests). The AUC ranges from 0.5 to 1 and indicates the proportion of times that the model discriminates between our 2 outcomes as better than random (Jiménez-Valverde 2012), or more specifically, the proportion of times a randomly chosen instance of dominance has a value larger than that for a randomly chosen instance of nondominance (Fielding and Bell 1997). Thus, higher AUC values indicate better models.

Finally, we compared our resultant tree distribution models with previous studies, general knowledge of the species' ranges, and trends of dominance per ecozone and ecoregion. In particular, we calculated the areal extent of each dominant species as predicted in our overall forest composition map per ecozone and ecoregion and ranked these in descending order. We repeated this calculation using homogenous polygons from NFI photo photo-plot data to assess agreement between model outputs and the training data over a broader scale. Similarly, we generated

regional summaries of species dominance by combining predictions of a relative basal area for the same species from a recent study by Beaudoin et al. (2014).

Assessing the Effects of the Image Composite on Spatial Patterns of the Models

To assess the impact that compositing had on predicting species distributions, we tested the hypothesis that the spatial pattern of predictions was similar for reference sample blocks and nonreference sample blocks. Reference blocks had pixel observations derived from a single year, a single DOY, and a single sensor, whereas nonreference blocks had pixel observations from multiple years, DOYs, and sensors. Spatial patterns can be quantified by a combination of composition and configuration. Whereas composition is *aspatial* and refers to the variety and (relative) abundance of different features (e.g., tree species), configuration refers both to the spatial characteristics of individual patches such as size and shape, as well as spatial relationships among neighboring patches or neighboring cells (Gustafson 1998). Although composition indicates what is present at any given location, configuration metrics provide

TABLE 4

Threshold used to translate probabilities of species dominance to a binary dominant or nondominant variable. The threshold selected minimized the distance on a plot of the ROC curve between the upper-left corner of the plot and the curve.

Species	Threshold Used to Create Binary Map[a]	OOB Error Rate[b] (for (for class "dominant")	Area Under the Curve (AUC)[c]
Betula papyrifera	0.7	0.24	0.99
Larix laricina	0.7	0.12	0.99
Picea glauca	0.7	0.24	0.99
Picea mariana	0.6	0.11	1
Pinus banksiana	0.7	0.14	1
Populus tremuloides	0.6	0.09	1

[a]According to the criteria of minimizing the distance on a plot of the ROC (receiver operating characteristic) curve between the upper-left corner of the plot and the curve.
[b]The Out of Bag (OOB) error indicates the total number of misclassified data points from within the out-of-bag sample (Breiman 2001).
[c]Values close to 1 indicate good model fit.

a context to local conditions and permit study of the manner in which spatial patterns are an expression of process (Fahrig 2005; Turner 1989).

The analysis was undertaken within sample blocks measuring 1020 m × 1020 m, distributed over a random 10% of the study area (for a total of 31,840 samples). An extent of 1020 m × 1020 m was chosen to ensure coverage of the data gaps (of 1 to 14 pixels in size, see Goward et al. 2010; Storey et al. 2005) resulting from Landsat 7 ETM+ Scan-Line Correction failure, while also being a number within which 30 m Landsat pixels could be equally divided. Three image composite characteristics were evaluated: acquisition year (2009, 2010, or 2011), sensor (TM or ETM+), and the number of days from the target DOY of August 1 (ranging from 0 to 30).

For each species, composition was quantified within each 1020 m × 1020 m block as the sum of pixels with predicted dominance for that species. All blocks had dominance of at least 1 species. Configuration was measured for each species using join counts within each 1020 m × 1020 m area. A join count test can be used to assess spatial autocorrelation in categorical, especially binary, variables such as dominance/nondominance (Boots 2006). Using a join count, the spatial configuration of a species can be quantified as clustered or dispersed, relative to complete spatial randomness (O'Sullivan and Unwin 2010). For binary data, the 2 categories are normally referred to as either "Black" (B) or "White" (W) (here, dominance or nondominance, respectively). For this analysis, we were interested only in the J_{BB} (dominance-dominance) join-count statistic:

$$J_{BB} = \frac{1}{2}\left(\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\delta_{ij}X_iX_j\right)\frac{1}{2}\left(\sum_{\substack{i=1 \\ i\neq j}}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}\delta_{ij}X_iX_j\right),$$

where *i* and *j* are the 2 sampling units being compared, x_i is the value of the sampling unit (1 or 0), and δij is the adjacency of *i* and *j* (1 when they are adjacent, 0 when they are

not). The expected values of joins are then calculated based on the proportion of each category and number of total joins in the study (which depends on how connectivity is defined), and the observed and expected values are then compared to assess the null hypothesis of complete spatial randomness (Fortin and Dale 2005). We computed a join count for each species, using the Rook's case definition of contiguity (4 neighbors). Each species had to be predicted within at least two 30-m cells within each 1020 m × 1020 m block to be included in the analysis. Missing data within each 1020 m × 1020 m block were reclassified as zeroes to allow calculation of the join-counts, while remaining statistically conservative.

To assess whether the use of an image composite affected our modeled species distributions, we compared values of composition and configuration between reference ($n = 23,581$) and nonreference ($n = 8259$) sample blocks. We calculated the frequency distribution of the composition and configuration values for each of the 6 species from the reference sample blocks, and extracted the 5th and 95th percentiles for each. The number of composition and configuration values in the nonreference sample blocks that fell below the 5th or above the 95th percentiles was then calculated. Blocks with these unexpected values were then mapped and a summary of their sensor type, year, and DOY characteristics was extracted.

RESULTS

Species Distribution Modeling

The ROC threshold optimization method resulted in a threshold of 0.7 for all species except for *Picea mariana* and *Populus tremuloides,* for which the optimum threshold was 0.6 (Table 4). These thresholds were used to map the distribution of each individual tree species (Figure 3). Combining the individual species maps creates an overall map of forest composition (Figure 4). *Pinus banksiana* was predicted to dominate over the largest spatial
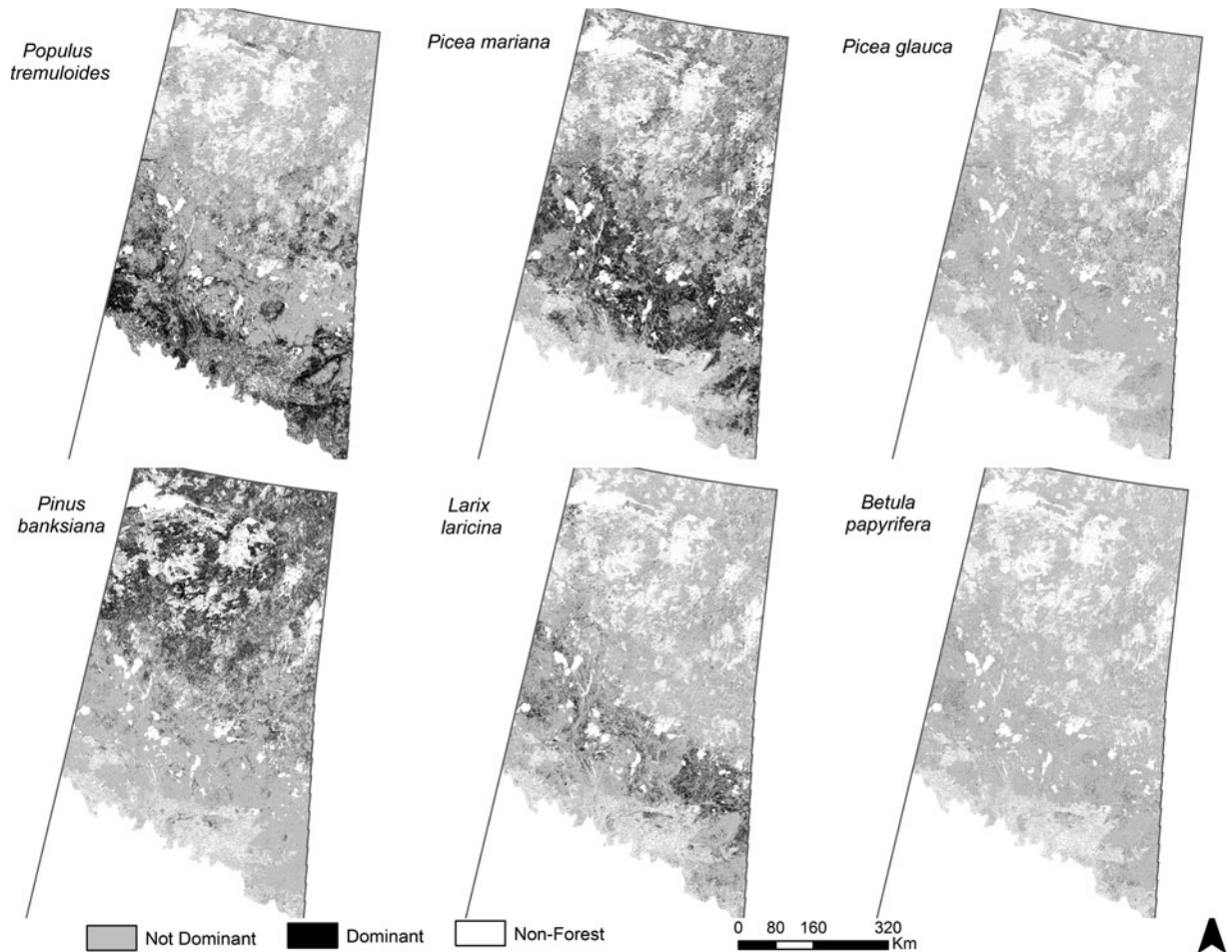
FIG. 3. Individual models of tree species dominance in Saskatchewan.

extent (8.9 million hectares in total), particularly at mid-to-high latitudes (Figure 4). *Populus tremuloides* was predicted to dominate in the southern extreme of the Boreal Plain ecozone, in the Aspen Parkland ecoregion, but is also found across the province even in the far north, with predicted dominance covering 7.3 million ha in total. *Picea mariana* was predicted to be the next most widespread species, dominating over 6.6 million ha, particularly at mid-latitudes, but being also widespread in the north. *Larix laricina* was predicted to be dominant at low-to-mid latitudes in the Boreal transition ecoregion and in lowland areas and known wet areas such as the Saskatchewan River delta, straddling the eastern border of Saskatchewan, covering 3.4 million ha in total. *Picea glauca* and *Betula papyrifera* were predicted to dominate with considerably less extent (approximately 1.3 million ha and 824,000 ha, respectively).

Model Evaluation

The ability of our models to correctly classify the dominance or nondominance of individual species varied from species to species (Table 4). Overall, OOB error rates were less than 25%, indicating reasonable model fit to the training data. Error rates were much lower for species with sufficient sample sizes. Specifically, at the selected thresholds, *Populus tremuloides* (with one of the highest sample sizes) had the lowest OOB error at 9%. *Betula papyrifera* and *Picea glauca* (with the 2 lowest sample sizes) had the highest OOB error rates at 24% each. AUC values were very high for all models (0.99 to 1), indicating good model performance.

Trends in the relative areal extent of species dominance for the province's forested ecozones and ecoregions predicted in this study are generally comparable to those in the NFI photo-plot database, as well as to those of species occurrence from Beaudoin et al. (2014) (Table 5). For example, all 3 studies/datasets indicated that both the Boreal Plain Ecozone and Boreal Transition Ecoregion are dominated by *Populus tremuloides*. All 3 datasets also suggest *Pinus banksiana* is dominant in the Athabasca Plain ecoregion. Some differences are also apparent. For instance, the current study predicts *Pinus banksiana* to be dominant across the largest proportion of the Taiga Shield, whereas the other datasets indicate that *Picea mariana* is most dominant.
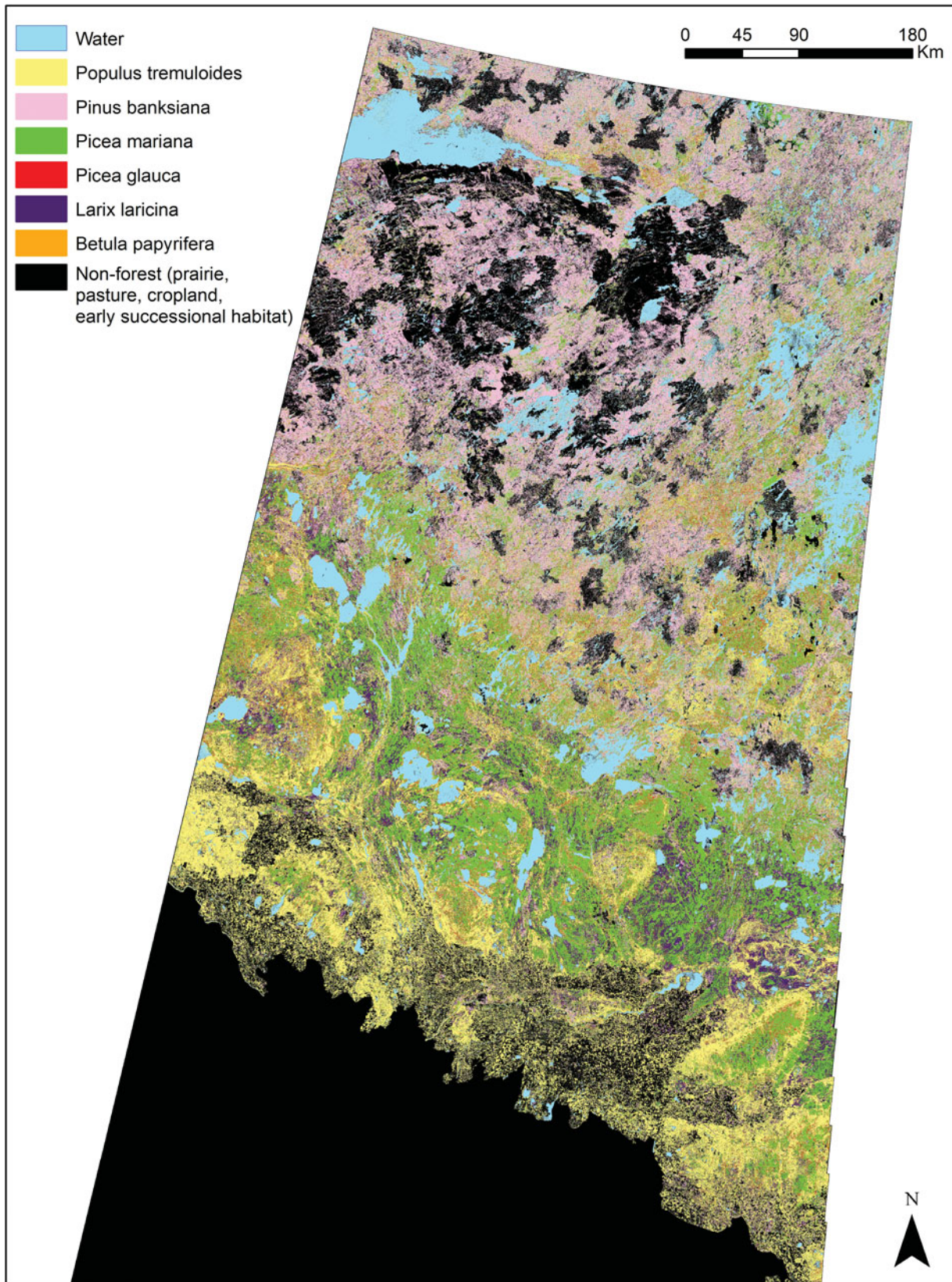
FIG. 4. Forest composition map showing tree species with highest predicted probability of dominance at each location.

TABLE 5
Spatial extent of dominant tree species in Saskatchewan's forested ecozones and ecoregions

		Most Spatially Expansive Species		
Ecozone	Ecoregion	NFI photo plot data[1]	Beaudoin et al. 2014[2]	This study
Taiga Shield		*Picea mariana* (70% of ecoregion)	*Picea mariana* (80% of ecoregion)	*Pinus banksiana* (61% of ecoregion)
	Tazin Lake Upland	*Picea mariana*	*Picea mariana*	*Pinus banksiana*
	Selwyn Lake Upland	*Picea mariana*	*Picea mariana*	*Pinus banksiana*
Boreal Shield		*Pinus banksiana* (48% of ecoregion)	*Picea mariana* (49% of ecoregion)	*Pinus banksiana* (46% of ecoregion)
	Athabasca Plain	*Pinus banksiana*	*Pinus banksiana*	*Pinus banksiana*
	Churchill River Upland	*Picea mariana*	*Picea mariana*	*Pinus banksiana*
Boreal Plain		*Populus tremuloides* (48% of ecoregion)	*Populus tremuloides* (48% of ecoregion)	*Populus tremuloides* (40% of ecoregion)
	Mid-Boreal Uplands	*Populus tremuloides*	*Picea mariana*	*Picea mariana*
	Boreal Transition	*Populus tremuloides*	*Populus tremuloides*	*Populus tremuloides*
	Mid-Boreal Lowlands	*Larix laricina*	*Picea mariana*	*Larix laricina*

[1]Calculations included only homogenous polygons (those of which ≥90% of the polygon comprises a single species).

[2]As in the generation of our forest composition map, each individual species distribution map from Beaudoin et al. 2014 was combined and assigned the value of the species with the highest predicted relative basal area. Calculations in this table are based on the combined map.

Assessing the Effects of the Image Composite on Spatial Patterns of the Models

The proportion of composition and configuration values within the nonreference sample blocks falling outside the expected distribution (5th to 95th percentiles of values of the reference blocks) was fairly low. Depending on the species, 3.7% to 9.6% of the blocks had unexpected composition values, whereas 7.1% to 16.5% had unexpected configuration values (Table 6). Given that we set the critical value of the statistical comparison to 0.10, we would expect around 10% of blocks to have unexpected composition and configuration. Only the configuration of *Populus tremuloides* had a higher than statistically expected number of unexpected blocks (16.5%). The spatial distribution of the sample blocks with unexpected composition and configuration values appears random (Figure 5). Further, the DOY, year, and sensor characteristics in these expected and unexpected regions were found to be very similar. For instance, all of these nonreference sample blocks, regardless of whether they had expected or unexpected values of species composition and configuration, comprised primarily imagery from 2010, with the difference in proportions of 2010 imagery between expected and unexpected blocks ranging from ∼2% to 9%, depending on the species. Likewise, all blocks contained imagery primarily from Landsat 5; blocks with unexpected values of composition and configuration differed in terms of sensor composition by no more than 9% from blocks with expected values. Mean DOY differed by no more than 2 days for blocks with expected and unexpected samples of species composition and configuration.

DISCUSSION

The use of satellite imagery for vegetation mapping is a desirable supplement to ground- or photo-based inventories because of the large spatial extents that can be covered by satellite

TABLE 6
Proportion of composition and configuration values within non-reference sample blocks *(n = 8259)** that are <5th or >95th percentile of values within reference sample blocks *(n = 23,581)*

	Betula papyrifera	*Larix laricina*	*Picea glauca*	*Picea mariana*	*Pinus banksiana*	*Populus tremuloides*
True sample size*	6177	7361	5588	7966	7446	7675
% unexpected Composition	3.71	9.96	5.71	9.63	7.13	8.33
True sample size*	4469	6517	4274	7682	6677	7481
% unexpected Configuration	7.14	9.56	7.25	8.69	8.62	16.53

*Note that the nonreference sample size was fewer than 8259 for each species because for each species, in turn, blocks had to contain at least 1 pixel of presence (dominance) for analysis of composition, and at least 2 pixels of presence (dominance) for analysis of configuration.
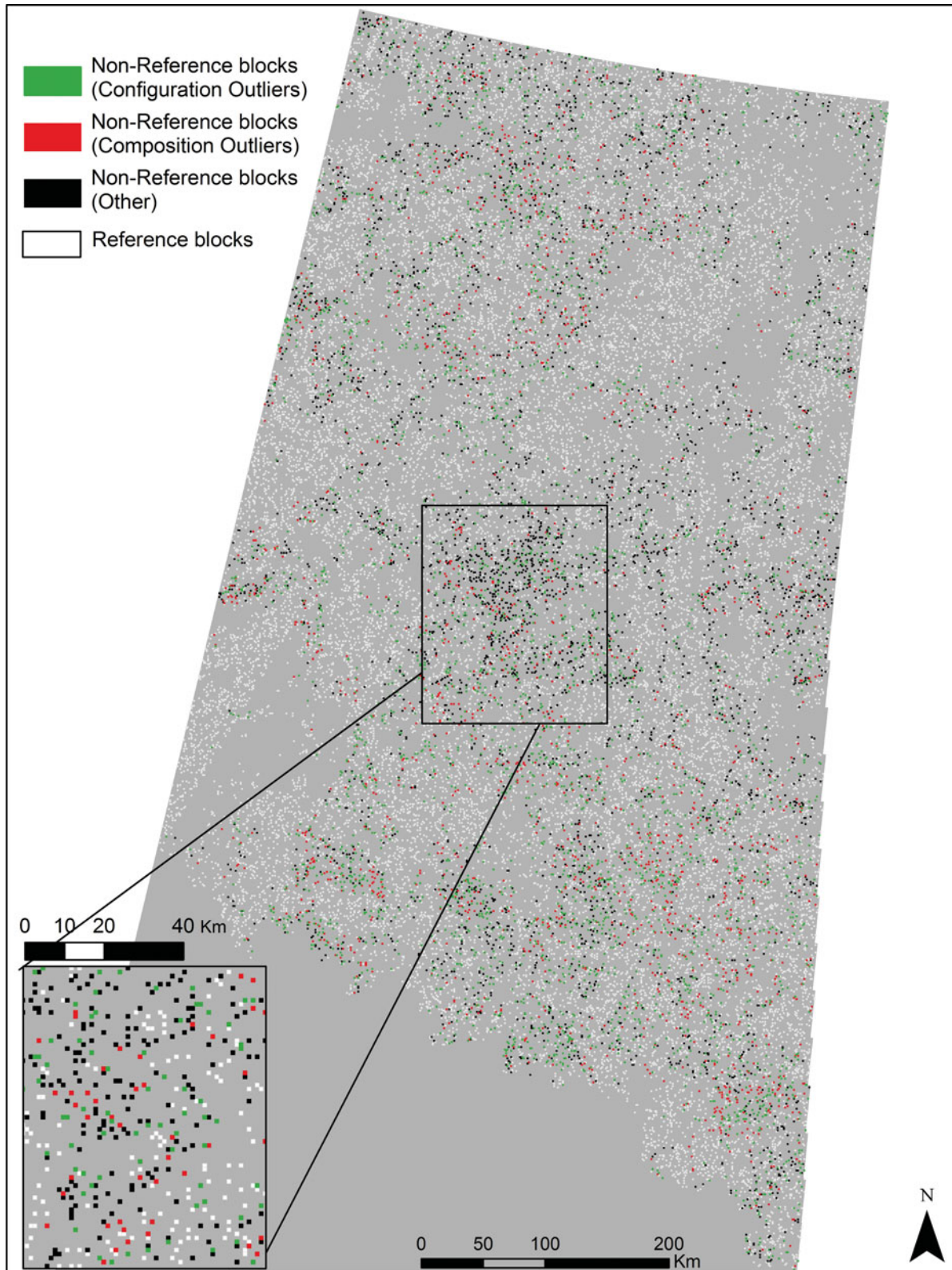
FIG. 5. Nonreference sample blocks (1020 m × 1020 m) are those comprising a mixture of sensor types, image years, or image DOY. Blocks are highlighted where species composition and configuration values were outside the 5th–95th percentile of values from reference sample blocks.

imagery, as well as the associated automated, repeated acquisition. Tree species distribution mapping based on satellite imagery involves the detection or classification of separate spectral reflectance signatures for each species (Bradter et al. 2011); however, tree species classification is difficult, because many vegetation species have overlapping spectral reflectance characteristics in the wavebands collected by typical multispectral sensors (Immitzer et al. 2012; van Aardt and Wynne 2001). Hyperspectral imagery, which collects reflectance in many narrow wavebands, is often needed to map species composition to a high, or even modest, degree of accuracy (e.g., Buddenbaum et al. 2005; Ustin and Xiao 2001). However, this type of imagery is not currently cost effective for inventorying large regions because, like high-spatial-resolution data, hyperspectral imagery are associated with small spatial extents, requiring multiple scenes or, more likely, airborne collections, to represent a given area, thereby increasing data costs and processing overhead. In this study we mapped the probable distribution of dominance of 6 tree species across a large region using freely available moderate spatial resolution multispectral imagery. Our model error rates (~10%–25%) were typical for forest species distribution modeling using this type of imagery (e.g., Evans and Cushman 2009).

Some error and uncertainty in our distribution models can be attributed to limited sample size. Sample size of the species data has been shown to affect the accuracy of predictive models in previous research (Stockwell and Peterson 2002; Wisz et al. 2008). Although machine learning and ensemble methods such as Random Forests can perform relatively well with small-to-moderate sample sizes, especially when absence information is available in addition to presence information (Elith et al. 2006; Guisan, Graham, et al. 2007), estimates of error in this study were nonetheless highest for the species with the lowest sample sizes (Table 1). Model accuracies could also have been affected by characteristics of the individual species modeled. Specifically, wide-ranging species are typically more challenging to model than species with more particular niches (Guisan, Zimmermann, et al. 2007; McPherson et al. 2004). The 6 tree species modeled in this study are all wide-ranging species, and are generally tolerant of a range of soils and parent materials (Farrar 1995). That our training data indicated where a species was and was not dominant was therefore likely particularly important. Indeed, distribution models are particularly robust when reliable absence data are available in addition to presence data (Brotons et al. 2004).

Another potential source of uncertainty in large-area mapping and modeling relates to the remotely sensed data itself. In multitemporal image analysis, differences in atmospheric conditions, and variability in phenology, sun angle, and view angle of imagery (Song and Woodcock 2003) can lead to some uncertainty. Moreover, relationships between species and image spectral reflectances will vary seasonally according to species phenology (Maeda et al. 2014). In this study, we explored the use of a multiyear BAP composite to generate a series of distribution

models for the 6 most common tree species in the forested area of Saskatchewan. We found that mean acquisition year, DOY, and sensor were similar regardless of the level of local complexity found in the composite. In other words, the variability of image characteristics across the BAP composite was actually very small, which was achievable due to the vast archive of open-source Landsat imagery (White and Wulder 2013), the rules used for compositing, and the preprocessing applied that converted the data to surface reflectance (White et al. 2014).

Composition and configuration of the predicted species are important characteristics to consider in species distribution modeling because spatial pattern is an expression of underlying spatial processes (Nelson and Boots 2008). Spatial pattern analysis is used to assess model error because it allows patterns in error or uncertainty to be detected and enables departures from random noise to be determined (Wulder et al. 2007). The approach used in this study allows mapping and detection of statistical departures in patterns of species distributions generated from nonreference imagery (Nelson and Boots 2005). We found that in all but 1 instance (*Populus tremuloides*) the composition and configuration of species distributions was not different among sample blocks with variable composite characteristics. This species in particular might stand out from the rest simply because of its overall prominence across the study area. Nonetheless, because the majority of our models were unaffected by the compositing, we are confident that the compositing rules used relating to sensor type, target DOY, and cloud contamination minimized illumination and phenological differences sufficiently across space. Overall, results indicate that the relation between predicted species distributions and important environmental processes are represented, rather than species' spatial patterns being the result of data artifacts in the composite. Overall, our individual maps of the probable distribution of tree species dominance across Saskatchewan meet expected trends as captured through 2 other independent data sources. For instance, broad-leafed species *Populus tremuloides* and *Betula papyrifera* were predicted to be dominant over a greater spatial extent in the south, relative to the north. The species predicted to be most widespread were *Picea mariana*, *Pinus banksiana*, and *Populus tremuloides*. These 3 species are typically dominant over a large number of ecosites in the province (McLaughlan et al. 2010). The primary difference among the 3 datasets is this study's predicted dominance of *Pinus banksiana* in the Taiga Shield, versus the dominance of *Picea mariana* in this region in the other datasets. We note that there are very few NFI plots in this ecozone, however, and that the differences between our results and those of Beaudoin et al. (2014) likely stem from the use of different remotely sensed data (MODIS) with a different spatial resolution (250 m), and a different modeling approach (kNN).

CONCLUSION

Regional-scale, spatially comprehensive maps of forest composition have traditionally been limited by the mismatch

between desired, versus available, spatial extent and spatial resolution of data. In this analysis, we have demonstrated the capacity to use 30-m Landsat data to map detailed tree species distributions over large areas by capitalizing on archived, multitemporal imagery composited using the BAP approach (White et al. 2014). The variability introduced by the BAP compositing was found to be minimal and resulted in mostly insignificant differences in this large-area mapping application. Future applications will benefit from analyses of the effects of BAP compositing in other geographic regions and for features of interest other than tree species. The potential of the BAP approach to provide source data for developing species distribution maps over large areas will continue to increase with Landsat continuity and the launch of complementary satellites such as Sentinel-2 in 2015 (Drusch et al. 2012; Roy et al. 2014). The constellations of these new satellites have been designed such that, taken together, the majority of Earth will be able to be imaged twice weekly (Wulder and Coops 2014) at a 30-m spatial resolution. Spatially continuous maps of tree species distributions over large areas will be useful for a variety of information needs, including forest management, carbon modeling, ecosystem service assessment, and conservation planning.

REFERENCES

Beaudoin, A., Bernier, P.Y., Guindon, L., Villemaire, P., Guo, X.J., Stinson, G., Bergeron, T., Magnussen, S., and Hall, R.J. 2014. "Mapping attributes of Canada's forests at moderate resolution through k NN and MODIS imagery." *Canadian Journal of Forest Research*, Vol. 44: pp. 521–532.

Beven, K.J., and Kirkby, M.J. 1979. "A physically based, variable contributing area model of basin hydrology." *Hydrological Sciences Bulletin*, Vol. 24: pp. 43–69.

Boots, B. 2006. "Local configuration measures for categorical spatial data: binary regular lattices." *Journal of Geographical Systems*, Vol. 8(No. 1): pp. 1–24.

Bradter, U., Thom, T.J., Altringham, J.D., Kunin, W.E., Benton, T.G. 2011. "Prediction of National Vegetation Classification communities in the British uplands using environmental data at multiple spatial scales, aerial images and the classifier random forest." *Journal of Applied Ecology*, Vol. 48: pp. 1057–1065.

Breiman, L. 2001. "Random forests." *Machine Learning*, Vol. 45(No. 1): pp. 5–32.

Brotons, L., Thuiller, W., Araújo, M., and Hirzel, A. 2004. "Presence-absence versus presence-only modelling methods for predicting bird habitat suitability." *Ecography (Cop.)*, Vol. 27: pp. 437–448.

Buddenbaum, H., Schlerf, M., and Hill, J. 2005. "Classification of coniferous tree species and age classes using hyperspectral data and geostatistical methods." *International Journal of Remote Sensing*, Vol. 26(No. 24): pp. 5453–5465.

Canadian Forest Service. 2013. *Canadian National Fire Database - Agency Fire Data. Natural Resources Canada, Canadian Forest Service, Northern Forestry Centre, Edmonton, Alberta.* http://cwfis.cfs.nrcan.gc.ca/en_CA/nfdb.

Chen, C., Liaw, A., and Breiman, L. 2004. *Using random forest to learn imbalanced data*. Technical Report 666, University of California, Berkeley, California.

Coops, N.C., and Waring, R.H. 2010. "A process-based approach to estimate lodgepole pine (*Pinus contorta*, Dougl.) distribution in the Pacific Northwest under climate change." *Climate Change*, Vol. 105: pp. 313–328.

Crist, E.P. 1985. "A TM tasseled cap equivalent transformation for reflectance factor data." *Remote Sensing of Environment*, Vol. 17(No. 3): pp. 301–306.

Crist, E.P., and Cicone, R.C. 1984. "Aphysically-based transformation of thematic mapper data - the TM tasseled cap." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-22: pp. 256–263.

Crist, E.P., Laurin, R., and Cicone, R.C. 1986. "Vegetation and soils information contained in transformed thematic mapper data." In *Proceedings of IGARSS' 86 Symposium. ESA SP-254*, pp. 1465–1470. Paris, France: European Space Agency.

De'ath, G., and Fabricus, K.E. 2000. "Classification and regression trees: a powerful yet simple technique for ecological data analysis." *Ecology*, Vol. 81: pp. 3178–3192.

Defries, R., Hansen, M., and Townshend, J., 1995. "Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data." *Remote Sensing of Environment*, Vol. 54(No. 3): pp. 209–222.

Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., and Bargellini, P. 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sensing of Environment*, Vol. 120: pp. 25–36.

Ecological Stratification Working Group. 1996. *A National Ecological Framework for Canada*. Ottawa/Hull. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research, and Environment Canada, State of the Environment Directorate, Ecozone Analysis Branch.

Elith, J., Graham, C., Anderson, R., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L., Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Tonwsend, A., Phillips, S.J., Richardson, K.S., Schachetti-Pereira, R., Schaprie, R., Soberon, J., Williams, S., Wisz, M., and Zimmermann, N.E. 2006. "Novel methods improve prediction of species' distributions from occurrence data." *Ecography (Cop.)*. Vol. 29: pp. 129–151.

Elith, J., Leathwick, J.R., and Hastie, T., 2008. "A working guide to boosted regression trees." *Journal of Animal Ecology*, Vol. 77: pp. 802–813.

Evans, J.S., and Cushman, S.A. 2009. "Gradient modeling of conifer species using random forests." *Landscape Ecology*, Vol. 24: pp. 673–683.

Fahrig, L. 2005. "When is a landscape perspective important?" In *Issues and Perspectives in Landscape Ecology*, edited by J. Wiens and M. Moss, pp. 3–10. Cambridge UK: Cambridge University Press.

Falkowski, M.J., Wulder, M.A., White, J.C., and Gillis, M.D. 2009. "Supporting large-area, sample-based forest inventories with very high spatial resolution satellite imagery." *Progress in Physical Geography*, Vol. 33: pp. 403–423.

Farrar, J. 1995. *Trees in Canada*. Ottawa: Fitzhenry and Whiteside Ltd.

Fielding, A.H., and Bell, J.F. 1997. "A review of methods for the assessment of prediction errors in conservation presence/absence models." *Environmental Conservation*, Vol. 24: pp. 38–49.

Fleishman, E., Noss, R., and Noon, B. 2006. "Utility and limitations of species richness metrics for conservation planning." *Ecological Indicators*, Vol. 6: pp. 543–553.

Fortin, M.-J., and Dale, M. 2005. *Spatial Analysis: A Guide for Ecologists*. Cambridge, UK: Cambridge University Press.

Franklin, J., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge UK: Cambridge University Press.

Freeman, E.A., and Moisen, G.G. 2008. "A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa." *Ecological Modelling*, Vol. 217: pp. 48–58.

García Márquez, J.R., Dormann, C.F., Sommer, J.H., Schmidt, M., Thiombiano, A., Da, S.S., Chatelain, C., and Dressler, S. 2012. "A methodological framework to quantify the spatial quality of biological databases." *Biodiversity & Ecology*, Vol. 4: pp. 25–39.

Gillis, M.D., Omule, A.Y., and Brierley, T. 2005. "Monitoring Canada's forests: The National Forest Inventory." *Forestry Chronicle*, Vol. 81(No. 2): pp. 214–221.

Goward, S., Williams, D., Arvidson, T., and Irons, J. 2010. "The future of Landsat-class remote sensing." In *Land Remote Sensing and Global Environmental Change: NASA's Earth Observing System and the Science of ASTER and MODIS*, edited by B. Ramachandran, C.O. Justice, and M. Abrams. New York: Springer.

Griffiths, P., Kuemmerle, T., Baumann, M., Radeloff, V.C., Abrudan, I.V., Lieskovsky, J., Munteanu, C., Ostapowicz, K., and Hostert, P. 2014. "Forest disturbances, forest recovery, and changes in forest types across the Carpathian ecoregion from 1985 to 2010 based on Landsat image composites." *Remote Sensing of Environment*, Vol. 151: pp. 72–88.

Griffiths, P., Linden, S. Van Der, Kuemmerle, T., and Hostert, P. 2007. "A pixel-based Landsat compositing algorithm for large area land cover mapping." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 6(No. 5): pp. 1–14.

Guisan, A., Graham, C., Elith, J., and Huettmann, F. 2007. "Sensitivity of predictive species distribution models to change in grain size." *Diversity and Distributions*, Vol. 13: pp. 332–340.

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

Gustafson, E.J. 1998. "Quantifying landscape spatial pattern: What is the state of the art?" *Ecosystems*, Vol. 1: pp. 143–156.

Hamann, A., and Wang, T. 2006. "Potential effects of climate change on ecosystem and tree species distribution in British Columbia." *Ecology*, Vol. 87: pp. 2773–2786.

Hansen, M.J., Franklin, S.E., Woudsma, C., and Peterson, M. 2001. "Forest structure classification in the North Columbia Mountains using the Landsat tm tasseled cap wetness component." *Canadian Journal of Remote Sensing*, Vol. 27(No. 1): pp. 20–32.

Hawkins, B.A. 2012. "Eight (and a half) deadly sins of spatial analysis." *Journal of Biogeography*, Vol. 39: pp. 1–9.

Hortal, J., Lobo, J.M., and Jiménez-Valverde, A. 2007. "Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands." *Conservation Biology*, Vol. 21: pp. 853–863.

Immitzer, M., Atzberger, C., and Koukal, T. 2012. "Tree species classification with random forest using very high spatial resolution 8-band WorldView-2 satellite data." *Remote Sensing*, Vol. 4(No. 9): pp. 2661–2693.

Jetz, W., Sekercioglu, C.H., and Watson, J.E.M. 2007. "Ecological correlates and conservation implications of overestimating species geographic ranges." *Conservation Biology*, Vol. 22(No. 1): pp. 110–119.

Jiménez-Valverde, A. 2012. "Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling." *Global Ecology and Biogeography*, Vol. 21: pp. 498–507.

Kauth, R., and Thomas, G. 1976. "The tasselled cap - a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat." In LARS Symposia, Paper 159, Proceedings of the Symposium on Machine Processing of Remotely Sensed Data. LARS, Purdue University, West Lafayette, Indiana.

Key, C.H., and Benson, N.C. 2006. "Landscape assessment (LA) sampling and analysis methods." In *FIREMON: Fire Effects Monitoring and Inventory System*. Gen. Tech. Rep. RMRS-GTR-164-CD: LA 1-51, Rocky Mountain Research Station, Ogden, UT: USDA Forest Service.

Kremen, C. 2005. Managing ecosystem services: what do we need to know about their ecology? *Ecology Letters*, Vol. 8: pp. 468–479.

Little, E.L. Jr. 1971. *Atlas of United States Trees, Vol.1, Conifers and Important Hardwoods*. Washington, D.C.: U.S. Department of Agriculture Miscellaneous Publication 1146.

Liu, C., Berry, P.M., Dawson, T.P., and Pearson, R.G. 2005. Selecting thresholds of occurrence in the prediction of species distributions." *Ecography*, Vol. 28(No. 3): pp. 385–393.

Loveland, T.R., Merchant, J.W., Ohlen, D.O., and Brown, J.F. 1991. "Development of a land-cover characteristics database for the conterminous U.S." *Photogrammetric Engineering & Remote Sensing*, Vol. 57(No. 11): pp. 1453–1463.

Maeda, E.E., Heiskanen, J., Thijs, K.W., and Pellikka, P.K.E. 2014. "Season-dependence of remote sensing indicators of tree species diversity." *Remote Sensing Letters*, Vol. 5: pp. 404–412.

McLaughlan, M.S., Wright, R.A., and Jiricka, R.D. 2010. *Field Guide to the Ecosites of Saskatchewan's Provincial Forests*. Prince Albert, Saskatchewan: Saskatchewan Ministry of Environment, Forest Service.

McPherson, J., Jetz, W., and Rogers, D. 2004. "The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?" *Journal of Applied Ecology*, Vol. 41: pp. 811–823.

McPherson, J.M., and Jetz, W. 2007. "Type and spatial structure of distribution data and the perceived determinants of geographical gradients in ecology: the species richness of African birds." *Global Ecology and Biogeography*, Vol. 16: pp. 657–667.

Nelson, T.A., and Boots, B. 2005. "Identifying insect infestation hot spots: an approach using conditional spatial randomization." *Journal of Geographical Systems*, Vol. 7(No. 3–4): pp. 291–311.

Nelson, T.A., and Boots, B. 2008. "Detecting spatial hot spots in landscape ecology." *Ecography (Cop.)*. Vol. 31: pp. 556–566.

Nenzén, H.K., and Araújo, M.B. 2011. Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling*, Vol. 222: pp. 3346–3354.

O'Sullivan, D., and Unwin, D. 2010. "Area objects and spatial autocorrelation." In *Geographic Information Analysis*. Hoboken, NJ: John Wiley & Sons.

Pastor, J., Mladenoff, D.J., Haila, Y., Bryant, J., and Payette, S. 1996. "Biodiversity and ecosystem processes in boreal regions." In *Functional Roles of Biodiversity: A Global Perspective*, edited by H.A. Mooney, J.H. Cushman, E. Medina, O.E. Sala, and E.-D. Schulze. New York: Wiley.

Pfeifer-Meister, L., Bridgham, S.D., Little, C.J., Reynolds, L.L., Goklany, M.E., and Johnson, B.R., 2013. "Pushing the limit: experimental evidence of climate effects on plant range distributions." *Ecology*, Vol. 94: pp. 2131–2137.

Prasad, A.M., Iverson, L.R., and Liaw, A. 2006. "Newer classification and regression tree techniques: bagging and random forests for ecological prediction." *Ecosystems* Vol. 9(No. 2): pp. 181–199.

Roberts, D.W., and Cooper, S.V. 1989. "Concepts and techniques of vegetation mapping." In *Land Classifications Based on Vegetation: Applications for Resource Management*. GTR INT-257, pp. 90–96. Ogden, UT: USDA Forest Service.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M., Loveland, T.R., Vermote, E., and Zhang, C. 2010. "Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States." *Remote Sensing of Environment*, Vol. 114: pp. 35–49.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., and Zhu, Z. 2014. "Landsat-8: Science and product vision for terrestrial global change research." *Remote Sensing of Environment*, Vol. 145: pp. 154–172.

Running, S., Loveland, T., and Pierce, L. 1994. "A vegetation classification logic based on remote sensing for use in global biogeochemical models." *Ambio*, Vol. 23(No. 1): pp. 77–81.

Saskatchewan Environment - Forest Service. 2004. *Saskatchewan Forest Vegetation Inventory—Forest Planning Manual*, Version 4.0. Author.

Saskatchewan Ministry of Environment. 2009. *Saskatchewan's 2009 State of the Environment Report*. Author.

Song, C., Schroeder, T., and Cohen, W. 2007. "Predicting temperate conifer forest successional stage distributions with multitemporal Landsat thematic mapper imagery." *Remote Sensing of Environment*, Vol. 106, 228–237.

Song, C., and Woodcock, C. 2003. "Monitoring forest succession with multitemporal Landsat images: factors of uncertainty." *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41: pp. 2557–2567.

Stockwell, D.R., and Peterson, A.T. 2002. "Effects of sample size on accuracy of species distribution models." *Ecological Modelling*, Vol. 148: pp. 1–13.

Storey, J., Scaramuzza, P., Schmidt, G., and Barsi, J. 2005. Landsat 7 scan line corrector-off gap-filled product development." Paper presented at Proceedings of 16th William T. Pecora Memorial Symposium: Global Priorities in Land Remote Sensing. Sioux Falls, SD, Oct. 23–27.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., and Prentice, I.C. 2005. "Climate change threats to plant diversity in Europe." *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102(23): pp. 8245–8250.

Turner, M. 1989. "Landscape ecology: the effect of pattern on process." *Annual Review of Ecology, Evolution, and Systematics*, Vol. 20: pp. 171–197.

Ustin, S.L., and Xiao, Q.F. 2001. Mapping successional boreal forests in interior central Alaska. *International Journal of Remote Sensing*, Vol. 22(No. 6): pp. 1779–1797.

Van Aardt, J.A.N., and Wynne, R.H., 2001. "Spectral separability among six southern tree species." *Photogrammetric Engineering & Remote Sensing*, Vol. 67: pp. 1367–1375.

Verbyla, D.L., and Hammond, T.O. 1995. "Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids." *International Journal of Remote Sensing*, Vol. 16(No. 3): pp. 581–587.

White, J.C., and Wulder, M.A. 2013. "The Landsat observation record of Canada: 1972–2012." *Canadian Journal of Remote Sensing*, Vol. 39(No. 6): pp. 455–467.

White, J.C., Wulder, M.A., Hobart, G.W., Luther, J.E., Hermosilla, T., Griffiths, P., Coops, N.C., Hall, R.J., Hostert, P., Dyk, A., and Guindon, L. 2014. "Pixel-based image compositing for large-area dense time series applications and science." *Canadian Journal of Remote Sensing*, Vol. 43: pp. 192–212.

Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., and Guisan, A. 2008. "Effects of sample size on the performance of species distribution models." *Diversity and Distributions*, Vol. 14: pp. 763–773.

Woodcock, C., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S.N., Helder, D., Hlemer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P., Vermote, E., Vogelmann, J., Wulder, M.A., and Wynne, R. 2008. "Free access to Landsat imagery." *Science*, Vol. 23(May): pp. 1011–1012.

Wulder, M.A., and Coops, N.C. 2014. "Make Earth observations open access." *Nature*, Vol. 513(No. 7516): pp. 30–31.

Wulder, M.A., Hall, R.J., Coops, N.C., and Franklin, S.E. 2004. "High spatial resolution remotely sensed data for ecosystem characterization." *Bioscience*, Vol. 54: pp. 511–521.

Wulder, M.A., Kurz, W.A., and Gillis, M. 2004. "National level forest monitoring and modeling in Canada." *Progress in Planning*, Vol. 61: pp. 365–381.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R., and Woodcock, C.E. 2012. "Opening the archive: How free data has enabled the science and monitoring promise of Landsat." *Remote Sensing of Environment*, Vol. 122: pp. 2–10.

Wulder, M.A., Skakun, R.S., Kurz, W.A., and White, J.C. 2004. "Estimating time since forest harvest using segmented Landsat ETM+ imagery." *Remote Sensing of Environment*, Vol. 93: pp. 179–187.

Wulder, M.A., White, J.C., Coops, N.C., Nelson, T.A., and Boots, B. 2007. "Using local spatial autocorrelation to compare outputs from a forest growth model." *Ecological Modelling*, Vol. 209: pp. 264–276.