NRC Research Press

ARTICLE

Tracking forest attributes across Canada between 2001 and 2011 using a *k* nearest neighbors mapping approach applied to MODIS imagery

A. Beaudoin, P.Y. Bernier, P. Villemaire, L. Guindon, and X. Jing Guo

Abstract: Mapping Canada's forests is a significant challenge given their extent and the interprovincial differences in forest inventories. We created new sets of nationally consistent forest attribute maps for the years 2001 and 2011 by building upon previously published work with the objective to determine if sequential maps of forest attributes could be used to quantify changes over time. We first refined our previously published methodology of using the *k* nearest neighbors (*k*NN) prediction method and MODIS spectral reflectance data as predictive variables. The maps were generated using an improved reference dataset and a new analytical *k*NN workflow. We then evaluated 2001 to 2011 changes in two key attributes, aboveground biomass and percent tree cover, on pixels identified from published sources as having undergone fire, harvest, or postdisturbance regrowth during that period. For all three change types, average changes in both aboveground biomass and percent tree cover between 2001 and 2011 matched expectations relative to the dynamics of Canadian forests. Our results support the use of sequential national maps of forest attributes for evaluating regionally aggregated disturbance-related changes in forest properties. The new forest attribute maps are available from Beaudoin et al. (2017; doi:10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990) at http://ouvert.canada.ca/data/fr/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990.

Key words: forest inventory, forest monitoring, biomass change, nonparametric estimation, disturbances.

Résumé : La cartographie des forêts canadiennes constitue un défi important compte tenu de leur étendue et des différences entre les inventaires forestiers des provinces. Nous avons créé de nouveaux jeux de cartes des attributs forestiers, cohérentes à l'échelle nationale, pour les années 2001 et 2011 en nous appuyant sur des travaux déjà publiés dans le but de déterminer si des cartes séquentielles des attributs forestiers pourraient être utilisées pour quantifier les changements au fil du temps. Nous avons d'abord affiné notre méthodologie publiée précédemment qui utilise la méthode de prédiction des *k* plus proches voisins (*k*NN) et les données de réflectance spectrale MODIS comme variables prédictives. Les cartes ont été générées en utilisant un ensemble de données de référence amélioré et un nouveau processus analytique de *k*NN. Nous avons ensuite évalué les changements de 2001 à 2011 pour deux attributs clés, la biomasse aérienne et le pourcentage de couvert arborescent, sur des pixels qui selon des sources publiées avaient subi durant cette période un feu ou une récolte, ou encore s'étaient régénérés après une perturbation. Pour les trois types de changement, les changements moyens de la biomasse aérienne et du pourcentage de couvert arborescent entre 2001 et 2011 correspondaient aux attentes par rapport à la dynamique des forêts canadiennes. Nos résultats appuient l'utilisation de cartes nationales séquentielles des attributs forestiers pour l'évaluation régionale des changements des propriétés de la forêt liés aux perturbations. Les nouvelles cartes d'attributs forestiers sont disponibles auprès de Beaudoin et al. (2017; doi : 10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990) à l'adresse : http://ouvert.canada.ca/data/fr/dataset/ec9e2659-1c29-4ddb-87a2-6aced147a990. [Traduit par la Rédaction]

Mots-clés : inventaire forestier, surveillance des forêts, changement de la biomasse, estimation non paramétrique, perturbations.

Introduction

In Canada, provincial and territorial agencies are responsible for forest management and therefore carry out forest inventories whose properties depend on their needs for such information. However, many drivers of forest dynamics operate at very large scales and can only be assessed using spatially continuous maps of forest attributes in which units and definitions are harmonized across jurisdictional boundaries. For this purpose, Beaudoin et al. (2014) produced the first Canada-wide maps of forest attributes at 250 m resolution. These maps were produced using *k* nearest neighbors (*k*NN) interpolation methodology, attribute values from the photo plots of Canada's National Forest Inventory (NFI) (Gillis et al. 2005; Stinson et al. 2016) as reference data, and 2001 MODIS imagery as the main source of predictive variables.

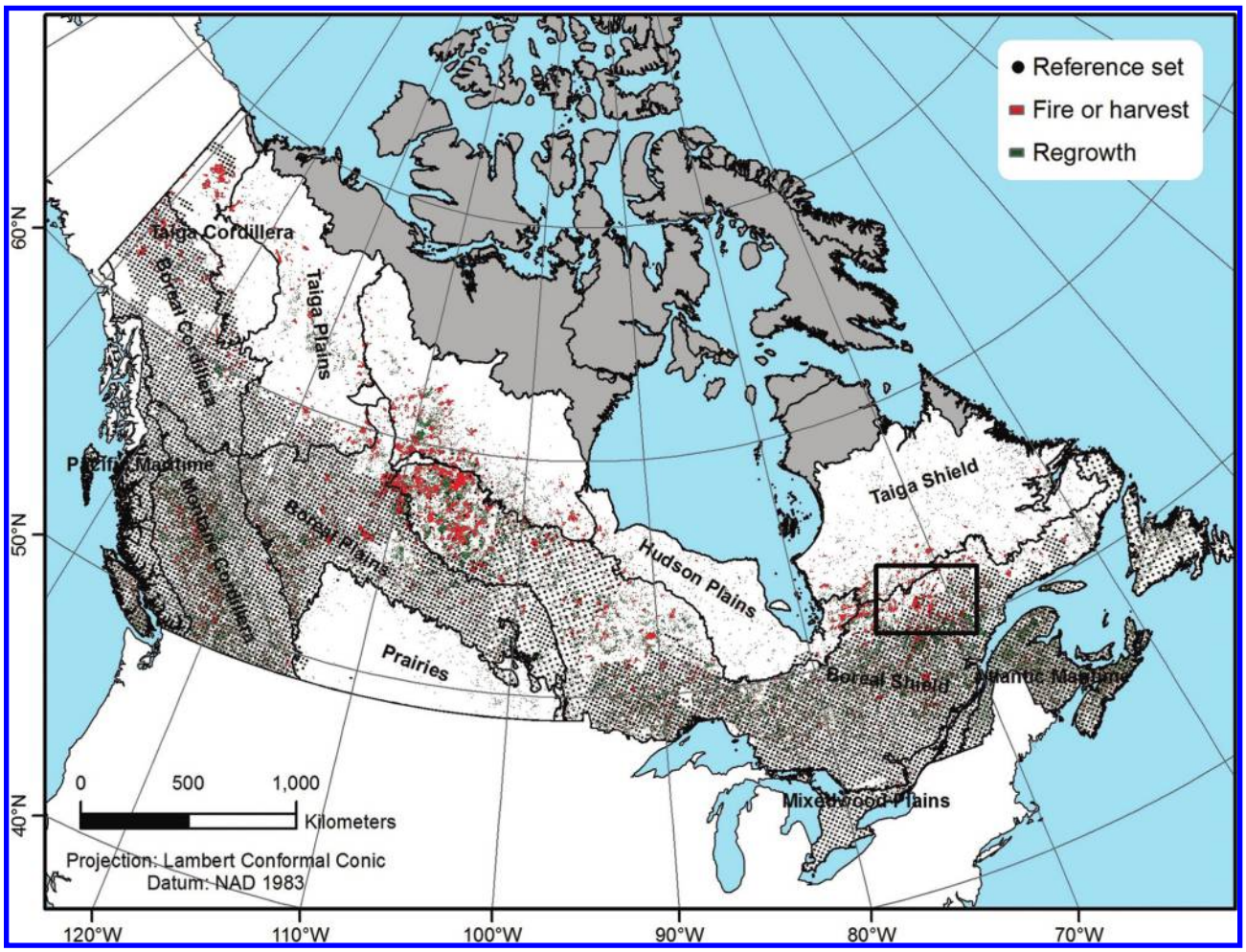
The 2001 maps have been released through the NFI web site (https://nfi.nfis.org/en/) and successfully exploited across a wide variety of application types such as mapping of treed peatlands (Thompson et al. 2016) and estimating postglacial afforestation dynamics (Blarquez and Aleman 2016). Moreover, the maps were designed to be spatiotemporally consistent with MODIS-based 2000–2011 yearly maps of fire and harvest across Canada (Guindon et al. 2014). The combination of these two products has enabled analyses such as the evaluation of timber supply vulnerability to projected changes in fire regime (Gauthier et al. 2015), the sustainable supply of biomass from postfire salvage logging (Mansuy

Received 11 May 2017. Accepted 16 November 2017.

A. Beaudoin, P.Y. Bernier, P. Villemaire, L. Guindon, and X.J. Guo. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, P.O. Box 10380 Stn. Sainte-Foy, Québec, QC G1V 4C7, Canada.

Corresponding author: A. Beaudoin (email: andre.beaudoin@canada.ca).

Fig. 1. Map of Canada's 12 treed ecozones showing the set of reference pixels (dots, n = 35 305) derived from NFI photo plots (Table 1). Areas in red indicate fires and harvest (Guindon et al. 2014) and areas in green indicate regrowth (Hansen et al. 2013) during the 2001–2011 time period. Black outlined box is a taiga–boreal transition area used in Fig. 4.



et al. 2015, 2017), the projection of forest composition under a changing climate (Boulanger et al. 2017), and the quantification of fire selectivity and avoidance as a function of forest composition, age, and biomass (Bernier et al. 2016).

Although the 2001 maps of forest attributes are already of great use, the production of a more recent set of forest attribute maps across Canada can offer further advantages. These include a more up-to-date portrait of the forest landbase but, more importantly, a potential capacity to track and evaluate over time temporal changes of attributes such as biomass and tree cover at the local, regional, or national scale. Because of the availability of year-specific MODIS imagery as a main source of time-dependant information, the mapping approach used by Beaudoin et al. (2014) offers the possibility to produce such a temporal update. However, the authors showed that the lack of agreement between their 2001 mapped estimates and validation data can be important, can vary regionally, and can be substantial as distances from inventoried areas increase. As a result, substantial and independent errors between sequential map products could make change estimation and trend analysis unsatisfactory.

Given the potential benefits related to the estimation of forest attribute change using sequential map products, the primary objective of this study was to evaluate if sequential maps of forest attributes can be used to estimate the magnitude of changes due to known stand-replacing disturbances (harvest, fire) and to regrowth between 2001 and 2011. As an additional objective, we set out to improve our *k*NN prediction procedure and address issues with the original reference dataset to improve the level of agreement with the reference data as compared with the original product of Beaudoin et al. (2014). We then used this improved procedure to generate an updated set of forest attributes maps for year 2001 and a new set for year 2011. The current mapping exercise covers all 12 treed ecozones of Canada (Ecological Stratification Working Group 1996), while the evaluation of the level of agreement with the reference data relate to the seven ecozones that are best covered by the NFI photo-plot network (Fig. 1; Table 1).

Materials and methods

The application of the *k*NN method entails identifying the k nearest reference pixels for each target pixel in the multidimensional space of predictive variables (X), also called the "feature space". Values of each response variable (Y), or the individual forest attributes in this case, for the k nearest reference pixels are then averaged and assigned to the target pixel (McRoberts 2012) as

$$\tilde{y}_i = \left(\sum_{j=1}^{k} w_j^i \right)^{-1} \sum_{j=1}^{k} w_j^i y_j^i \quad (1)$$

with

Table 1. Reference pixels across the seven treed ecozones well inventoried by NFI photo plots (expressed as % of total count; n = 35 305) by forest and nonforest classes and 1% sample set of target pixels used for change analysis (expressed as % of total count by ecozone and change type; n = 437 957).

		(A) *k*NN reference pixels (% of total count)		(B) Target pixels for change analysis (% of total count by ecozone)				
Treed ecozones	Area,* 10^6 km^2	Nonforested (*nf*)	Forested (*f*)	No change (%)	Harvest (%)	Fire (%)	Regrowth (%)	Total count
Boreal Shield	1.921	5.2	31.2	81.6	1.7	5.0	11.6	225 225
Boreal Plains	0.741	5.7	13.4	86.8	1.8	2.7	8.7	72 280
Montane Cordillera	0.490	3.7	10.9	74.6	4.5	1.7	19.2	44 884
Boreal Cordillera	0.471	4.0	6.1	92.1	0.0	4.6	3.3	49 205
Pacific Maritime	0.209	2.3	3.6	88.8	1.2	0.1	9.9	19 034
Atlantic Maritime	0.202	0.6	5.4	72.6	4.4	0.0	23.0	20 678
Mixedwood Plains	0.169	2.4	1.0	94.8	0.1	0.0	5.0	6 651
Total	4.203	23.8	71.5	83.0	1.9	3.7	11.0	437 957
Five other ecozones	2.785	1.1	3.5					

*Source: https://nfi.nfis.org/publications/standard_reports/NFI_T1_LC_AREA_en.html.

$$(2) \qquad w_j^i = \frac{1}{d_{ij}^t}$$

where $\{y_j^i; j = 1, ..., k\}$ is the set of observations for the k reference pixels nearest in the feature space to target pixel i, as calculated using a given distance metric; w_j^i is the weight for each of the k neighbors based on the distance in feature space d_{ij} between target pixel i and the nearest reference pixel j; and exponent t usually takes on values of 0 for simple mean or 1 for inverse distance weighting. In the current work, because preliminary tests using more complex distance metrics or weighted averaging (t = 1) provided negligible improvements (results not shown), we opted to use the Euclidian distance metric and unweighted averaging (t = 0) as in Beaudoin et al. (2014).

Improvements to the *k*NN prediction process evaluated below involved the implementation of procedures to increase the quantity and quality of reference information, the testing of *k*NN predictions stratified by forest and nonforest covers, and the selection of an optimal value of k. In addition, we used published remote sensing products to create the ancillary datasets required for improving the reference set, performing stratified *k*NN predictions, and evaluating the estimated 2001–2011 changes in forest attributes (Supplementary Table S1[1]).

Ancillary datasets

We used yearly maps (2001–2011) of within-pixel fractional changes (percent area of pixel affected by changes) from fires and harvest at the 250 m pixel resolution (hereafter fractional disturbance) from the Canada-wide MODIS forest change product of Guindon et al. (2014). We also derived binary values of treed or nontreed status from the EOSD 2000 Canada-wide Landsat-based 25 m resolution land cover product (Wulder et al. 2008). Finally, we extracted values of tree cover proportion (percent, year 2000) and binary identification of forest cover loss (yearly, 2000–2014) and gain (periodic, 2000–2012) from a global Landsat-based 30 m resolution forest cover product (Hansen et al. 2013).

The Canada-wide datasets at 30 m resolution from the EOSD product (Wulder et al. 2008) and from the global forest cover product of Hansen et al. (2013) were aggregated independently to the 250 m resolution of the MODIS grid to create two maps of percent tree cover for year 2000. These two maps were to be used in the identification of unsuitable reference photo plots, as described below in the section on the reference set. We also used the Landsat-resolution forest cover gain and loss layers from the global forest cover product to create MODIS-resolution maps of fractional tree cover losses resulting from stand-replacing disturbances that occurred during the 2000 to 2011 period (hereafter fractional loss), as well as of fractional tree cover gains resulting from pre-2000 stand-replacing disturbances (hereafter fractional gain).

The fractional loss and gain maps and the MODIS-based fractional disturbance maps of Guindon et al. (2014) were then used to identify pixels for which values of fractional disturbance, loss, and gain were all equal to zero during the 2001 to 2011 study interval to create a no-change map. The fractional maps were also used to update the tree cover information from its original base year of year 2000 to our target years of 2001 and 2011 (Supplementary Table S1[1]). A forest–nonforest binary mask was then created for each target year by applying a 25% threshold to the updated tree cover maps. These two binary masks were to be used in the new stratified *k*NN prediction workflow described in the section on *k*NN predictions.

Response variables

The values of forest attributes used as response variables in the original reference set had been initially extracted by Beaudoin et al. (2014) from the photo-plot product of Canada's NFI. Photo plots are 2 km × 2 km areas located on the nodes of a Canada-wide 20 × 20 km grid within which forest stands polygons are delineated by photo interpretation. For each polygon, estimates had been provided for a standardized suite of forest attributes using a combination of photo interpretation and modelling as part of the National Forest Inventory procedure (Gillis et al. 2005; Boudewyn et al. 2007; Stinson et al. 2016). Beaudoin et al. (2014) rasterized the photo plots on the MODIS 250 m × 250 m grid and estimated the value of attributes within each resulting pixel as the spatial average of values found in the underlying polygons.

Because of issues with coherence of information across jurisdictions and the small sample sizes for less abundant species, the original set of 102 attributes representing percent compositions of individual tree species were aggregated into 78 layers of species or genus. The full set of 93 attributes to be mapped thus included, in addition to the 78 composition layers, the original four land covers and 11 structure variables (Supplementary Tables S2 and S3[1]). For the sake of simplicity, results are reported below only for three key attributes: proportion of tree cover (TREED, %), proportion of needle-leaved species (NLS, %), and live aboveground biomass (AGB, tonnes (t)·ha^{-1}). Maps of all 93 forest attributes are available from Beaudoin et al. (2017).

[1]Supplementary material is available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfr-2017-0184.

Predictive variables

We first considered 13 static and 18 dynamic predictive variables from the set used by Beaudoin et al. (2014) (Supplementary Table S4[1]). The set of static variables was composed of nine climatic variables derived from spatially interpolated 1970–2000 climate normals (McKenney et al. 2011) and four topographic variables derived from a SRTM-based 90 m digital elevation model (Farr et al. 2007). The set of dynamic predictive variables was composed of MODIS-based spectral features derived from summer and winter yearly (2001–2011) surface reflectance composites (Pouliot et al. 2009) that had been temporally normalized (Fernandes and Leblanc 2005). Temporal normalization of MODIS images allows the stabilization of the relationships between spectral features and forest attributes across years and thus makes the reference sets time invariant.

From this initial set of 31 predictive variables, we first selected the 27 best ones for year 2001 using a *k*NN-based multivariate iterative selection procedure (varSelection) implemented within the yaImpute package in R (Crookston et al. 2015). Correlation among predictive variables, as well as between predictive and response variables, was dealt with by using the most similar neighbor (MSN) distance metric based on canonical correlation analysis (Packalén and Maltamo 2007) (Supplementary Fig. S1[1]). Then, using a subset of densely treed reference pixels (n = 13 578) flagged as unchanged, we rejected seven dynamic spectral variables that were too temporally unstable in spite of the availability of temporally normalized MODIS images. The rejection decision was based on thresholds of 10% and 25% relative change for the mean and the mean plus one standard deviation, respectively, across the 2001 to 2011 period (Supplementary Fig. S2[1]). The final set of 20 predictive variables was composed of nine MODIS spectral features and 11 static features (Supplementary Table S4[1]) that were then standardized to values between –1 and 1 to avoid scale effects in *k*NN predictions (LeMay and Temesgen 2005).

Reference set

In an ideal reference set, the spectral features used as predictive variables and the forest attribute measurements used as response variables would have been acquired at the same time to guarantee the best possible relationship between predictive and response variables. However, this was generally not the case as 69% of the photo plots had been established using aerial photos acquired before 2001, the first year of reliable MODIS imagery. We dealt with the resulting temporal mismatch between the reported forest attributes and the MODIS spectral features as follows. We first rejected the photo plots established prior to 1987 (1.2% of all photo plots). For photo plots established between 1987 and 2000 inclusively (68% of all photo plots), we eliminated those that had changed substantially, i.e., in which the tree cover estimated at year 2000 from our two Landsat-based ancillary datasets was 50% greater (12.6% of all photo plots) or smaller (6.6% of all photo plots) than the reported tree cover at the establishment year. For photo plots established from 2001 onward (31%), we used values of spectral features from MODIS images of matching years. However, we eliminated from that set the photo plots that had been disturbed on the year of establishment (0.73% of all photo plots) because we could not know if the airborne images used for photo-plot establishment had been taken before or after the disturbance.

Beaudoin et al. (2014) used the four corner pixels of each photo plot for their reference set. For the current analysis, using only pixels from the photo plots retained in the selection process described above, we added the centre pixel as a fifth sample per photo plot. Prior analysis had revealed that this addition did not cause any substantial spatial autocorrelation among the samples. We also filtered out a small number of pixel outliers (1.3% of pixels) defined as those for which the absolute value of standardized residuals of *k*NN AGB predictions are greater than 3.0 (McRoberts 2009). We carried out this filtering process using the NNDiag package in R (from https://cran.r-project.org/src/contrib/Archive/nnDiag/).

Finally, we replaced the 30% sample set aside for accuracy assessments used by Beaudoin et al. (2014) with a fivefold cross-validation process, thereby increasing the size of the reference set by retaining all selected samples. The resulting reference sample set (n = 35 305) was then split into nonforest (*nf*, n = 8804) and forest (*f*, n = 26 501) subsets based on the photo-plot TREED attribute using a 25% tree cover threshold (Table 1) to be used in the stratified *k*NN prediction process described below.

*k*NN predictions

We produced the maps of predicted forest attributes with the in-house *k*NNMapping C++ software for both target years (2001 and 2011) using the improved reference set. Maps were produced with or without forest–nonforest stratification for values of k ranging from 1 to 15. The stratification of target pixels into forest (*f*) and nonforest (*nf*) subsets was performed using the binary forest–nonforest masks derived from the ancillary data. The masks for 2001 and 2011 were 92.7% and 96.0% accurate, respectively, relative to the NFI photo-plot TREED attribute split in forest–nonforest subsets. We then performed separate *k*NN predictions in each subset for each target year using the corresponding reference sample subsets (Table 1) and spatially recombined the separate forest and nonforest output maps to create continuous maps for each of the 93 response variables.

Evaluation of static *k*NN predictions

Measurements from regular field plots established by provincial forest agencies and by the NFI could not be used for the validation of static predictions, mostly because of their much smaller size compared with the MODIS pixels (0.04 ha and 6.25 ha, respectively). An additional obstacle to the use of field plot measurements for validation was the challenge of creating a uniform, spatially and temporally representative validation set from disparate provincial inventories. We therefore used attribute values from the 250 m × 250 m pixels within the reference set to estimate the level of agreement of the resulting *k*NN predictions, as in Beaudoin et al. (2014). Because the NFI photo-plot attributes on which the reference set is based are derived either from photo interpretation or from models with unknown error structures, we used the error metrics derived below only in a relative manner to quantify the prediction improvements relative to the predictions of Beaudoin et al. (2014).

We estimated values of four error metrics for each of the three test attributes (TREED, NLS, AGB) for unstratified and stratified *k*NN predictions across a range of k values (1 to 15): T^2 (similar to R^2; McRoberts 2012), RMSD (root mean square deviation), and MD (mean deviation) as in Beaudoin et al. (2014):

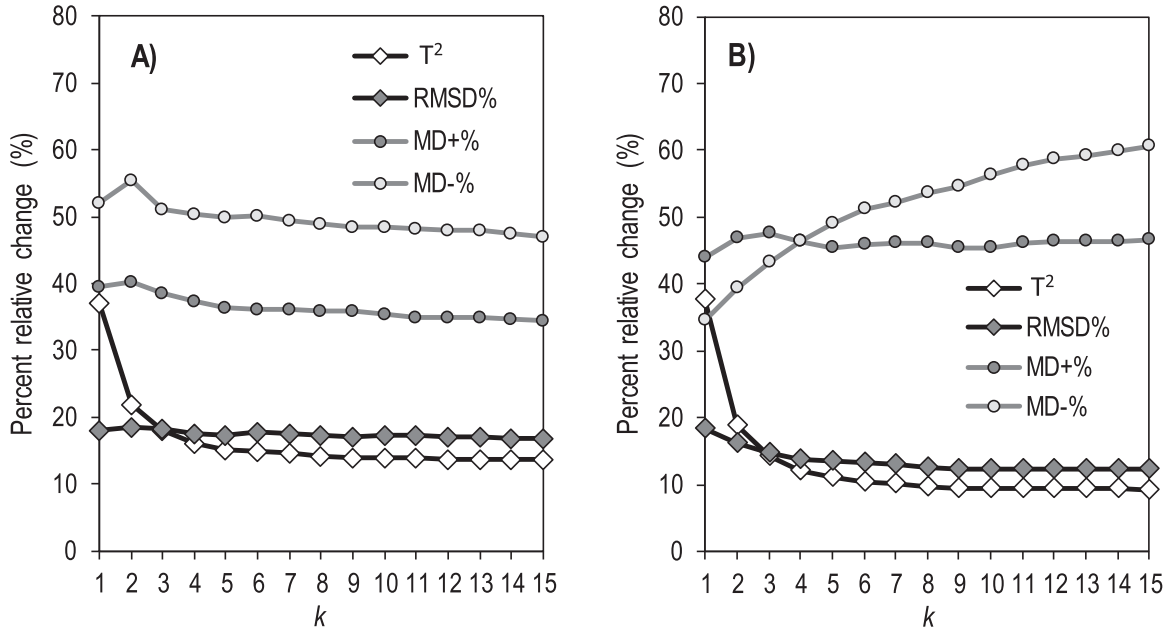
$$(3a) \qquad T_m^2 = \frac{\sum_{i=1}^{n} (\bar{y}_{mi} - y_{mi})^2 - \sum_{i=1}^{n} (\tilde{y}_{mi} - y_{mi})^2}{\sum_{i=1}^{n} (\bar{y}_{mi} - y_{mi})^2}$$

$$(3b) \qquad \mathrm{RMSD}_m = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{mi} - y_{mi})^2}$$

$$(3c) \qquad \mathrm{MD}_m = \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{mi} - y_{mi})$$

where m refers to a given cross-validation fold m = 1 to 5 each containing n pixels indexed i; and $\tilde{y}_i$, y_i, and $\bar{y}_i$ are the predicted, observed, and mean observed values of a given test attribute,

Fig. 2. Percent relative change (%) as a function of *k* for four multivariate error metrics derived from the new 2001 *k*NN process relative to the initial 2001 process from Beaudoin et al. (2014): (A) unstratified predictions and (B) stratified predictions. Values of percent relative change are positive for an increase in T^2 or a decrease in the other three measures.



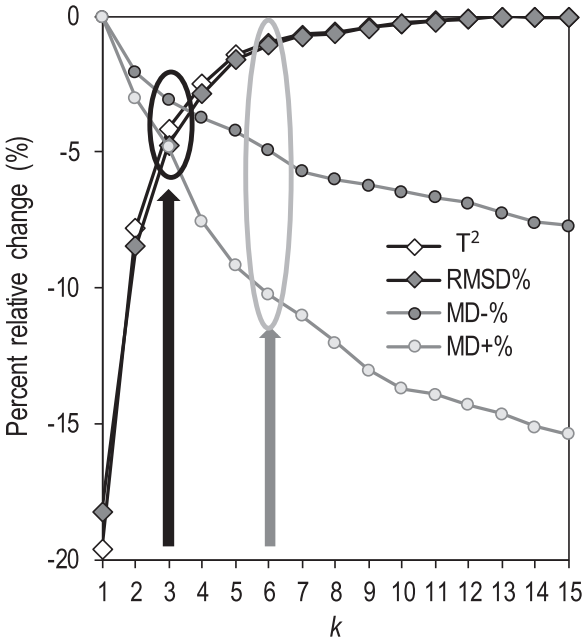
respectively. Values of MD in eq. 3*c* were computed on the lower and upper 25% values of forest attribute distribution (hereafter MD– and MD+, adapted from McRoberts 2009) to characterize the systematic overestimation and underestimation at lower and upper distribution tails documented for the original 2001 product. For each of the three key forest attributes, the estimated values of the four univariate error metrics are the average of values calculated within a fivefold cross-validation analysis. These metrics were used to evaluate the improvements of new 2001 maps relative to the initial 2001 maps from Beaudoin et al. (2014) for each of our three test attributes and to assess the prediction stability of new *k*NN maps between the 2001 and 2011 target years. Values of the univariate accuracy metrics for all 93 attributes can be found in Supplementary Table S3[1].

For each of the four error metrics, we then combined the univariate values from the three key forest attributes into a single multivariate error metric. The combination involved the expression of the univariate RMSD and MD values as percentages of the mean observed attribute value (RMSD%, MD%). Then, RMSD%, MD%, and T^2 values were averaged across the three key attributes. We used the relative percent change of the four resulting multivariate error metrics between the new and the original *k*NN predictions to quantify the impact of the improved reference set and the stratified predictions by forest and nonforest covers on the adjustment to the reference data relative to that obtained using the original *k*NN process of Beaudoin et al. (2014). We also selected the optimal *k* value based on the values of the relative percent change of the multivariate error metrics across a range of *k* values.

Evaluation of 2001–2011 change from *k*NN predictions

An ideal evaluation of change estimates would have been performed using independent field-based information on change in forest attributes at the appropriate spatial and temporal scales and with an appropriate Canada-wide distribution. However, assembling such a dataset would have been a daunting task. In addition, in the absence of photo-plot re-measurement data, we could not use this source of information to evaluate our change estimates. Instead, we performed a qualitative evaluation in which we assessed calculated changes in AGB and TREED predictions between the two target years against expectations of changes based on our knowledge of disturbance dynamics and forest properties.

Fig. 3. Percent relative change (%) across *k* values for each of the four multivariate error metrics in the new 2001 *k*NN process relative to their optimal values (relative change = 0%), showing the current selection of *k* = 3 (black arrow) compared with the value of *k* = 6 (grey arrow) used in Beaudoin et al. (2014).

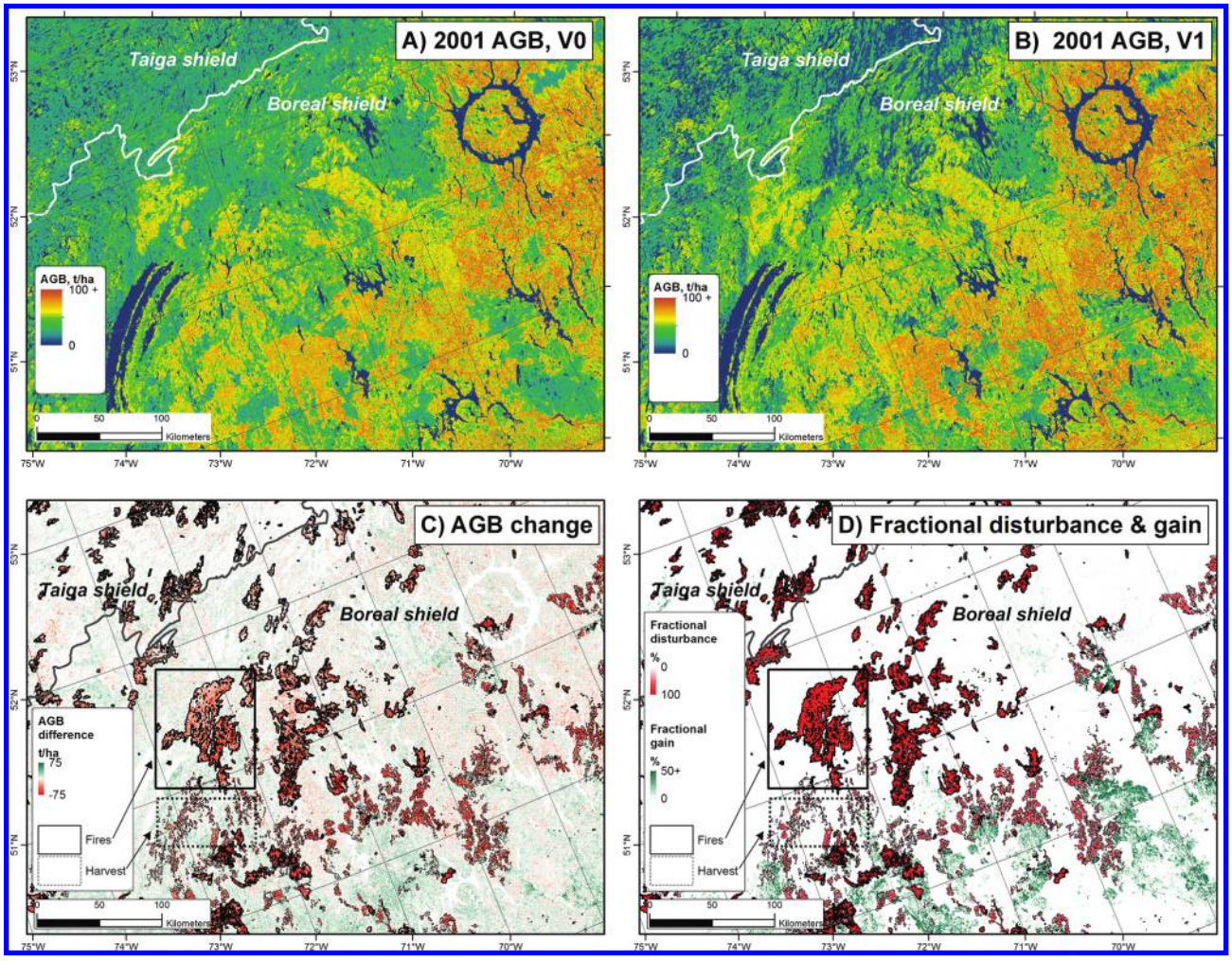


We evaluated whether changes in AGB and TREED due to fire, harvest, or regrowth met expectations in terms of either the differences between the change types or their relationship to the values of fractional disturbance of fractional gain. Our expectations with respect to postdisturbance changes in AGB and TREED were that the mean difference between 2011 and 2001 values

Table 2. Summary of the four univariate error measures (eqs. 3*a* to 3*c*) for aboveground biomass (AGB, t·ha^{-1}), treed proportion (TREED, %), and needle-leaf species proportion (NLS, %) for V0 (the 2001 map of Beaudoin et al. (2014)) and V1 (the current maps for 2001 and 2011).

	AGB (t·ha^{-1})			TREED (%)			NLS (%)		
Error measure	2001, V0	2001, V1	2011, V1	2001, V0	2001, V1	2011, V1	2001, V0	2001, V1	2011, V1
T^2	0.62	0.68	0.63	0.60	0.68	0.66	0.58	0.62	0.61
RMSD	44.47	40.30	43.70	25.41	23.22	24.00	25.79	25.19	25.60
MD+	−34.08	−20.01	−26.84	−13.44	−8.17	−8.08	−15.51	−9.88	−9.26
MD−	15.18	8.34	9.25	21.93	12.33	13.18	21.08	13.21	13.90

Fig. 4. Predicted aboveground biomass (AGB) in a boreal–taiga transition zone of eastern Canada (box in Fig. 1): (A) in the initial (V0) 2001 map from Beaudoin et al. (2014); (B) in the current (V1) 2001 map from this study; (C) AGB change as difference between current 2011 and 2001 predictions; and (D) within-pixel fractional disturbance (Guindon et al. 2014) from fire and harvest (red) and fractional tree cover gain from regrowth (green), as well as no change (white).



would be centered on 0 for "no-change" pixels, would be positive for regrowth pixels, and would be negative for fire and harvest pixels. We also expected changes in AGB to be larger for harvest than for fire. This expectation was based on the highly selective nature of harvest towards fully stocked stands on productive sites in contrast to the more random nature of fire across age and productivity classes and its greater occurrence at higher latitudes where productivity is inherently lower (Van Wagner 1978; Bernier et al. 2016; Guindon et al. 2014).

Finally, we expected changes in both AGB and TREED to be proportional to the within-pixel fractional disturbance or fractional gain. This expectation was based on the observation that the mean reflectance of a whole pixel composed of different surface types (end members) such as harvested versus nonharvested trees could be explained as a linear mixing of spectral reflectance from end members (e.g., Kuusinen et al. 2015). Given the lack of independent change estimates at comparable temporal and spatial scales for rigorous validation, we felt that this analysis would at least provide a qualitative evaluation of adequacy of the change estimates extracted from bi-temporal *k*NN map products.

The evaluation was carried out on a 1% systematic sample of vegetated pixels drawn from our continuous maps of forest prop-

erties across the seven ecozones best sampled by NFI photo plots (n = 437 957; Table 1). Using the disturbances maps of Guindon et al. (2014) described earlier, we then assigned to each pixel the appropriate change type (fire, harvest, regrowth, or no-change) and its associated within-pixel fractional disturbance or fractional gain (fractional disturbance and fractional gain = 0 for no-change pixels).

Results and discussion

Evaluation of static *k*NN predictions

The improvements to the reference set and the better selection of predictive features provided the bulk of the increase in the level of agreement to the reference data as compared with the original process of Beaudoin et al. (2014). Overall, gains were larger than 14% in all four multivariate error metrics, while ranging from 36% to 56% for the two MD% error metrics (MD–% and MD+%, respectively, for the lower and upper ends of distribution of test variables) (Fig. 2A). Additional improvements of about 10% were obtained through stratified predictions by forest and nonforest covers for MD+% and MD–% measures at the cost of slightly lower gains for T^2 and RMSD% (Fig. 2B). Finally, the selection of k = 3 within a priori stratified *k*NN workflow ensured that all four multivariate error metrics remained within 5% of their best values, as compared with within 11% of their best values when using k = 6 as in Beaudoin et al. (2014) (Fig. 3). The resulting implementation of the stratified *k*NN workflow and the lower k value provided gains in univariate error metrics in the 2001 values of our three test variables (Table 2). Values of T^2 substantially increased by 7% to 13%, whereas RMSD decreased by –2% to –9%. Major gains were observed for MD– and MD+ measures, with decreases in overestimations of –31% to –41% and in underestimations of –37% to –45%.

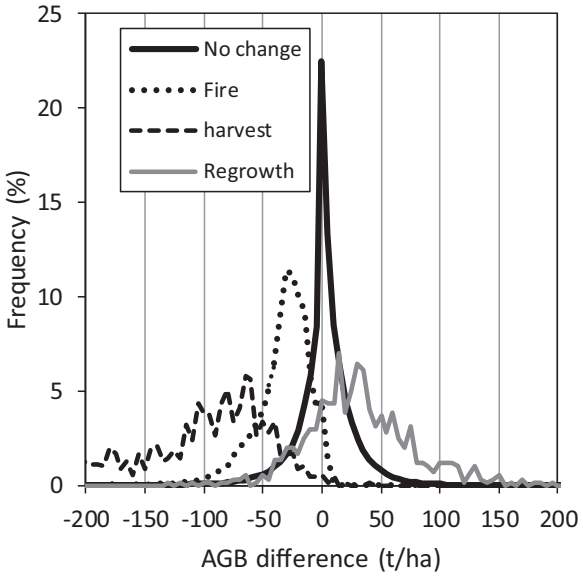
Overall, the continued use of the Euclidian distance and t = 0 and the selection of k = 3 are in line with reported practices in a recent meta-analysis of *k*NN studies (Chirici et al. 2016). Low values of k have been found to yield good error measures in other inventory-based applications of the *k*NN (Halperin et al. 2016). Compared with the original 2001 maps, the new 2001 maps are more contrasted, particularly in areas of sparse tree cover such as the prairie–forest ecotone of central Canada and the boreal–taiga transition where patterns of attributes such as AGB are better spatially defined (Figs. 4A and 4B). The substantial reduction in overestimation and underestimation at lower and upper AGB range, respectively (Table 2), is reflected in the more frequent occurrence of low and high biomass values in the maps (Figs. 4A and 4B).

Estimated values of univariate error metrics show a slight degradation of adjustment to the reference set in the 2011 map as compared with the 2001 map in unchanged pixels, but differences in univariate error estimates between the 2001 and 2011 predictions do not exceed 7% for TREED and NLS attributes and 11% for AGB (Table 2). Although the changes in the error estimates are small, they likely reflect inability of our various procedures to totally eliminate time-related effects. Possible error sources include imperfect spectral normalization of interannual variability in MODIS composites, as well as imperfect selection of dynamic spectral variables. Another possible source of error is the larger uncertainty in the forest–nonforest mask of 2011 relative to the one from 2001 resulting from the updating process based on available ancillary datasets (Supplementary Table S1[1]).

Evaluation of 2001–2011 changes from *k*NN predictions

Estimated changes in AGB between 2011 and 2001 for test pixels that had been fully disturbed (fractional disturbance > 90%) by harvest or fire or that within which fractional gain had been identified over more than 90% of their surface show expected behavior. Overlaps in frequency distributions of AGB change among these change classes and with the no-change pixels are relatively limited (Fig. 5). The near-zero mean (0.18 t·ha^{-1}) change in AGB between 2001 and 2011 for no-change pixels meets our expectations, while the standard deviation (±26.3 t·ha^{-1}) reflects the combined 2001 and 2011 prediction errors and the possible effects of partial disturbances or limited growth not identified in any of the ancillary datasets.

Fig. 5. Frequency distributions of estimated change in AGB as difference between 2011 and 2001 predictions (t·ha^{-1}) due to fire, harvest, and regrowth, as well as for no-change pixels (n = 407 333) within pixels with values of fractional disturbance or fractional gain ≥ 90% (n = 34 124) from the 1% sample set used for change analysis (Table 1).

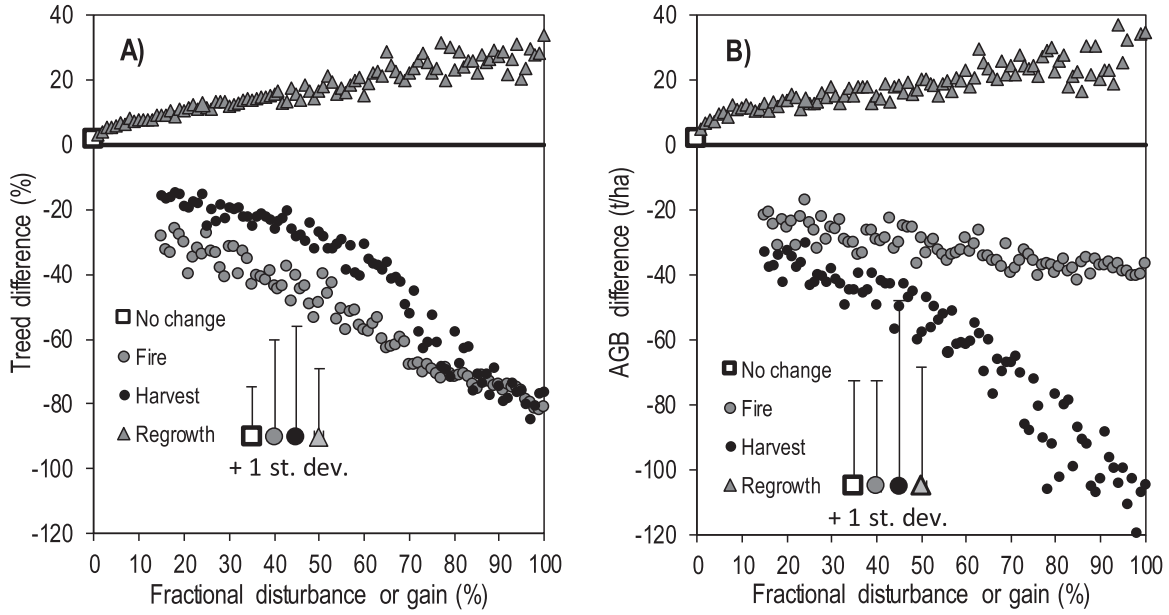


Pixels having undergone harvest or fire show expected negative changes in AGB in 99.2% and 98.2% of the cases in our sample, respectively. In addition, changes are greater for harvested pixels (–103.7 ± 59.5 t·ha^{-1}) than for burned pixels (–35.7 ± 26.4 t·ha^{-1}). Again, this result is in agreement with our expectations as it reflects the systematic selection of mature, well-stocked stands for harvesting in contrast to the more random and northerly occurrence of fires. Moreover, the mean harvest-related loss of about 104 t·ha^{-1} translates roughly to a stem-only volume of 210 m^3·ha^{-1} for softwood species (conversion factors: stem wood = 80% of total biomass, wood density = 0.4 t·m^{-3}). This value is close to a mean of 205 m^3·ha^{-1} that can be calculated from reported annual values of total volume harvested and of total area harvested across Canada (Natural Resources Canada 2015).

Change in AGB from 2001 to 2011 in pixels having undergone regrowth (29.9 ± 50.9 t·ha^{-1}) is positive as expected in 74% of our 1% sample set, but with greater overlap with the no-change AGB difference distribution compared with the fire and harvest classes. The occurrence of negative values in 26% of the test pixels is contrary to our expectations but may result, in part, from *k*NN prediction errors and from the inherent difficulty in the identification of pixels with forest cover gains in the original product of Hansen et al. (2013). Nevertheless, the larger proportion of pixels with positive AGB changes is in agreement with our expectations. In addition, the resulting mean yearly biomass rate of 3.6 t·ha^{-1}·year^{-1} over the 10-year period is compatible with the faster rates of juvenile growth across productive forests stands in Canada, as documented in operational timber yield tables (e.g., Pothier and Savard 1998), where forest cover gains would have been detected by Hansen et al. (2013).

Values of AGB and TREED in pixels affected by harvest or fire drop between target years 2001 and 2011 in proportion to the

Fig. 6. Mean 2001 to 2011 change estimates in (A) treed proportion (TREED, %) and (B) aboveground biomass (AGB, t·ha^{-1}) for no-change pixels and as binned pixels by classes of 1% fractional disturbance from fire or harvest or fractional gain from regrowth (st.dev., standard deviation).



fractional disturbance, a result that is in line with our expectations (Figs. 6A and 6B). The changes in AGB due to regrowth also behave according to expectations, with pixels showing gains in both AGB and TREED proportional to the fractional gain (Figs. 6A and 6B). Finally, when mapped across the landscape, AGB change values are quite similar for fire and harvest (Fig. 4C). This again meets our expectations. Although the average value of AGB loss is larger for harvest than for fire on a per-hectare basis, values of fractional change are larger for fire than for harvest (Fig. 4D) (Guindon et al. 2014). As a result, the realized pixel-level losses (AGB loss per hectare × fractional change) are quite similar for both disturbances.

The new maps of forest attributes for 2001 and 2011 share a reference dataset presumed to be time invariant and therefore applicable to any particular year of the MODIS time series. Time invariance is a general principle used in all MODIS-based products that rely on a fixed set of relationships between spectral features and properties at the surface of the Earth to map these properties across years (Pouliot et al. 2009). In this study, time invariance was sought by using MODIS images that were normalized across years and by removing predictive variables that still varied across years over stable dense forests in spite of the normalization. However, we may expect a certain degree of estimate uncertainty to be generated by the imperfect time invariance of our reference set.

Although the changes made to the *k*NN prediction process listed above have improved the overall level of agreement across all error measures, there are still substantial uncertainties associated with these predictions. Potential sources include a remaining impact of time mismatch of 2001 MODIS images and the year of aerial photo capture for the pre-2001 photo plots retained for this analysis (McRoberts et al. 2016) and the errors associated with the use of photo interpretation and modelling for creating the photo-plot dataset (Magnussen and Russo 2012). Ongoing re-measurement of the photo plots will alleviate some of these problems as these data become available, but a more fundamental shift in approach to the current analysis may be required to achieve substantial gains in accuracy.

Our estimates of changes between 2001 and 2011 are based on national-level samples assembled by classes of fractional disturbance or fractional gain and thus represent aggregate values from pixels with no systematic spatial affiliation. Because of a lack of true validation data for change estimates, what we cannot evaluate at this stage is the appropriate area over which pixel-level results have to be aggregated to correctly represent quantitative change in forest attributes following regrowth or losses due to disturbances. Beaudoin et al. (2014) found that aggregation of pixel-level biomass estimates to 1 km^2 improved values of T^2 by 25% and of RMSD by 40%. Because estimates of change involve a difference between two yearly estimates, we suggest that aggregation would need to be performed to units larger than 1 km^2 for similar improvements of error measures. Also, Beaudoin et al. (2014) found that the quality of estimates degrades in areas with more complex topography or that are under-represented within our reference dataset through lack of forest inventory. The same limitations will apply to change estimates.

Conclusion

There has been a number of national or regional forest attributes maps already produced using the nonparametric *k*NN method (e.g., Wilson et al. 2013; McRoberts et al. 2007; Bernier et al. 2010). However, possibly the only national inventory programs within which the *k*NN method is being applied are those of Finland (Tomppo 2006), which has been operational since 1990, and Sweden. In the Finnish program, re-measured field plots and Landsat 30 m spectral reflectance data are combined with other data sources to yield new map coverage on a 5-year basis as NFI by-products. Reproducing such a data-intensive program at the scale of Canada would be a significant challenge. However, the original 2001 250 m resolution maps of forest attributes across Canada's well-inventoried forests have already enabled unique analyses to be carried out, and the current update of this original base year will further improve its use. We also expect that the addition of the 2011 maps of forest attributes, and thus the ability to perform bi-temporal analyses, will create significant opportunities for strategic analyses at the regional and national scales across Canada. The new version of the 2001 maps and the new 2011 maps for all 93 predicted forest attributes are available for public use (Beaudoin et al. 2017).

Acknowledgements

This project would have not been possible without the high-quality MODIS mosaics provided graciously by Darren Pouliot and Rasim Latifovic from the Canada Centre for Mapping and Earth Observation of Natural Resources Canada. In addition to our regular operating funds, this project was supported by the Forest Change initiative of the Canadian Forest Service (CFS), whose unwavering confidence in our capacity was deeply appreciated, as well as by Canada's National Forest Inventory and its provincial partner agencies that provided the data foundation for this research. We thank Canadian Forest Service specialist Rémi St-Amant for the creation of the fast and efficient C++ *k*NNMapping code, without which the task of performing Canada-wide analysis would have been a significantly greater challenge. Finally, we thank Steen Magnussen and Graham Stinson, both from CFS, for their constructive review and useful comments on earlier versions of this manuscript.

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 imputation and MODIS imagery. Can. J. For. Res. **44**: 521–532. doi:10.1139/cjfr-2013-0401.

Beaudoin, A., Bernier, P.Y., Villemaire, P., Guindon, L., and Guo, X.J. 2017. Species composition, forest properties and land cover types across Canada's forests at 250m resolution for 2001 and 2011. Natural Resources Canada, Canadian Forest Service, Laurentian Forestry Centre, Quebec, Canada. doi:10.23687/ec9e2659-1c29-4ddb-87a2-6aced147a990.

Bernier, P.Y., Daigle, G., Rivest, L.-P., Ung, C.-H., Labbé, F., Bergeron, C., and Patry, A. 2010. From plots to landscape: a k-NN-based method for estimating stand-level merchantable volume in the Province of Québec, Canada. For. Chron. **86**: 461–468. doi:10.5558/tfc86461-4.

Bernier, P.Y., Gauthier, S., Jean, P.-O., Manka, F., Boulanger, Y., Beaudoin, A., and Guindon, L. 2016. Mapping local effects of forest properties on fire risk across Canada. Forests, **7**: 157. doi:10.3390/f7080157.

Blarquez, O., and Aleman, J.C. 2016. Tree biomass reconstruction shows no lag in postglacial afforestation of eastern Canada. Can. J. For. Res. **46**: 485–498. doi:10.1139/cjfr-2015-0201.

Boudewyn, P., Song, X., Magnussen, S., and Gillis, M.D. 2007. Model-based, volume-to-biomass conversion for forested and vegetated land in Canada. Canadian Forest Service, Pacific Forestry Centre, Victoria, B.C., Inf. Rep. BC-X-411.

Boulanger, Y., Taylor, A.R., Price, D.T., Cyr, D., McGarrigle, E., Rammer, W., Sainte-Marie, G., Beaudoin, A., Guindon, L., and Mansuy, N. 2017. Climate change impacts on forest landscapes along the Canadian southern boreal forest transition zone. Landscape Ecol. **32**: 1415–1431. doi:10.1007/s10980-016-0421-7.

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., Travaglini, D., and McRoberts, R.E. 2016. A meta-analysis and review of the literature on the *k*-Nearest Neighbors technique for forestry applications that use remotely sensed data. Remote Sens. Environ. **176**: 282–294. doi:10.1016/j.rse.2016.02.001.

Crookston, N.L., Finley, A.O., and Coulston, J. 2015. Nearest Neighbor observation imputation and evaluation tools [online]. Available from https://cran.r-project.org/web/packages/yaImpute/yaImpute.pdf [accessed 14 December 2016].

Ecological Stratification Working Group. 1996. A national ecological framework for Canada. Agriculture and Agri-Food Canada, Research Branch, Centre for Land and Biological Resources Research and Environment Canada, State of the Environment Directorate, Ottawa.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D. 2007. The Shuttle Radar Topography Mission. Rev. Geophys. **45**: RG2004. doi:10.1029/2005RG000183.

Fernandes, R.A., and Leblanc, S.G. 2005. Parametric (modified least squares) and non-parametric (Theil–Sen) linear regressions for predicting biophysical parameters in the presence of measurement errors. Remote Sens. Environ. **95**: 303–316. doi:10.1016/j.rse.2005.01.005.

Gauthier, S., Bernier, P.Y., Boulanger, Y., Guo, J., Guindon, L., Beaudoin, A., and Boucher, D. 2015. Vulnerability of timber supply to projected changes in fire regime in Canada's managed forests. Can. J. For. Res. **45**: 1439–1447. doi:10.1139/cjfr-2015-0079.

Gillis, M.D., Omule, A.Y., and Brierley, T. 2005. Monitoring Canada's forests: the National Forest Inventory. For. Chron. **81**: 214–221. doi:10.5558/tfc81214-2.

Guindon, L., Bernier, P.Y., Beaudoin, A., Pouliot, D., Villemaire, P., Hall, R.J., Latifovic, R., and St-Amant, R. 2014. Annual mapping of large forest disturbances across Canada's forests using 250 m MODIS imagery from 2000 to 2011. Can. J. For. Res. **44**: 1545–1554. doi:10.1139/cjfr-2014-0229.

Halperin, J., LeMay, V., Coops, N., Verchot, L., Marshall, P., and Lochhead, K. 2016. Canopy cover estimation in miombo woodlands of Zambia: comparison of Landsat 8 OLI versus RapidEye imagery using parametric, nonparametric, and semiparametric methods. Remote Sens. Environ. **179**: 170–182. doi:10.1016/j.rse.2016.03.028.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., and Townshend, J.R.G. 2013. High-resolution global maps of 21st-century forest cover change. Science, **342**: 850–853. doi:10.1126/science.1244693. PMID:24233722. Data available online from http://earthenginepartners.appspot.com/science-2013-global-forest.

Kuusinen, N., Stenberg, P., Tomppo, E., Bernier, P., and Berninger, F. 2015. Variation in understory and canopy reflectance during stand development in Finnish coniferous forests. Can. J. For. Res. **45**: 1077–1085. doi:10.1139/cjfr-2014-0538.

LeMay, V., and Temesgen, H. 2005. Comparison of nearest neighbor methods for estimating basal area and stems per hectare using aerial auxiliary variables. For. Sci. **51**(2): 109–119.

Magnussen, S., and Russo, G. 2012. Uncertainty in photo-interpreted forest inventory variables and effects on estimates of error in Canada's National Forest Inventory. For. Chron. **88**: 439–447. doi:10.5558/tfc2012-080.

Mansuy, N., Thiffault, E., Lemieux, S., Manka, F., Paré, D., and Lebel, L. 2015. Sustainable biomass supply chains from salvage logging of fire-killed stands: a case study for wood pellet production in eastern Canada. Appl. Energy, **154**: 62–73. doi:10.1016/j.apenergy.2015.04.048.

Mansuy, N., Paré, D., Thiffault, E., Bernier, P.Y., Cyr, G., Manka, F., Lafleur, B., and Guindon, L. 2017. Estimating the spatial distribution and locating hotspots of forest biomass from harvest residues and fire-damaged stands in Canada's managed forests. Biomass Bioenergy, **97**: 90–99. doi:10.1016/j.biombioe.2016.12.014.

McKenney, D.W., Hutchinson, M.F., Papadopol, P., Lawrence, K., Pedlar, J., Campbell, K., Milewska, E., Hopkinson, R.F., Price, D., and Owen, T. 2011. Customized spatial climate models for North America. Bull. Am. Meteorol. Soc. **92**: 1611–1622. doi:10.1175/2011BAMS3132.1.

McRoberts, R.E. 2009. Diagnostic tools for nearest neighbors techniques when used with satellite imagery. Remote Sens. Environ. **113**: 489–499. doi:10.1016/j.rse.2008.06.015.

McRoberts, R.E. 2012. Estimating forest attribute parameters for small areas using nearest neighbors techniques. For. Ecol. Manage. **272**: 3–12. doi:10.1016/j.foreco.2011.06.039.

McRoberts, R.E., Tomppo, E.O., Finley, A.O., and Heikkinen, J. 2007. Estimating areal means and variances of forest attributes using the *k*-Nearest Neighbors technique and satellite imagery. Remote Sens. Environ. **111**: 466–480. doi:10.1016/j.rse.2007.04.002.

McRoberts, R.E., Næsset, E., and Gobakken, T. 2016. The effects of temporal differences between map and ground data on map-assisted estimates of forest area and biomass. Ann. For. Sci. **73**: 839–847. doi:10.1007/s13595-015-0485-6.

Natural Resources Canada. 2015. The State of Canada's Forests: Annual Report 2015. Government of Canada, Ottawa.

Packalén, P., and Maltamo, M. 2007. The k-MSN method for the prediction of species-specific stand attributes using airborne laser scanning and aerial photographs. Remote Sens. Environ. **109**: 328–341. doi:10.1016/j.rse.2007.01.005.

Pothier, D., and Savard, F. 1998. Actualisation des tables de production pour les principales essences forestières du Québec. Quebec Ministry of Natural Resources, Québec, Canada.

Pouliot, D., Latifovic, R., Fernandes, R., and Olthof, I. 2009. Evaluation of annual forest disturbance monitoring using a static decision tree approach and 250 m MODIS data. Remote Sens. Environ. **113**: 1749–1759. doi:10.1016/j.rse.2009.04.008.

Stinson, G., Magnussen, S., Boudewyn, P., Eichel, F., Russo, G., Cranny, M., and Song, A. 2016. Canada. *In* National forest inventories: assessment of wood availability and use. *Edited by* C. Vidal, I. Alberdi, L. Hernández Mateo, and J. Redmond. Springer, Cham, Switzerland. pp. 233–247. doi:10.1007/978-3-319-44015-6_12.

Thompson, D.K., Simpson, B.N., and Beaudoin, A. 2016. Using forest structure to predict the distribution of treed boreal peatlands in Canada. For. Ecol. Manage. **372**: 19–27. doi:10.1016/j.foreco.2016.03.056.

Tomppo, E. 2006. The Finnish National Forest Inventory. *In* Forest inventory: methodology and applications. *Edited by* A. Kangas and M. Maltamo. Springer, Dordrecht, Netherlands. pp. 179–194. doi:10.1007/1-4020-4381-3_11.

Van Wagner, C.E. 1978. Age-class distribution and the forest fire cycle. Can. J. For. Res. **8**: 220–227. doi:10.1139/x78-034.

Wilson, B.T., Woodall, C.W., and Griffith, D.M. 2013. Imputing forest carbon stock estimates from inventory plots to a nationally continuous coverage. Carbon Balance Manage. **8**: 1. doi:10.1186/1750-0680-8-1.

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G., and Dechka, J.A. 2008. Monitoring Canada's forests. Part 1: completion of the EOSD land cover project. Can. J. Remote Sens. **34**(6): 549–562. doi:10.5589/m08-066.