Article

Use of Sentinel-2 Data to Improve Multivariate Tree Species Composition in a Forest Resource Inventory

Jay R. Malcolm ^{1,*}, Braiden Brousseau ² , Trevor Jones ³ and Sean C. Thomas ¹

¹ Institute of Forestry and Conservation, University of Toronto, 33 Willcocks St., Toronto, ON M5S 3B3, Canada; sc.thomas@utoronto.ca

² Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada; braiden.brousseau@gmail.com

³ Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, 1219 Queen St. East, Sault Ste Marie, ON P6A 2E5, Canada; trevor.jones3@canada.ca

* Correspondence: jay.malcolm@utoronto.ca; Tel.: +1-647-706-2042

Abstract: Aerial-photo interpreted inventories of forest resources, including tree species composition, are valuable in forest resource management, but are expensive to create and can be relatively inaccurate. Because of differences among tree species in their spectral properties and seasonal phenologies, it might be possible to improve such forest resource inventory information (FRI) by using it in concert with multispectral satellite information from multiple time periods. We used Sentinel-2 information from nine spectral bands and 12 dates within a two-year period to model multivariate percent tree species composition in >51,000 forest stands in the FRI of south-central Ontario, Canada. Accuracy of random forest (RF) and convolutional neural network (CNN) predictions were tested using species-specific basal area information from 155 0.25-ha field plots. Additionally, we created models using the Sentinel-2 information in concert with the field data and compared the accuracy of these models and the FRI-based models by use of basal areas from a second (13.7-ha) field data set. Based on average R^2 values across species in the two field data sets, the Sentinel-FRI models outperformed the FRI, showing 1.5- and 1.7-fold improvements relative to the FRI for RF and 2.1- and 2.2-fold improvements for CNN (mean R^2 : 0.141–0.169 (FRI); 0.217–0.295 (RF); 0.307–0.352 (CNN)). Models created with the field data performed even better: improvements relative to the FRI were 2.1-fold for RF and 2.8-fold for CNN (mean R^2 : 0.169 (FRI); 0.356 (RF); 0.469 (CNN)). As predicted, R^2 values between FRI- and field-trained predictions were higher than R^2 values with the FRI. Of the 21 tree species evaluated, 8 relatively rare species had poor models in all cases. Our multivariate approach allowed us to use more FRI stands in model creation than if we had been restricted to stands dominated by single species and allowed us to map species abundances at higher resolution. It might be possible to improve models further by use of tree stem maps and incorporation of the effects of canopy disturbances.

Keywords: remote sensing; multivariate tree species composition; Sentinel-2; forest resource inventory; random forest; convolutional neural network; forest management

Citation: Malcolm, J.R.; Brousseau, B.; Jones, T.; Thomas, S.C. Use of Sentinel-2 Data to Improve Multivariate Tree Species Composition in a Forest Resource Inventory. *Remote Sens.* **2021**, *13*, 4297. <https://doi.org/10.3390/rs13214297>

Academic Editor: Krzysztof Stereńczak

Received: 17 September 2021

Accepted: 22 October 2021

Published: 26 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Spatially extensive and up-to-date information on forest attributes, including tree species composition, play vital roles in forest management and in many other environmental fields. In Canada, for example, forest resource inventories (FRIs) created from interpreted aerial photos provide stand-level estimates of tree species composition, tree height, tree density, and site quality [1]. These attributes in turn play key roles in tactical and strategic management of wood fibre resources. Species composition is particularly important because it is used to define major forest types, which in turn are used to operationalize many aspects of forest management, including growth-and-yield calculations and forecasts of timber supply [2,3]. Forest management in Germany similarly makes use of

databases of stand-level attributes derived from a mix of aerial photography, field sampling, and expert judgement (Peerenboom 2003, cited by [4]). In Norway, manual interpretation of orthophotos or stereo photogrammetry is commonly used [5]. In Finland, recent FRIs (also termed forest management inventories) use spectral and texture information from large-format digital cameras in combination with LiDAR information to estimate stand attributes, including attributes of dominant species and major forest types [6]. Such forest resource inventories are also widely used in a variety of other applications; for example, in quantifying wildlife habitats [1,7,8] and carbon stocks (e.g., [9,10]).

Unfortunately, although valuable, forest resource inventories are expensive and laborious to create and the subjective photo-interpretation can be relatively inaccurate, especially with regards to tree species composition. Potential sources of error are many, including photo-interpreter skills, image quality, forest complexity, and inventory definitions [2,11]. For example, Thompson et al. [12] used 50–100 ground sampling points in each of 129 boreal stands and found that percent species abundances in the FRI were accurate only about 50% of the time for common species and <25% of the time for rare species (where accuracy was defined as either $\leq 20\%$ off or $\leq 10\%$ off when FRI abundance was zero). Maxie et al. [13] reported compositional accuracies of 29–88% across six studies and found only 55–62% agreement between field and FRI data when coarse forest-type classifications were compared. Especially high error rates have been reported for mixed- compared to single-species stands [12,14]. In some regions, such as the boreal forest, relatively homogenous areas of one or a few tree species are common; however, in other regions, including northern temperate forests of North America, forest stands often are spatially heterogenous species mixes. In addition to being error prone, FRIs also are typically at multiple-hectare scales. The ability to map species composition at finer spatial scales could prove useful in many applications; for example, to facilitate more detailed understanding of wildlife-habitat relationships [13], to aid in cut-block design, to map small-scale disturbances, or to better mitigate effects of invasive forest pests.

Continuing advances in remote sensing offer considerable potential to improve estimates of tree community composition. Because of differences among tree species in their spectral characteristics and phenologies, of particular interest are multi- or hyper-spectral data from multiple time periods. Such information at high resolution has been used to identify individual tree species with considerable success, especially when combined with high-resolution information on tree structure from LiDAR (e.g., [15,16]). However, high-resolution multi-spectral data can be expensive and may not be widely available. Copernicus Sentinel-2 satellites, which were launched in 2015 and 2017 and whose data are freely available, are potentially useful in that they provide multiple spectral bands in the visible, near and short-wave infrared parts of the spectrum at moderate resolution (10–60 m) (e.g., [17]). Importantly, the satellites have nominal return times of 2–5 days, which allows for the possibility of multiple cloud-free images during the growing season. Several studies have reported success in identifying tree species at the stand level using these data, especially when images from several dates were included. For example, Immitzer et al. [17] summarized information from seven studies and reported overall accuracies of 76–92% in identifying tree species and forest types. In their own study they achieved an overall accuracy of 89% for 12 species using multiple Sentinel-2 scenes.

Some Sentinel-2 studies to date have used regional or national FRIs to identify areas that were heavily dominated by single species, and then tested the predictive power of Sentinel-2 models against these same data (see also [4]). This begs the question though: can such models be used to improve the FRI itself? In particular, can Sentinel-2 imagery in concert with the FRI be used to improve the FRI? The reasoning here is that because of species-specific phenological variation, the satellite information in concert with the FRI might increase the accuracy of the FRI, resulting in better estimation of tree species composition than in the original training data. Such an approach might also lend itself to higher resolution mapping of species composition, given that the satellite information is

primarily at 20-m resolution compared to the multiple-hectare size of forest stands in FRIs. This approach might also help to simplify and improve the process of creating FRI updates.

A notable challenge, however, is that even single Sentinel-2 pixels, and certainly many FRI stands, include more than one tree species [17,18]. For example, in the FRI used in the present analysis (see below), only 5% of stands in the study area contained only one species and in 35% of stands, even the single-most abundant species comprised <45% of the forest canopy. Instead of focusing on areas heavily dominated by a single species, an interesting alternative might be to use a multivariate approach; that is, to model the vector of species abundances in each stand. Such an approach could make use of more FRI information, rather than just that from monodominant (or near-monodominant) stands and might allow mapping of species abundances rather than simple presence or absence. We are unaware of any tests of such an approach in the literature, although multivariate methods have been used to impute forest attributes for dominant species and major forest types from a combination of aerial and LiDAR information (e.g., [6]) or from a combination of field and satellite information [19,20].

Here, we investigate the utility of Sentinel-2 data in combination with FRI information to predict multivariate species composition in an area of the Great Lakes—St. Lawrence forest of south-central Ontario, Canada. We were especially interested to see if Sentinel-2 predictions improved on the FRI itself, which we tested by use of independent, field-based information on species-specific basal areas from 155 0.25-ha field plots. We were also interested to see how well the models performed against ones that were trained with field data rather than FRI information; therefore, we compared the performance of both FRI- and field-trained models using a second field-based data set (a 13.7-ha, exhaustively sampled area). Under the hypothesis that the FRI in combination with Sentinel information provides more accurate information on tree species composition than the FRI itself, we predicted that correlations among FRI- and field-trained predictions would be higher than their correlations with the FRI itself. Finally, in creating our models, we compared two statistical modelling methods. One was random forest, which is a relatively simple approach that has proven successful in many studies (e.g., [21,22]). However, several authors reported better prediction with multi-layer neural networks (e.g., [16,23]). Hence, as a second method we used convolutional neural networks.

2. Materials and Methods

2.1. Study Area

The study area consisted of an approximately 1.3 million ha area in south-central Ontario, Canada that included Algonquin Provincial Park (APP) to the north and adjoining areas to the south, including the privately-owned Haliburton Forest & Wild Life Reserve Ltd. (Figure 1a). The area falls within Rowe's [24] Great Lakes—St. Lawrence forest zone. This moderately-dissected region of the Canadian Shield has an elevation range of approximately 160–590 m, which in concert with variation in soils results in considerable diversity in local site conditions and tree communities. White pine and other conifers are especially prevalent on glacial outwash plains to the northeast and sugar maple and other hardwoods on the higher-elevation headwater regions to the southwest. In the Forest Resource Inventory (FRI) for the study area, 23 tree taxa were represented: sugar maple (26.6% of the canopy), *Populus* spp. (11.6%), white pine (8.0%), yellow birch (7.3%), eastern hemlock (7.0%), red maple (6.8%), balsam fir (6.5%), white birch (5.8%), white spruce (4.4%), black spruce (3.2%), red pine (2.8%), beech (2.5%), red oak (2.5%), eastern white cedar (2.3%), black ash (0.6%), basswood (0.6%), jack pine (0.5%), white ash (0.3%), larch (0.2%), ironwood (0.1%), black cherry (0.1%), red spruce (<0.1%), and white elm (<0.1%; see Table 1 for scientific names). Forests were diverse at the stand scale, with 95% of stands comprised of more than one tree species. Most of the area is managed for timber resources. Partial harvesting systems dominate, especially single tree selection in maple-dominated stands and shelterwood silviculture in pine-dominated stands. Mean monthly temperatures for

January and July, respectively, were -10 and 19 °C, and annual precipitation was 1078 mm (1981–2010 normals; [25]).

Figure 1. Map of the study area in south-central Ontario, Canada, showing geographical coordinates (UTM 17N, WGS84), major roads (black), Algonquin Provincial Park (APP), Haliburton Forest (HF), map location within Ontario (inset), and Forest Resource Inventory information (green) that was used in combination with Sentinel-2 satellite information to model percent tree species composition (a). White areas are lakes and areas outside the study area. Also shown are locations of field-measured 0.25-ha plots (“plus” symbols; (b)) and tree stems ≥ 10 cm DBH mapped in the “Megaplot” (dots) with superimposed 0.24-ha rectangular plots (black lines; (c)).

2.2. The Forest Resource Inventory

In the Ontario FRI, 1:10,000 or 1:20,000 analogue aerial photos that were panchromatic and stereoscopic were interpreted to delineate forests stands (relatively homogeneous areas with respect to tree species composition, age, arrangement, or condition) and to estimate associated forest attributes such as tree species composition, average tree height, relative tree density, and site quality.

Field cruises and other sources of information were used in some cases to aid in interpretation [2,26]. The FRIs that we used were based on photography taken in 2000 (APP) or 2007–2008 (areas outside of APP). More recent FRIs based on interpretation of higher-resolution colour imagery were available for some of the study area, but not all, hence we used the older FRI. Also, three forest managers in the region expressed reservations to us about the accuracy of the newer FRI compared to the older one. The disparity in age between the FRI data and the Sentinel-2 imagery is addressed in the discussion. In the FRI, photo interpreters estimated percent crown occupancy of tree species to the nearest 10% and included all species contributing at least 10% [2]. As noted above, 23 native tree taxa were represented in the FRI for the study area (22 species plus one taxon at the genus level (*Populus tremuloides*, *P. balsamifera*, plus *P. grandidentata*)). Other species groups and non-native species were rare and were excluded from consideration.

2.3. Sentinel-2 Data

Because of the relatively diverse species pool represented in the study area and evidence that multi-temporal imagery is useful when identifying multiple species or forest types (e.g., [17,27]), we used a relatively large set of cloud-free scenes between the approximate start and end of the growing season. Sentinel-2 data became available only recently for our study area, hence we were able to obtain 12 growing-season images with <10% cloud cover during 2018 and 2019. Most of the images were already transformed to bottom-of-atmosphere (BOA) reflectance (Sentinel product 2A); however, for two we used *sen2cor* in R to convert them from top-of-atmosphere (Sentinel product 1C) to BOA [28]. Scene dates were: 7 April 2018, 12 April 2018, 17 April 2018, 6 June 2019, 11 June 2018, 21 June 2018, 26 July 2019, 10 August 2018, 20 August 2019, 19 September 2019, 24 September 2018, and 9 October 2019. We used the 9 spectral bands provided in the download product at 20-m resolution: namely, B02, B03, B04, B05, B06, B07, B8A, B11, and B12. In hindsight, it might also have been useful to have included B08 (10-m resolution), but it was excluded.

2.4. Field Data

We used information from two sets of field plots georeferenced at sub-metre accuracy. Both were in Haliburton Forest in the southwestern part of the study area (Figure 1).

The first consisted of 155 0.25-ha circular plots sampled in 2008–2011 in which all trees ≥ 8 cm were identified and their diameter at breast height (DBH) measured (Figure 1b; see [29]). A relatively diverse tree community was represented in these plots, including 21 of the 23 FRI species. Percent basal areas of stems ≥ 10 cm DBH were: sugar maple (35.8%), eastern hemlock (16.5%), beech (11.0%), red maple (8.4%), yellow birch (6.2%), balsam fir (4.8%), red oak (4.0%), white spruce (3.1%), white birch (2.0%), *Populus* spp. (2.0%), white ash (1.4%), eastern white cedar (1.1%), white pine (0.8%), ironwood (0.8%), black spruce (0.6%), black cherry (0.5%), white elm (0.4%), basswood (0.3%), black ash (0.1%), red spruce (0.1%), and eastern larch (<0.1%). Basal area is defined as the cross-sectional area at breast height of stems per unit of land area.

The second consisted of a 13.7-ha “Megaplot” sampled in 2007–2008 located on a peninsula on the northern shore of Havelock Lake (Figure 1c). All stems ≥ 1 cm DBH were mapped, identified, and their DBHs measured. This plot is part of the Smithsonian Forest-GEO network and followed sampling protocols common to the network [30,31]. In the Megaplot, conifers dominated along the lake shore, whereas hardwoods (especially sugar maple) dominated inland areas. Again, although sugar maple was relatively abundant, a relatively diverse tree community was represented, including 15 of the 23 FRI species. Percent basal areas for stems ≥ 10 cm DBH were: sugar maple (46.8%), eastern hemlock (15.3%), beech (10.7%), red maple (8.4%), balsam fir (3.6%), red oak (2.9%), white ash (2.9%), yellow birch (2.3%), eastern white cedar (2.0%), white birch (1.4%), white pine (1.2%), white spruce (1.1%), black cherry (1.1%), ironwood (0.3%), and basswood (<0.1%).

2.5. Data Pre-Processing

We created models to predict multivariate tree species abundances from the Sentinel-2 scenes using two sets of training data: the FRI (termed Sentinel-FRI models) and the 0.25-ha field plots (termed Sentinel-field models; see Figure 2 for a summary of the project workflow and Supplementary Materials Table S1 for a summary of model creation and validation).

Prior to creating Sentinel-FRI models, we excluded any FRI stands that included species or species groups other than the 23 listed above. We excluded defective, cloud, cloud-shadow, snow, and “dark area” Sentinel-2 pixels by making use of accompanying scene classification and cloud and snow probabilities (also at 20-m resolution). Specifically, we excluded any pixels that for one or more images had a probability of cloud or snow of >50% and we included only scene classification types 4 (vegetation) or 5 (non-vegetated land). The FRI stand polygons were overlaid onto the Sentinel images and those that included at least 2 ha of Sentinel imagery were used ($n = 51,751$ polygons; mean

area = 11.3 ha; range = 2.2–26.7 ha). For the Sentinel-2 pixels in each polygon, we calculated the mean reflectance for each of the 108 combinations of scene and band. Each combination was normalized across stands to range between zero and one by subtracting the minimum and dividing by the maximum. These same constants were used to normalize all other reflectances mentioned below.

Figure 2. Workflow in which Random Forest (RF) and convolutional neural networks (CNN) that used Sentinel-2 satellite information were trained using percent tree canopy composition from one of two data sets in central Ontario, Canada: (1) Forest Resource Inventory (FRI) information for the whole study area extent (termed “Sentinel-FRI” models) or (2) field-based measurements from 155 0.25-ha plots (termed “Sentinel-field” models). Prediction success of the first was evaluated using field data from 155 0.25-ha plots, whereas prediction success of both were evaluated using field data from 51 0.24-ha plots in the “Megaplot”. See text for details.

In order to create Sentinel-field models, we calculated percent basal area of each tree species in each 0.25-ha plot. We used only trees ≥ 10 cm DBH because they were more likely to be represented in the canopy. Among deciduous trees sampled by Thomas [32] in Haliburton Forest, 10 cm was close to the lower DBH threshold at which trees typically reached a “co-dominant” size class where $>50\%$ of the canopy was exposed. In his sample, 95% of trees with $>50\%$ upper crown exposure had DBHs ≥ 10 cm (Thomas, unpubl. data). Plot boundaries were overlaid onto the Sentinel imagery, and mean reflectance for each combination of image and band were calculated based on weighted averages of Sentinel pixel areas within the plot boundaries. For comparative purposes, we also calculated FRI species composition for the field plots; again, weighted averages of the FRI polygon areas within the plot boundaries were used.

2.6. Model Creation

We used two modelling approaches: random forest (RF) and convolutional neural networks (CNN). The former is a regression-tree (or classification-tree) method in which random subsets of data and predictors are used to create multiple trees, with the final model representing a consensus among them [33]. It is non-parametric and implicitly incorporates interactions among predictors. We used the multivariate implementation of RF provided by R function *rfsrc* and used default parameter settings [34,35]. Several

authors have found that RF models built with subsets of predictors performed better than those built with all predictors (e.g., [17,18,36]), hence we tested a variable selection strategy in which forward selection in a redundancy analysis (RDA) was used to pick best subsets of predictors. This is similar to the “greedy-Wilks” strategy used by Lim et al. [37]. Because of its linear nature, we used arcsine-transformed percent tree composition in the RDA. Ordination and forward selection were accomplished using *rda* and *ordiR2step* in the R library *vegan* [38]. For the Sentinel-FRI model, we undertook a single 80:20 split of the data into training and validation data sets, respectively, and tried the best 20, 40, 60, etc. variables. We observed that model performance increased with the number of variables (as judged by mean R^2 values between observed and predicted values across species) for both the training and validation data; accordingly, we used all variables (and data) in the final model. For the Sentinel-field model, because sample sizes were much smaller, we undertook 30 random 80:20 splits, and for each tried the best 10, 20, 30, etc. variables. In this case, the highest mean R^2 values were obtained with models with ≥ 90 variables, so again we used all variables (and data) in the final model.

Neural networks use layers of interconnected “neurons” in which the connections among them are weighted and iteratively improved to maximize prediction of the dependent data (the observed percent canopy composition) from the independent data (the Sentinel-2 imagery). Convolutional neural networks (CNN) are optimized to find spatial relationships in data and are widely used in the field of computer vision for both regression and classification tasks [39–41]. CNNs also have been used in many domains that process temporal information [42]; for example, via transformation of audio information into spectrogram images [43]. Here, we envisioned the 108 predictors as a 9 by 12 “image” with rows as spectral bands and columns as acquisition dates.

Unlike RF models, there are a plethora of possible CNN architectures and architecture-specific parameterizations. Typical architectures consist of a pipeline of standard computation layers, including: convolution layers, in which sliding filter windows are applied across the input image; pooling layers, in which the size of incoming data is reduced by averaging or taking the maximum of neighbouring pixels; batch normalization layers, in which output feature maps of hidden layers are normalized; dense layers, in which all-to-all connected sets of neurons are used to analyse features generated from the convolutional layers; regularization layers, in which typical L1 and/or L2 penalties are applied; and dropout layers, in which random proportions of the previous layers outputs are ignored to reduce overfitting. Additional parameterizations include choices of activation functions between layers, number of input samples used to train each optimization step, the final output loss function used to guide the optimization, and the learning rate associated with each optimization step. The base architecture used in our CNN analyses is shown in Figure 3.

Given the high dimensionality of the parametrization space, we used Bayesian hyper-parameter searches to find best-performing models by comparing mean R^2 values among predictions for 30% validation datasets. In these searches, we allowed parameters to vary within predefined limits, including the optimizer, number of convolutional and dense layers, number of kernels per convolutional layer, the filter radius of each convolutional layer, number of neurons at each dense layer, pooling, dropout percentage, L1 and L2 regularization strength at each layer, and activation and loss functions. Networks were trained for an unlimited number of epochs with a decaying learning rate and were stopped once validation loss stopped improving. Final activation was through a softmax layer to ensure that the output distribution (i.e., proportional species composition) summed to 1 (see Supplementary Materials Table S2 for search parameters and final parameterizations for the Sentinel-FRI and Sentinel-field models). We used TensorFlow v. 2.1 [45] to undertake the CNN modelling; hyper-parameter searches made use of the HyperOpt library ([46]). All other data processing and modelling was undertaken in R (v. 4.0.2; [47]).

Figure 3. Base architecture of the convolutional neural networks in which Sentinel-2 information was used to model tree species community composition in Forest Resource Inventory or field data from south-central Ontario, Canada. A series of standard computational layers were pipelined, including convolution, pooling, normalization, dense, and dropout layers, and with various possible activation functions between layers (see text for details). Final model architectures and parameterizations were based on Bayesian hyper-parameter searches (see Supplementary Materials Table S2). Modified from Figure 4 in [44].

2.7. Model Testing

Predictive capabilities of the Sentinel-FRI model (and of the FRI itself) were evaluated using the 0.25-ha field plots. Predictive capabilities of both the Sentinel-FRI and Sentinel-field models (and of the FRI) were evaluated using information from the Megaplot at both plot and stand levels. In all cases, predictions were compared against the actual species composition (percent basal areas (BA) of trees with DBH ≥ 10 cm). In the Megaplot plot-level approach, we used plot sizes that approximated the size of the 0.25-ha plots used to train the Sentinel-field models; specifically, we subdivided the Megaplot into 51 non-overlapping 40-by-60 m (0.24 ha) plots (Figure 1c). Plot boundaries were defined by Sentinel pixels. Because they were contiguous, information from these plots is presumably spatially autocorrelated; however, given that our purpose was primarily to compare model performances, spatial autocorrelation was ignored. For each plot, we calculated the mean reflectance for each of the 108 combinations of Sentinel image and band. As before, in calculating the FRI composition of the plots, we used weighted averages of the FRI polygon areas within the plot boundaries. In the Megaplot, stand-level approach, we used the four FRI polygons (stands) that overlapped parts of the Megaplot (areas of 3.9, 0.8, 6.0, and 3.1 ha). Corresponding field-based percent BA of the three leading tree species in the four stands were: (1) 42% eastern hemlock, 16% red maple, 7% eastern white cedar; (2) 44% American beech, 27% sugar maple, 20% eastern hemlock; (3) 67% sugar maple, 11% American beech, 6% eastern hemlock, and (4) 54% sugar maple, 12% American beech, 12% red maple. For each of the four stands, we calculated the mean reflectance for each of the 108 combinations of Sentinel-2 date and band based on weighted averages of Sentinel pixel areas within the stand boundaries. For each stand, we compared observed, FRI, and predicted species composition by calculating pair-wise Euclidean distances (ED):

$$ED = \left(\sum_{i=1}^n (p_i - q_i)^2 \right)^{\frac{1}{2}}$$

where n is the number of species and p_i and q_i are the relative abundances of species i in respective vectors p and q .

In a final comparison of model performance, we used a 3634-ha area in the Depot Lake region of Haliburton Forest. We subdivided it into 40-by-60 m contiguous plots (with

boundaries defined by Sentinel pixels) and for each plot calculated the mean reflectance for each of the 108 combinations of Sentinel date and band. As before, we used 40-by-60 m (0.24 ha) plots to approximate the size of the 0.25-ha plots used to train the Sentinel-field models. We examined plot-level agreement among model predictions (and with the FRI) by calculating coefficients of determination (R^2). FRI values for each plot were based on weighted averages of plot areas within FRI stands. Calculations were undertaken for only those plots that had Sentinel information for 5 or more of the 6 Sentinel pixels per plot ($n = 13,374$ plots).

3. Results

Averaged across R^2 values for the 21 tree species in the 0.25-ha field plots, the Sentinel-FRI model created using RF outperformed the FRI by a factor of 1.5 (average R^2 values of 0.217 and 0.141, respectively). The Sentinel-FRI model created using CNN performed even better, showing a 2.2-fold increase over the FRI (average R^2 value of 0.307; Table 1). For species that showed an R^2 value of at least 0.2, RF predictions outperformed the FRI for all species except red oak, and CNN predictions outperformed both RF predictions and the FRI for all species. Eight species showed poor R^2 values (<0.2) in all cases (eastern white cedar, red maple, black spruce, eastern larch, basswood, black cherry, white elm, and red spruce). FRI and model predictions plotted against field observations for eastern hemlock and sugar maple illustrated the tighter relationships obtained for the Sentinel models compared to the FRI, but also showed that both models (but especially RF) tended to underestimate relatively high abundances (Figure 4).

Similar results were obtained when using the Sentinel-FRI model to estimate percent abundances of the 15 species in the Megaplot. At the plot level, based on average R^2 values across species, the RF model showed an average 1.7-fold improvement over the FRI and the CNN model showed a 2.1-fold improvement over the FRI (average R^2 values were 0.169 for the FRI, 0.295 for RF, and 0.352 for CNN; Table 2). For the 12 species with an R^2 value >0.2 , the FRI had the highest R^2 value for two species, RF had the highest for two species, and CNN had the highest for eight species. Three species had no R^2 values >0.2 (ironwood, black cherry, and basswood). At the stand level, RF predictions slightly outperformed CNN predictions, showing a 35% improvement relative to the FRI, versus 33% for CNN (Euclidian distances from field observations averaged across the four stands were 6.8 for the FRI, 4.4 for RF, and 4.6 for CNN; Table 3).

When sentinel-based models were trained using field data (155 0.25-ha plots) instead of the FRI, they showed even better predictive performance for the Megaplot. For RF at the plot level, instead of a 1.7-fold improvement, we found a 2.1-fold improvement (Table 2). Similarly for CNN at the plot level, instead of the 2.2-fold improvement, we found a

Table 1. Coefficients of determination (R^2 and corresponding p values) between percent basal area of stems >10 cm DBH in 155 0.25-ha field plots and: (1) percent canopy composition in the Forest Resource Inventory (FRI), (2) random forest (RF) predictions, and (3) convolutional neural network (CNN) predictions. The last two were from Sentinel-2-based models trained with FRI information for the entire study area. Species order is based on average R^2 values (highest to lowest).

Species	FRI ¹		RF Sentinel-FRI		CNN Sentinel-FRI	
	R^2	p	R^2	p	R^2	p
Sugar maple (<i>Acer saccharum</i> Marsh.)	0.494	<0.001	0.803	<0.001	0.805	<0.001
Red oak (<i>Quercus rubra</i> L.)	0.655	<0.001	0.633	<0.001	0.723	<0.001
Eastern hemlock (<i>Tsuga canadensis</i> (L.) Carrière)	0.396	<0.001	0.617	<0.001	0.783	<0.001
Poplar (<i>Populus</i> spp.)	0.172	<0.001	0.398	<0.001	0.476	<0.001
Yellow birch (<i>Betula alleghaniensis</i> Britt.)	0.190	<0.001	0.271	<0.001	0.433	<0.001
Balsam fir (<i>Abies balsamea</i> (L.)	0.133	<0.001	0.275	<0.001	0.375	<0.001
White pine (<i>Pinus strobus</i> L.)	0.099	<0.001	0.113	<0.001	0.562	<0.001
White spruce (<i>Picea glauca</i> (Moench) Voss)	0.075	<0.001	0.300	<0.001	0.341	<0.001
White birch (<i>Betula papyrifera</i> Marsh.)	0.005	ns ²	0.221	<0.001	0.416	<0.001
White ash (<i>Fraxinus americana</i> L.)	0.066	0.001	0.229	<0.001	0.230	<0.001
Ironwood (<i>Ostrya virginiana</i> (Mill.) K. Koch)	0.005	ns	0.151	<0.001	0.359	<0.001

Table 1. Cont.

Species	FRI ¹		RF Sentinel-FRI		CNN Sentinel-FRI	
	R ²	p	R ²	p	R ²	p
American beech (<i>Fagus grandifolia</i> Ehrh.)	0.041	0.011	0.124	<0.001	0.226	<0.001
Black ash (<i>Fraxinus nigra</i> Marsh.)	0.043	0.010	0.002	ns	0.290	<0.001
Eastern white cedar (<i>Thuja occidentalis</i> L.)	0.049	0.006	0.160	<0.001	0.096	<0.001
Red maple (<i>Acer rubrum</i> L.)	0.018	ns	0.108	<0.001	0.079	<0.001
Black spruce (<i>Picea mariana</i> (Mill.) BSP)	0.044	0.009	0.075	<0.001	0.048	0.006
Eastern larch (<i>Larix laricina</i> (Du Roi) K. Koch)	0	-	0.011	ns	0.108	<0.001
Basswood (<i>Tilia americana</i> L.)	0	-	0.031	0.029	0.068	0.001
Black cherry (<i>Prunus serotina</i> Ehrh.)	0	-	0.024	ns	0.034	0.022
White elm (<i>Ulmus americana</i> L.)	0.046	0.007	0.004	ns	<0.001	ns
Red spruce (<i>Picea rubens</i> Sarg.)	0.002	ns	0.001	ns	<0.001	ns
Mean	0.141		0.217		0.307	

¹ Three species were absent from the FRI for the field plots; hence their R² values were set to zero. ² ns = not significant ($p > 0.05$).

Figure 4. Percent canopy composition of eastern hemlock and sugar maple from: (1) the Forest Resource Inventory (FRI), (2) Sentinel-FRI random forest (RF) predictions, and (3) Sentinel-FRI convolutional neural network (CNN) predictions plotted against percent basal area (BA) of stems ≥ 10 cm diameter in 155, 0.25-ha field plots. See text for details.

Table 2. As Table 1 except that the coefficients of determination are with percent basal area of stems >10 cm DBH in 51 0.24-ha field plots from the “Megaplot” in Haliburton Forest. The last two columns are for Sentinel-2 models (termed “Sentinel-field”) trained with information from 155 0.25-ha field plots in Haliburton Forest. Again, species order is based on average R^2 values (highest to lowest).

Species	FRI ¹		RF Sentinel-FRI		CNN Sentinel-FRI		RF Sentinel-Field		CNN Sentinel-Field	
	R^2	p	R^2	p	R^2	p	R^2	p	R^2	p
Sugar maple (<i>Acer saccharum</i>)	0.659	<0.001	0.782	<0.001	0.864	<0.001	0.810	<0.001	0.873	<0.001
Eastern hemlock (<i>Tsuga canadensis</i>)	0.666	<0.001	0.640	<0.001	0.827	<0.001	0.528	<0.001	0.692	<0.001
Red oak (<i>Quercus rubra</i>)	0	-	0.641	<0.001	0.714	<0.001	0.402	<0.001	0.648	<0.001
Red maple (<i>Acer rubrum</i>)	0.245	<0.001	0.224	<0.001	0.469	<0.001	0.528	<0.001	0.681	<0.001
Balsam fir (<i>Abies balsamea</i>)	0	-	0.395	<0.001	0.534	<0.001	0.492	<0.001	0.698	<0.001
Yellow birch (<i>Betula alleghaniensis</i>)	0.160	0.004	0.313	<0.001	0.421	<0.001	0.465	<0.001	0.594	<0.001
American beech (<i>Fagus grandifolia</i>)	0.003	ns ²	0.149	0.005	0.225	<0.001	0.452	<0.001	0.903	<0.001
White spruce (<i>Picea glauca</i>)	0	-	0.360	<0.001	0.045	ns	0.659	<0.001	0.596	<0.001
Eastern white cedar (<i>Thuja occidentalis</i>)	0.467	<0.001	0.168	0.003	0.205	<0.001	0.264	<0.001	0.474	<0.001
White birch (<i>Betula papyrifera</i>)	0	-	0.441	<0.001	0.328	<0.001	0.403	<0.001	0.089	0.033
White pine (<i>Pinus strobus</i>)	0.289	<0.001	0.250	<0.001	0.187	0.002	0.117	0.014	0.223	<0.001
White ash (<i>Fraxinus americana</i>)	0	-	0.015	ns	0.417	<0.001	0.143	0.006	0.395	<0.001
Ironwood (<i>Ostrya virginiana</i>)	0.044	0.141	0.012	ns	0.036	ns	0.044	ns	0.162	0.003
Black cherry (<i>Prunus serotina</i>)	0	-	0.022	ns	0.006	ns	0.029	ns	0.007	ns
Basswood (<i>Tilia americana</i>)	0	-	0.014	ns	0.005	ns	0.005	ns	0.004	ns
Mean	0.169		0.295		0.352		0.356		0.469	

¹ Seven species were absent from the FRI for the Megaplot; hence their R^2 values were set to zero. ² ns = not significant ($p > 0.05$).

Table 3. Euclidian distances between field observations and FRI or model predictions for four stands in the Haliburton Forest “Megaplot” in south-central Ontario. Tree communities in the field plots were quantified by species-specific percent basal area (stems ≥ 10 cm diameter at breast height).

Comparison ¹	Stand 1 (3.9 ha)	Stand 2 (0.8 ha)	Stand 3 (6.0 ha)	Stand 4 (3.1 ha)
Field vs. FRI	6.95	12.12	3.83	4.27
Field vs. RF Sentinel-FRI	4.43	8.96	1.93	2.30
Field vs. CNN Sentinel-FRI	4.06	8.69	1.98	3.48
Field vs. RF Sentinel-field	3.60	6.13	1.74	1.69
Field vs. CNN Sentinel-field	4.72	2.84	1.21	2.70

¹ FRI = Forest Resource Inventory; RF Sentinel-FRI = random forest predictions from Sentinel-2-based models trained with FRI information for the entire study area; CNN Sentinel-FRI = as previous, but using a convolutional neural network; RF Sentinel-field = random forest predictions from Sentinel-2-based models trained with field-sampled information from 155, 0.25-ha plots; and CNN Sentinel-field = as previous, but using a convolutional neural network.

2.8-fold improvement. Across all comparisons for species with at least one R^2 value >0.2 , Sentinel-field CNN models had the highest R^2 for six species, Sentinel-FRI CNN models had the highest for three species, and the remaining models and the FRI each had one highest R^2 value. Again, no R^2 values >0.2 were observed for ironwood, black cherry, or basswood. At the stand level for the Megaplot, the Sentinel-field CNN model showed a 58% improvement relative to the FRI, whereas the Sentinel-field RF model showed a 52% improvement (respective mean Euclidian distances were 3.3 and 2.9; Table 3). In the plots of FRI and field observations and model predictions for sugar maple, tighter relationships for model predictions compared to the FRI were evident (Figure 5). Again, all models underestimated abundances when they were at their highest, although underestimation was least pronounced for the Sentinel-field models.

In a final comparison, we plotted the FRI and predicted abundances of red oak in the Depot Lake region of Haliburton Forest (Figure 6). The ability of the Sentinel-based models to provide higher-resolution mapping compared to the FRI was evident. For

example, some areas of homogeneous abundances in the FRI showed a finer-grain pattern of higher and lower abundances in the models. Also, red oak was absent from the FRI in the north-western quadrant of the region, whereas the models predicted several areas of moderate abundances. Compared to CNN models, RF models appeared to show a more restricted ranges of values, overestimating relatively low abundances and underestimating relatively high abundances. When we looked at plot-level R^2 values among the various data sets, as predicted we observed that R^2 values among the various model predictions were higher than R^2 values between the model predictions and the FRI (ranges of 0.24–0.44 vs. 0.11–0.16; Table 4).

Table 4. Average species-specific coefficients of determination (R^2) between the FRI and model predictions of percent species composition for the Depot Lake region in Haliburton Forest, Ontario. The 3634-ha study site was subdivided into 40 by 60 m plots; see Table 3 for descriptions of the various predictions.

	FRI	RF Sentinel-FRI	CNN Sentinel-FRI	RF Sentinel-Field	CNN Sentinel-Field
FRI	-	0.163	0.178	0.107	0.114
RF Sentinel-FRI		-	0.441	0.365	0.239
CNN Sentinel-FRI			-	0.237	0.253
RF Sentinel-field				-	0.334
CNN Sentinel-field					-

Figure 5. Percent canopy composition of sugar maple from: (1) the Forest Resource Inventory (FRI), (2) Sentinel-FRI random forest (RF) predictions, (3) Sentinel-FRI convolutional neural network (CNN) predictions, (4) Sentinel-field RF predictions, and (5) Sentinel-field CNN predictions plotted against percent basal area (BA) of stems ≥ 10 cm diameter in 51 0.24-ha field plots in the “Megaplot”. See text for details.

Figure 6. Map of the Depot Lake region of Haliburton Forest (UTM 17N, WGS84) showing percent abundance of red oak from the forest resource inventory (FRI) and from Sentinel-based predictions using either random forest (RF) or convolutional neural network (CNN) models trained with either FRI or field-based information (155 0.25-ha plots). White areas indicate lakes and wetlands.

4. Discussion

As tested against two independent field data sets, we could use Sentinel-2 information in concert with the FRI to better estimate tree species composition than in the FRI itself. Based on average R^2 values, improvements were approximately 1.5- and 2.1-fold using RF and CNN models, respectively. The FRI's poor performance at the relatively small scales that we used (c. 0.25 ha) is perhaps not surprising given that it was designed to be used at stand (multiple-ha) scales; however, our test at the stand scale also showed improvements over the FRI of 33–35%, although based on a small sample of four stands. We also found that predictions from FRI-trained models agreed more closely with predictions from field-data-trained models than with the FRI itself, supporting the hypothesis that the Sentinel-2 models extracted accurate information from the FRI. The use of Sentinel-2 information also meant that we could map species composition at higher spatial resolution than in the FRI. Additionally, our multivariate approach allowed us to use a much larger set of FRI stands for model training than if we had relied upon stands heavily dominated by single tree species. For example, in our study area, this meant that we could use >50,000 stands for model training instead of the 4153 stands dominated by a single species (90% abundance or more). It also allowed us to model abundances of a relatively large number of species, which was a key future research need identified by Immitzer et al. [17].

Although they had improved performance relative to the FRI, our models still performed poorly for approximately 8 of 21 tree species. Differences in classification success among species using Sentinel-2 data have been noted by several authors and have been attributed to short-term changes in tree phenology missed by the satellite scenes available, imbalanced representation of tree species in the training data, small training sample sizes, and similar spectral signatures among species [17,27,48]. Sample size evidently was a major factor in our analyses: when we plotted model success (as judged by R^2 values) against

percent abundances in the testing data, a strong effect of sample size was observed, with R^2 values increasing with abundance (Figure 7). For both the FRI and field-based training data, but especially for the latter, model fits showed rapid improvement over a relatively small range of abundances. For the field-based training data, any species with abundance of <4% showed relatively poor R^2 values (<0.4), whereas more abundant species showed systematically higher R^2 values (>0.5). In some cases, for these rarer species, the few tree canopies available may not have permitted accurate Sentinel-2 signatures to be created. Part of the failure may also have represented biases in the models towards maximizing fits for the most abundant species, holding out the possibility that cost-sensitive learning or re-balancing efforts might improve model performance [49]. Interestingly, even for relatively abundant species, the considerable variability shown among FRI-trained model fits was not mirrored in the field-trained fits (Figure 7). For example, despite similar abundances, FRI-trained fits for red maple were relatively poor and those for red oak and eastern hemlock were relatively good. These same relative differences were not evident for field-trained models, suggesting that the differences were not driven by difficulties in discriminating these species based on spectral or phenological characteristics. Instead, they may have reflected species identification errors in the FRI. Under this hypothesis, the abundance of red maple was poorly estimated by photo interpreters, whereas red oak and eastern hemlock abundances were relatively accurately assessed. Thompson et al. [12] found evidence of systematic biases in FRI species accuracy between two boreal study areas, suggesting interpreter errors.

In general, one might expect field-trained models to outperform FRI-trained ones, however the improved performance that we observed was somewhat surprising given the much larger sample size available in the FRI compared to the field data (>50,000 stands of FRI information vs. 38.75 ha of field data). This appears to be a testament to the overall poor quality of the FRI information and suggests that field-based data should be used whenever possible in model creation, especially to avoid possible human identification errors. The field data in this case represented a relatively large effort (23,202 stems ≥ 10 cm DBH); such data may not always be available. Also, we used relatively large field plots (0.25 ha), which have special value because they encompass multiple remotely sensed pixels and have low perimeter:area ratios and hence a lesser influence of neighbouring (unmeasured) trees [50]. FRI information, on the other hand, is already available for the entire managed forest area of Ontario and in many other jurisdictions as well. Although beyond the scope of this study, an interesting research question concerns the quantity of field data required to improve upon simple use of the FRI, and the relative costs and benefits of doing so. The value of field data in training also might be increased without much additional effort. We simply included any stem ≥ 10 cm in diameter at breast height in the training data: actual canopy representation as seen by the satellites might be more accurately estimated by using stem maps in concert with estimates of tree canopy size (e.g., [51]) or by using LiDAR.

The nature of training data also raises the interesting question of approaches that might be used to periodically update or create FRIs. An obvious shortcoming of the present effort is the long temporal gap between the Sentinel-2 imagery and the FRI and field information, which was usually close to 10 years, but in the case of the FRI for Algonquin Provincial Park was nearly 20 years. Although trees in the study region are relatively slow-growing, a smaller temporal gap can be expected to lead to improved model performance. At the same time, such gaps may not be unrealistic given the relative infrequency with which the FRI is updated; for example, in Ontario such updates occur approximately every 20 years [2]. The ability of the Sentinel-2 models to outperform the FRI even given the considerable temporal gap highlights a potentially useful aspect of our approach as applied to FRI updates. As new satellite imagery for a region becomes continuously available, evidently even quite old FRI or field-based information can be used to train models, which then present a snapshot of tree species information at the time the imagery was collected. That is, the satellites are automatically collecting new data all of time, whereas the FRI or

plot-based information could be updated at less frequent intervals. Retraining of models presumably would be needed for each new set of satellite imagery because of year-to-year variability in scene availability and tree phenology. Here, methods for efficient use of large collections of satellite images, including those with considerable cloud cover, might be especially useful (e.g., [52,53]). Our methods might also find application in national forest inventories, where satellite information is often used in combination with large networks of field plots to derive forest attributes, although in this case by imputing information from neighbouring plots (e.g., [19,20]). Another potential use of our approach might be to test for biases of interpreter errors [11].

Figure 7. Average Sentinel-based model R^2 values plotted against species abundances from corresponding field test datasets in Haliburton Forest, Ontario ((a) = percent basal area of stems ≥ 10 cm DBH in 155 0.25-ha plots; (b) = percent basal area of stems ≥ 10 cm DBH in 51 0.24-ha plots in the Megaplot). Models in (a) were trained using the forest resource inventory (Sentinel-FRI; see Table 1) and in (b) using field data (Sentinel-field; see Table 2). R^2 values for RF and CNN models were averaged. Species acronyms are the first two letters of the genus and species.

An additional opportunity to improve model performance might be to incorporate information on disturbance events such as forest harvesting. Stand and species spectral properties can be influenced by many factors, including tree age, local site conditions, shadowing effects, and crown health [4,48,54]. We did not have comprehensive information on harvesting activities hence we excluded them; we also expected such effects to be of lesser importance in this region where partial harvesting systems dominate compared to more intensive methods such as clearcutting. Presumably, information on gap fractions and gap

ages could be used to further improve model performance, perhaps by treating expected understory species and shade as additional “species” in a multivariate approach. Another approach might be to screen out disturbed stands a priori; for example, Immitzer et al. [17] used NDVI changes from year to year to exclude disturbed stands.

In general, CNN models outperformed RF models by factors of 1.1–1.4 depending on the specific training and testing data sets. The exception was the Sentinel-FRI model at the stand scale, where RF slightly outperformed CNN. The generally higher performance of CNN models highlights the considerable potential of neural network approaches, which have received support in other studies as well (e.g., [16,23,41]). We did not incorporate vegetation indices into our analysis, which might have improved our RF models (e.g., [16,17]), but might have decreased performance in the CNN models [16]. Modifications of RF might also provide better performance; for example, optimized AdaBoosted RF [55] or Cascaded RF [56]. Certainly, RF was much easier to implement than CNN, although we hope this will change with continuing research on neural networks and their parameterization, especially as applied to remote sensing. We also did not incorporate environmental information such as topography into our species modelling, which has led to improved forest classification in several Sentinel-2 studies (e.g., [22,57]). This was purposeful however: although such approaches can improve model performance because of habitat associations of the various tree species, they might preclude extrapolation to situations where habitat associations have been altered; for example, due to management activities or climate change.

5. Conclusions

In conclusion, our use of multiple Sentinel-2 scenes from a two-year period in combination with local FRI information allowed us to map tree species abundances more accurately and at higher spatial resolution than in the original FRI. Use of information from 155 0.25-ha field plots instead of the FRI resulted in even more accurate models, as did CNN compared to RF models. Importantly, our multivariate approach meant that we could use information from any forested area to train models, irrespective of whether it was dominated by a single tree species or not. Such approaches may prove useful in updating and validating FRI information. Areas for future research include use of tree stem maps in improving field-trained models and incorporation of the effects of canopy disturbances.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/rs13214297/s1>, Table S1: Summary of model creation and testing, including model names, statistical methods, training data, validation data, and tables with R^2 values; Table S2: Neural network architecture and parameter values used in Bayesian hyper-parameter searches and final values used in Sentinel-FRI and Sentinel-field models. References [58–64] are cited in the supplementary materials.

Author Contributions: Conceptualization, J.R.M. and B.B.; Methodology, J.R.M., B.B., S.C.T. and T.J.; Software, J.R.M. and B.B.; Validation, J.R.M., B.B., S.C.T. and T.J.; Formal Analysis, J.R.M. and B.B.; Investigation, J.R.M. and B.B.; Resources, J.R.M., B.B., S.C.T. and T.J.; Data Curation, J.R.M., B.B., S.C.T. and T.J.; Writing—Original Draft Preparation, J.R.M.; Writing—Review & Editing, J.R.M., B.B. and S.C.T.; Visualization, J.R.M. and B.B.; Supervision, J.R.M. and B.B.; Project Administration, J.R.M.; Funding Acquisition, J.R.M., S.C.T. and T.J. All authors have read and agreed to the published version of the manuscript.

Funding: Funding was from the Natural Sciences and Engineering Research Council of Canada (Discovery Grants to J.R.M. and S.C.T.; Canada Research Chair to S.C.T.), the Ontario Ministry of Natural Resources, and Ontario Power Generation.

Acknowledgments: We are indebted to G. Joannis and J. Fink, who suggested the use of Sentinel-2 to us, and J. Rose, who suggested the collaboration between J.M. and B.R. M. Cockwell and A. Gorgolewski provided comments on an earlier draft of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- McDermid, G.J.; Hall, R.J.; Sanchez-Azofeifa, G.A.; Franklin, S.E.; Stenhouse, G.B.; Kobliuk, T.; LeDrew, E.F. Remote sensing and forest inventory for wildlife habitat assessment. *For. Ecol. Manag.* **2009**, *257*, 2262–2269. [\[CrossRef\]](#)
- Pinto, F.; Rouillard, D.; Sobze, J.M.; Ter-Mikaelian, M. Validating tree species composition in forest resource inventory for Nipissing Forest, Ontario, Canada. *For. Chron.* **2007**, *83*, 247–251. [\[CrossRef\]](#)
- Penner, M. Yield prediction for mixed species stands in boreal Ontario. *For. Chron.* **2008**, *84*, 46–52. [\[CrossRef\]](#)
- Stoffels, J.; Mader, S.; Hill, J.; Werner, W.; Ontrup, G. Satellite-based stand-wise forest cover type mapping using a spatially adaptive classification approach. *Eur. J. For. Res.* **2012**, *131*, 1071–1089. [\[CrossRef\]](#)
- Maltamo, M.; Packalen, P.; Kangas, A. From comprehensive field inventories to remotely sensed wall-to-wall stand attribute data—A brief history of management inventories in the Nordic countries. *Can. J. For. Res.* **2021**, *51*, 257–266. [\[CrossRef\]](#)
- Maltamo, M.; Packalen, P. Species-specific management inventory in Finland. In *Forestry Applications of Airborne Laser Scanning: Concepts and Case Studies*; Maltamo, M., Naesset, E., Vauhkonen, J., Eds.; Managing Forest Ecosystems; Springer: Dordrecht, The Netherlands, 2014; Volume 27, pp. 241–252. [\[CrossRef\]](#)
- Rettie, W.J.; Sheard, J.W.; Messier, F. Identification and description of forested vegetation communities available to woodland caribou: Relating wildlife habitat to forest cover data. *For. Ecol. Manag.* **1997**, *93*, 245–260. [\[CrossRef\]](#)
- Malcolm, J.R.; Campbell, B.D.; Kuttner, B.G.; Sugar, A.; Malcolm, J.R.; Kuttner, B.G. Potential indicators of the impacts of forest management on wildlife habitat in northeastern Ontario: A multivariate application of wildlife habitat suitability matrices. *For. Chron.* **2004**, *80*, 91–106. [\[CrossRef\]](#)
- Hennigar, C.R.; MacLean, D.A.; Amos-Binks, L.J. A novel approach to optimize management strategies for carbon stored in both forests and wood products. *For. Ecol. Manag.* **2008**, *256*, 786–797. [\[CrossRef\]](#)
- Malcolm, J.R.; Holtsmark, B.; Piascik, P.W. Forest harvesting and the carbon debt in boreal east-central Canada. *Clim. Chang.* **2020**, *161*, 433–449. [\[CrossRef\]](#)
- Magnussen, S.; Russo, G. Uncertainty in photo-interpreted forest inventory variables and effects on estimates of error in Canada's national forest inventory. *For. Chron.* **2012**, *88*, 439–447. [\[CrossRef\]](#)
- Thompson, I.D.; Maher, S.C.; Rouillard, D.P.; Fryxell, J.M.; Baker, J.A. Accuracy of forest inventory mapping: Some implications for boreal forest management. *For. Ecol. Manag.* **2007**, *252*, 208–221. [\[CrossRef\]](#)
- Maxie, A.J.; Hussey, K.F.; Lowe, S.J.; Middel, K.R.; Pond, B.A.; Obbard, M.E.; Patterson, B.R. A comparison of forest resource inventory, provincial land cover maps and field surveys for wildlife habitat analysis in the Great Lakes—St. Lawrence forest. *For. Chron.* **2010**, *86*, 77–86. [\[CrossRef\]](#)
- Potvin, F.; Belanger, L.; Lowell, K. The validity of forest maps for the description of wildlife habitats on the local level—A case study in the Abitibi-Temiscamingue region. *For. Chron.* **1999**, *75*, 851–859. [\[CrossRef\]](#)
- Féret, J.B.; Asner, G.P. Semi-supervised methods to identify individual crowns of lowland tropical canopy species using imaging spectroscopy and Lidar. *Remote Sens.* **2012**, *4*, 2457–2476. [\[CrossRef\]](#)
- Hartling, S.; Sagan, V.; Sidike, P.; Maimaitijiang, M.; Carron, J. Urban tree species classification using a Worldview-2/3 and LiDAR data fusion approach and deep learning. *Sensors* **2019**, *19*, 1284. [\[CrossRef\]](#)
- Immitzer, M.; Neuwirth, M.; Böck, S.; Brenner, H.; Vuolo, F.; Atzberger, C. Optimal input features for tree species classification in central Europe based on multi-temporal Sentinel-2 data. *Remote Sens.* **2019**, *11*, 2599. [\[CrossRef\]](#)
- Bolyn, C.; Michez, A.; Gaucher, P.; Lejeune, P.; Bonnet, S. Forest mapping and species composition using supervised per pixel classification of Sentinel-2 imagery. *Biotech. Agron. Soc.* **2018**, *22*, 172–187. [\[CrossRef\]](#)
- Tomppo, E.; Olsson, H.; Ståhl, G.; Nilsson, M.; Hagner, O.; Katila, M. Combining national forest inventory field plots and remote sensing data for forest databases. *Remote Sens. Environ.* **2008**, *112*, 1982–1999. [\[CrossRef\]](#)
- McRoberts, R.E.; Chen, Q.; Walters, B.F. Multivariate inference for forest inventories using auxiliary airborne laser scanning data. *For. Ecol. Manag.* **2017**, *401*, 295–303. [\[CrossRef\]](#)
- Belgiu, M.; Drăgu, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm.* **2016**, *114*, 24–31. [\[CrossRef\]](#)
- Liu, Y.; Gong, W.; Hu, X.; Gong, J. Forest type identification with Random Forest using Sentinel-1A, Sentinel-2A, multi-temporal Landsat-8 and DEM data. *Remote Sens.* **2018**, *10*, 946. [\[CrossRef\]](#)
- Rezaee, M.; Zhang, Y.; Mishra, R.; Tong, F.; Tong, H. Using a VGG-16 Network for individual tree species detection with an object-based approach. In Proceedings of the 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing, PRRS 2018, Beijing, China, 19–20 August 2018; IEEE: Piscataway, NJ, USA, 2018. [\[CrossRef\]](#)
- Rowe, J.S. *Forest Regions of Canada*; Canadian Forestry Service: Ottawa, ON, Canada, 1972.
- Environment Canada, Canadian Climate Normals 1981–2010, Haliburton, Ontario. Available online: https://climate.weather.gc.ca/climate_normals/ (accessed on 15 August 2020).
- OMNR. *Forest Information Manual*; Queen's Printer for Ontario: Peterborough, ON, Canada, 2001.
- Persson, M.; Lindberg, E.; Reese, H. Tree species classification with multi-temporal Sentinel-2 data. *Remote Sens.* **2018**, *10*, 1794. [\[CrossRef\]](#)
- Ranghetti, L.; Boschetti, M.; Nutini, F.; Busetto, L. “sen2r”: An R toolbox for automatically downloading and preprocessing Sentinel-2 satellite data. *Comput. Geosci.* **2020**, *139*, 104473. [\[CrossRef\]](#)

29. Spriggs, R.S.; Vanderwel, M.C.; Jones, T.A.; Caspersen, J.P.; Coomes, D.A. A simple area-based model for predicting airborne LiDAR first returns from stem diameter distributions: An example study in an uneven-aged, mixed temperate forest. *Can. J. For. Res.* **2015**, *45*, 1338–1350. [\[CrossRef\]](#)
30. Condit, R.S. *Tropical Forest Census Plots—Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots*; Springer: Berlin/Heidelberg, Germany; R G. Landes Company: Georgetown, TX, USA, 1998.
31. Anderson-Teixeira, K.J.; Davies, S.J.; Bennett, A.C.; Gonzalez-Akre, E.B.; Muller-Landau, H.C.; Wright, S.J.; Abu Salim, K.; Almeyda Zambrano, A.M.; Alonso, A.; Baltzer, J.L.; et al. CTFS-ForestGEO: A worldwide network monitoring forests in an era of global change. *Glob. Chang. Biol.* **2015**, *21*, 528–549. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Thomas, S.C. Photosynthetic capacity peaks at intermediate size in temperate deciduous trees. *Tree Physiol.* **2010**, *30*, 555–573. [\[CrossRef\]](#)
33. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
34. Segal, M.; Xiao, Y. Multivariate random forests. *WIREs Data Min. Knowl.* **2011**, *1*, 80–87. [\[CrossRef\]](#)
35. Ishwaran, H.; Kogalur, U.B. Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC) 2020. R Package Version 2.9.3. Available online: <https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf> (accessed on 16 January 2020).
36. Shang, C.; Treitz, P.; Caspersen, J.; Jones, T. Estimation of forest structural and compositional variables using ALS data and multi-seasonal satellite imagery. *Int. J. Appl. Earth Obs.* **2019**, *78*, 360–371. [\[CrossRef\]](#)
37. Lim, J.; Kim, K.M.; Jin, R. Tree species classification using Hyperion and Sentinel-2 data with machine learning in South Korea and China. *ISPRS Int. Geo-Inf.* **2019**, *8*, 150. [\[CrossRef\]](#)
38. Oksanen, J.; Blanchet, F.G.; Friendly, M.; Kindt, R.; Legendre, P.; McGlinn, D.; Minchin, P.R.; O'Hara, R.B.; Simpson, G.L.; Solymos, P.; et al. Vegan: Community Ecology Package 2019. R Package Version 2.5-6. Available online: <https://cran.r-project.org/web/packages/vegan/vegan.pdf> (accessed on 16 January 2020).
39. Khan, S.; Hossein, R.; Syed, A.S.; Mohammed, B.; Gerard, M.; Dickinson, S. *A Guide to Convolutional Neural Networks for Computer Vision*; Morgan & Claypool: San Rafael, CA, USA, 2018.
40. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
41. Hamraz, H.; Jacobs, N.B.; Contreras, M.A.; Clark, C.H. Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees. *ISPRS J. Photogramm.* **2019**, *158*, 219–230. [\[CrossRef\]](#)
42. Wang, S.; Cao, J.; Yu, P. Deep learning for spatio-temporal data mining: A survey. *IEEE Trans. Knowl. Data Eng.* **2020**. [\[CrossRef\]](#)
43. Ping, W.; Peng, K.; Gibiansky, A.; Arik, S.Ö.; Kannan, A.; Narang, S.; Raiman, J.; Miller, J. Deep voice 3: 2000-speaker neural text-to-speech. *arXiv* **2018**, arXiv:1710.07654.
44. Brousseau, B.; Rose, J.; Eizenman, M. Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model. *Sensors* **2020**, *20*, 543. [\[CrossRef\]](#)
45. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. Tensorflow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
46. Bergstra, J.; Yamins, D.; Cox, D.D. Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. In Proceedings of the 12th Python in Science Conference, Austin, TX, USA, 24–29 June 2013.
47. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020; Available online: <https://www.R-project.org/> (accessed on 18 December 2019).
48. Grabska, E.; Hostert, P.; Pflugmacher, D.; Ostapowicz, K. Forest stand species mapping using the Sentinel-2 time series. *Remote Sens.* **2019**, *11*, 1197. [\[CrossRef\]](#)
49. Chen, C.; Liaw, A.; Breiman, L. Using Random Forest to Learn Imbalanced Data. pp. 1–12. Available online: <https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf> (accessed on 22 April 2021).
50. Rejou-Mechain, M.; Muller-Landau, H.C.; Detto, M.; Thomas, S.C.; le Toan, T.; Saatchi, S.S.; Barreto-Silva, J.S.; Bourg, N.A.; Bunyavejchewin, S.; Butt, N.; et al. Local spatial structure of forest biomass and its consequences for remote sensing of carbon stocks. *Biogeosciences* **2014**, *11*, 6827–6840. [\[CrossRef\]](#)
51. Bohlman, S.; Pacala, S. A forest structure model that determines crown layers and partitions growth and mortality rates for landscape-scale applications of tropical forests. *J. Ecol.* **2012**, *100*, 508–518. [\[CrossRef\]](#)
52. Axelsson, A.; Lindberg, E.; Reese, H.; Olsson, H. Tree species classification using Sentinel-2 imagery and Bayesian inference. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *100*, 102318. [\[CrossRef\]](#)
53. Pan, L.; Xia, H.; Yang, J.; Niu, W.; Wang, R.; Song, H.; Guo, Y.; Qin, Y. Mapping cropping intensity in Huaihe basin using phenology algorithm, all Sentinel-2 and Landsat images in Google Earth Engine. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102376. [\[CrossRef\]](#)
54. Leckie, D.G.; Tinis, S.; Nelson, T.; Burnett, C.; Gougeon, F.A.; Cloney, E.; Paradine, D. Issues in species classification of trees in old growth conifer stands. *Can. J. Remote Sens.* **2005**, *31*, 175–190. [\[CrossRef\]](#)
55. Isaac, E.; Easwarakumar, K.S.; Isaac, J. Urban landcover classification from multispectral image data using optimized adaboosted random forests. *Remote Sens. Lett.* **2017**, *8*, 350–359. [\[CrossRef\]](#)

-
56. Zhang, Y.; Cao, G.; Li, X.; Wang, B. Cascaded Random Forest for hyperspectral image classification. *IEEE J. Sel. Top. Appl.* **2018**, *11*, 1082–1094. [[CrossRef](#)]
 57. Hościło, A.; Lewandowska, A. Mapping forest type and tree species on a regional scale using multi-temporal Sentinel-2 data. *Remote Sens.* **2019**, *11*, 929. [[CrossRef](#)]
 58. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010.
 59. Clevert, D.-A.; Unterthiner, T.; Hochreiter, S. Fast and accurate deep network learning by exponential linear units (elus). *arXiv* **2016**, arXiv:1511.07289.
 60. Kingma, D.P.; Ba, J.L.B. ADAM: A Method for stochastic optimization. *arXiv* **2015**, arXiv:1412.6980.
 61. Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.
 62. Tieleman, T.; Hinton, G. Lecture 6.5-Rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA Neural Netw. Mach. Learn.* **2012**, *4*, 26–31.
 63. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*; Lechevallier, Y., Saporta, G., Eds.; Physica-Verlag HD: Heidelberg, Germany, 2010. [[CrossRef](#)]
 64. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]