




## Article

# Characterization of Vegetation Dynamics on Linear Features Using Airborne Laser Scanning and Ensemble Learning

Narimene Braham <sup>1,\*</sup>,<sup>†</sup> , Osvaldo Valeria <sup>1,2</sup>  and Louis Imbeau <sup>1</sup> 

<sup>1</sup> Institut de Recherche sur les Forêts, Université du Québec en Abitibi-Témiscamingue, 445 Boulevard de l'Université, Rouyn-Noranda, QC J9X 5E4, Canada

<sup>2</sup> Hémera Centro de Observación de la Tierra, Escuela de Ingeniería Forestal, Facultad de Ciencias, Universidad Mayor, Camino La Pirámide, Santiago 5750, Chile

\* Correspondence: narimene.braham@uqat.ca

† This manuscript is part of a M.Sc. thesis by the first author, available online at (<https://depositum.uqat.ca/>).

**Abstract:** Linear feature networks are the roads, trails, pipelines, and seismic lines developed throughout many commercial boreal forests. These linear features, while providing access for industrial, recreational, silvicultural, and fire management operations, also have environmental implications which involve both the active and non-active portions of the network. Management of the existing linear feature networks across boreal forests would lead to the optimization of maintenance and construction costs as well as the minimization of the cumulative environmental effects of the anthropogenic linear footprint. Remote sensing data and predictive modelling are valuable support tools for the multi-level management of this network by providing accurate and detailed quantitative information aiming to assess linear feature conditions (e.g., deterioration and vegetation characteristic dynamics). However, the potential of remote sensing datasets to improve knowledge of fine-scale vegetation characteristic dynamics within forest roads has not been fully explored. This study investigated the use of high-spatial resolution (1 m), airborne LiDAR, terrain, climatic, and field survey data, aiming to provide information on vegetation characteristic dynamics within forest roads by (i) developing a predictive model for the characterization of the LiDAR-CHM vegetation cover dynamic (response metric) and (ii) investigating causal factors driving the vegetation cover dynamic using LiDAR (topography: slope, TWI, hillshade, and orientation), Sentinel-2 optical imagery (NDVI), climate databases (sunlight and wind speed), and field inventory (clearing width and years post-clearing). For these purposes, we evaluated and compared the performance of ordinary least squares (OLS) and machine learning (ML) regression approaches commonly used in ecological modelling—*multiple linear regression (mlr)*, *multivariate adaptive regression splines (mars)*, *generalized additive model (gam)*, *k-nearest neighbors (knn)*, *gradient boosting machines (gbm)*, and *random forests (rf)*. We validated our models' results using an error metric—root mean square error (RMSE)—and a goodness-of-fit metric—coefficient of determination ( $R^2$ ). The predictions were tested using stratified cross-validation and were validated against an independent dataset. Our findings revealed that the *rf* model showed the most accurate results (cross-validation:  $R^2 = 0.69$ , RMSE = 18.69%, validation against an independent dataset:  $R^2 = 0.62$ , RMSE = 20.29%). The most informative factors were clearing width, which had the strongest negative effect, suggesting the underlying influence of disturbance legacies, and years post-clearing, which had a positive effect on the vegetation cover dynamic. Our long-term predictions suggest that a timeframe of no less than 20 years is expected for both wide- and narrow-width roads to exhibit ~50% and ~80% vegetation cover, respectively. This study has improved our understanding of fine-scale vegetation dynamics around forest roads, both qualitatively and quantitatively. The information from the predictive model is useful for both the short- and long-term management of the existing network. Furthermore, the study demonstrates that spatially explicit models using LiDAR data are reliable tools for assessing vegetation dynamics around forest roads. It provides avenues for further research and the potential to integrate this quantitative approach with other linear feature studies. An improved knowledge of vegetation dynamic patterns on linear features can help support sustainable forest management.



**Citation:** Braham, N.; Valeria, O.; Imbeau, L. Characterization of Vegetation Dynamics on Linear Features Using Airborne Laser Scanning and Ensemble Learning. *Forests* **2023**, *14*, 511. <https://doi.org/10.3390/f14030511>

Academic Editors: José Ciro Hernández-Díaz, José Ángel Prieto-Ruiz, Christian Wehenkel and Emilio Badalamenti

Received: 5 January 2023

Revised: 16 February 2023

Accepted: 28 February 2023

Published: 5 March 2023



**Copyright:** © 2023 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/>).

**Keywords:** airborne light detection and ranging LiDAR; linear features; forest roads; road network; forest management; random forests; ensemble learning; vegetation dynamics; boreal forest

## 1. Introduction

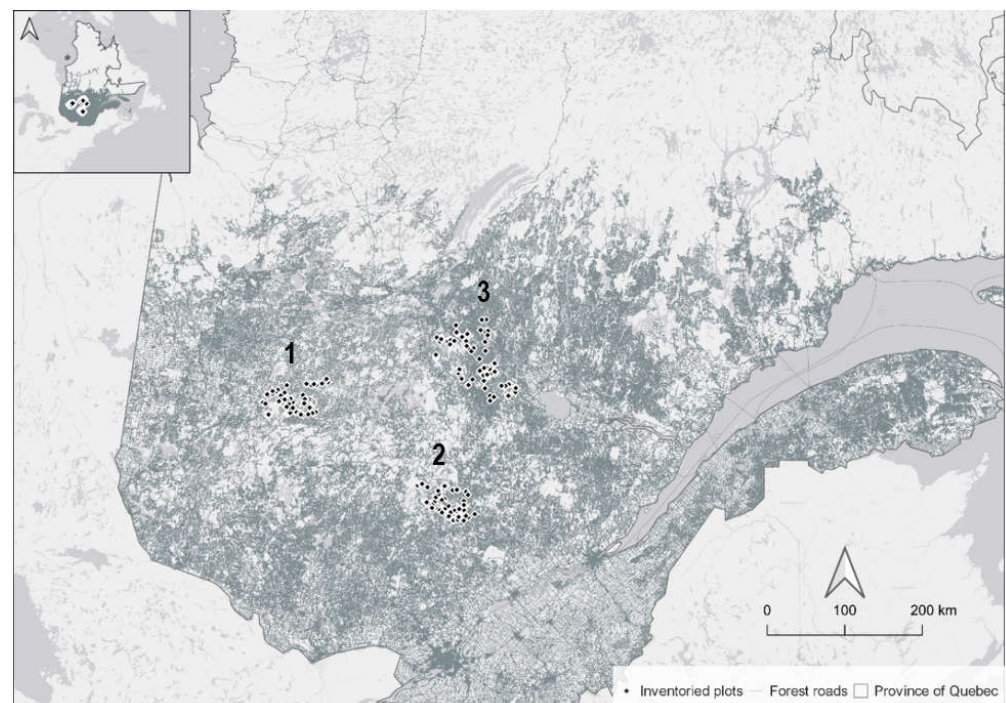
Anthropogenic linear features are forest access infrastructure, namely forest roads and seismic lines, and are essential for boreal forest natural resource provisioning and transportation. These features may have distinct morphological characteristics and functions, but their similar geometry and spatial patterns result in analogous environmental effects which allow their approximation. Particularly, linear features (LFs) are similar in terms of their disturbance legacies as they require the use of machines that result in the compaction of the surface layer through construction operations and consistent traffic intensity [1–6]. A consequence of these legacies is prolonged post-clearing vegetation growth. LFs also play a major role in expanding forest cover discontinuity as they represent an extensive crisscross in terms of their spatial distribution. In terms of their geometry, LFs have higher perimeter-to-area ratios and higher edge-to-area ratios [7,8]. Even if some of these LFs are temporary or deemed to have a “low-impact” [9,10], they contribute to fragmentation, with the majority (70%) of the world’s forests being within 1 km of a forest edge, leading to diminished habitat suitability adjacent to LFs caused by edge effects [11,12]. Moreover, LFs have direct effects on wildlife species [11,13–16], soil [5,17–19], seed dispersal and the spread of wind-dispersed invasive species [20], abiotic conditions [21,22], forest structure and composition, and their adjacent environment [23,24]. Since the most prevalent linear anthropogenic feature in many regions of eastern boreal forest are forest roads, the management of this vast network to minimize the associated linear footprint on biodiversity and wildlife habitat, requires an understanding of forest road vegetation characteristic dynamics [25,26]. However, vegetation patterns around forest roads need to be further explored: previous studies have shown that the growth process around linear features is complex and slow [27,28]. Furthermore, fine-scale knowledge on growth mechanisms around forest roads and the application of this knowledge to management of the linear footprint is based on limited spatial levels and time scales. Previous studies assessing the post-clearing, forest canopy spatio-temporal dynamic showed that the growth process is conditioned by disturbance factors, site conditions, and location [29,30]. Moreover, in natural canopy openings, factors such as light, nutrients, and water, have been shown to contribute and interact to affect the growth of individual trees and saplings [31]. Abib et al. [1] and Franklin et al. [21] confirmed this relationship for LFs and showed that variations in vegetation growth are explained by LF attributes (i.e., LF width and orientation), local environmental factors (i.e., sunlight availability and the potential for the accumulation of surface water) as well as terrain conditions. However, vegetation dynamics around forest roads require more research for a better understanding of the conditioning factors. The analysis of vegetation characteristic dynamics can be challenging if in situ measurements are used to acquire the information needed because forest roads are extensive throughout the landscape and have variable clearing widths which are permanently fluctuating over time due to vegetation growth in the immediate surroundings. Moreover, in situ measurements are restricted to a limited number of data points (high precision measurements from a few small plots) instead of continuous data, and require additional human resources to perform the field surveys. For this task, up-to-date, spatially explicit, and continuous information about vegetation three-dimensional characteristics (e.g., height and cover of the trees and shrubs, presence or absence of strata, canopy closure, gap fraction) is essential [32,33]. Remote sensing techniques can reliably expand the measurement possibilities of vegetation characteristics, across multiple levels (e.g., plot, landscape, region) and multiple time intervals. Particularly, LiDAR data can be used to accurately quantify a variety of metrics describing vegetation [34] as well as subcanopy topography [35]. Coupled with the fact that this information can be derived across a range of spatial scales from fine (e.g., ~1 m<sup>2</sup>)

to coarse (e.g.,  $\sim 100 \text{ km}^2$ ) [36], the use of LiDAR data should provide a way to advance the high-resolution quantification of vegetation and terrain characteristics around forest roads. For instance, high-resolution LiDAR data, in conjunction with various sources of ancillary data, have been recently incorporated into the modelling of fine-scale forest road deterioration [37,38]. LiDAR structural metrics related to height, density, and complexity are relevant for research on forest structural characteristics [39]. In our study, we considered a density-related LiDAR-Canopy Height Model (CHM) metric to derive the percentage of vegetation returns with a  $\geq 1.3$  height threshold [40]. This metric provides a measurement of the road surface covered in vegetation. The potential factors conditioning the vegetation cover response were selected to be available across the study area, consistent with the spatial resolution of available LiDAR data, and with the published literature assessing the influence on vegetation dynamics. In particular, the size of canopy openings [29,41], years post-clearing, disturbance history [29,30], topography and climate [42] were the main factors that have been shown to influence forest structural characteristic dynamics. Forests' structural characteristics are also determined by site conditions [43–46], species composition [47], and successional status [48]. Previous studies showed that these aforementioned candidate factors are relevant for the characterization of vegetation cover on LFs [21,27,28,49–51]. The extraction of ecologically relevant information on forest road vegetation characteristics requires the processing of canopy height model (CHMs) data into suitable metrics such as height metrics (e.g., the mean and maximum percentiles of height) and density metrics (e.g., percentage of vegetation returns  $\geq$  a given height threshold) [52,53]. These metrics are then used to develop products related to environmental modelling and forest management (i.e., a predictive model or a set of predictive models). For this purpose, machine learning (ML) approaches are usually the selected tool in forestry applications in the form of both classification and regression tasks due to the absence of distributional assumptions and the ability to fit nonlinear and complex relationships characterizing environmental and ecological data. Examples of predictive approaches include nearest neighbor methods (*knn*), e.g., [54–58], and multivariate adaptive regression splines (*mars*), e.g., [59–61]. In particular, ensembles approaches, e.g., *gradient boosting machines* (*gbm*) and *random Forests* (*rf*), are the tools of choice in forestry [62–65], and in forestry modelling applications with airborne LiDAR [1,66,67]. The widely used *rf* tree-based ensemble approach [68] is based on an aggregation of decision trees and uses several methods to introduce added randomness: i) through resampling, i.e., each tree is grown on a subset of the training points, and ii) through factor restriction (i.e., each decision tree uses a randomly selected subset of both the available factors and observations). At each step of the decision tree building process, a subset of the factors is randomly chosen, and the best factor and split point is chosen from that reduced set of factors. The average of decision trees is used to predict new observations. Other characteristics of *rf* are the reduced number of parameters to calibrate, and the choice of these parameters generally having very little influence on the accuracy of the results [69]. Although the *rf* approach is sufficiently versatile and widely used for such modeling, often the predictive capability of other ML techniques is not explored. The *gbm* is another tree-based ensemble approach and is a recent advance in predictive modeling [70]. The decision trees are sequentially built from the residuals of the preceding tree(s) and iteratively perform boosting through choosing, at each step, an arbitrary sample of the data, ultimately causing a progressive improvement in the model's performance [71,72]. However, *gbm* has yet to be tested to predict vegetation characteristics. To optimize accuracy and avoid overfitting using ML approaches, model parameter specifications are an important step. They usually involves a number of interacting parameters that have to be calibrated (i.e., regularized) in order to achieve optimal results [73]. Our primary aim is: (i) to investigate the predictive performance of six modelling approaches (*mlr*, *gam*, *mars*, *knn*, *rf*, and *gbm*) for the characterization of within-forest road vegetation cover dynamics, and (ii) to provide information on the underlying factors conditioning vegetation cover dynamics. We assumed that machine learning (ML) approaches would have better accuracies than ordinary least squares (OLS)

approaches. More specifically, tree-based approaches would show improved vegetation cover predictions. The evaluated approaches were constructed using ancillary geoclimatic as well as field inventory data. The required parameters for model fitting were set by using a 10-fold stratified cross-validation with 20 repetitions. For the final fitted model, parameters with the lowest error metric (root mean square error) were used and accuracy measures and analyses were conducted using both cross-validation and an independent validation dataset. The combined use of LiDAR measurements and predictive modelling allow for a fine-scale and representative measurement of forest road vegetation cover dynamics. This would also enhance our ability to precisely predict this dynamic along a spatial continuum and over extended timeframes.

## 2. Study Sites

For this study, we retrieved forest road clearing width data from the field, across three study areas representative of Canadian forestry activity, between 47° and 49° N and 72° and 78° W, in the mixed and coniferous boreal forest of Quebec (Canada) (Figure 1). The field data were collected in August 2019, as described in Girardin et al. [37]. The climate across our study areas is typically boreal, with very cold winters and short cool summers. The temperatures change according to latitude and altitude, with the southernmost and northernmost sites being the warmest and the coldest, respectively, and the sites at higher altitudes being the coldest in winter and the least warm in summer. Precipitation also varies along the latitudinal gradient, with drier conditions toward the North. The mean annual temperatures range between  $-5.9$  and  $4.2$  °C and total precipitation ranges between 650 to 1424 mm. The May–September mean temperatures range between  $9.1$  and  $17.7$  °C. The study areas are characterized by a gently rolling topography, with the highest mountains concentrated in the southern part, and thick and undifferentiated glacial deposits [37,74–76]. Table 1 provides a brief description of our study sites.



**Figure 1.** Overview of forest road network (dark grey polylines) and distribution of sampled field plots (black dots) within the three respective study areas (1–3) in the province of Quebec in eastern Canada.



**Table 1.** Properties of forest roads and their bioclimatic data, grouped by study area (1–3). The information in the table is in part adapted from [37,76–78].

Characteristic	Study Area 1	Study Area 2	Study Area 3
Location	Northeastern Abitibi-Témiscamingue region	Mauricie region	Northeast of the Saguenay-Lac-Saint-Jean region
Latitude/Longitude	(48.42° N, 77.23° W)	(47.51° N, 72.78° W)	(48.89° N, 72.23° W)
Mean elevation of sampled roads (m)	393	430	407
Total number of sampled plots	84	73	84
Cumulative length of sampled roads (km)	4.2	3.65	4.2
Mean clearing width measured in the field (m)	8.59	7.74	8.55
Mean years post-clearing (years)	9.23	6.83	6.17
Mean slope (%)	5.10	5.58	4.27
On-road mean vegetation coverage * measured in the field (m)	0.47	0.41	0.46
On-road mean tree height measured in the field (m)	4.22	6.08	5.22
On-road mean shrub height measured in the field (m)	1.24	2.87	2.19
Average annual temperature (°C)	1.5	3.8	1
Annual precipitation (mm)	875	928	999
Bioclimatic domain/Vegetation type	Balsam fir [ <i>Abies balsamea</i> (L.) Mill.]—White birch ( <i>Betula papyrifera</i> Marsh.)	Balsam fir—Yellow birch ( <i>Betula alleghaniensis</i> Britton)	Black spruce <i>Picea mariana</i> (Mill.)—Moss domain and Balsam fir—White birch

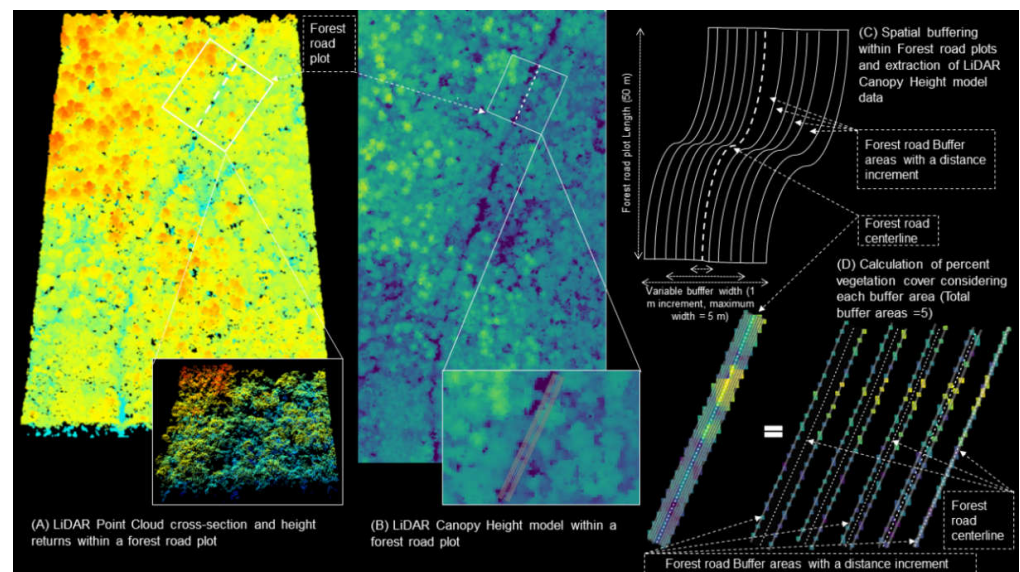
\* Vegetation coverage measured as the ratio of the mean width of the road covered in vegetation to the original width of the road, both measured in the field.

### 3. Data

#### 3.1. Reference Data

We used 240 rectangular field plots (50 m length) which were at least 250 m apart from one another. These field plots were randomly sampled among a selection of forest road stratified by the clearing width class (three classes: narrow, medium, and wide), years post-clearing (YPC) class (two classes: short-term and long-term timeframes), and slope class (two classes: low and high longitudinal slope, range: 0%–16%), following Girardin et al. [37]. Clearing width varied between 4 and 14.4 m and included winter only roads and all-weather gravel roads. Paved highways were not considered. YPC ranged between 0 and 46 years and was estimated based on the time elapsed since the last clearing (maintenance or construction). Maintenance activities usually consist of culvert repairs, surfacing, layer gravelling, and vegetation clearing. These reference data were used for the retrieval of geospatial information spanning from the road centerline, as described in the data extraction step (Section 3.2) and Figure 2.

For visualization purposes, clearing widths were binned into narrow forest roads (total narrow forest roads = 96), which were  $\leq 7$  m wide, and wide forest roads, which were  $>7$  m wide (total wide forest roads = 144) [6].



**Figure 2.** Visualization of LiDAR-based data. (A) 3D point cloud. (B) Canopy height model (CHM) over a forest road. (C) Extraction of forest road plot-level vegetation cover (%) using the CHM. (D) Calculation of mean vegetation cover, continuously, within the five multi-buffer areas (length = 50 m, and width increment = 1 m).

### 3.2. Forest Road Data Extraction

#### 3.2.1. Digital Surface Models (DSM) and Canopy Height Models (CHM)

LiDAR point clouds feature the heights of objects on the ground. Digital surface models (DSM), canopy height models (CHM), and digital terrain models (DTM) are common layers derived from point clouds after the classification of individual LiDAR points. DSM and CHM feature the highest elevation of ALS returns. DTM represents the elevation of the ground. In vegetated areas, the CHM represents the heights of the trees on the ground. It can be derived by subtracting the ground elevation (represented by the DTM) from the elevation of the top of the surface, or the tops of the trees (represented by the DSM). Different methods exist to create gridded DSMs and CHMs [79]. In this study, we used LiDAR-based gridded products provided by the government of Quebec [80].

#### 3.2.2. Spatial Buffering for Data Extraction

To recreate the footprint polygons of forest roads from field-inventoried centerlines, we first delineated and digitized the centerlines using GPS coordinates (Trimble GNSS Hand-held Geo7X, provided by Trimble Inc., Westminster, CO, USA) (three sampling locations for the edges and midpoint of the 50 m centerline) (Supplementary Material, Figure S1). To ensure the proper alignment of the digitized centerlines, we used the LiDAR datasets provided by the Government of Quebec's airborne LiDAR surveys, consisting of 1 m × 1 m grids [80], collected under growing season conditions between 2016 and 2020, with a mean pulse density of 2–4 pulse/m<sup>2</sup> [81]. More specifically, we derived the topographic position index (TPI) from the digital terrain model (DTM) (spatial resolution of 1 m) to locate topographic breaks and inspect roadside geomorphological attributes (i.e., the drainage structures or ditches). We then performed a buffer analysis to partition the geographic space around the digitized centerlines into multi-buffers with similar areas (1 m increment). This spatial buffering step resulted in 1 m wide “hollow” multi-buffers that extend over 5 m, which we used to compute our input dataset (vegetation cover response and causal factors) for the characterization of the vegetation dynamics (Figure 2). All data processing, modelling, as well as validation were performed in the R project for statistical computing, software environment (Version 4.1, R Core Team) [82]. All regression models were produced using the *caret* library 6.0 [83].

### 3.2.3. Causal Factor Data Computation

We established a framework that used data from multiple sources, including airborne LiDAR and geo-climatic data. We extracted these data from the 240 field plots using the multi-buffer delineation approach described in Section 3.2.1 and Figure 2. The proposed approach had a fixed length (50 m) and a variable width extending over 5 m, which allowed us to derive our data with a distance increment from the road centerline. All training data were extracted within the boundaries of the delineated multi-buffer areas, annotated 1 to 5, indicating the buffer width. For buffer areas more than one meter wide, data were extracted within hollow bands to exclude data points from the other buffers.

Specifically, we used the LiDAR-CHM data to measure the vegetation cover response and LiDAR-based terrain data (1 m resolution) to compute: (i) Slope, in degrees. (ii) orientation (northernness) transformed to a continuous factor ranging between  $-1$  and  $1$  (The northernness values closer to  $-1$  are southwards and those closer to  $+1$  are northwards) [28]. Orientation is typically transformed into a continuous factor because it is circular (large values may be very close to small values). (iii) The topographic wetness index (TWI) is used as a proxy for soil moisture. It provides information on the potential for water accumulation over the land as a function of slope and accumulation at a given pixel. More specifically, TWI integrates the water supply from an upslope catchment area and downslope water drainage for each cell in a digital terrain model [84]. (iv) Hillshade is a proxy for the shadow based on the surface elevation [85,86].

NDVI (normalized difference vegetation index) extracted from Sentinel-2, resampled to 1 m resolution, provides a measure of the difference between the reflectance of wavelengths emitted by the sunlight in the near infrared (PIR) and in the visible red band [87,88].

Climate data were obtained from WorldClim (Version 2.1) for the time period 1970–2000 [89,90]. This dataset is based on historical climate records at a resolution of 30 s. The available monthly climate data of precipitation, incident sunlight (in units of  $\text{kJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$  wind speed ( $\text{m}\cdot\text{s}^{-1}$ ), total precipitation (mm), and minimum, mean, and maximum temperature ( $^{\circ}\text{C}$ ), were used to compute the growing season climate dataset, resampled to a 1 m resolution. Only two growing season averaged climatic factors, namely incident sunlight and wind speed, were retained for further analyzes, because a high correlation between the initial variables was found in Pradhan and Setyawan, 2021 [91]. Particularly, sunlight is a proxy for vegetation growth as it moderates the available photosynthetically active radiation. Sunlight and wind speed are proxies for the potential for in situ evapotranspiration due to locally warmer/drier or cooler/shaded conditions, as suggested in Stern et al. [22], and van Rensen et al. [28].

Prior to the modelling analysis, we checked for outliers using the interquartile range and removed all values above the 95th and below the 5th percentile, as well as collinearity (relationships between more than two covariates), and correlation (linear relationships between two covariates), following Zuur et al. [92]. All uninformative metrics that showed a variance inflation factor greater than 3 or were highly correlated with one another ( $|r \text{ Pearson}| > 0.7$ ) were excluded from the analysis. We summarize in Table 2, the various factors examined and their description.

**Table 2.** Overview of the factors used in the modelling of vegetation cover. Geospatial layers had a cell resolution of 1 m or were resampled to 1 m prior to the modelling step, for all the factors.

Data Source(s)	Factor(s)	Unit	Description	Spatial/Temporal Resolution	
LiDAR-based, CHM	Vegetation cover (response)	%	Mean vegetation cover (height above 1.3 m) within the buffer area	1 m	-

Table 2. Cont.

Data Source(s)	Factor(s)	Unit	Description	Spatial/Temporal Resolution	
LiDAR-based, Terrain	(i) Slope	%	Mean slope within the buffer area	1 m	-
	(ii) Orientation (Northernness)	Unitless	Mean northernness index within the buffer area	1 m	-
	(iii) TWI	Unitless	Mean TWI index within the buffer area	1 m	-
	(iv) Hillshade	Unitless	Mean hillshade index within the buffer area	1 m	-
Climate	NDVI	Unitless	Mean NDVI index within the buffer area	1 m	-
	Solar radiation	Kj·m <sup>-2</sup> ·day <sup>-1</sup>	Mean solar radiation within the buffer area	1 m	30 s
	Wind speed	m·s <sup>-1</sup>	Mean wind speed within the buffer area	1 m	30 s
Linear feature attributes	Clearing width	m	Line width derived from three measurement plots along the 50 m plot	-	-
	Years since last clearing (clearing) (YSC)	years	Time since last clearing (establishment or maintenance)	-	-

## 4. Methods

### 4.1. Statistical Approaches

To provide an optimal predictive model for the estimation of vegetation cover on forest roads, we compared the performance of the following OLS regression approaches: (i) *multiple linear regression (mlr)*, (ii) *multivariate adaptive regression splines (mars)*, (iii) *generalized additive model (gam)*, (iv) *k-nearest neighbors (knn)*, (v) *random forests (rf)*, and (vi) *gradient boosting machines (gbm)*.

*mlr* was assessed for its straightforwardness and simplicity and was extended to *gam*, a flexible approach used to identify and characterize non-linear regression effects [93].

*gam* was included because it presents an advantage over predefined basis functions to achieve nonlinearities and is relatively easy to interpret [94].

The parsimony of *mlr* and *gam* approaches were assessed with the Akaike information criterion (AIC) [95]. All possible combinations of factors and interaction effects were analyzed with the *MuMIn* library in R [96]. This step was essential because the inclusion of uninformative factors in parametric and semi-parametric models (i.e., *mlr* and *gam*) can reduce their overall predictive performance.

*mars* is also regarded as an extension of linear models and is an adaptive non-linear estimation method that can present interaction between influencing attributes without any assumptions about input data distribution [97]. It structures a relation from established basis functions and coefficients, which are generally determined from the regression information [98]. The construction phase of a *mars* model involves adding and removing of basic functions. *mars* is considered as a modification of the classification and regression tree (CART) method, to improve the latter's performance in a regression setting, owing to *mars'* ability to capture additive effects [93]. Therefore, *mars* could simplify the challenges of solving non-linear relationships, compared to other non-parametric approaches [98].

We used the basic *knn* method [99], a simple and intuitive approach in which each observation is predicted based on its similarity to other observations [69]. More specifically, the prediction of new observations values uses the sampled observations from a training



data set that are the closest (nearest neighbor(s)) to each new observation. The similarity between new and training samples is based on a Euclidean distance metrics (or other related metrics) [100]. *knn* is considered a simple approach as there is no model to be fit and the prediction results depend on feature scaling, measurement of similarity, and the value of *k*. Other advantages include decent predictive power, especially when the response is dependent on the local structure of the features [100], flexible assumptions regarding normality and homoscedasticity required by parametric methods, and the preservation of much of the covariance structure among the metrics that define the response and factors' vectors [99].

*rf* is tree-based ensemble which builds a large collection of independent decision trees to further improve predictive performance by averaging individual predictions. More specifically, *rfs* use a combination of bagging, which randomly selects factors with replacement as training for growing the trees, which makes it robust against overfitting [101]. The training is carried out on datasets created from a random resampling on the training set itself, which adds an extra layer of randomness [68,101].

*gbm* is another recent tree-based ensemble which builds a base model (i.e., trees with only a few splits) [102] and the additional trees iteratively correct mistakes made by the previous trees, which progressively improves prediction accuracy. Particularly, *gbm* sequentially generates base models from a weighted version of the training data to find the optimal combination of trees and optimize predictive performance [69,103].

Both *rf* and *gbm* present the numerous advantages of tree-based ensemble methods, accommodating different types of factors and efficiently dealing with missing data and outliers. They have no need for prior data transformation, can fit complex non-linear hierarchical relationships, and automatically handle interaction effects between the factors [94].

#### 4.2. Model Parameter Tuning

ML model performance can benefit significantly from tuning as it may reduce overfitting [73,104]. The *caret* library [83] was used to execute a grid search for each model where we assessed every combination of parameters of interest. More specifically, for *mars*, relevant model parameters were related to the number of retained terms (*nprune*) and the degree of interactions (*degree*) [69,105]. The implementation and performance of *knn* approaches required choices for three parameters: the value for *k*, the number of nearest neighbors (in a regression setting, for *k* = *n*, the average is used across all training samples as the predicted value), a scheme for weighting neighbors when calculating predictions (kernel function), and a similarity metric (distance). The prediction performance of *rf* is influenced mainly by three model parameters: correlation between individual trees, the performance of each tree, and the total number of trees [106]. Hence, we executed a grid search to evaluate: *ntree*, which is the number of trees in a forest, and *mtry*, which defines the number of random factors at each split [69]. For *gbm*, we performed sensitivity analyses on tree complexity (interaction depth), learning rate (shrinkage), and the minimum number of observations in nodes (*minobs*) [69,105]. During the tuning phase, a stratified 10-fold cross-validation resampling method allowed us to partition the training set for each fold. Model performances of every parameter combination were computed at the tuning level and averaged across all folds. The parameter combination with the lowest RMSE was used to train our model during the performance assessment phase. Details about the parameter values and combinations that optimized the RMSE for our data can be found in Supplementary Material Table S1.

#### 4.3. Model Performance, Comparison, and Diagnostics Using Cross-Validation and Independent Dataset

Inheriting spatial information from dependent observations is one of the main challenges of spatial statistical modeling using ML techniques [73,107–110]. In this regard, to account for spatial dependencies in our spatially explicit data and reduce prediction bias, the choice of cross-validation (resampling technique) emerged as an important step in

the implementation of our approaches [73,109,111]. Therefore, we performed a stratified 10-fold cross-validation, with the forest road identifiers being the stratifying factor. This allowed the condition of equal distribution of our stratified samples between (i) training, testing, and validation samples, and (ii) the cross-validation folds to be met (e.g., [112]), which showed that dividing by strata produces similar distributions between training and testing sets for the majority of validation folds. The stratified partitioning was conducted prior to modelling and the samples were randomized with respect to the established strata. It is suggested that when the set of factors affect the response in different ways (positive/negative and/or linear/non-linear) and a model's output is transferred to unsampled locations, more rigorous validation is necessary [113]. We conducted a 60%–40% training–validation combination to evaluate our model's performance. In addition, to avoid skewed results, each model was run 20 times (20 repetitions). Both stratified cross-validation and independent validation (using the hold-out 40% of our data) performance were evaluated with the RMSE and the mean absolute error (MAE) metric to assess the accuracy. The  $R^2$  metric was used to evaluate the goodness-of-fit. Model performance metrics were taken as the mean from the number of repeats. After the models were trained and compared, we assessed visual diagnostics and factor importance computed from the fitted model that yielded optimal results (i.e., *rf*). *rf* typically includes a permutation-based importance measure which assesses the decrease in accuracy averaged over all the trees for each factor. The factors with the largest average decrease in accuracy across all trees are considered the most important [69]. The factor importance computation was implemented using the *varImpPlot* function in the *Random Forest* library [114]. Partial dependence plots (PDPs) are especially useful for visualizing the relationships discovered from ML approaches by isolating the effect of a single factor on the response [115]. We evaluated the partial dependence from our fitted *rf* model using two functions *partial* and *plotpartial* [116] as there are advantages for model specific interpretations such as a close relation to the model performance and an accurate incorporation of the correlation structure between factors [115].

## 5. Results

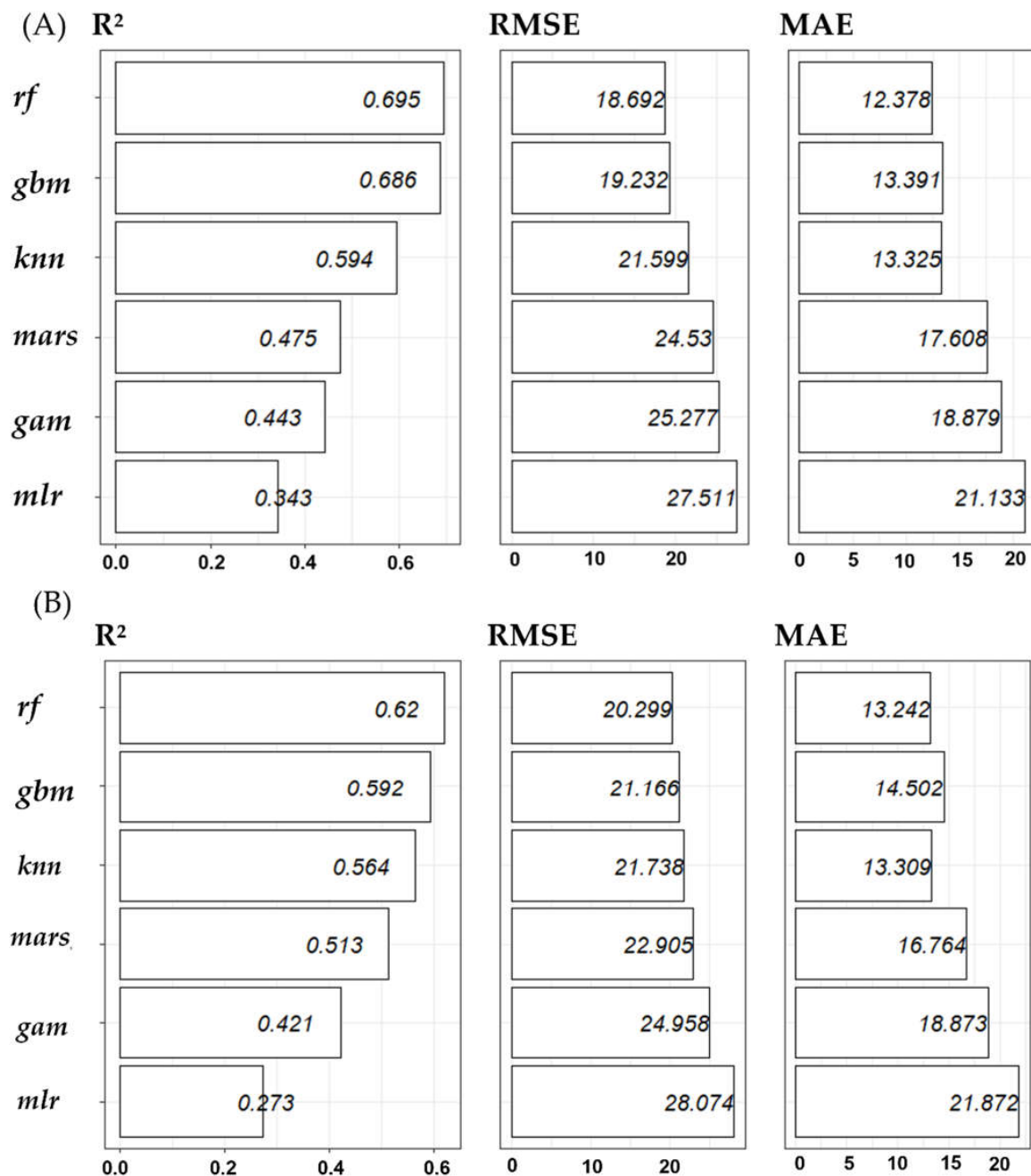
### 5.1. Modelling Approaches' Performance

For the study, vegetation cover (LiDAR-measured vegetation cover (%)), forest road attributes (clearing width (m) and years post-clearing (years)) by means of in situ measurements, climatic factors (sunlight ( $\text{kJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ ) and wind speed ( $\text{m}\cdot\text{s}^{-1}$ )), terrain factors (slope (%), northernness (index), TWI (index) and shade (index)) were computed. An overview and the distribution of these input data are summarized in Table 3.

**Table 3.** Distribution of model input data for the characterization of vegetation cover dynamics on forest roads.

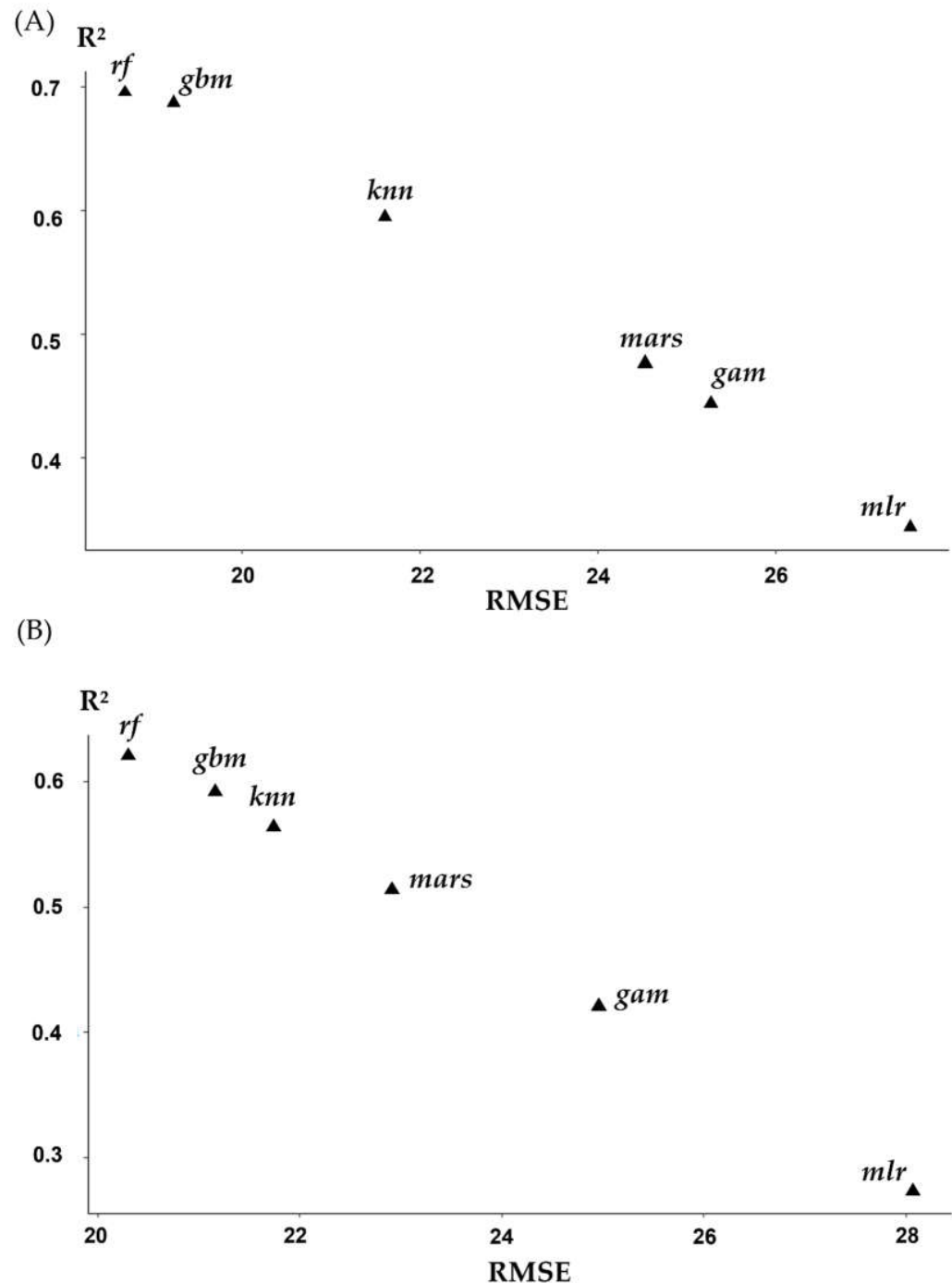
Input (s)	Min	Max	Range	Median	Mean	Standard Deviation
LiDAR measured vegetation cover (%)	0	100	100	0	22.07	33.36
Slope (%)	0	27.73	27.73	6.71	7.94	5.41
Northernness (index)	−0.55	0.46	1	−0.01	−0.03	0.2
TWI (index)	1.72	16.46	14.74	6.52	6.88	2.81
Hillshade (index)	139.68	202.97	63.29	178.82	177.7	9.85
NDVI (index)	0.12	0.89	0.77	0.66	0.62	0.19
Sunlight ( $\text{kJ}\cdot\text{m}^{-2}\cdot\text{day}^{-1}$ )	17,228.74	17,729.8	501.06	17,598.99	17,545.63	136.74
Wind Speed ( $\text{m}\cdot\text{s}^{-1}$ )	2.2	2.88	0.68	2.34	2.45	0.2
Clearing width (m)	4	14.47	10.47	7.4	8.24	2.48
Years post-clearing (years)	0	39	39	7	7.79	8.35

The predictive performance of ML approaches (*rf*, *gbm*, *knn*, and *mars*) and OLS (*gam* and *mlr*) approaches using stratified cross-validation and independent datasets are shown in Figure 3A,B, respectively. ML approaches consistently had higher testing and validation RMSE and higher  $R^2$  values than OLS approaches. The greatest accuracy was obtained with the *rf* approach (RMSE ranging from 18.69% to 20.29% and  $R^2$  ranging from 0.69 to 0.62), followed by *gbm* (RMSE ranging from 19.23% to 21.16% and  $R^2$  ranging from 0.68 to 0.59), and finally *knn* (RMSE ranging from 21.59% to 21.73% and  $R^2$  ranging from 0.59 to 0.56).



**Figure 3.**  $R^2$ , RMSE, and MAE for ML and OLS approaches for the characterization of vegetation cover dynamics obtained from (A) 10-fold stratified cross-validation (results from 20 repetitions were considered) and (B) an independent validation dataset. *rf* = random forests, *gbm* = gradient boosting machines, *knn* = *k*-nearest-neighbors, *mars* = multivariate adaptive regression splines, *gam* = generalized additive model, *mlr* = multiple linear regression.

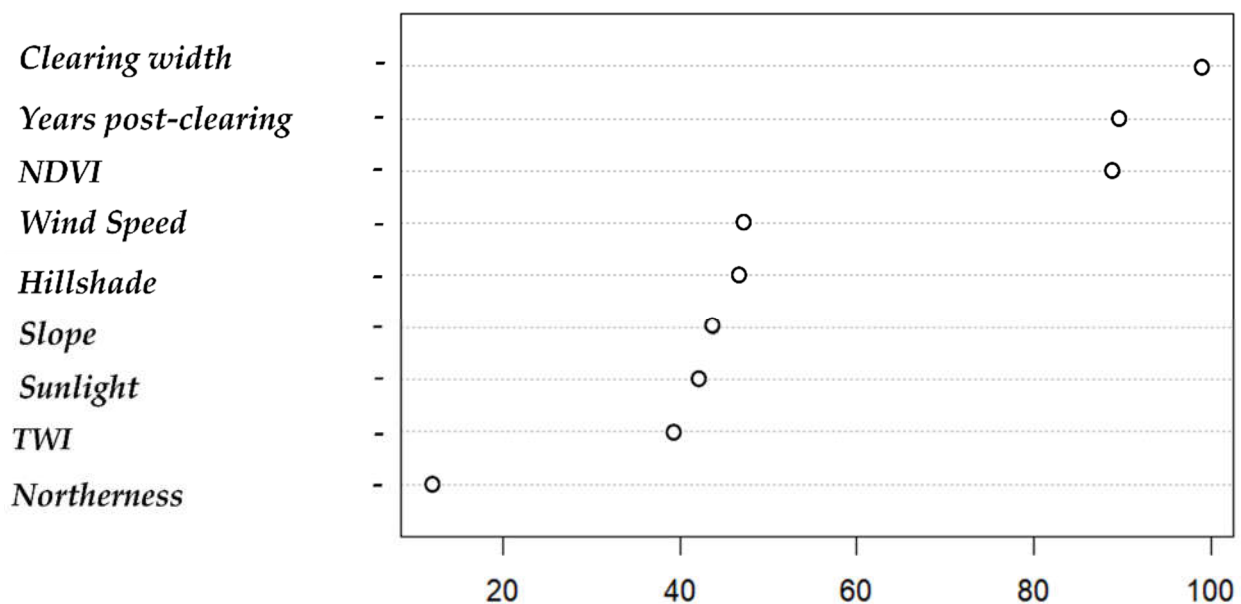
Assessed using RMSE and  $R^2$  (Figure 4A,B), the highest relative improvement in predictive performance was found using tree-based ensemble approaches (i.e., *rf* and *gbm*). Particularly, *rf* and *gbm* were similar in terms of predictive capability; they showed the highest predictive accuracy. *knn* and *mars* approaches showed slight reductions in the predictive capability compared with the *rf* and *gbm*, and significant reductions were obtained with the *mlr* approach compared with *rf*.



**Figure 4.** Predictive performance of ML and OLS for the characterization of vegetation cover dynamic using (A) 10-fold cross-validation approaches, (B) An independent validation dataset. *rf* = random forests, *gbm* = gradient boosting machines, *knn* = k-nearest-neighbors, *mars* = multivariate adaptive regression splines, *gam* = generalized additive model, *mlr* = multiple linear regression.



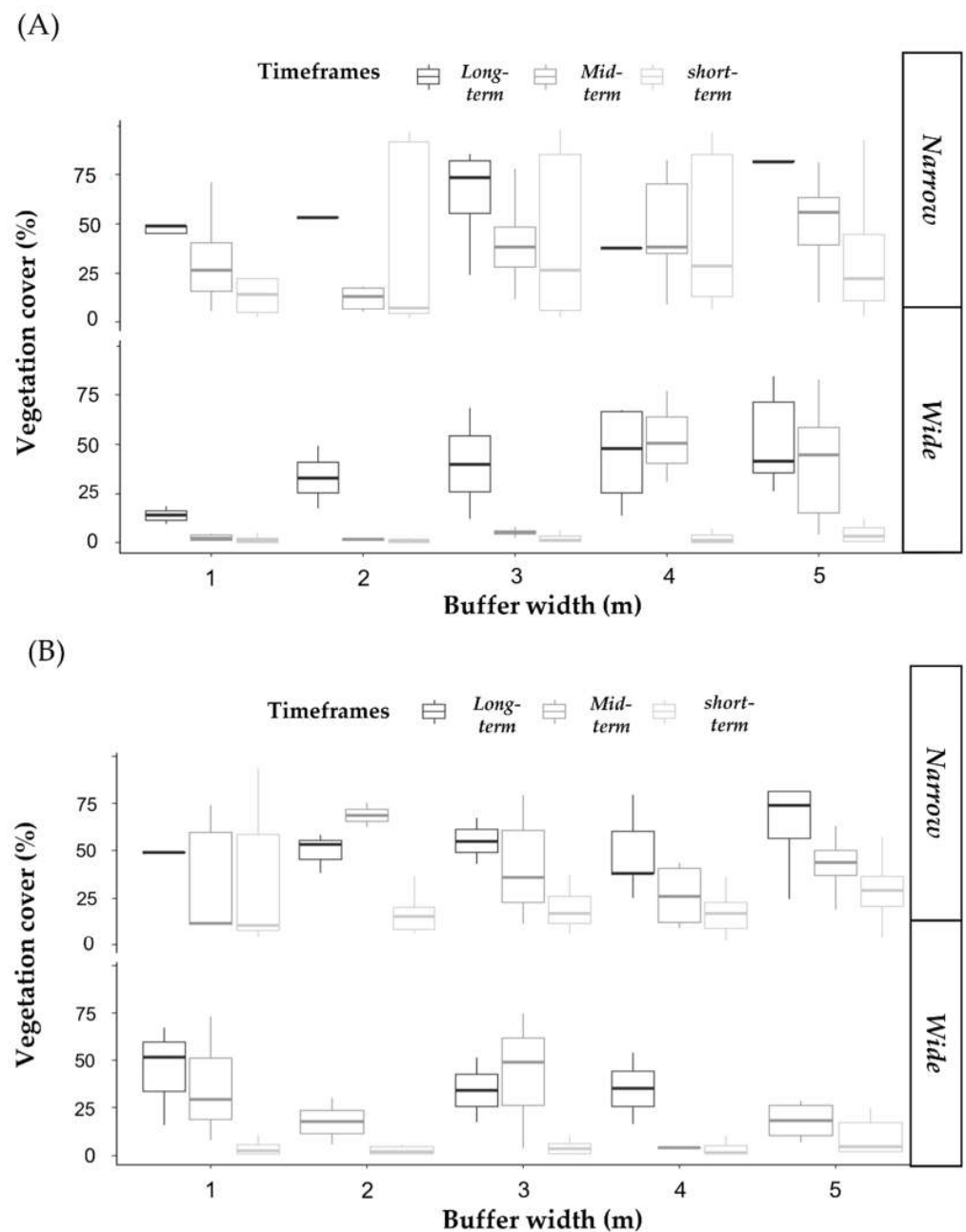
The causal factors which contributed most to the accuracy of vegetation cover characterization using *rf* are shown in Figure 5. Because *rf* generally provided optimal performance results, factor ranking was derived using this approach. Clearing width was the most important factor explaining vegetation cover dynamics around forest roads. The importance of all the other factors was lower: years post-clearing (YPC), NDVI, as well as geoclimatic (wind speed, sunlight, slope) and shade factors were of intermediate importance. The PDPs of the *rf* regression revealed a general downward trend of vegetation cover with increasing clearing width, sunlight, hillshade and TWI as well as a general upward trend with increasing years post-clearing, wind speed, slope, northernness and NDVI. PDPs for clearing width show that vegetation cover drops substantially as the clearing width increases until the width was approximately 6 m (Supplementary Material, Figure S3).



**Figure 5.** *rf*-based factor importance by permutation accuracy. A higher average importance of the variable (X-axis) indicates a greater contribution of this individual variable in explaining within-forest road vegetation cover dynamic. A ranking of all factors is included.

## 5.2. Characterization of Vegetation Cover Dynamic around Forest Roads

*rf*-based vegetation cover dynamics grouped by buffers extending from the road centerline (1–5 m), timeframe (short-, mid- and long-term), and clearing width (narrow and wide) are shown in Figure 6A using the cross-validation predictions, and Figure 6B using the independent dataset predictions. Overall, vegetation cover predictions were greater within the buffers furthest from the centerline. For the short-, mid- and long-term timeframes, the patterns were consistent: vegetation cover increased with YPC, with vegetation cover predictions on narrow forest roads slightly exceeding those on wide forest roads. Particularly, predictions grouped by timeframe showed that long-term vegetation cover (>20 YPC timeframe) exceeded those experienced in the mid- ([10–20] YPC timeframe) and short-term ([0–10] YPC timeframe), indicating a positive effect of YPC. Vegetation cover varied also across forest road types: narrow forest roads exhibited higher predictions over time across all five buffers with a higher range and higher mean predictions. The lowest prediction (~1.6%) was shown for wide roads for the short-term timeframe and the highest (~82.3%) for narrow forest roads for long-term timeframes. Wide forest roads showed an average vegetation cover of ~3%–53% and ~14%–52% in the mid- and long-term, respectively. Narrow forest roads showed an average of ~17%–51% and ~40%–82%, in the mid- and long-term, respectively (Supplementary Material, Figure S2A,B).



**Figure 6.** (A) Boxplots representing cross-validated *rf* model predictions ( $R^2 = 0.69$ , RMSE = 18.69%) of vegetation cover recorded within the multi-buffers extending from the road centerline, across forest road types (wide and narrow roads) for the post-clearing timeframes: >20 YPC (long-term, black boxes), [10–20] YPC (mid-term, dark grey boxes), and [0–10] YPC (short-term, light grey boxes). (B) Boxplot of vegetation cover predictions values from the *rf* model ( $R^2 = 0.62$ , RMSE = 20.29%) considering the independent validation dataset. The X axis indicates the width of every individual buffer. Boxplots present the median (dark black line),  $\pm 1$  standard deviation (rectangle) and maximum-minimum value (vertical lines or whiskers).

As shown in Figure 3A,B, the stratified cross-validation testing dataset had a higher accuracy of prediction than the independent validation dataset. Both testing (cross-validation) and independent validation datasets were considered as stratified random samples, but the testing dataset had a closer relationship with the training dataset (reference population), as records from all strata were included in both the training and the test subsets (Figure 6A,B).

In general, we found the ML approaches evaluated here to be useful tools for improving predictions of vegetation cover dynamics on forest roads.

## 6. Discussion

### 6.1. Modelling Approaches' Performance

The performance results of OLS and ML approaches demonstrated that *rf* was the most reliable model, exhibiting the best prediction accuracy rates among the *gbm*, *knn*, *mars*, and *gam* approaches. The least accurate model was *mlr*. These results suggest that using ML approaches was appropriate for the characterization of vegetation cover dynamics around forest roads. Furthermore, compared to *rf* and *gbm*, *knn*, *mars*, and *gam* showed minimal accuracy reductions. Conversely, *mlr* performed poorly. The significant performance difference between *mlr* and *rf* can be explained by the limitation in handling non-linear relationships between the vegetation cover response and causal factors, as well as model assumptions about the non-linear distribution of input data: *rf* better accommodates non-linear relationships between factors that *mlr* could not adequately solve [117–119]. Consistent with our hypothesis, tree-based ensemble approaches outperformed their nonensemble counterparts. *rf* and *gbm* are extremely randomized trees and are both based on ensemble learning theory. The ensemble—aggregation of decision trees [117]—considerably improves the accuracy and certainty of the predictions by suppressing the weaknesses and disadvantages of each individual decision tree, and by taking advantage of the responses of the combined decision trees [66,68,120,121]. ML approaches require the setting of parameter specifications prior to modeling to reduce overfitting and enhance performance. For this reason, the use of *rf* can be more straightforward because of its ability to yield accurate results when default parameters are used [122]. These findings, and previous results, suggest that no single ML algorithm might serve best for every task and that many models should be calibrated to identify the most accurate model for a given prediction task [55,104,113,118,123,124].

### 6.2. Factors Conditioning Vegetation Cover Dynamic around Forest Roads

#### 6.2.1. Factors Associated with Vegetation Dynamics

Our results identified that the most influential factors that explained significant vegetation cover variations were clearing width and years post-clearing (YPC). Particularly, vegetation cover was greatest in samples with narrow widths and long post-clearing time frames. NDVI, terrain (i.e., slope, hillshade, TWI, and northernness) and climatic factors (i.e., wind speed and sunlight) ranked lower. The samples where vegetation cover was most advanced had higher NDVI values, steeper slopes, higher orientation values, higher levels of wind speed, lower incident sunlight, shade, and TWI levels. Abib et al. [1], and Franklin et al. [21], showed that variations in proximity-based vegetation cover are explained by LF attributes (i.e., LF width and orientation) and local environmental factors (i.e., incident sunlight and the potential for accumulation of surface water). More evidence comes from van Rensen et al. [28], where clearing width was a strong predictor of growth occurrence within LFs (>3 m height cut-off was applied as a criterion for growth occurrence). It was suggested that clearing width implicitly reflects the severity of soil disturbance moisture supplies. Additionally, the ecosite type was the most important factor associated with growth (LF lines in bogs and fens were less likely to experience growth than those in drier conditions). Similarly, Finnegan et al. [125], suggested that soil wetness, nutrients, and adjacent stand affected growth levels. LFs in wet areas were least likely to promote vegetation growth and wet seismic LFs that were adjacent to more open forest stands were more likely to promote the occurrence of disturbance-tolerant taxa.

#### 6.2.2. Clearing Width and Its Relationship to Disturbance Legacies

Narrow-width forest roads experienced higher levels of vegetation cover, likely because of reduced disturbance (i.e., use of machinery in the construction phase and continuous vehicular traffic), supporting findings from the LF literature [27,28]. Particularly, LF

construction and design specifications can differ with respect to their characteristics (e.g., bearing capacity) and moisture conditions [126]. These differences are reflected in their trafficability, frequency, and intensity of use [127]. For instance, coarse material with higher levels of granular content (coarse gravel and/or crushed rock) is frequently used as a top layer on wide LFs to ensure higher bulk density and bearing capacity [127,128]. Due to their high trafficability, wide LFs are also prone to experience an increased intensity of use by heavy machinery (heavy vehicles inflict more damage to the surface layer than lighter vehicles), trucks, and off-road vehicles, which lead to severe disturbance of the top surface over longer time frames [6,38,127,129,130]. A consequence of compaction is the alteration of the hydro-physical properties in the surface layer. Therefore, it is likely that increased trafficability results in higher levels of compaction, which reduce porosity and infiltration, increase pore water pressure in the road material, and lead to long-term restricted water exchanges, flow, and moisture storage capacity. Gartzia-Bengoetxea et al. [131], showed that soil compaction caused by shearing and ripping persisted for 15 years. In addition, water holding capacity was lower in mechanically prepared plots 15 years after site preparation. Cambi et al. [132], showed that except for coarse textured excessively drained soils, soil compaction reduces oxygen and water availability to roots and microorganisms. Zang and Ding [51], suggested that compaction potentially interferes with the establishment of woody species on the surface of the LFs by reducing water infiltration, soil moisture availability, aeration, and rooting space, and by increasing the physical resistance for plant root growth which result in increased recruitment difficulty [133–135]. Unlike wide LFs, the surface layer of narrow LFs consists of material excavated from ditches, and a thin layer of construction material aggregates. The poor physical condition of the surface layer and low bearing capacity interfere with narrow LFs' intensity of use [6,127,130]. Hence, it is very likely that the integration of LF clearing width captured underlying differences in hydrological conditions such as water and nutrient availability, driven by compaction and construction substrate type. Additionally, due to uneven vehicular activities, different traffic intensity patterns on wide and narrow LFs likely explain variation in vegetation cover levels between forest road types.

#### 6.2.3. Clearing width and Its Relationship to Local Environmental Conditions

The advanced vegetation cover levels on narrow-width forest roads can be attributed to a combination of limited disturbance and favorable growing conditions. Our data support that a range of vegetation covers can be observed, depending on variations in incident sunlight, shade, and wind conditions. Evidence on wind and incident sunlight patterns on LFs come from Stern et al. [22], where LF openings exhibited double incident sunlight intensity and double maximum wind speed compared to the adjacent forests. The abiotic conditions were different between LFs with different clearing widths: wide LFs exhibited increased sunlight penetration that extended into the forest. Centers of wide seismic lines were characterized by >1.5 times higher sunlight intensity than those of narrow seismic lines. These results corroborate the findings in Franklin et al. [21], showing that the microclimatic conditions in the middle of LFs were generally intermediate between the interior forest and anthropogenic infrastructures, such as well pads, with narrow seismic lines more similar to the interior forest and wide seismic lines more similar to well pads. The width and orientation of LFs also influenced growth trends, as shown in Franklin et al. [21], by changing the abiotic environment: regeneration density on seismic lines increased by 5.8 times for each 10-fold increase in sunlight intensity. Our findings showed that wide forest roads experienced lower vegetation cover levels compared to narrow forest roads. Sunlight was a limiting factor and higher wind speed promoted higher levels of vegetation development. These results are not contrary to the findings in Franklin et al. [21], as their sampled wide LFs were older than the narrow LFs and therefore had more time for tree establishment and growth. Moreover, given the ranking of our factors conditioning vegetation cover, it is very likely that the clearing width moderates the changes in abiotic conditions leading to significant variations in vegetation cover levels between



forest road types. Conceptually, clearing width influence various processes: on wide LFs, greater sunlight availability could result in higher temperature and lower moisture levels (warmer and drier conditions near the ground on wide lines) [9,21]. On narrow LFs, however, significant shading from the adjacent canopy provides more favorable conditions for vegetation cover. This supports the assumption that the clearing width is a modulator of online abiotic conditions including sunlight, wind, and moisture [28,136]. Hence, research on the abiotic environment within LFs is needed to provide insight into potential explanations for abiotic–biotic associated patterns. Additionally, the floristic aspect of online communities should be considered for an integrative investigation of vegetation characteristics within LFs [27,137]. Forest roads with low NDVI levels exhibited limited vegetation cover, likely because low NDVI values indicate less or no vegetation. Contrary to van Rensen et al. [28], YPC was among the most influential factors, and it is possible that our continuous factor better accounted for the variation in vegetation cover. Steeply sloped forest roads (i.e., slopes greater than 15%) experienced advanced vegetation cover. A likely explanation for this is that steeper slopes provide favorable subsurface water exchanges and flow, which promote drier terrain conditions. This is supported by the TWI data indicating that increased water accumulation reduces vegetation cover on forest roads.

### 6.3. Characterization of Vegetation Cover Dynamic around Forest Roads

Our model predictions showed that for extended timeframes (>2 decades post-clearing), vegetation cover sustained an overall upward trend; however, slight variations occurred between wide and narrow forest roads, meeting our expectations of a more advanced cover on narrow-width forest roads. Early studies assessing vegetation cover were carried out in Latin America [138,139], South East Asia [51,134] and Central Africa [140]. They provided evidence of the increased disturbance on wide LFs, as well as variations in density, diversity, and vegetation structure across the LF surface and their proximal environments (edge and adjacent forest). These results and findings in Lee and Boutin [27], allowed us to compare our results with respect to the factors associated with vegetation growth and further confirm that disturbance legacies on wide LFs can persist for decades in boreal forests. A characterization of post-clearing vegetation growth patterns within LFs across the range of forest ecosystems is still in development and different definitions of vegetation growth have been proposed in the forest and LF literature (e.g., spectral indices [141], structure: closure through both height (regeneration), and lateral growth [142–144]). These notable limitations in previous studies and data availability over long timeframes constrained our quantitative analysis. Our ability to compare vegetation cover predictions was further constrained by the small number of studies available: many individual studies have not been conducted over the longer timeframes necessary to detect vegetation growth, or growth has not been properly defined to efficiently compare patterns across forest ecosystems, or across different forest regions in Canada [29]. A quantitative study in a Central African forest [145] demonstrated the potential for vegetation growth on abandoned LFs (logging roads) through natural processes: for an average of a 20 m clearing width, twenty-five years following abandonment, canopy closure recovered to 83% (very close to the value in the adjacent forest in their study area). In our study, wide forest roads showed an average vegetation cover of ~3%–53% and ~14%–52% for the mid- ([10–20] YPC) and long-term (>20 YPC) timeframes, respectively. Narrow forest roads showed an average cover of 17%–51% and 40%–82%, in the mid- and long-term, respectively. The differences could be attributed to forest ecosystem specifications (e.g., vegetation and soil conditions), the metric used to quantify vegetation characteristics on the roads, or road construction specifications (e.g., clearing widths). Findings in Lee and Boutin [27], for the boreal forest ecosystem showed low woody vegetation growth increments thirty-five years post-clearing: most LFs in the study (i.e., ~65% of total LFs) remained in a cleared state with a cover of low forbs, and only 8.2% of LFs across all forest types had exhibited more than 50% woody vegetation growth. LF vegetation predictions in Finnegan et al. [125], showed a 1–2 m

height growth increment 10 years post-clearing, with low lateral cover, and it was mostly disturbance-tolerant taxa. Further evidence comes from Revel et al. [146], where the growth increment of saplings was low with most saplings less than 2 m tall 10 years post-clearing. These quantitative measures for LFs highlight the importance of a unified protocol for the study of vegetation growth within LFs, which better standardize the spatiotemporal component to allow for comparisons. This would require the establishment of a coordinated long-term network of monitoring sites within the existing LF network. Moreover, the use of LiDAR data to estimate post-clearing growth patterns would be more straightforward if LFs were stratified by number of years/decades post-clearing. This would help integrate more structure into the sampling scheme and compensate for the large extent of the road network which can make the monitoring task difficult. The examination of growth patterns following fire or harvest in plot-level studies across forest ecosystems showed variable annual increments [29]. The timeframe is five years for cleared areas to attain a benchmark canopy cover of 10% post-fire, compared to 10 years to attain 10% of canopy cover post-harvest. Furthermore, Senf et al. [30], provided a direct quantification of post-clearing vegetation growth increments; the average is 84% of the disturbed areas reaching recovery benchmarks (i.e., a minimum tree cover of 40% and minimum stand height of 5 m), 30 years post-clearing. While comparisons with post-harvest and post-fire growth increments allow us to contextualize and evaluate our findings, some key differences should be noted. For example, linear (e.g., forest roads) and polygonal (e.g., cutblocks) openings differ with respect to spatial footprint, canopy clearing technique, and disturbance legacies.

#### 6.4. Research Limitations

The prediction accuracy of the *rf* approach can benefit from the inclusion of additional factors such as transport flux, compaction levels, and specifications on the construction materials. From the comparison results, ensemble approaches such as *rf* and *gbm* showed low error rates. However, additional model calibration and testing are needed to further validate these findings and evaluate the generalization capabilities of these approaches. Additionally, other techniques for factor importance and ML interpretation should also be tested. Similar to the proximity-based analysis in Abib et al., 2019 [1], both cross-validated and independently validated *rf* results satisfied the accuracy and goodness-of-fit criteria. Since repeated measurements provide additional information, it is important that dependencies in the input data are accounted for. For this purpose, stratified random sampling is used when there are strata that need to be considered in the analysis: it reproduces characteristics in the samples that are representative of the strata. Estimates generated within strata are more accurate than those from random sampling because dividing the input data into homogeneous strata often reduces sampling error and increases precision. Nonetheless, we suggest that spatial autocorrelation should be a factor of further analysis in this spatial application. Future studies could further assess model performance in the context of clustered data [147]. In general, the main disadvantages with ML approaches compared to OLS approaches are: (i) simple linear functions are highly approximated; (ii) for certain data sets, it is difficult to constrain the model by selecting the optimum parameters through cross-validation; and (iii) the output can be unstable, for example, small changes in data can produce highly divergent trees for example [119]. In this study, ML approaches, compared to OLS approaches yielded satisfactory accuracy results for the prediction of vegetation dynamics, but there are limitations concerning the generalization of the results of this study. The models were calibrated and tested with samples collected from a range of forest road sizes (i.e., clearing width) and over a bounded years post-clearing interval. Moreover, the samples were taken from three study areas which share common soil and climatic properties. This means that the predictive models could not be generalized for the prediction of the same characteristics in any unsampled location or within-forest roads with different specifications. Because large-area generalization (e.g., regional, national) depends on the variability of the training and test samples, more observations are needed. This would require a greater range of geoclimatic

conditions within forest roads as well as a higher diversity of forest road specifications. Our findings are consistent with recent LiDAR-based studies in the boreal region which have shown that the post-clearing vegetation dynamic is complex and growth increments are low. Our long-term predictions suggest that a timeframe of no less than 20 years must be expected for both wide and narrow LFs to exhibit ~50% and ~80% of vegetation cover, respectively. Future studies could compare growth patterns and evaluate whether the differences between polygonal features (resulting from fire and harvest) and LFs lead towards distinct successional trajectories [133,148]. Another consideration can emerge from this comparison and is related to the linear aspect of anthropogenic infrastructures which makes the application of chrono-sequence approaches difficult [149]. In our analysis, our plots represent points along a spatial continuum; however, the temporal component was constrained to specific data points in time. Therefore, it is important to predict post-clearing growth patterns along a temporal continuum.

## 7. Conclusions

In this study, we characterized within-forest road vegetation cover dynamics for boreal forest ecosystems using LiDAR-based CHM data and predictive modelling. Our predictive accuracy findings demonstrated that the ML approaches performed better than OLS approaches, with the *rf* model providing a better fit over that obtained with other OLS and ML models (RMSE ranging from 18.69% to 20.29% and  $R^2$  ranging from 0.69 to 0.62, using stratified cross-validation and independent datasets, respectively). The *rf* model was closely followed by *gbm*, which suggests that tree-based ensemble approaches can improve prediction accuracy. The inability of OLS approaches to handle non-linear relationships between the vegetation cover response and the causal factors is the main limitation for an accurate characterization of forest road vegetation cover dynamics. Clearing width was found to be the most important factor and was followed by years post-clearing, NDVI, shade, and climatic variables in predicting vegetation cover at a fine scale. Vegetation cover varied by forest road type, with narrow-width roads having higher mean vegetation cover predictions (~17%–51% and ~40%–82% across all five buffers extending from the road centerline, for the mid- and long-term timeframes, respectively) compared to wide roads (~3%–53% and ~14%–52% across all five buffers extending from the road centerline, for the mid- and long-term timeframes, respectively). The *rf* prediction capability, though satisfactory, requires further testing for large-area generalization. Additionally, transport flux and volumes, compaction levels, and the construction materials are among the potential factors that could be included to evaluate possible decreases in model error. With the increasing availability of remote sensing datasets, there is potential for broad-scale mapping of vegetation dynamics around forest roads (landscape or regional level). Further investigations are also required to improve the temporal resolution of vegetation measurements with LiDAR.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/f14030511/s1>, Figure S1: Field inventory plot design used to reconstruct the centerlines (Dimensions: 50 m\*clearing width). Table S1: Hyperparameters (ranges and types) and their definitions. Figure S2: (A) Summary of vegetation cover predictions (means and means +/- standard deviation error bars) grouped by different forest road categories and timeframes, from cross-validated *rf* model ( $R^2 = 0.69$ , RMSE = 18.69%) recorded within the multi-buffers around the road centerlines, across forest road types (wide roads and narrow forest roads) for the post-clearing timeframes: >20 YPC (long-term, black boxes), [10–20] YPC (mid-term, dark grey boxes), and [0–10] YPC (short-term, light grey boxes). (B) Vegetation cover mean predictions using independently-validated *rf* model ( $R^2 = 0.62$ , RMSE = 20.29%) across forest road types and post-clearing timeframes. Figure S3: *rf*-based Partial dependence plots (black curves) showing impacts of single factor on vegetation cover when all remaining factors are constant. Smooth curves are shown in blue.

**Author Contributions:** Conceptualization, N.B. and O.V.; methodology, N.B., O.V. and L.I.; software, N.B. and O.V.; validation, N.B., O.V. and L.I.; formal analysis, N.B. and O.V.; investigation, N.B., O.V. and L.I.; resources, O.V.; data curation, N.B. and O.V.; writing—original draft preparation, N.B. and O.V.; writing—review and editing, N.B., O.V. and L.I.; visualization, N.B., O.V. and L.I.; supervision, O.V. and L.I.; project administration, O.V.; funding acquisition, O.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Natural Sciences and Engineering Research Council of Canada (NSERC), grant number RDCPJ 543921-19 under the project “Subvention de recherche et développement coopérative—projet (RDCPJ)” “Analyse de la végétalisation des chemins forestiers et de leur utilisation par les prédateurs et compétiteurs du caribou des bois dans le nord du Québec”, “avec Eacom Timber Corporation, Matériaux innovants Rayonier, Ministère des Forêts, de la Faune et des Parcs” with the collaboration of Pierre Drapeau as PI for the funding.

**Data Availability Statement:** The data presented in this study are openly available in <https://depositum.uqat.ca/> (accessed on 1 January 2023).

**Acknowledgments:** We thank Maxence Martin for his valuable help in the statistical analysis and Thomas Maxime for his help during the field campaign. We thank all the individuals who have been involved in the project for their expertise and assistance in all aspects of our study and for their help in developing and reviewing the manuscript.

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

## References

1. Abib, T.H.; Chasmer, L.; Hopkinson, C.; Mahoney, C.; Rodriguez, L.C. Seismic line impacts on proximal boreal forest and wetland environments in Alberta. *Sci. Total Environ.* **2019**, *658*, 1601–1613. [\[CrossRef\]](#)
2. Filicetti, A.T. Fire and Tree Recovery on Seismic Lines. Doctoral Thesis, University of Alberta, Edmonton, AB, Canada, 2021.
3. Hornseth, M.L.; Pigeon, K.E.; MacNearney, D.; Larsen, T.A.; Stenhouse, G.; Cranston, J.; Finnegan, L. Motorized Activity on Legacy Seismic Lines: A Predictive Modeling Approach to Prioritize Restoration Efforts. *Environ. Manag.* **2018**, *62*, 595–607. [\[CrossRef\]](#)
4. Lovitt, J.; Rahman, M.M.; Saraswati, S.; McDermid, G.J.; Strack, M.; Xu, B. UAV remote sensing can reveal the effects of low-impact seismic lines on surface morphology, hydrology, and methane (CH<sub>4</sub>) release in a boreal treed bog. *J. Geophys. Res. Biogeosci.* **2018**, *123*, 1117–1129. [\[CrossRef\]](#)
5. Pigeon, K.E.; Anderson, M.; MacNearney, D.; Cranston, J.; Stenhouse, G.; Finnegan, L. Toward the Restoration of Caribou Habitat: Understanding Factors Associated with Human Motorized Use of Legacy Seismic Lines. *Environ. Manag.* **2016**, *58*, 821–832. [\[CrossRef\]](#) [\[PubMed\]](#)
6. *Guide d'Application du Règlement sur l'Aménagement Durable des Forêts du Domaine de l'état*; Ministère des Ressources Naturelles et des Forêts: Quebec, QC, Canada, 2021.
7. Vepakomma, U.; Kneeshaw, D.D.; De Grandpré, L. Influence of Natural and Anthropogenic Linear Canopy Openings on Forest Structural Patterns Investigated Using LiDAR. *Forests* **2018**, *9*, 540. [\[CrossRef\]](#)
8. Zhou, T.; Luo, X.; Hou, Y.; Xiang, Y.; Peng, S. Quantifying the effects of road width on roadside vegetation and soil conditions in forests. *Landsc. Ecol.* **2019**, *35*, 69–81. [\[CrossRef\]](#)
9. Dabros, A.; Hammond, H.J.; Pinzon, J.; Pinno, B.; Langor, D. Edge influence of low-impact seismic lines for oil exploration on upland forest vegetation in northern Alberta (Canada). *For. Ecol. Manag.* **2017**, *400*, 278–288. [\[CrossRef\]](#)
10. Kansas, J.L.; Charlebois, M.L.; Skatter, H.G. Vegetation recovery on low impact seismic lines in Alberta's oil sands and visual obstruction of wolves (*Canis lupus*) and woodland caribou (*Rangifer tarandus caribou*). *Can. Wildl. Biol. Manag.* **2015**, *4*, 137–149.
11. Forman, R.T.T. Estimate of the Area Affected Ecologically by the Road System in the United States. *Conserv. Biol.* **2000**, *14*, 31–35. [\[CrossRef\]](#)
12. Haddad, N.M.; Brudvig, L.A.; Clobert, J.; Davies, K.F.; Gonzalez, A.; Holt, R.D.; Lovejoy, T.E.; Sexton, J.O.; Austin, M.P.; Collins, C.D.; et al. Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.* **2015**, *1*, e1500052. [\[CrossRef\]](#)
13. Fisher, J.T.; Burton, A.C. Wildlife winners and losers in an oil sands landscape. *Front. Ecol. Environ.* **2018**, *16*, 323–328. [\[CrossRef\]](#)
14. Mahon, C.L.; Holloway, G.L.; Bayne, E.M.; Toms, J.D. Additive and interactive cumulative effects on boreal landbirds: Winners and losers in a multi-stressor landscape. *Ecol. Appl.* **2019**, *29*, e01895. [\[CrossRef\]](#)
15. Moreau, G.; Fortin, D.; Couturier, S.; Duchesne, T. Multi-level functional responses for wildlife conservation: The case of threatened caribou in managed boreal forests. *J. Appl. Ecol.* **2012**, *49*, 611–620. [\[CrossRef\]](#)
16. Sun, C.; Beirne, C.; Burgar, J.M.; Howey, T.; Fisher, J.T.; Burton, A.C. Simultaneous monitoring of vegetation dynamics and wildlife activity with camera traps to assess habitat change. *Remote Sens. Ecol. Conserv.* **2021**, *7*, 666–684. [\[CrossRef\]](#)
17. Barber, Q.E.; Bate, C.W.; Dabros, A.; Pinzon, J.; Nielsen, S.E.; Parisien, M.-A. Persistent impact of conventional seismic lines on boreal vegetation structure following wildfire. *Can. J. For. Res.* **2021**, *51*, 1581–1594. [\[CrossRef\]](#)



18. Toivio, J.; Helmisaari, H.-S.; Palviainen, M.; Lindeman, H.; Ala-Ilomäki, J.; Sirén, M.; Uusitalo, J. Impacts of timber forwarding on physical properties of forest soils in southern Finland. *For. Ecol. Manag.* **2017**, *405*, 22–30. [\[CrossRef\]](#)
19. Zenner, E.K.; Fauskee, J.T.; Berger, A.L.; Puettmann, K.J. Impacts of Skidding Traffic Intensity on Soil Disturbance, Soil Recovery, and Aspen Regeneration in North Central Minnesota. *North. J. Appl. For.* **2007**, *24*, 177–183. [\[CrossRef\]](#)
20. Roberts, D.; Ciuti, S.; Barber, Q.E.; Willier, C.; Nielsen, S.E. Accelerated seed dispersal along linear disturbances in the Canadian oil sands region. *Sci. Rep.* **2018**, *8*, 4828. [\[CrossRef\]](#)
21. Franklin, C.M.; Filicetti, A.T.; Nielsen, S.E. Seismic line width and orientation influence microclimatic forest edge gradients and tree regeneration. *For. Ecol. Manag.* **2021**, *492*, 119216. [\[CrossRef\]](#)
22. Stern, E.R.; Riva, F.; Nielsen, S.E. Effects of Narrow Linear Disturbances on Light and Wind Patterns in Fragmented Boreal Forests in Northeastern Alberta. *Forests* **2018**, *9*, 486. [\[CrossRef\]](#)
23. Davidson, S.J.; Goud, E.M.; Malhotra, A.; Estey, C.O.; Korsah, P.; Strack, M. Linear Disturbances Shift Boreal Peatland Plant Communities Toward Earlier Peak Greenness. *J. Geophys. Res. Biogeosci.* **2021**, *126*, e2021JG006403. [\[CrossRef\]](#)
24. Eldegard, K.; Totland, Ø.; Moe, S.R. Edge effects on plant communities along power line clearings. *J. Appl. Ecol.* **2015**, *52*, 871–880. [\[CrossRef\]](#)
25. Bourgeois, L.; Kneeshaw, D.; Boisseau, G. Les routes forestières au Québec: Les impacts environnementaux, sociaux et économiques. *Vertig. Rev. Électron. Sci. L'Environ.* **2005**, *6*. [\[CrossRef\]](#)
26. Clawges, R.; Vierling, K.; Vierling, L.; Rowell, E. The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sens. Environ.* **2008**, *112*, 2064–2073. [\[CrossRef\]](#)
27. Lee, P.; Boutin, S. Persistence and developmental transition of wide seismic lines in the western Boreal Plains of Canada. *J. Environ. Manag.* **2006**, *78*, 240–250. [\[CrossRef\]](#)
28. Van Rensen, C.K.; Nielsen, S.E.; White, B.; Vinge, T.; Lieffers, V.J. Natural regeneration of forest vegetation on legacy seismic lines in boreal habitats in Alberta's oil sands region. *Biol. Conserv.* **2015**, *184*, 127–135. [\[CrossRef\]](#)
29. Bartels, S.F.; Chen, H.Y.; Wulder, M.A.; White, J.C. Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. *For. Ecol. Manag.* **2016**, *361*, 194–207. [\[CrossRef\]](#)
30. Senf, C.; Müller, J.; Seidl, R. Post-disturbance recovery of forest cover and tree height differ with management in Central Europe. *Landsc. Ecol.* **2019**, *34*, 2837–2850. [\[CrossRef\]](#)
31. Oliver, C.D.; Larson, B.C. *Forest Stand Dynamics*, updated ed.; John Wiley and Sons: Hoboken, NJ, USA, 1996.
32. Atkins, J.W.; Bohrer, G.; Fahey, R.T.; Hardiman, B.S.; Morin, T.H.; Stovall, A.E.L.; Zimmerman, N.; Gough, C.M. Quantifying vegetation and canopy structural complexity from terrestrial LiDAR data using the forest r package. *Methods Ecol. Evol.* **2018**, *9*, 2057–2066. [\[CrossRef\]](#)
33. Pérez-Luque, A.J.; Benito, B.M.; Bonet-García, F.J.; Zamora, R. Ecological Diversity within Rear-Edge: A Case Study from Mediterranean *Quercus pyrenaica* Willd. *Forests* **2020**, *12*, 10. [\[CrossRef\]](#)
34. Wulder, M.A.; White, J.; Bater, C.W.; Coops, N.; Hopkinson, C.; Chen, G. Lidar plots—A new large-area data collection option: Context, concepts, and case study. *Can. J. Remote Sens.* **2012**, *38*, 600–618. [\[CrossRef\]](#)
35. Brubaker, K.M.; Myers, W.L.; Drohan, P.J.; Miller, D.A.; Boyer, E.W. The Use of LiDAR Terrain Data in Characterizing Surface Roughness and Microtopography. *Appl. Environ. Soil Sci.* **2013**, *2013*, 1–13. [\[CrossRef\]](#)
36. Vierling, K.T.; Vierling, L.A.; Gould, W.A.; Martinuzzi, S.; Clawges, R.M. Lidar: Shedding new light on habitat characterization and modeling. *Front. Ecol. Environ.* **2008**, *6*, 90–98. [\[CrossRef\]](#)
37. Girardin, P.; Valeria, O.; Girard, F. Measuring Spatial and Temporal Gravelled Forest Road Degradation in the Boreal Forest. *Remote Sens.* **2022**, *14*, 457. [\[CrossRef\]](#)
38. Alavi, S.J.; Najafi, A.; Alavi, S. Pavement deterioration modeling for forest roads based on logistic regression and artificial neural networks. *Croat. J. For. Eng.* **2018**, *39*, 271–287.
39. Wetz, M.A.; Ritchie, J.C.; Fox, H.D. Comparison of laser and field measurements of vegetation height and canopy cover. *Water Resour. Res.* **1994**, *30*, 1311–1319. [\[CrossRef\]](#)
40. Næsset, E. Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* **2001**, *80*, 88–99. [\[CrossRef\]](#)
41. Hart, S.A.; Chen, H. Understory Vegetation Dynamics of North American Boreal Forests. *Crit. Rev. Plant Sci.* **2006**, *25*, 381–397. [\[CrossRef\]](#)
42. Hansen, A.J.; Phillips, L.B.; Dubayah, R.; Goetz, S.; Hofton, M. Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. *For. Ecol. Manag.* **2014**, *329*, 214–226. [\[CrossRef\]](#)
43. Boucher, D.; Gauthier, S.; De Grandpré, L. Structural changes in coniferous stands along a chronosequence and a productivity gradient in the northeastern boreal forest of Québec. *Écoscience* **2006**, *13*, 172–180. [\[CrossRef\]](#)
44. Mansuy, N.; Gauthier, S.; Robitaille, A.; Bergeron, Y. Regional patterns of postfire canopy recovery in the northern boreal forest of Quebec: Interactions between surficial deposit, climate, and fire cycle. *Can. J. For. Res.* **2012**, *42*, 1328–1343. [\[CrossRef\]](#)
45. Thompson, I.; Mackey, B.; McNulty, S.; Mosseler, A. (Eds.) *Forest Resilience, Biodiversity, and Climate Change*; Montreal Technical Series no 43 1-67; Secretariat of the Convention on Biological Diversity: Montreal, QC, Canada, 2009.
46. Weiskittel, A.; Crookston, N.L.; Radtke, P. Linking climate, gross primary productivity, and site index across forests of the western United States. *Can. J. For. Res.* **2011**, *41*, 1710–1721. [\[CrossRef\]](#)
47. Ilisson, T.; Chen, H.Y.H. The direct regeneration hypothesis in northern forests. *J. Veg. Sci.* **2009**, *20*, 735–744. [\[CrossRef\]](#)

48. Swanson, M.E.; Franklin, J.F.; Beschta, R.L.; Crisafulli, C.M.; DellaSala, D.A.; Hutto, R.L.; Lindenmayer, D.B.; Swanson, F.J. The forgotten stage of forest succession: Early-successional ecosystems on forest sites. *Front. Ecol. Environ.* **2011**, *9*, 117–125. [\[CrossRef\]](#)
49. Filicetti, A.T.; Nielsen, S.E. Tree regeneration on industrial linear disturbances in treed peatlands is hastened by wildfire and delayed by loss of microtopography. *Can. J. For. Res.* **2020**, *50*, 936–945. [\[CrossRef\]](#)
50. St-Pierre, F.; Drapeau, P.; St-Laurent, M.-H. Drivers of vegetation regrowth on logging roads in the boreal forest: Implications for restoration of woodland caribou habitat. *For. Ecol. Manag.* **2020**, *482*, 118846. [\[CrossRef\]](#)
51. Zang, R.; Ding, Y. Forest recovery on abandoned logging roads in a tropical montane rain forest of Hainan Island, China. *Acta Oecologica* **2009**, *35*, 462–470. [\[CrossRef\]](#)
52. Koma, Z.; Seijmonsbergen, A.C.; Kissling, W.D. Classifying wetland-related land cover types and habitats using fine-scale lidar metrics derived from country-wide Airborne Laser Scanning. *Remote Sens. Ecol. Conserv.* **2020**, *7*, 80–96. [\[CrossRef\]](#)
53. Martinuzzi, S.; Vierling, L.A.; Gould, W.A.; Falkowski, M.J.; Evans, J.S.; Hudak, A.T.; Vierling, K.T. Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. *Remote Sens. Environ.* **2009**, *113*, 2533–2546. [\[CrossRef\]](#)
54. Chirici, G.; Mura, M.; McInerney, D.; Py, N.; Tomppo, E.O.; Waser, L.T.; Travaglini, D.; McRoberts, R.E. 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.* **2016**, *176*, 282–294. [\[CrossRef\]](#)
55. Cosenza, D.N.; Korhonen, L.; Maltamo, M.; Packalen, P.; Strunk, J.L.; Næsset, E.; Gobakken, T.; Soares, P.; Tomé, M. Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. *For. Int. J. For. Res.* **2020**, *94*, 311–323. [\[CrossRef\]](#)
56. Finley, A.O.; McRoberts, R.E. Efficient k-nearest neighbor searches for multi-source forest attribute mapping. *Remote Sens. Environ.* **2008**, *112*, 2203–2211. [\[CrossRef\]](#)
57. Franco-Lopez, H.; Ek, A.R.; Bauer, M.E. Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. *Remote Sens. Environ.* **2001**, *77*, 251–274. [\[CrossRef\]](#)
58. McRoberts, R.E. Estimating forest attribute parameters for small areas using nearest neighbors techniques. *For. Ecol. Manag.* **2012**, *272*, 3–12. [\[CrossRef\]](#)
59. Leathwick, J.; Elith, J.; Hastie, T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecol. Model.* **2006**, *199*, 188–196. [\[CrossRef\]](#)
60. Moisen, G.G.; Frescino, T.S. Comparing five modelling techniques for predicting forest characteristics. *Ecol. Model.* **2002**, *157*, 209–225. [\[CrossRef\]](#)
61. Yang, L.; Liang, S.; Zhang, Y. A New Method for Generating a Global Forest Aboveground Biomass Map From Multiple High-Level Satellite Products and Ancillary Information. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2587–2597. [\[CrossRef\]](#)
62. Abdi, A.M. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GISci. Remote Sens.* **2019**, *57*, 1–20. [\[CrossRef\]](#)
63. Schönauer, M. Prediction of Forest Soil Trafficability by Topography-Based Algorithms and In-Situ Test Procedures. Doctoral Dissertation, Georg-August-Universität Göttingen, Göttingen, Germany, 2022.
64. Zhang, Y.; Ma, J.; Liang, S.; Li, X.; Li, M. An Evaluation of Eight Machine Learning Regression Algorithms for Forest Aboveground Biomass Estimation from Multiple Satellite Data Products. *Remote Sens.* **2020**, *12*, 4015. [\[CrossRef\]](#)
65. Matasci, G.; Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W.; Zald, H.S.J. Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. *Remote Sens. Environ.* **2018**, *209*, 90–106. [\[CrossRef\]](#)
66. Ahmed, O.S.; Franklin, S.E.; Wulder, M.A.; White, J.C. Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. *ISPRS J. Photogramm. Remote Sens.* **2015**, *101*, 89–101. [\[CrossRef\]](#)
67. Venier, L.A.; Swystun, T.; Mazerolle, M.J.; Kreutzweiser, D.P.; Wainio-Keizer, K.L.; McIlwrick, K.A.; Woods, M.E.; Wang, X. Modelling vegetation understory cover using LiDAR metrics. *PLoS ONE* **2019**, *14*, e0220096. [\[CrossRef\]](#)
68. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [\[CrossRef\]](#)
69. Boehmke, B.; Greenwell, B. *Hands on Machine Learning with R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2019.
70. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [\[CrossRef\]](#)
71. Forkuor, G.; Hounkpatin, O.K.L.; Welp, G.; Thiel, M. High Resolution Mapping of Soil Properties Using Remote Sensing Variables in South-Western Burkina Faso: A Comparison of Machine Learning and Multiple Linear Regression Models. *PLoS ONE* **2017**, *12*, e0170478. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Martin, M.; Orton, T.; Lacarce, E.; Meersmans, J.; Saby, N.; Paroissien, J.; Jolivet, C.; Boulonne, L.; Arrouays, D. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. *Geoderma* **2014**, *223–225*, 97–107. [\[CrossRef\]](#)
73. Schratz, P.; Muenchow, J.; Iturrutxa, E.; Richter, J.; Brenning, A. Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. *Ecol. Model.* **2019**, *406*, 109–120. [\[CrossRef\]](#)
74. Robitaille, A.; Saucier, J. *Paysages Régionaux du Québec Méridional*; Direction de la gestion des stocks forestiers et Direction des Relations Publiques; Ministère des Ressources Naturelles du Québec: Quebec, QC, Canada, 1998.

75. Rossi, S.; Cairo, E.; Krause, C.; DesLauriers, A. Growth and basic wood properties of black spruce along an alti-latitudinal gradient in Quebec, Canada. *Ann. For. Sci.* **2014**, *72*, 77–87. [\[CrossRef\]](#)
76. Ministère de l'Environnement et de la Lutte Contre les Changements Climatiques. *Les Provinces Naturelles. Niveau I du Cadre Écologique de Référence du Québec (Natural Province. Level I of the Quebec Reference Ecological Framework)*; Ministère de l'Environnement et de la Lutte Contre les Changements Climatiques: Québec, QC, Canada, 1999.
77. Blouin, J.; Berger, J. *Guide de Reconnaissance des Types Écologiques de la Région Écologique 5b Coteaux du Réservoir Gouin*; Ministère des Ressources Naturelles: Québec, QC, Canada, 2001.
78. Gosselin, J.; Berger, J.-P. *Guide de Reconnaissance des Types Écologiques: Région Écologique 4b: Coteaux du Réservoir Cabonga: Région Écologique 4c: Collines du Moyen-Saint-Maurice*; Ministère des Ressources Naturelles: Québec, QC, Canada, 2002.
79. Roussel, J.-R.; Auty, D.; Coops, N.C.; Tompalski, P.; Goodbody, T.R.; Meador, A.S.; Bourdon, J.-F.; de Boissieu, F.; Achim, A. lidR: An R package for analysis of Airborne Laser Scanning (ALS) data. *Remote Sens. Environ.* **2020**, *251*, 112061. [\[CrossRef\]](#)
80. Ministère des Ressources naturelles et des Forêts. *Guide d'Utilisation des Produits Dérivés du LiDAR*; Ministère des Ressources naturelles et des Forêts: Québec, QC, Canada, 2020.
81. Ministère des Ressources naturelles et des Forêts. *Métadonnées des Acquisitions LiDAR*; Ministère des Ressources naturelles et des Forêts: Québec, QC, Canada, 2022.
82. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020.
83. Kuhn, M. Caret: Classification and Regression Training; Astrophysics Source Code Library. Available online: <https://ascl.net/1505.003> (accessed on 1 May 2021).
84. Kopecký, M.; Macek, M.; Wild, J. Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. *Sci. Total. Environ.* **2020**, *757*, 143785. [\[CrossRef\]](#)
85. Hong, T.; Lee, M.; Koo, C.; Jeong, K.; Kim, J. Development of a method for estimating the rooftop solar photovoltaic (PV) potential by analyzing the available rooftop area using Hillshade analysis. *Appl. Energy* **2017**, *194*, 320–332. [\[CrossRef\]](#)
86. Piedallu, C.; Gégout, J.-C. Multiscale computation of solar radiation for predictive vegetation modelling. *Ann. For. Sci.* **2007**, *64*, 899–909. [\[CrossRef\]](#)
87. Carlson, T.N.; Ripley, D.A. On the relation between NDVI, fractional vegetation cover, and leaf area index. *Remote Sens. Environ.* **1997**, *62*, 241–252. [\[CrossRef\]](#)
88. Tarpley, J.D.; Schneider, S.R.; Money, R.L. Global Vegetation Indices from the NOAA-7 Meteorological Satellite. *J. Clim. Appl. Meteorol.* **1984**, *23*, 491–494. [\[CrossRef\]](#)
89. Poggio, L.; Simonetti, E.; Gimona, A. Enhancing the WorldClim data set for national and regional applications. *Sci. Total. Environ.* **2018**, *625*, 1628–1643. [\[CrossRef\]](#)
90. WorldClim. WorldClim Version 2. 2017. Available online: <http://www.worldclim.com/version2> (accessed on 1 August 2020).
91. Pradhan, P.; Setyawan, A.D. Filtering multi-collinear predictor variables from multi-resolution rasters of WorldClim 2.1 for Ecological Niche Modeling in Indonesian context. *Asian J. For.* **2021**, *5*.
92. Zuur, A.F.; Ieno, E.N.; Elphick, C.S. A protocol for data exploration to avoid common statistical problems. *Methods Ecol. Evol.* **2010**, *1*, 3–14. [\[CrossRef\]](#)
93. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer: Berlin/Heidelberg, Germany, 2001.
94. Elith, J.; Leathwick, J.R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813. [\[CrossRef\]](#)
95. Burnham, K.P.; Anderson, D.R. *A Practical Information-Theoretic Approach. Model Selection and multimodel Inference*; Springer: Berlin/Heidelberg, Germany, 2002.
96. Barton, K. MuMIn: Multi Model Inference. 2009. Available online: <https://cran.r-project.org/web/packages/MuMIn/MuMIn.pdf> (accessed on 1 August 2020).
97. Vu, D.T.; Tran, X.-L.; Cao, M.-T.; Tran, T.C.; Hoang, N.-D. Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline. *Measurement* **2020**, *164*, 108066. [\[CrossRef\]](#)
98. Lay, U.S.; Pradhan, B.; Bin Yusoff, Z.; Bin Abdallah, A.F.; Aryal, J.; Park, H.-J. Data Mining and Statistical Approaches in Debris-Flow Susceptibility Modelling Using Airborne LiDAR Data. *Sensors* **2019**, *19*, 3451. [\[CrossRef\]](#)
99. Crookston, N.L.; Finley, A.O. yaImpute: An R Package for kNN Imputation. *J. Stat. Softw.* **2008**, *23*. [\[CrossRef\]](#)
100. Bruce, P.; Bruce, A.; Gedeck, P. *Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python*; O'Reilly Media: Sebastopol, CA, USA, 2020.
101. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
102. Friedman, J.H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **2002**, *38*, 367–378. [\[CrossRef\]](#)
103. Natekin, A.; Knoll, A. Gradient boosting machines, a tutorial. *Front. Neurobot.* **2013**, *7*, 21. [\[CrossRef\]](#)
104. Martínez-Santos, P.; Renard, P. Mapping Groundwater Potential Through an Ensemble of Big Data Methods. *Groundwater* **2019**, *58*, 583–597. [\[CrossRef\]](#)
105. Liu, D. *Work in Process Decision Support System with Predictive Modeling in the Food Manufacturing Industry*; The George Washington University: Washington, DC, USA, 2020.
106. Zhang, Y.; Haghani, A. A gradient boosting method to improve travel time prediction. *Transp. Res. Part C: Emerg. Technol.* **2015**, *58*, 308–324. [\[CrossRef\]](#)



107. Meyer, H.; Pebesma, E. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol. Evol.* **2021**, *12*, 1620–1633. [\[CrossRef\]](#)
108. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **2018**, *101*, 1–9. [\[CrossRef\]](#)
109. Pohjankukka, J.; Pahikkala, T.; Nevalainen, P.; Heikkonen, J. Estimating the prediction performance of spatial models via spatial k-fold cross validation. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2001–2019. [\[CrossRef\]](#)
110. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [\[CrossRef\]](#)
111. Geib, C.; Pelizari, P.A.; Schrade, H.; Brenning, A.; Taubenbock, H. On the Effect of Spatially Non-Disjoint Training and Test Samples on Estimated Model Generalization Capabilities in Supervised Classification With Spatial Features. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2008–2012. [\[CrossRef\]](#)
112. Garbajevski, O.M.; Schmiedt, J.E.; Verma, T.; Lefter, I.; Altes, W.K.K.; Droin, A.; Schiricke, B.; Wurm, M. Spatial factors influencing building age prediction and implications for urban residential energy modelling. *Comput. Environ. Urban Syst.* **2021**, *88*, 101637. [\[CrossRef\]](#)
113. Kosicki, J.Z. Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. *Environ. Ecol. Stat.* **2020**, *27*, 273–292. [\[CrossRef\]](#)
114. Liaw, A.; Wiener, M. The randomforest package. *R News* **2002**, *2*, 18–22.
115. Molnar, C. Interpretable Machine Learning. 2020. Available online: <https://christophm.github.io/interpretable-ml-book/> (accessed on 1 December 2022).
116. Greenwell, B.M. pdp: An R package for constructing partial dependence plots. *R J.* **2017**, *9*, 421. [\[CrossRef\]](#)
117. Hassan, M.A.; Khalil, A.; Kaseb, S.; Kassem, M. Exploring the potential of tree-based ensemble methods in solar radiation modeling. *Appl. Energy* **2017**, *203*, 897–916. [\[CrossRef\]](#)
118. Morellos, A.; Pantazi, X.-E.; Moshou, D.; Alexandridis, T.; Whetton, R.; Tziotziou, G.; Wiebensohn, J.; Bill, R.; Mouazen, A.M. Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. *Biosyst. Eng.* **2016**, *152*, 104–116. [\[CrossRef\]](#)
119. Prasad, A.M.; Iverson, L.R.; Liaw, A. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* **2006**, *9*, 181–199. [\[CrossRef\]](#)
120. Stojanova, D.; Panov, P.; Gjorgjioski, V.; Kobler, A.; Džeroski, S. Estimating vegetation height and canopy cover from remotely sensed data with machine learning. *Ecol. Informatics* **2010**, *5*, 256–266. [\[CrossRef\]](#)
121. Ahmad, M.W.; Mourshed, M.; Rezgui, Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* **2018**, *164*, 465–474. [\[CrossRef\]](#)
122. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of machine-learning classification in remote sensing: An applied review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [\[CrossRef\]](#)
123. Hultquist, C.; Chen, G.; Zhao, K. A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. *Remote Sens. Lett.* **2014**, *5*, 723–732. [\[CrossRef\]](#)
124. Nawar, S.; Mouazen, A.M. Comparison between Random Forests, Artificial Neural Networks and Gradient Boosted Machines Methods of On-Line Vis-NIR Spectroscopy Measurements of Soil Total Nitrogen and Total Carbon. *Sensors* **2017**, *17*, 2428. [\[CrossRef\]](#)
125. Finnegan, L.; Pigeon, K.E.; MacNearney, D. Predicting patterns of vegetation recovery on seismic lines: Informing restoration based on understory species composition and growth. *For. Ecol. Manag.* **2019**, *446*, 175–192. [\[CrossRef\]](#)
126. O'Mahony, M.; Ueberschaer, A.; Owende, P.; Ward, S. Bearing capacity of forest access roads built on peat soils. *J. Terramechanics* **2000**, *37*, 127–138. [\[CrossRef\]](#)
127. Kaakkurivaara, T.; Vuorimies, N.; Kolisoja, P.; Uusitalo, J. Applicability of portable tools in assessing the bearing capacity of forest roads. *Silva Fenn.* **2015**, *49*. [\[CrossRef\]](#)
128. Ministère des Ressources naturelles et des Forêts. *Guide d'Application du Règlement sur l'Aménagement Durable des Forêts du Domaine de l'état. Annexe 4—Caractéristiques des Chemins selon leur Classement*; Ministère des Ressources naturelles et des Forêts: Quebec, QC, Canada, 2021.
129. Rummer, B.; Wear, D. *Forest Operations Technology. Southern Forest Resource Assessment General Technical Report SRS-53*; USDA-Forest Service, Southern Research Station: Asheville, NC, USA, 2002; pp. 341–353.
130. Waga, K. Unpaved forest road quality assessment using airborne LiDAR data. *Diss. For.* **2021**, 2021. [\[CrossRef\]](#)
131. Gartzia-Bengoetxea, N.; de Arano, I.M.; Arias-González, A. Forest productivity and associated soil ecosystem services remain altered 15 years after mechanized site preparation for reforestation with *Pinus radiata*. *Soil Tillage Res.* **2021**, *213*, 105150. [\[CrossRef\]](#)
132. Cambi, M.; Certini, G.; Neri, F.; Marchi, E. The impact of heavy traffic on forest soils: A review. *For. Ecol. Manag.* **2015**, *338*, 124–138. [\[CrossRef\]](#)
133. Dabros, A.; Pyper, M.; Castilla, G. Seismic lines in the boreal and arctic ecosystems of North America: Environmental impacts, challenges, and opportunities. *Environ. Rev.* **2018**, *26*, 214–229. [\[CrossRef\]](#)
134. Pinard, M.; Barker, M.; Tay, J. Soil disturbance and post-logging forest recovery on bulldozer paths in Sabah, Malaysia. *For. Ecol. Manag.* **2000**, *130*, 213–225. [\[CrossRef\]](#)

135. Startsev, A.D.; McNabb, D.H. Effects of compaction on aeration and morphology of boreal forest soils in Alberta, Canada. *Can. J. Soil Sci.* **2009**, *89*, 45–56. [[CrossRef](#)]
136. Filicetti, A.T.; Nielsen, S.E. Fire and forest recovery on seismic lines in sandy upland jack pine (*Pinus banksiana*) forests. *For. Ecol. Manag.* **2018**, *421*, 32–39. [[CrossRef](#)]
137. Lázaro-Lobo, A.; Ervin, G.N. A global examination on the differential impacts of roadsides on native vs. exotic and weedy plant species. *Glob. Ecol. Conserv.* **2019**, *17*, e00555. [[CrossRef](#)]
138. Guariguata, M.R.; Dupuy, J.M. Forest Regeneration in Abandoned Logging Roads in Lowland Costa Rica. *Biotropica* **1997**, *29*, 15–28. [[CrossRef](#)]
139. Olander, L.P.; Scatena, F.; Silver, W.L. Impacts of disturbance initiated by road construction in a subtropical cloud forest in the Luquillo Experimental Forest, Puerto Rico. *For. Ecol. Manag.* **1998**, *109*, 33–49. [[CrossRef](#)]
140. Malcolm, J.R.; Ray, J.C. Influence of Timber Extraction Routes on Central African Small-Mammal Communities, Forest Structure, and Tree Diversity. *Conserv. Biol.* **2000**, *14*, 1623–1638. [[CrossRef](#)]
141. White, J.C.; Saarinen, N.; Kankare, V.; Wulder, M.A.; Hermosilla, T.; Coops, N.C.; Pickell, P.D.; Holopainen, M.; Hyypä, J.; Vastaranta, M. Confirmation of post-harvest spectral recovery from Landsat time series using measures of forest cover and height derived from airborne laser scanning data. *Remote Sens. Environ.* **2018**, *216*, 262–275. [[CrossRef](#)]
142. Finnegan, L.; Pigeon, K.E.; Cranston, J.; Hebblewhite, M.; Musiani, M.; Neufeld, L.; Schmiegelow, F.; Duval, J.; Stenhouse, G.B. Natural regeneration on seismic lines influences movement behaviour of wolves and grizzly bears. *PLoS ONE* **2018**, *13*, e0195480. [[CrossRef](#)]
143. Matasci, G.; Hermosilla, T.; Wulder, M.A.; White, J.C.; Coops, N.C.; Hobart, G.W.; Bolton, D.K.; Tompalski, P.; Bater, C.W. Three decades of forest structural dynamics over Canada’s forested ecosystems using Landsat time-series and lidar plots. *Remote Sens. Environ.* **2018**, *216*, 697–714. [[CrossRef](#)]
144. Vepakomma, U.; St-Onge, B.; Kneeshaw, D. Response of a boreal forest to canopy opening: Assessing vertical and lateral tree growth with multi-temporal lidar data. *Ecol. Appl.* **2011**, *21*, 99–121. [[CrossRef](#)]
145. Kleinschroth, F.; Healey, J.R.; Sist, P.; Mortier, F.; Gourlet-Fleury, S. How persistent are the impacts of logging roads on Central African forest vegetation? *J. Appl. Ecol.* **2016**, *53*, 1127–1137. [[CrossRef](#)]
146. Revel, R.D.; Dougherty, T.D.; Downing, D.J. *Forest Growth and Revegetation along Seismic Lines*; University of Calgary Press: Calgary, AB, USA, 1984.
147. Hajjem, A.; Bellavance, F.; Larocque, D. Mixed-effects random forest for clustered data. *J. Stat. Comput. Simul.* **2012**, *84*, 1313–1328. [[CrossRef](#)]
148. Finnegan, L.; MacNearney, D.; Pigeon, K.E. Divergent patterns of understory forage growth after seismic line exploration: Implications for caribou habitat restoration. *For. Ecol. Manag.* **2018**, *409*, 634–652. [[CrossRef](#)]
149. Norden, N.; Chazdon, R.L.; Chao, A.; Jiang, Y.-H.; Vilchez-Alvarado, B. Resilience of tropical rain forests: Tree community reassembly in secondary forests. *Ecol. Lett.* **2009**, *12*, 385–394. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.