UNIVERSITÉ DU QUÉBEC EN ABITIBI-TÉMISCAMINGUE

CARACTÉRISATION DE LA DYNAMIQUE DE VÉGÉTATION DES STRUCTURES LINÉAIRES À L'AIDE DU SYSTÈME DE BALAYAGE LASER AÉROPORTÉ

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN ÉCOLOGIE

PAR

NARIMENE BRAHAM

DÉCEMBRE 2022

UNIVERSITÉ DU QUÉBEC EN ABITIBI-TÉMISCAMINGUE

CHARACTERIZATION OF VEGETATION DYNAMICS ON LINEAR FEATURES USING AIRBORNE LASER SCANNING SYSTEM

DISSERTATION SUBMITTED
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF ECOLOGY

BY

NARIMENE BRAHAM

DECEMBER 2022

ACKNOWLEDGMENTS

The completion of this dissertation would have not been possible without the support and patience of a few individuals. I have learned such an incredible amount from the time spent in the forest research institute.

I would like to thank my supervisor and my co-supervisor for all their time and patience, and for providing valuable guidance, ideas, and support. Thank you to Osvaldo, for helping me kick this journey off with the right support, for mentoring me throughout this research process, and for the insights and humor that have made my experience in the program valuable and enjoyable.

Being in the GIS lab, I had the chance of meeting and working with exceptional students who had always been so engaged and passionate about several topics we share mutual interest in. I really enjoyed the conversations on research and their willingness to help. I would like to thank all members of the lab whose help in has made the last two years easier. The members of the lab have been great desk neighbors and I appreciate their technical help and support. Thank you to all members of the lab for taking the time to carefully examine dissertation drafts. To Maxence and Carlos, thank you for the insights and for being such valuable assets for graduate students. Thank you for the support throughout my time in the program and for giving this dissertation a good first read, the more than helpful recommended edits on this document, had undoubtedly made this dissertation a better version.

I would also like to thank my family for being so supportive throughout this project.

I thank all individuals outside of the conventional real-live exchange dynamic, the collective entity that has helped expand my repertoire of both statistical and scientific methods.

FOREWORD

This master dissertation is composed of two chapters. Chapter I serves as a general introduction to the study, presenting the background, state of knowledge, objectives, and hypotheses. Chapter II presents the study in the form of a research paper, including the problem, methodology, and results, as well as a discussion of the study's findings, and limitations. The dissertation concludes with an extended conclusion.

TABLE OF CONTENTS

LIST OF FIGURES AND ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

CHM	Canopy Height Model
DEM	Digital Elevation Model
DTM	Digital Terrain Model
gam	Generalized additive model
gbm	Gradient boosting machines
knn	k-nearest-neighbors
LFs	Linear Features
LiDAR	*Light Detection and Ranging*
MAE	Mean Absolute Error
mars	Multivariate Adaptive Regression Splines
ML	Machine learning
mlr	Multiple Linear Regression
NDVI	Normalized Difference Vegetation Index
OLS	Ordinary Least Squares
PDPs	Partial Dependence Plots
R^2	Coefficient of determination
rf	Random Forests
RMSE	Root Mean Square Error
TWI	Topographic Wetness Index
YPC	Years post-clearing

RÉSUMÉ

Le réseau de structures linéaires est constitué de routes, de sentiers, de pipelines et de lignes sismiques aménagés dans une grande partie de la forêt boréale commerciale. Ces structures linéaires, fournissent un accès pour les opérations industrielles, récréatives, sylvicoles et de gestion des incendies, mais ont également des incidences économiques et environnementales qui impliquent à la fois les parties actives et non actives du réseau (e.g., les coûts d'entretien, érosion de la biodiversité, dégradation de l'habitat de la faune qui dépend de la forêt). Par conséquent, pour prévenir tous ces inconvénients, il est nécessaire de comprendre la dynamique des caractéristiques de végétation des structures linéaires et particulièrement des chemins forestiers.

Les données de télédétection et la modélisation prédictive sont des outils utiles en fournissant des informations quantitatives précises et détaillées visant l'évaluation de l'état des structures linéaires (par exemple, la détérioration de la surface ou la dynamique des caractéristiques de végétation), et ce à plusieurs niveaux (paysage, région). Cependant, le potentiel des données de télédétection pour améliorer notre connaissance des caractéristiques de végétation à fine échelle sur les chemins forestiers n'a pas été entièrement exploré.

Cette étude a examiné l'utilisation de données LiDAR aéroporté à haute résolution spatiale (1 m), de données climatiques et de terrain dans le but de fournir une meilleure compréhension de la dynamique de végétation des chemins forestiers: i) en développant un modèle prédictif pour l'estimation de la couverture végétale dérivée du modèle de hauteur de canopée (métrique de réponse), ii) en examinant les facteurs ayant un effet sur la couverture végétale en utilisant les mesures LiDAR (topographie: pente, TWI, ombrage et orientation), de l'imagerie optique Sentinel-2 (NDVI), des bases de données climatiques (ensoleillement et vitesse du vent) et de l'inventaire de terrain (largeur de l'ouverture du chemin et le temps depuis la construction ou entretien majeur). Nous avons évalué et comparé les performances des approches de régression par la méthode des moindres carrés et par apprentissage automatique couramment utilisées en modélisation écologique – régression linéaire multiple (*mlr*), méthode des splines de régression adaptative multivariée (*mars*), modèle additif généralisé (*gam*), méthode du plus proche voisin (*knn*), méthode d'arbres de régression boostés (*gbm*) et la méthode des forêts aléatoires (*rf*) –. Nous avons validé les résultats de nos approches en utilisant une métrique d'erreur – erreur quadratique moyenne (RMSE) – et une métrique de qualité d'ajustement – coefficient de détermination (R^2) –. Les prédictions ont été testées par validation croisée et validées par rapport à un jeu de données indépendant. Nos résultats ont révélé que le modèle *rf* a montré les résultats les plus précis (validation croisée: R^2=0.69, RMSE=18.69%, validation par un jeu de données indépendant: R^2=0.62, RMSE=20.29%) et que les facteurs les plus informatifs étaient la largeur de l'ouverture du chemin qui avait l'effet négatif le plus marqué sur la couverture végétale, suggérant l'influence sous-jacente des perturbations antérieures, et le temps écoulé depuis la construction ou dernier entretien du chemin, qui avaient un effet positif sur l'augmentation de la couverture végétale. Les prédictions à long terme suggèrent qu'il faudra au moins 20 ans pour que les routes larges et étroites présentent respectivement ~50% et ~80% de couverture végétale.

Cette étude a permis d'améliorer notre compréhension de la dynamique de végétation des chemins forestiers à fine échelle, tant sur le plan qualitatif que quantitatif. Les informations issues du modèle prédictif sont utiles pour la gestion à court et à long terme du réseau existant. De plus, le présent mémoire démontre que les modèles spatialement explicites utilisant des données LiDAR sont des outils fiables pour évaluer la dynamique de végétation des chemins forestiers et fournit des pistes pour de futures recherches et la possibilité d'intégrer cette approche quantitative à d'autres études de structures linéaires. Une meilleure connaissance des modèles de dynamique des caractéristiques de végétation sur les chemins forestiers, sur de grandes zones géographiques, peut aider à soutenir la gestion durable des forêts par la modulation de l'impact environnemental associé à l'empreinte linéaire.

Mots clés : Chemins forestiers, LiDAR (Light Detection and Ranging) aéroporté, strctures linéaires, chemins forestiers, réseau routier, aménagement forestier, forêts aléatoires, dynamique de végétation, forêt boréale.

ABSTRACT

Linear features network are roads, trails, pipelines, and seismic lines developed throughout much of the commercial boreal forest. 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 features network across the boreal forest would lead to the optimization of maintenance and construction costs as well as the minimisation of the cumulative environmental effects of the anthropogenic linear footprint. Remote sensing data and predictive modelling are valuable support tools for multi-level management of this network by providing accurate and detailed quantitative information aiming to assess linear features conditions (e.g., deterioration and vegetation characteristics dynamic). However, the potential of remote sensing datasets to improve knowledge of fine-scale vegetation characteristics dynamic within forest roads have 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 characteristics dynamic within forest roads by i) developing a predictive model for the characterization of LiDAR-CHM vegetation cover dynamic (response metric), ii) investigating causal factors driving 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 by stratified cross-validation and 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 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% of vegetation cover, respectively.

This study has improved our understanding of fine-scale vegetation dynamic within forest roads, both qualitatively and quantitatively. The information from the predictive model is useful for both short- and long-term management of the existing network. Further, the study demonstrates that spatially-explicit models using LiDAR data are reliable tools for assessing vegetation dynamics within forest roads. It provides avenues for further research and the potential of integrating this quantitative approach with other linear feature studies. Improved knowledge of vegetation dynamic patterns on linear features can help support sustainable forest management.

Key words: Forest roads, airborne Light Detection and Ranging LiDAR, linear features, forest roads, road network, forest management, Random Forests, vegetation dynamic, boreal forest.

CHAPTER I

INTRODUCTION

1.1 Forest linear features and their effects

In forested areas throughout the world, industrial activities (e.g., logging, mining, and oil and gas exploration) are highly dependent on supply and provisioning services. When such industrial activities are mapped, they simultaneously produce polygonal features resulting from resource extraction (e.g., cutblocks, open pit mines, well sites), as well as linear features (LFs) resulting from a dense network of roads and seismic lines used for exploration and transportation of the resource (e.g., roads, pipelines, seismic lines, and communication infrastructure) (Bourgeois et al., 2005; Dabros et al., 2018; Laurance et al., 2009); Musetta-Lambert et al. (2019). In the boreal forest ecosystem of Canada, LFs have become widespread landscape elements (Pasher et al., 2013; Pattison et al., 2016). For instance in Alberta, the linear polygonal footprint of seismic lines exceeds 1.5 million km with a density greater than 10 linear km per km^2 (Abib, 2018), and often reach densities of 40 km/km^2 (Filicetti et al., 2019). British Columbia has an estimated total of 400,000–800 000 km of unpaved forest road network, 74% of them are connected with the forest industry. An estimated 4000–5000 km of forest roads is established annually in Quebec, with a forest road network covering 600 000 km (St-Pierre et al., 2021; Vepakomma et al., 2018; Waga et al., 2020). In addition to economic and industrial activities, considerations such as cost effectiveness, environmental impact of the linear footprint, and climate change, have created an increased need for quality control and assessment of the conditions on the existing LFs network. Particularly, dense LFs networks can result in the degradation of ecosystem functioning (Bennett, 2017) and often broad fragmentation of the landscape (Eldegard et al., 2015; Harper et al., 2005; Harper et al., 2015; Pattison et al., 2016) which ultimately affect vegetation structural diversity (Guo et al., 2017), and biodiversity (Fahrig, 2003; Moudrý et al., 2021; Venier et al., 2014; Whitehurst, 2014). In Canada, there is particular concern about adverse effects on forest-dwelling species, such as woodland caribou (Rangifer Tarandus), for which interior habitat conditions are minimized in managed forests compared to roadless parts of the landscape (Lapointe, 2019; Ray, 2014; St-Laurent et al., 2014). Furthermore, LFs have no natural linear opening homologues due to compaction and continuous motorized use (Dabros et al., 2022). The minimum time required for LFs vegetation to transition to a state comparable to their adjacent forests is yet to be established (Abib et al., 2019; Finnegan et al., 2018a). The possible causes for the slow transition are complex and are related to permanent vehicle traffic resulting in mechanical stress as well as disturbance legacies which include the construction of LFs using machinery that flatten terrain relief, simplify microtopography (loss of microsites), and depress and compact the soil (Filicetti and Nielsen, 2020; Stevenson et al., 2019). Both the literature on the densification and the long-lasting effect of disturbance legacies on LFs provided important context for the current study on LFs vegetation characteristics dynamic: i) Terrain conditions and the effects of compaction: traffic by heavy machines and vehicles has significant impacts on LFs soil properties and functions resulting in alterations of soil physical properties such as increased bulk density, decreased porosity which affect soil moisture content, water infiltration, exchange and aeration. Changes in levels of available resources within compacted LFs results in decreased water and nutrients absorption by vegetation (Toivio et al., 2017) which creates a response towards changes in vegetation growth patterns. ii) Abiotic conditions: LFs openings locally affect sunlight and evapotranspiration rates. These underlying environmental factors affect growth increments, with different vegetation species traits suited for various tolerance ranges (e.g., for soil moisture, temperature, and incident sunlight levels). Moreover, variations in environmental factors vary with LFs attributes (e.g., orientation). For example, van Rensen et al. (2015) reported that the shade of east-west lines might reduce competition from shade-intolerant shrubs and herbs, thereby promoting higher growth levels. iii) Vegetation structure and composition: even on narrow LFs, vegetation growth can still be limited (Dabros et al., 2018) which would further increase evapotranspiration rates, leading to drier surface conditions (Abib et al., 2019; Groot et al., 1997). Moreover, LFs typically cause changes in the local microclimate by creating discontinuities in the canopy, resulting in heterogeneous conditions allowing species with different requirements to establish. Higher sunlight levels and changes in site conditions result in changes in vegetation species composition, diversity, and richness. Particularly, shade-intolerant species are considered "pioneer' species" because of their ability to colonize disturbed sites following the removal of the canopy leading to the proliferation of disturbance-tolerant communities richer in juveniles and low-stature trees (Finnegan et al., 2018a; Rioux, 2018). iv) Clearing width is an informative factor: LFs with different sizes would produce different characteristics of environmental and biological effects. For example, in Zhou et al. (2020), narrow LFs had no measurable effects on the biological environmental conditions of the adjacent forest, and the differences between wide and narrow LFs were attributed to clearing width which integrates multiple factors (frequency of vehicle use, traffic volumes, construction materials).

1.2 Linear features morphological characteristics, classification criteria and disturbance legacies

The functional criterion used in provincial classifications of forest roads distinguishes between two broad categories which differ with regards to their bearing capacity and compaction levels. High-grade roads (primary and secondary) are wide forest roads designed for permanent all season use, where coarse aggregates, gravel, and sand are commonly used to enhance the bearing capacity. Low-grade roads have reduced width and provide access to timber harvesting sites and include temporary roads which are useful during winter season (Girardin et al., 2022; Government of Alberta, 2016; Ministère des Ressources naturelles et des Forêts, 2021b; Waga, 2021).

The geometry criterion is integrated within the functional criterion and allows the differentiation of two kinds of forest roads: wide and narrow. Wide roads (≥7.5–8 m) provide general access for heavy forest vehicles and are typically surfaced with coarse allochthonous material, such as gravel and sand. The narrow roads are usually unpaved, and less robust (Government of Alberta, 2016; Mercier et al., 2019; Ministère des Ressources naturelles et des Forêts, 2021a). For example, construction specifications for forest roads in Quebec determines clearing width with the widest roads built with high-standard specifications to ensure robustness and long-term functioning. The surface structure consists of highly compacted coarse material. In contrast, the surface structure of narrow roads (<7.5–8 m) typically consists of mineral and/or organic soil and/or woody debris (Ministère des Ressources naturelles et des Forêts, 2021b).

Forest roads and seismic lines are similar with reference to their construction specifications and the impact sustained due to recurrent use. For example, low-impact seismic lines are widespread across Alberta and have reduced clearing widths (~2–3 m) compared to conventional seismic lines (~8–10 m). Their establishment requires canopy removal, soil flattening and compaction (Dabros et al., 2018; Lovitt et al., 2018). One different aspect should be noted and is in relation to their spatial association with other polygonal features such as cutblocks: seismic lines are constructed in dense grid networks (~50 m intervals) across vast areas, whereas forest roads are often associated with harvested patches (cutblocks) (Filicetti, 2021; Lovitt et al., 2018). However, we are not aware of comparative studies within these two types of linear features and their approximation in our study is based on a geometry criterion (similarities in construction specifications, i.e., clearing width). We adopted the geometry criterion for the discrimination of forest road types – wide (≥7 m) and narrow

(<7 m) – because clearing width is a comprehensive indicator expressing disturbance legacies and a driver of various local environmental effects. This discrimination criterion is coherent with previous studies integrating LiDAR based information on LFs in the boreal forest and would allow for standardization of LFs construction specification.

1.3 Methods to assess vegetation dynamic

Lee and Boutin (2006) and MacFarlane (2004) are among the first studies to provide insights on vegetation growth patterns exhibited on LFs in the boreal forest. Their findings showed slow growth on wide lines (5–8 m) over three-and-a-half-decade timeframe. Although conventional in-situ measurements of vegetation characteristics provide accurate growth increments, the broad-scale application of such tools is demanding in terms of time and human resources. Moreover, it is challenging to monitor slow processes such as vegetation growth and to characterize vegetation dynamic through conventional field surveys, especially within an extensive network of linear features.

With the development of remote sensing techniques, active remote-sensing platforms such as LiDAR have conferred large efficiency advantages to forest monitoring (Bour et al., 2021; Leitold et al., 2021) as it can generate numerous metrics characterizing vegetation including canopy height and cover (Barber et al., 2021; Martin and Valeria, 2022), as well as understory structure (Venier et al., 2019). Further, LiDAR is a promising data source for enhancing accurate and detailed vegetation characteristics mapping within linear features as it: i) contains canopy as well as under-canopy terrain information, ii) has large spatial extent (province-wide) consequently allowing for large-area monitoring of linear features, iii) has the potential to produce continuous wall-to-wall databases both in time and space and iv) allows for varying spatial and temporal resolutions depending on the purpose of the study and the level of detail required (e.g., low or high spatial resolution, multi-temporal).

Barber et al. (2021), Bayne et al. (2011), Dickie et al. (2017), Finnegan et al. (2018b), and van Rensen et al. (2015), represent the first applications of LiDAR to study vegetation characteristics within LFs by considering different LiDAR-measured metrics such as height growth using bi-temporal data and density of successional vegetation species. The need to satisfy spatially-continuous datasets requirements over the extensive and dense LFs network has also led to an increasingly use of LiDAR. Particularly, Abib et al. (2019) represents a proximity-based and spatial application of airborne LiDAR to characterize the variability in height and cover residuals within LFs and their proximal environments along a distance gradient.

ML approaches (e.g., nearest neighbour, ensembles such as random forests) have been extensively used in forestry remote sensing applications (Dalla Corte et al., 2020; Gleason and Im, 2012; Guo et al., 2020; Torre-Tojal et al., 2022; Zhao et al., 2011). The standard ML task is often to learn a predictive model that uses a remotely sensed dataset as input for the purpose of predicting the value of forest characteristics for unobserved cases/ at unsampled locations (Gangappa et al., 2017; Meyer and Pebesma, 2021). ML approaches have numerous advantages including the absence of assumptions about the structure of the input data, the ability to infer complex relationships, insensitivity to correlations among variables, as well as automatic modelling of interactions between input predictors (Torre-Tojal et al., 2022; Venier et al., 2019). Particularly, these non-linear approaches have been introduced in the literature as the best option to model complex relationships characterizing ecological and environmental data (Cosenza et al., 2021; García-Gutiérrez et al., 2015; Kalantar et al., 2020; Liu et al., 2022; Liu et al., 2018; Shin et al., 2016; Stojanova et al., 2010).

In recent years, various quantitative methods have been proposed and explored comparisons between conventional Ordinary Least Squares (OLS) approaches and regression approaches in machine learning (ML) (Bera et al., 2021; García-Gutiérrez et al., 2015; Gleason and Im, 2012; Liu et al., 2022; Stojanova et al., 2010). The advantages of ML in comparison to OLS approaches include flexibility and scalability, which makes them deployable for several tasks (Ahmed et al., 2015; Koma et al., 2021; Maxwell et al., 2018). While OLS approaches are aimed at inferring relationships between causal factors (e.g., multiple linear regression has been recognized as one of the most effective because of its ease of interpretation), ML approaches are focused on making improved predictions (Kalantar et al., 2020).

For instance, the Random Forest ensemble approach has showed improved predictions over other ML models in numerous comparative studies (Nawar and Mouazen, 2017; Stojanova et al., 2010). Furthermore, it can also overcome the overfitting problem seen with decision tree approaches (Cosenza et al., 2021; Forkuor et al., 2017; Kalantar et al., 2020). The Gradient boosting machine, suggested by Friedman (2001) is another recent ensemble approach in predictive modeling. The regression trees are strategically built from the residuals of the preceding decisions tree(s) and iteratively perform boosting through choosing, at each step, an arbitrary sample of the data ultimately leading to a progressive improvement of performance (Forkuor et al., 2017; Martin et al., 2014). The Gradient boosting machine approach has not been extensively explored and has yet to be tested for predicting vegetation characteristics (Bagherzadeh et al., 2021; Yang et al., 2020; Zhang and Haghani, 2015). Recent efforts are being focused on improving the interpretability of ML approaches, which will advance the capability to produce ecologically interpretable relationships. Particularly, the interpretability of tree-based ensemble approaches would enable a better understanding of the output of the predictive model and is critical in analyzing relationships between the response and the factors conditioning this response (Boehmke and Greenwell, 2019a; Greenwell, 2017; Molnar, 2020).

1.4 Objectives and hypothesis

Detailed spatially-explicit information on the three-dimensional characteristics of vegetation on forest roads are essential for a better understanding of growth mechanisms. However, studies examining spatial and temporal vegetation patterns on forest roads are scarce. We therefore extended on previous studies using airborne LiDAR remote sensing data for analyzing forest road vegetation characteristics dynamic. Because province-wide LiDAR canopy height models (CHM) data are now increasingly becoming available, we used openly accessible LiDAR data provided by the government of Quebec (Canada) (point density of 2–4 points/m^2) (Ministère des Ressources naturelles et des Forêts, 2022). We derived a LiDAR-CHM vegetation cover metric (response metric) by quantifying and extracting the mean of vegetation returns that were above 1.3 m height threshold. This metric determines the occurrence of vegetation within the road and represents the percentage of vegetation returns that are above a 1.3 m height cut-off, it describes the vertical projection of shrub and trees from all strata onto a horizontal surface representing the ground surface.

Our main objective was to investigate the performance of relevant regression approaches to determine a predictive model that achieves optimal accuracy for the characterization of within forest road vegetation cover dynamic. More specifically, we aimed to first, compare the performance of conventional OLS regression and ML approaches – i.e., multiple linear regression (*mlr*), generalized additive model (*gam*), multivariate adaptive regression splines (*mars*), k-nearest neighbors (*knn*), Random Forests (*rf*) and gradient boosting machine (*gbm*) – to provide a proximity-based predictive modelling framework for vegetation cover dynamic on forest roads. It is a surrogate for the surface of the forest road that experienced vegetation growth. Second, we evaluate the factors conditioning vegetation cover dynamic within forest roads.

To test our hypotheses (Table 1.1), we evaluated vegetation cover dynamic within a range of forest roads, in the managed forest of Quebec, Canada.

Table 1.1 Working hypotheses and predictions proposed to investigate vegetation cover dynamic on forest roads

Objective /Hypothesis	Description	Predictions
Objective i) Hypothesis a)	Tree-based ensemble approaches (gradient boosting machines and randoms forests) improve predictions of vegetation dynamic on forest roads.	Tree-based ensemble approaches (gradient boosting machines and random forests) outperform OLS and nonensemble modelling approaches for the characterization of forest road vegetation cover dynamic: tree-based ensemble approaches, combine multiple decisions trees (base models) to optimize the predictive performance, instead of fitting a single "best" model. This would lead to significant accuracy improvement over OLS and nonensemble approaches.
Objective ii) Hypothesis a)	History of forest roads best explains vegetation dynamic.	(Positive effect of time elapsed since the last clearing) Vegetation growth on forest roads is incremental with years post-clearing (construction or maintenance)
Objective ii) Hypothesis b)	The width attribute of forest roads best explains vegetation dynamic.	(Negative effect of forest road size, i.e., clearing width) Vegetation growth on forest roads is conditioned by clearing width: lower grade LFs would exhibit greater levels of vegetation cover.

Table 1.1 Continued

		Low-grade LFs are designed with lower initial construction costs which typically consists of the removal of the canopy without the addition of coarse granular material on the surface. Moreover, because the release of resources should be directly proportional to forest road clearing width, low-grade LFs should show reduced changes of microenvironmental conditions (e.g., incident sunlight) and are characterized by lower traffic rates.

Table 1.1 Continued

Objective / Hypothesis	Description	Predictions
Objective ii) Hypothesis c)	Local topography and vegetation attributes on forest roads best explain vegetation dynamic.	(Positive effect of steeper slopes) Vegetation cover levels are more advanced on steeper slope well-drained forest road. The slope controls the drainage behavior of the road. It is also a natural means to avoid the accumulation of water under the effect of gravity. (Negative effect of higher wetness index) Vegetation cover levels are more advanced on drier forest roads. (Positive effect of higher NDVI index)
Objective ii) Hypothesis d)	Proximity to the centreline of the road best explains vegetation dynamic.	(Positive effect of the distance from the center of the road) Vegetation cover patterns with regards to distance from the center are different between natural (e.g., gaps) and road openings because of traffic. In the center of natural canopy openings such as gaps, the distance from the center of the opening influence resource availability and micro-climatic conditions and increase vegetation cover in the center. A different pattern is expected on forest roads because of vehicle traffic (i.e., unlike natural openings, vegetation cover levels are likely to increase with distance from the road centerline).

CHAPTER II

CHARACTERIZATION OF VEGETATION DYNAMICS ON LINEAR FEATURES USING AIRBORNE LASER SCANNING AND ENSEMBLE LEARNING

Narimene BRAHAM[1], Osvaldo VALERIA [1], Louis IMBEAU [1]

[1]Institut de recherche sur les forêts, Université du Québec en Abitibi-Témiscamingue, 445 boul. de l'Université, Rouyn-Noranda, QC J9X 5E4

ABSTRACT

Forest roads are extensive throughout Quebec's boreal forest and are established for a wide array of activities: industrial (e.g., supply and distribution), recreational, silvicultural operations, environmental monitoring, and wildfire management. A consequence of this broad access network is its close implication in expanding forest matrix discontinuity resulting in the fragmentation of the landscape and degradation of forest dwelling species habitat. Accurate and fine-scale spatial information on vegetation characteristics play an important role in informing on vegetation dynamics on forest roads and has the potential to provide a better understanding of its spatial patterns. However, the dynamics of vegetation on linear features in general, and on forest roads have not been extensively explored. Conventionally, the measurement of vegetation characteristics is performed manually by means of field surveys. Light Detection and Ranging (LiDAR) technology can provide fine-scale continuous measurements of various vegetation characteristics because of its accuracy, acquisition flexibility and ability to generate wall-to-wall measurements. Predictive modelling, particularly regression approaches in ML, can increase the efficiency of LiDAR for broad-scale measurements, and can be used, under certain conditions, for the prediction of vegetation characteristics in unsampled locations. This study investigates the performance of Ordinary Least Squares (OLS) and Machine Learning (ML) approaches – namely, multiple linear regression (*mlr*), generalised additive model (*gam*), multivariate adaptive regression splines (*mars*), k-nearest neighbors (*knn*) gradient boosting machines (*gbm*), and random forests (*rf*) – for the characterization of vegetation cover dynamic within forest roads using an experimental set-up of 240 sample plots distributed over the managed forest of Quebec in eastern Canada. The *rf* tree-based ensemble approach yielded optimal performance results over that obtained using the other evaluated models (cross-validation: R^2=0.69, RMSE=18.69%, validation against an independent dataset: R^2=0.62, RMSE=20.29%) and was further used to derive factor importance. Consistent with our hypotheses, vegetation cover showed an upward trend with increasing years post-clearing and a downward trend with increasing clearing width. Our long-term predictions also suggest that a timeframe of no less than 20 years is expected for both wide- and narrow-width roads to exhibit ~50% and ~80% of vegetation cover, respectively.

The approach presented herein provides an effective assessment of fine-scale forest road vegetation cover dynamic. It demonstrates the value of using forest road specifications such as clearing width and years post-clearing to inform on vegetation cover dynamic and obtain improved estimates using the *rf* ensemble approach. The predictive framework is a versatile tool for forest road network planning and layout and management applications as it could improve spatially-explicit characterization of vegetation dynamics, wildlife habitat, and could be further tested and improved for the estimation of vegetation dynamics at the landscape and/or regional level.

Key words: Forest roads, airborne Light Detection and Ranging LiDAR, linear features, forest roads, road network, forest management, Random Forests, vegetation dynamic, boreal forest.

2.1 Introduction

Anthropogenic linear features (LFs) 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, LFs are similar in terms of their disturbance legacies as they require the use of machines that result in compaction of the surface layer through construction operations and consistent traffic intensity (Abib et al., 2019; Filicetti, 2021; Hornseth et al., 2018; Lovitt et al., 2018; Ministère des Ressources naturelles et des Forêts, 2021b; Pigeon et al., 2016). A consequence of these legacies is the 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 ratio (Vepakomma et al., 2018; Zhou et al., 2020). Even if some of these LFs are temporary or deemed to have a "low-impact" (Dabros et al., 2017; Kansas et al., 2015), they contribute to fragmentation with the majority (70%) of the world's forests being within 1 km of a forest edge (Forman, 2000; Haddad et al., 2015), with diminished habitat suitability adjacent to LFs caused by edge effects. Moreover, LFs have direct effects on wildlife species (Fisher and Burton, 2018; Forman, 2000; Mahon et al., 2019; Moreau et al., 2012; Sun et al., 2021), soil (Barber et al., 2021; Pigeon et al., 2016; Toivio et al., 2017; Zenner et al., 2007), seed dispersal and spread of wind-dispersed invasive species (Roberts et al., 2018), abiotic conditions (Franklin et al., 2021; Stern et al., 2018), forest structure and composition, both on LFs and their adjacent environment (Davidson et al., 2021; Eldegard et al., 2015). Since the most prevalent linear anthropogenic feature in many regions of eastern boreal forest are forest roads, management of this vast network to minimize the associated linear footprint on biodiversity and wildlife habitat, requires understanding of within forest road vegetation characteristics dynamic (Bourgeois et al., 2005; Clawges et al., 2008). However, vegetation patterns within forest roads need to be further explored: previous studies have shown that the growth process within linear features is complex and slow (Lee and Boutin, 2006; van Rensen et al., 2015). Furthermore, fine-scale knowledge on growth mechanisms within forest roads and application of this knowledge to management of the linear footprint is based on limited spatial levels and time scales. Previous studies assessing post-clearing, forest canopy spatio-temporal dynamic showed that the growth process is conditioned by disturbance factors, site conditions and location (Bartels et al., 2016; Senf et al., 2019). Moreover, in natural canopy openings, factors like light, nutrients, and water, have been showed to contribute and interact to affect growth of individual trees and saplings (Oliver and Larson, 1996). Abib et al. (2019) and Franklin et al. (2021) confirmed this relationship for LFs and showed that variations in vegetation growth is explained by LFs attributes (i.e., LFs width and orientation), local environmental factors (i.e., sunlight availability and the potential for accumulation of surface water) as well as terrain conditions. However, vegetation dynamic within forest roads requires more research for a better understanding of the conditioning factors.

The analysis of vegetation characteristics dynamic 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 (Atkins et al., 2018; Perez-Luque et al., 2020). Remote sensing techniques can reliably expand 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 (Wulder et al., 2012) as well as subcanopy topography (Brubaker et al., 2013). Coupled with the fact that these information can be derived across a range of spatial scales from fine (e.g.,~1 m^2) to coarse (e.g.,~100 km^2) (Vierling et al., 2008). The use of LiDAR data should provide a way to advance high-resolution quantification of vegetation and terrain characteristics within 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 (Girardin et al., 2022; Heidari et al., 2018).

LiDAR structural metrics related to height, density and complexity are relevant for research in forest structural characteristics (Weltz et al., 1994). In our study, we considered a density related LiDAR-CHM metric – sensu (Næsset, 2002) – to derive the percentage of vegetation returns ≥ 1.3 height threshold. 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 had published literature assessing their influence on vegetation dynamic. In particular, the size of canopy openings (Bartels et al., 2016; Hart and Chen, 2006), years post-clearing, disturbance history (Bartels et al., 2016; Senf et al., 2019), topography and climate (Hansen et al., 2014) were the main factors that have been shown to influence forest structural characteristics dynamic. Forests' structural characteristics are also determined by site conditions (Boucher et al., 2006; Mansuy et al., 2012; Thompson et al., 2009; Weiskittel et al., 2011), species composition (Ilisson and Chen, 2009) and successional status (Swanson et al., 2011). Previous literature showed that these aforementioned candidate factors are relevant for the characterization of vegetation cover on LFs (Filicetti and Nielsen, 2020; Franklin et al., 2021; Lee and Boutin, 2006; St-Pierre et al., 2021; van Rensen et al., 2015; Zang and Ding, 2009).

The extraction of ecologically relevant information on forest road vegetation characteristics requires the processing of Canopy Height Models (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 ≥ a given height threshold) (Koma et al., 2021; Martinuzzi et al., 2009). 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, 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 ability to fit nonlinear and complex relationships characterizing environmental and ecological data. Examples of predictive approaches include nearest neighbor methods (knn) – e.g.,(Chirici et al., 2016; Cosenza et al., 2021; Finley and McRoberts, 2008; Franco-Lopez et al., 2001; McRoberts, 2012) –, multivariate adaptive regression splines (mars) – e.g.,(Leathwick et al., 2006; Moisen and Frescino, 2002; Yang et al., 2020) –. Particularly, ensembles approaches – e.g., Gradient boosting machines (gbm) and Random Forests (rf) – are tools of choice in forestry (Abdi, 2020; Matasci et al., 2018b; Schönauer, 2022; Zhang et al., 2020), and in forestry modelling applications with airborne LiDAR (Abib et al., 2019; Ahmed et al., 2015; Venier et al., 2019). The widely-used rf tree-based ensemble approach (Breiman, 2001) 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 decision tree building, 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 is the reduced number of parameters to calibrate, and the choice of these parameters generally have very little influence on the accuracy of the results (Boehmke and Greenwell, 2019a). 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, suggested by Friedman (2001) and is a recent advance in predictive modeling. 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 of the model performance (Forkuor et al., 2017; Martin et al., 2014). However, gbm has yet to be tested for predicting vegetation characteristics. To optimize accuracy and avoid overfitting using ML approaches, model parameter specifications is an important step. It usually involves a number of interacting parameters that have to be calibrated (i.e. regularized) in order to achieve optimal results (Schratz et al., 2019).

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 dynamic and ii) to inform on the underlying factors conditionning vegetation cover dynamic. 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 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 independent validation dataset.

The combined use of LiDAR measurements and predictive modelling allow for a fine-scale and representative measurement of forest roads vegetation cover dynamic. This would also enhance our ability to precisely predict this dynamic along a spatial continuum and over extended timeframes.

2.2 Study sites

For this study, we retrieved forest roads clearing width data from the field across three study areas – field data collected in August 2019 (Girardin et al., 2022) –, representative of Canadian forestry activity, spanning between 47 and 49°N, -72 and -78°W in the mixed and coniferous boreal forest of Quebec (Canada) (Figure 2.1). 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, and the sites at higher altitudes being the coldest in winter and the least warm in summer. Precipitations also vary along the latitudinal gradient with drier conditions toward the North. The mean annual temperatures range between -5.9 and 4.2°C and total precipitations range between 650 to 1424 mm. 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 (Girardin et al., 2022; Ministère de l'Environnement et de la Lutte contre les changements climatiques, 1999; Robitaille and Saucier, 1998; Rossi et al., 2015).

Table 2.1 Properties of forest roads and their bioclimatic data, grouped by study area (1−3) (Blouin and Berger, 2001; Girardin et al., 2022; Gosselin and Berger, 2002; Ministère de l'Environnement et de la Lutte contre les changements climatiques, 1999).

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 [*Abies balsamea (L.) Mill*.] – white birch (*Betula papyrifera Marsh.)*	balsam fir – yellow birch (*Betula alleghaniensis Britton*)	black spruce *Picea mariana* (Mill.) – moss domain andbalsam 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.

2.3 Data

2.3.1 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 i) 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., 2022). 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. For visualization purposes, clearing widths were binned into: narrow forest roads (total narrow forest roads = 96), and were ≤7 m wide, whereas wide forest roads were >7 m wide (total wide forest roads = 144) (Ministère des Ressources naturelles et des Forêts, 2021b).

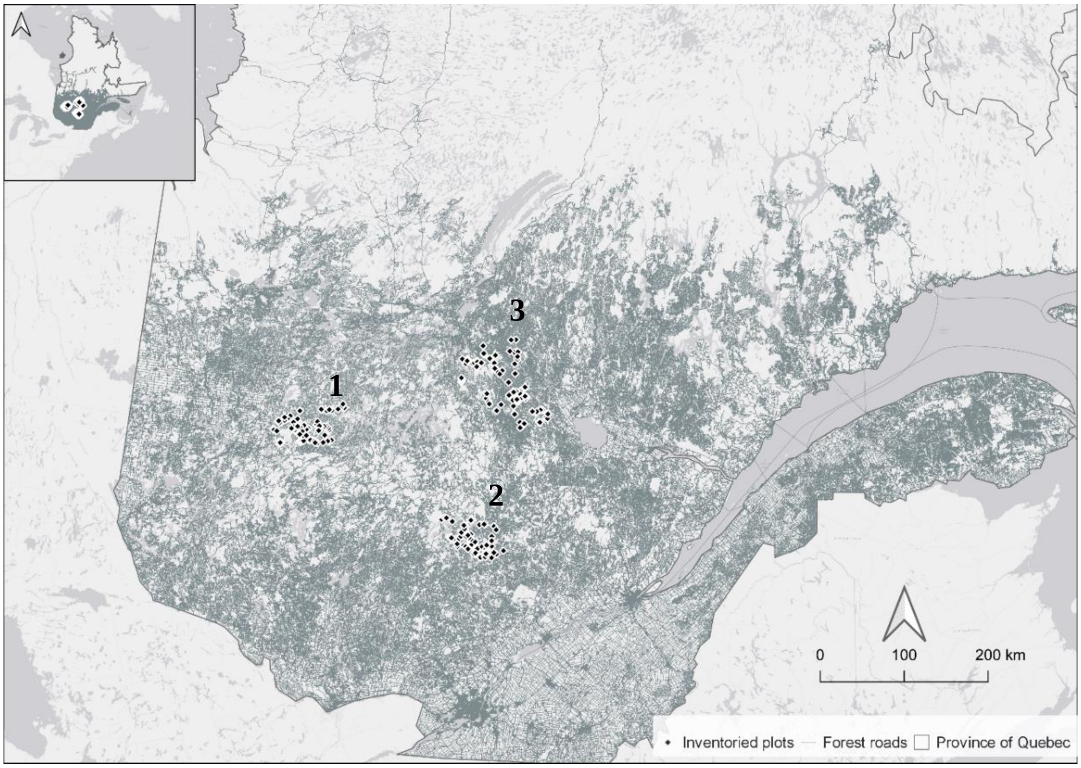


Figure 2.1 Overview of forest roads network (dark grey polylines) and distribution of sampled field plots (black dots) within the three respective study areas (1−3) in the province of Quebec in eastern Canada.

2.3.2 Forest road data: spatial buffering and causal factor data computation

To recreate the footprint polygons of forest roads from field-inventoried centerlines, we first delineated and digitized the centerlines using GPS coordinates (Trimble GNSS Handheld Geo7X) (three sampling locations for the edges and midpoint of the 50 m centerline) (Supplementary material Figure 1). To ensure proper alignment of the digitized centerlines, we used the LiDAR datasets provided by the Government of Quebec airborne LiDAR surveys, consisting in 1 m × 1 m grids (Ministère des

Ressources naturelles et des Forêts, 2020), collected between 2016 and 2020 with a mean pulse density of 2−4 pulse/m^2 (Martin et al., 2021). 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 and which we used to compute our input dataset (vegetation cover response and causal factors) for the characterization of vegetation dynamic (Figure 2.2). All data processing, modelling as well as validation were performed in the R programming environment Version 4.1 (R Core Team, 2020). All regression models were produced using the *caret* package 6.0 (Kuhn, 2015).

We established a framework that used data from multiple sources, including, airborne LiDAR, and geo-climatic ancillary data. We extracted these data from the 240 field plots using the multi-buffer delineation approach described in Figure 2.2. 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-buffers 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 data (1 m resolution) to compute: i) Slope, in degrees. ii) Orientation (Northerness) transformed to a continuous factor ranging between -1 and 1 (The Northerness values closer to -1 are southwards and those closer to +1 are northwards) (van Rensen et al., 2015). Orientation is typically transformed into a continuous factor because it is circular (large values may be very close to small values). iii) 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 upslope catchment area and downslope water drainage for each cell in a digital terrain model (Kopecký et al., 2021). iv) Hillshade is a proxy for the shadow based on the surface elevation (Hong et al., 2017; Piedallu and Gégout, 2007). 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 (Carlson and Ripley, 1997; Tarpley et al., 1984). Climate data were obtained from WorldClim version 2.1 for the time period 1970–2000 (Poggio et al., 2018; WorldClim, 2017). This dataset is based on historical climate records at a resolution of 30 seconds. The available monthly climate data of precipitation, incident sunlight (in units of kJ m^{-2} day^{-1}), wind speed ($m\,s^{-1}$), total precipitation (mm) and minimum, mean, and maximum temperature (°C), were used to compute growing season climate dataset, resampled to 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). 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 by Stern et al. (2018) and van Rensen et al. (2015). Before analysing 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., 2010). All uninformative metrics that showed a variance inflation factor greater than 3 or were highly correlated with one another ($|r_{Pearson}| > 0.7$) were excluded from the analysis.

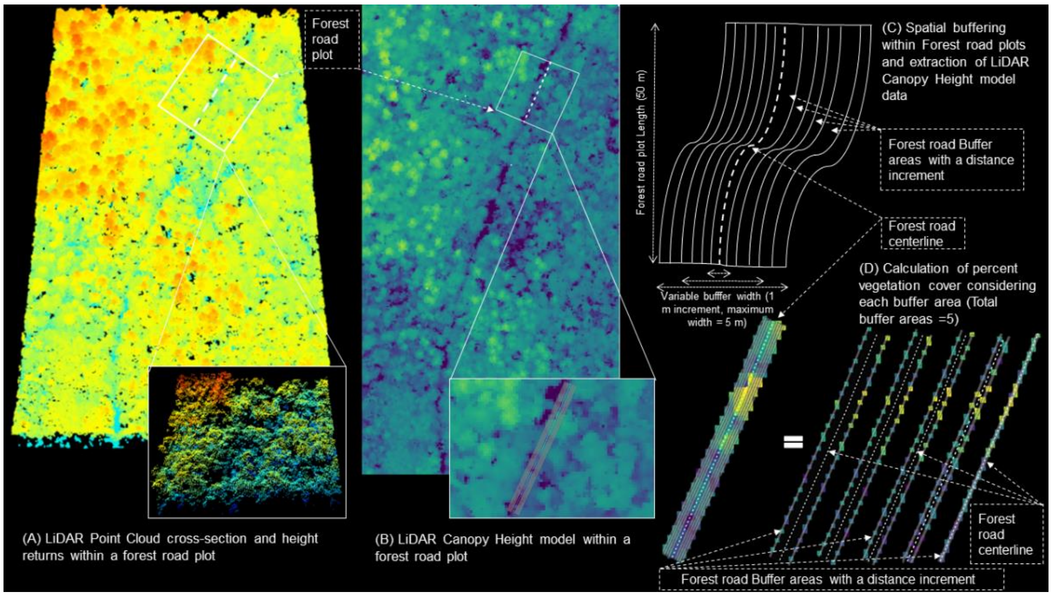


Figure 2.2 Visualization of LiDAR (A) 3D point cloud (B) Canopy Height Model (CHM) over a forest road. Extraction of forest road plot-level vegetation cover (%) using the CHM. Mean vegetation cover was derived continuously, and with a distance increment, within the five multi-buffer areas (Length = 50 m, and width increment = 1 m).

2.4 Methods

2.4.1 Statistical approaches

To provide an optimal predictive model for the estimation of vegetation cover on forest roads we compare the performance of the following OLS regression approaches: i) *mlr*, ii) *gam*, and four ML regression approaches including iii) *mars*, iv) *knn*, vi) *rf*, and vi) *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 (Friedman et al., 2001). *gam* was included because it presents an advantage over predefined basis functions to achieve nonlinearities and is relatively easy to interpret (Elith et al., 2008). The parsimony of *mlr* and *gam* approaches were assessed with the Akaike information criterion (AIC) (Burnham and Anderson, 2002). All possible combinations of factors and interaction effects were analysed with the *MuMIn* library in R (Barton, 2009). 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 (Vu et al., 2020). It structures a relation from established basis functions and coefficients, which are generally determined from the regression information (Lay et al., 2019).

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 (Friedman et al., 2001). Therefore, *mars* could simplify the challenges of solving non-linear relationships, compared to other non-parametric approaches (Lay et al., 2019).

We used the basic *knn* method (Crookston and Finley, 2008), a simple and intuitive approach in which each observation is predicted based on its similarity to other observations (Boehmke and Greenwell, 2019b). 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) (Bruce et al., 2020). *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 (Bruce et al., 2020), flexible assumptions regarding normality and homoscedasticity required by parametric methods, and preservation of much of the covariance structure among the metrics that define the response and factors' vectors (Crookston and Finley, 2008).

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, *rf* use a combination of bagging, which randomly selects factors with replacement as training for growing the trees which makes it robust against overfitting (Liaw and Wiener, 2002a). The training is carried out on datasets created from a random resampling on the training set itself which adds an extra layer of randomness (Breiman, 2001; Liaw and Wiener, 2002a). *gbm* is another recent tree-based ensemble which builds a base model (i.e., trees with only a few splits) (Friedman, 2002) and the additional trees iteratively correct mistakes made by the previous trees which progressively improves prediction accuracy. Particularly, *gbm* sequentially generate base models from a weighted version of the training data to find the optimal combination of trees and optimize predictive performance (Boehmke and Greenwell, 2019b; Natekin and Knoll, 2013). Both *rf* and *gbm* present 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 (Elith et al., 2008).

2.4.2 Models parameter tuning

ML model performance can benefit significantly from tuning as it may reduce overfitting (Martínez-Santos and Renard, 2020; Schratz et al., 2019). The caret library (Kuhn, 2015) 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*) (Boehmke and Greenwell, 2019a; Liu, 2020). 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 (Zhang and Haghani, 2015). Hence, we executed a grid search to evaluate: *ntree* which is the number of trees in a forest, *mtry* which defines the number of random factors at each split (Boehmke and Greenwell, 2019a). For *gbm*, we performed sensitivity analyses on tree complexity (*interaction depth*), learning rate (shrinkage), and minimum number of observations in nodes (*minobs*) (Boehmke and Greenwell, 2019a; Liu, 2020). During the tuning phase, a stratified 10-fold cross-validation resampling method allowed 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 1.

2.4.3 Models performance, comparison, and diagnostics using cross-validation and independent dataset

Inherit spatial information among dependent observations is one of the main challenges of spatial statistical modeling using ML techniques (Meyer and Pebesma, 2021; Meyer et al., 2018; Pohjankukka et al., 2017; Roberts et al., 2017; Schratz et al., 2019). 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 (Geiß et al., 2017; Pohjankukka et al., 2017; Schratz et al., 2019). 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, ii) the cross-validation folds, to be met (e.g., Garbasevschi et al. (2021) 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. Kosicki (2020) suggested that when the set of factors affect the response in different ways (positive/negative and/or linear/non-linear), and model's output is transferred to unsampled locations, more rigorous validation is necessary. Thus, we conducted a 60%–40% training-validation combination for evaluating our models' 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 factor importance computed from the fitted model that yielded optimal results (i.e., *rf)*. *rf* typically include 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 most important (Boehmke and Greenwell, 2019b). The factor importance computation was implemented using the *varImpPlot* function in the *randomForest* library (Liaw and Wiener, 2002b). Partial Dependence Plots (PDPs) are especially useful for visualizing the relationships discovered by ML approaches by isolating the effect of a single factor on the response (Molnar, 2020). We evaluated the partial dependence from our fitted *rf* model using two functions *partial* and *plotPartial* (Greenwell, 2017) 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 (Molnar, 2020).

2.5 Results

2.5.1 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 (kJ m-2 day-1) and Wind speed (m s−1)), terrain factors (Slope (%), Northerness (index), TWI (index) and Shade (index)) were computed. Overview and distribution of these input data are summarized in Table 2.2.

Table 2.2 Distribution of models input data for the characterization of vegetation cover dynamic on forest roads

Input(s)	Min	Max	Range	Median	Mean	std. dev*
LiDAR measured vegetation cover (%)	0	100	100	0	22.07	33.36
Years post-clearing (years)	0	39	39	7	7.79	8.35
Clearing width (m)	4	14.47	10.47	7.4	8.24	2.48
Sunlight (kJ m-2 day-1)	17228.74	17729.8	501.06	17598.99	17545.63	136.74
Wind Speed (m s−1)	2.2	2.88	0.68	2.34	2.45	0.2
Slope (%)	0	27.73	27.73	6.71	7.94	5.41
Northerness (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
Shade (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

*Standard deviation

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 2.3 (A) and 2.3 (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).

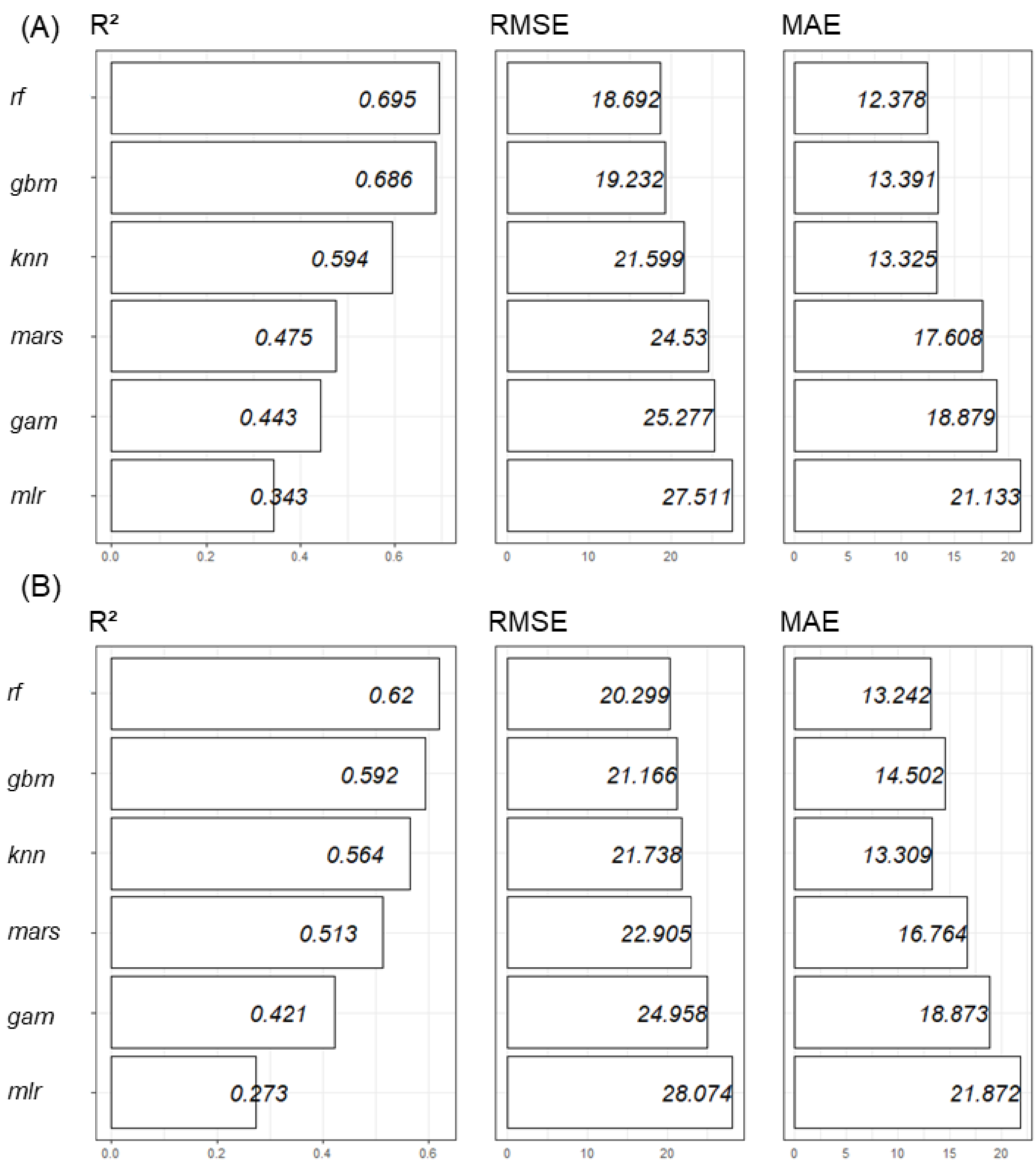


Figure 2.3 R², RMSE, and MAE for ML and OLS approaches for the characterization of vegetation cover dynamic obtained from (A) 10-fold stratified cross-validation

(results from 20 repetitions were considered), (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 by RMSE and R^2 (Figure 2.4 (A) and Figure 2.4 (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*.

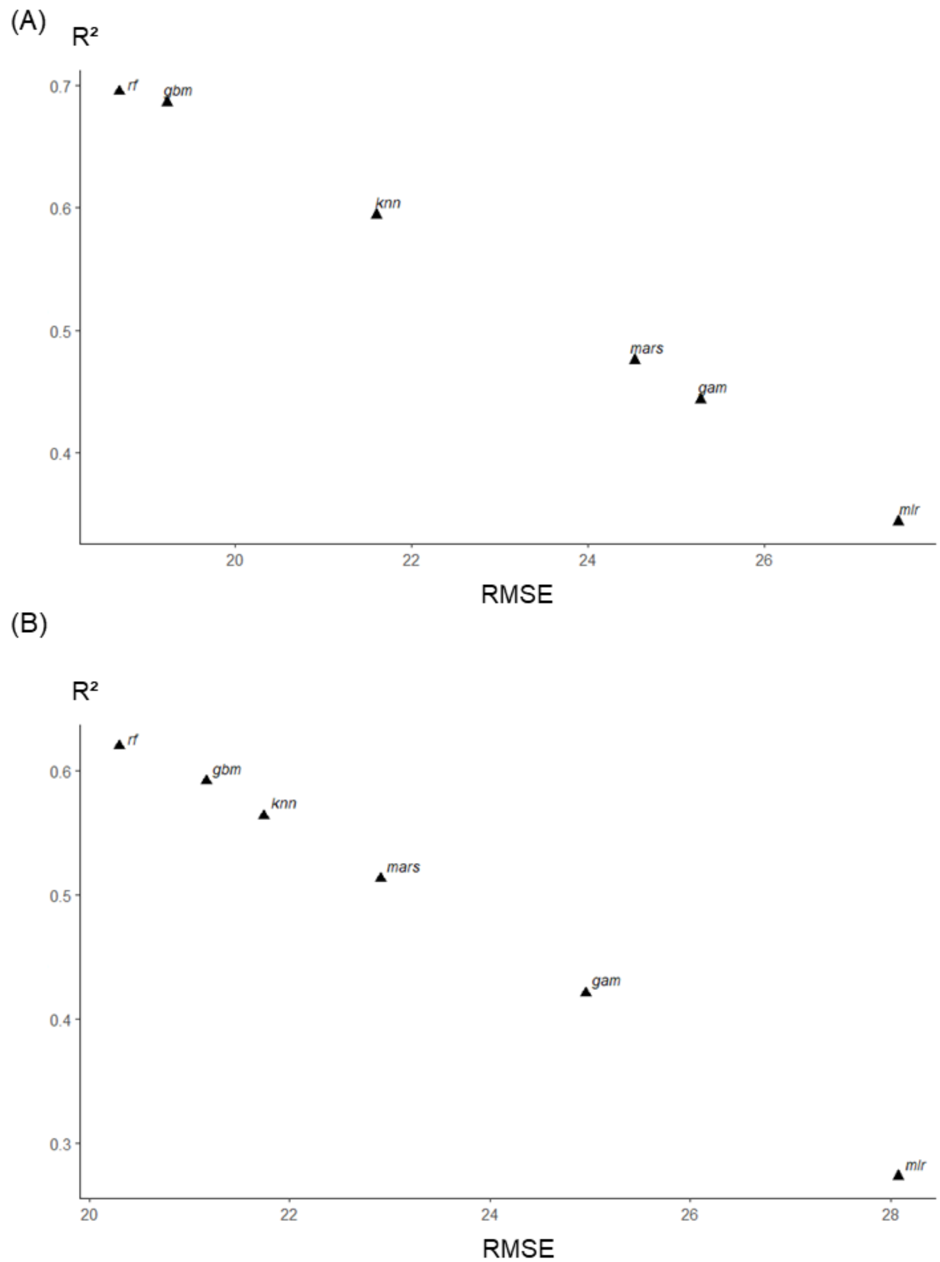


Figure 2.4 Predictive performance of ML and OLS for the characterization of vegetation cover dynamic using (A) 10-fold cross-validation approaches, (B) An independent validation dataset. *rf* = random forests, *gbm* = gradient boosting machines,

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 2.5. Because *rf* generally provided optimal performance results, factors ranking was derived using this approach. Clearing width was the most important factor explaining vegetation cover dynamic within 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 had 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, Northerness 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 4).

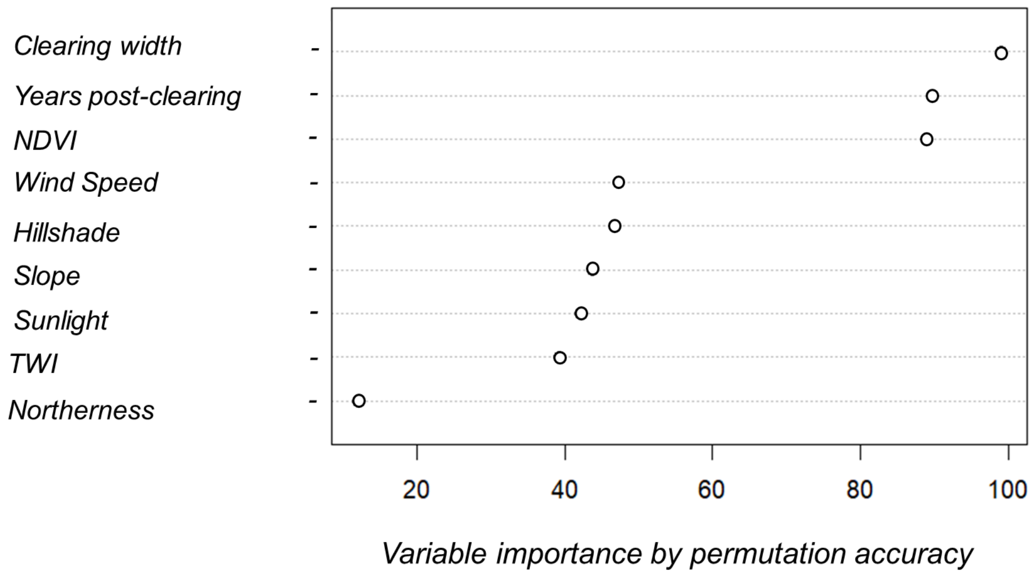
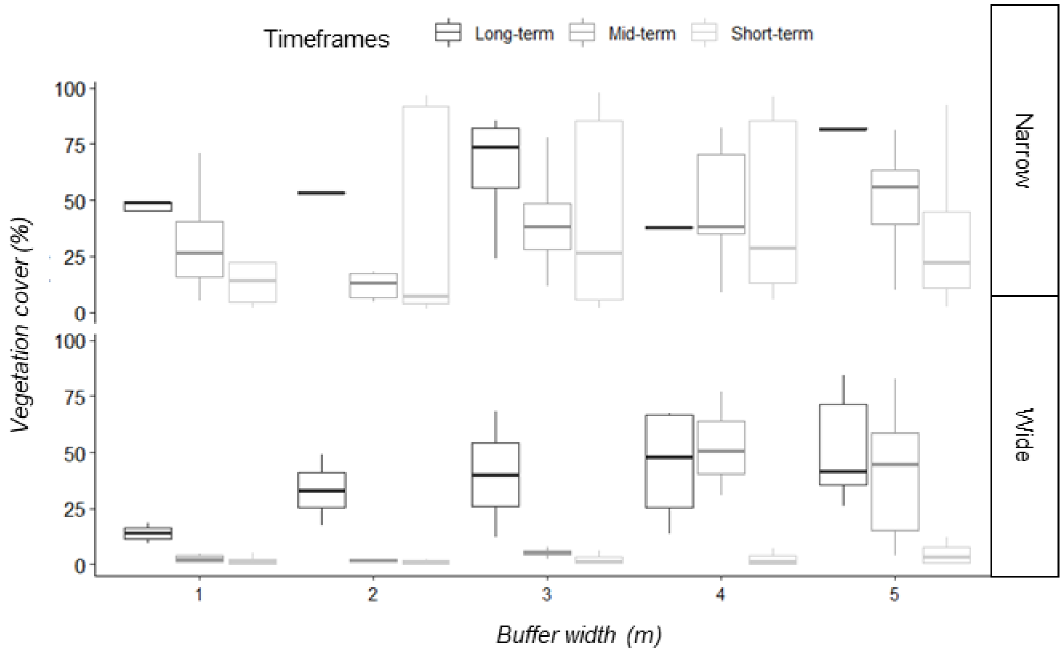


Figure 2.5 *rf*-based factor importance by permutation accuracy. A higher average importance of the variable (x-axis) indicates a greater contribution of this individual variable in explaining within forest road vegetation cover dynamic. A ranking of all factors is included.

2.5.2 Characterization of vegetation cover dynamic within forest roads

rf-based vegetation cover dynamic 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 2.6 (A) using the cross-validation predictions, and Figure 2.6 (B) 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 3 (A) and (B)).

(A)


Timeframes
Long-term
Mid-term
Short-term
Vegetation cover (%)
100
75
50
25
0
Narrow
Wide
1
2
3
4
5
Buffer width (m)

(B)

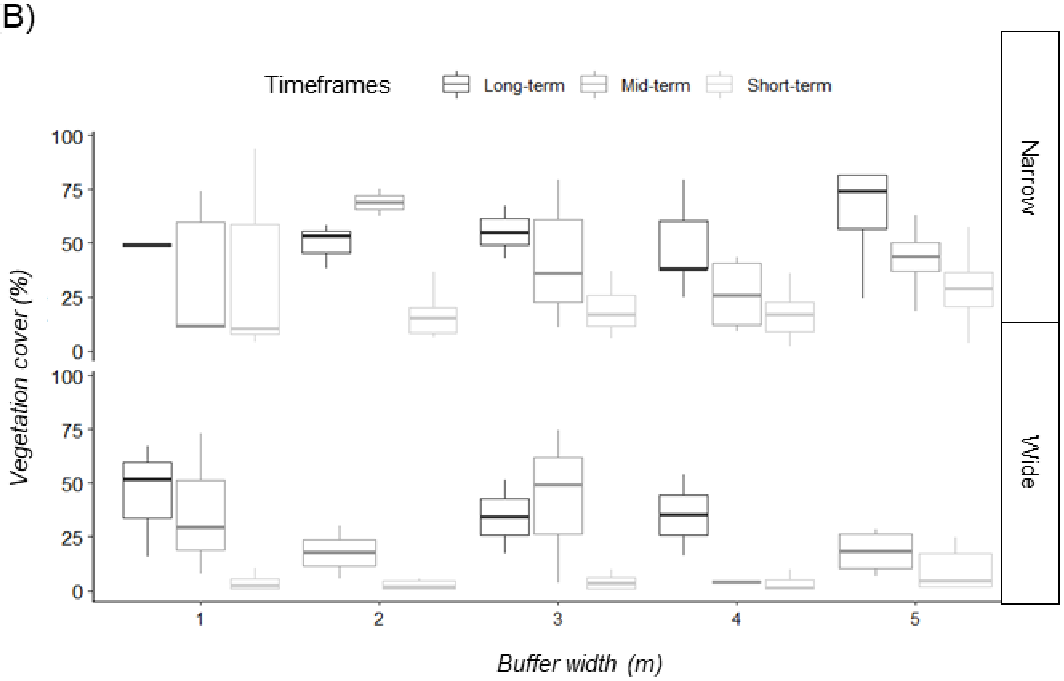

Timeframes
Long-term
Mid-term
Short-term
Vegetation cover (%)
100
75
50
25
0
Narrow
Wide
1
2
3
4
5
Buffer width (m)

Figure 2.6 (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), ±1 standard deviation (rectangle) and maximum-minimum value (vertical lines or whiskers).

As shown in Figure 2.3 (A) and (B), 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 2.6 (A) and (B)). In general, we found the ML approaches evaluated here to be useful tools for improving predictions of vegetation cover dynamic on forest roads.

2.6 Discussion

2.6.1 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 *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 dynamic within forest roads. Further, 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 models assumptions about the non-linear distribution of input data: *rf* better accommodates nonlinear relationships between factors that *mlr* could not adequately solve (Hassan et al., 2017; Morellos et al., 2016; Prasad et al., 2006).

Consistent with our hypothesis, tree-based ensemble approaches outperformed their nonensemble counterparts. *rf* and *gbm* are extremely randomised trees and are both based on ensemble learning theory. The ensemble – aggregation of decision trees (Hassan et al., 2017) – 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 (Ahmad et al., 2018; Ahmed et al., 2015; Breiman, 2001; Stojanova et al., 2010).

ML approaches require the setting of parameters specification 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 (Maxwell et al., 2018). 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 (Cosenza et al., 2021; Hultquist et al., 2014; Kosicki, 2020; Martínez-Santos and Renard, 2020; Morellos et al., 2016; Nawar and Mouazen, 2017).

2.6.2 Factors conditioning vegetation cover dynamic within forest roads

Factors associated with vegetation dynamics

Our results identified that the topmost 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 Northereness) 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. (2019) and Franklin et al. (2021) showed that variations in proximity-based vegetation cover is explained by LFs 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. (2015) 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 (LFs lines in bogs and fens were less likely to experience growth than those in drier conditions). Similarly, Finnegan et al. (2019) 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 occurrence of disturbance-tolerant taxa.

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 LFs literature (Lee and Boutin, 2006; van Rensen et al., 2015). Particularly, LFs construction and design specifications can differ with respect to their characteristics (e.g., bearing capacity) and moisture conditions (O'mahony et al., 2000). These differences are reflected in their trafficability, frequency and intensity of use (Kaakkurivaara et al., 2015). 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 (Kaakkurivaara et al., 2015; Ministère des Ressources naturelles et des Forêts, 2021a). Due to their high trafficability, wide LFs are also prone to experience 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 (Heidari et al., 2018; Kaakkurivaara et al., 2015; Ministère des Ressources naturelles et des Forêts, 2021b; Rummer and Wear, 2002; Waga, 2021). 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. (2021) demonstrated 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. (2015) showed that except for coarse textured excessively drained soils, soil compaction reduces oxygen and water availability to roots and microorganisms. Zang and Ding (2009) 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 (Dabros et al., 2018; Pinard et al., 2000; Startsev and McNabb, 2009). 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, low bearing capacity interfere with narrow LFs intensity of use (Kaakkurivaara et al., 2015; Ministère des Ressources naturelles et des Forêts, 2021b; Waga, 2021). Hence, it is very likely that the integration of LFs clearing width captured underlying differences in hydrological conditions such as water and nutrients 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 among forest road types.

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 cover 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. (2018) where LFs openings exhibited double incident sunlight intensity and double maximum wind speed compared to the adjacent forests. The abiotic conditions were different among LFs with different clearing widths: wide LFs exhibited increased sunlight penetration that extended into the forest. Centres of wide seismic lines were characterized by >1.5 times sunlight intensity than those of narrow seismic lines. These results corroborate the findings in Franklin et al. (2021) showing that the microclimatic conditions in the middle of LFs were generally intermediate between interior forest and anthropogenic infrastructures such as well pads, with narrow seismic lines more similar to interior forest and wide seismic lines more similar to well pads. The width and orientation of LFs also influenced growth trends in Franklin et al. (2021) 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. (2021) as their sampled wide LFs were older than the narrow LFs and had therefore 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 among 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) (Dabros et al., 2017; Franklin et al., 2021). On narrow LFs, however, significant shading from the adjacent canopy, provides more favourable conditions for vegetation cover. This is in favor of the assumption that the clearing width is a modulator of online abiotic conditions including sunlight, wind, and moisture (Filicetti and Nielsen, 2018; van Rensen et al., 2015). Hence, research on the abiotic environment within LFs is needed to provide insight into potential explanations for abiotic-biotic associated patterns. Also, the floristic aspect of online communities should be considered for an integrative investigation of vegetation characteristics within LFs (Lázaro-Lobo and Ervin, 2019; Lee and Boutin, 2006).

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. (2015), YPC was among the most influential factors, and it is possible that our continuous factor better accounted for the variation of 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 TWI data indicating increased water accumulation reduce vegetation cover on forest roads.

2.6.3 Characterization of vegetation cover dynamic within 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 (Guariguata and Dupuy, 1997; Olander et al., 1998), South-East Asia (Pinard et al., 2000; Zang and Ding, 2009) and Central Africa (Malcolm and Ray, 2000). They provided evidence of increased disturbance on wide LFs, as well as variations in density, diversity, and vegetation structure across the LFs surface and their proximal environments (edge and adjacent forest). These results and findings in Lee and Boutin (2006) 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 the boreal forest. 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 LFs literature (e.g., spectral indices (White et al., 2018), structure: closure through both height (regeneration) and lateral growth (Finnegan et al., 2018b; Matasci et al., 2018a; Vepakomma et al., 2011). 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 (Bartels et al., 2016).

A quantitative study in the Central African forest (Kleinschroth et al., 2016) demonstrated the potential for vegetation growth on abandoned LFs (logging roads) through natural processes: for an average of 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). Three decades post-clearing allowed the convergence of canopy closure litter layer depth and herb cover converged notably to levels that were comparable to the adjacent forest. However, the thirty-year timeframe was not sufficient to recover the same biomass on roads as in the adjacent logged forest. 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 of 17–51% and 40–82%, in the mid- and long-term, respectively. The differences could be attributed to forest ecosystems specifications (e.g., vegetation and soil conditions), the metric used to quantify vegetation characteristics on the roads, road construction specification (e.g., clearing widths). Lee and Boutin (2006) findings 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% of woody vegetation growth. LFs vegetation predictions in Finnegan et al. (2019), showed 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. (1984) 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 LFs 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 (Bartels et al., 2016). 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. Further, Senf et al. (2019) 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.

2.6.4 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 Abib et al. (2019)’s proximity-based analysis, 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 (Hajjem et al., 2014).

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, iii) the output can be unstable: for example, small changes in data can produce highly divergent trees for example (Prasad et al., 2006). In this study, ML approaches, compared to OLS approaches yielded satisfactory accuracy results for the prediction of vegetation dynamic but there are limitations concerning the generalisation 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 generalised for the prediction of the same characteristics in any unsampled location or within forest roads with different specifications. Because large-area generalisation (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 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 (Dabros et al., 2018; Finnegan et al., 2018a). 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 (Norden et al., 2009). 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.

2.7 Conclusion

In this study, we characterized within forest road vegetation cover dynamic 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 accurate characterization of forest road vegetation cover dynamic. 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 generalisation. Additionally, transport flux and volumes, compaction levels, and the construction materials are among the potential factors that could be included to evaluate possible decrease in model error. With the increasing availability of remote sensing datasets, there is potential for broad scale mapping of vegetation dynamics within forest roads (landscape or regional level). Further investigations are also required in improving the temporal resolution of vegetation measurements with LiDAR.

GENERAL CONCLUSION

Spatially-explicit detailed vegetation information on forest roads is essential for a better understanding of growth mechanisms, assessment of forest road quality and management of forest porosity caused by linear anthropogenic openings. Our study extends on earlier research focused on anthropogenic linear features quality assessment and aimed to investigate fine-scale vegetation cover dynamic using remote sensing data derived from high spatial resolution LiDAR-CHM (1 m). For this purpose, we used our field inventory as a reference to guide our experimental set-up around the centerline of forest roads. Using the same set-up, we quantified the area of the road that is covered in vegetation (height cut-off > 1.3 m) within multi-buffered areas (5 m) around the centerline. This approach also allowed us to compute our vegetation metric (i.e., vegetation cover response) and causal factors (i.e., topographic (Slope, shade, TWI), vegetation (NDVI), climatic (incident sunlight levels and wind speed) using airborne LiDAR and Sentinel-2 data.

The performance accuracy of conventional OLS and ML statistical approaches – multiple linear regression (*mlr*) generalized additive model (*gam*), (*mars*), random forest regression (*rf*), k-nearest neighbors (*knn*), gradient boosting machines (*gbm*) – were evaluated and compared. The predictions were validated using a stratified cross-validation scheme and against an independent set of samples. Our findings demonstrate that the *rf* approach had optimal predictive accuracy (RMSE ranging from 18.69 to 20.29 and R^2 ranging from 0.69 to 0.62, using stratified cross-validation and the independent dataset, respectively) due to its i) ability to model complex non-linear relationships between the factors and the vegetation response, ii) ability to reduce possible errors and over-fitting and form an average optimal model with improved predictions. The *rf* regression model results indicated that clearing width and years post-clearing were the most informative causal factors. Thereby, vegetation cover predictions would require these abiotic data inputs to accurately estimate vegetation levels on forest roads. Future studies could expand on the input dataset and include transport flux and volumes, compaction levels, and the construction materials for potential improvement in the prediction capability of the predictive model. Moreover, further research is required to explore the floristic component and differences in growth patterns between vegetation types and to examine growth mechanisms on linear openings (i.e., height and cover increments and their contribution to the increase of the vegetation fraction within anthropogenic linear features). In addition, the potential of temporal analysis approaches in improving prediction accuracy are worth investigating for a better understanding of spatio-temporal vegetation dynamics within linear features.

Our finding improved our understanding of the relevant factors underlying vegetation dynamic on forest road. The proposed proximity-based approach for spatially-annotated predictions surrounding forest roads and could thus allow for flexibility in future studies focused on linear features and their environments (immediate or proximal). Moreover, the combined use of remotely-sensed data, (i.e., LiDAR DEMs and CHMs) and ML prediction framework are efficient for the extraction of extensive and continuous information on vegetation characteristics dynamic which can assist in the management of the linear footprint of forest roads in the boreal forest. Particularly, quantitative methods for road quality assessment can provide decision support tools for multi-scale forest road network management.

SUPPLEMENTARY MATERIAL

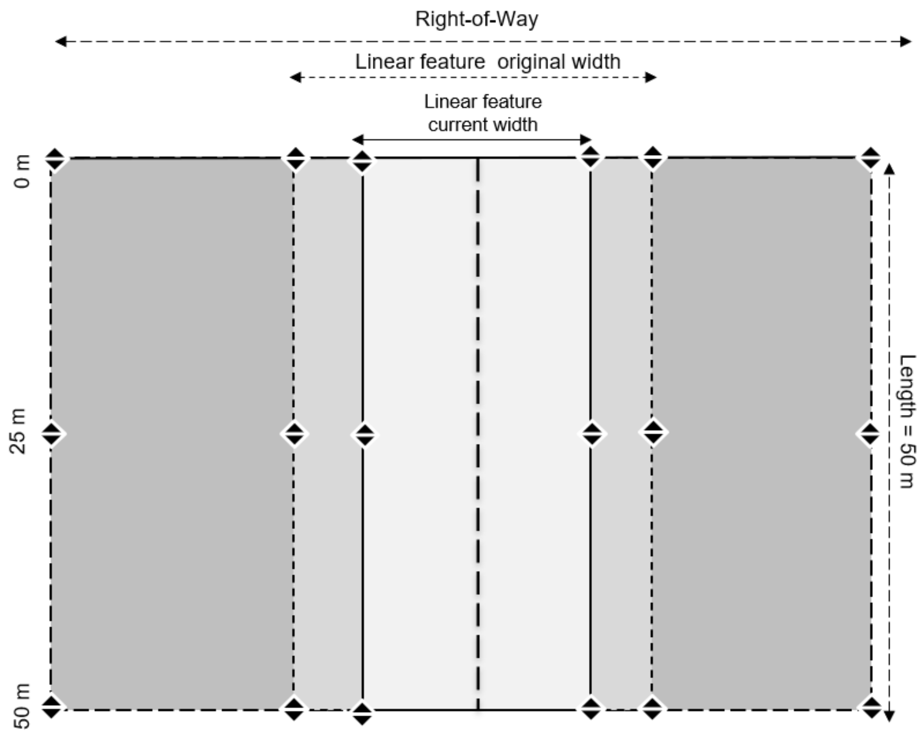


Figure 1 Field inventory plot design used to reconstruct the centerlines (Dimensions: 50m*clearing width).

Table 1 Hyperparameters (ranges and types) and their definitions

	Hyperparameter	Description	Type	Start	End	Default
mars	*degree*	Number of retained terms	Integer	1	3	1
	nprune	Degree of interactions	Integer	2	100	8
knn	*k*	Number of neighbors	Integer	5	20	7
	distance	Metrics to measure the distance between observations	Integer	1	5	2
	kernel	types of kernel functions	Nominal	Rank, cos, inv, Gaussian, optimal.	-	-
rf	m_{try}	Number of variables to randomly sample as candidates at each split	Integer	1	3	$\sqrt{p}$
	ntree	Number of trees	Integer	250	3000	500
gbm	*shrinkage*	Learning rate of greedy search	Numeric	0.001	0.1	0.001
	interaction.depth	Maximal tree depth	Integer	1	9	1
	n.minobsinnode	Minimal terminal node size	Integer	1	10	20

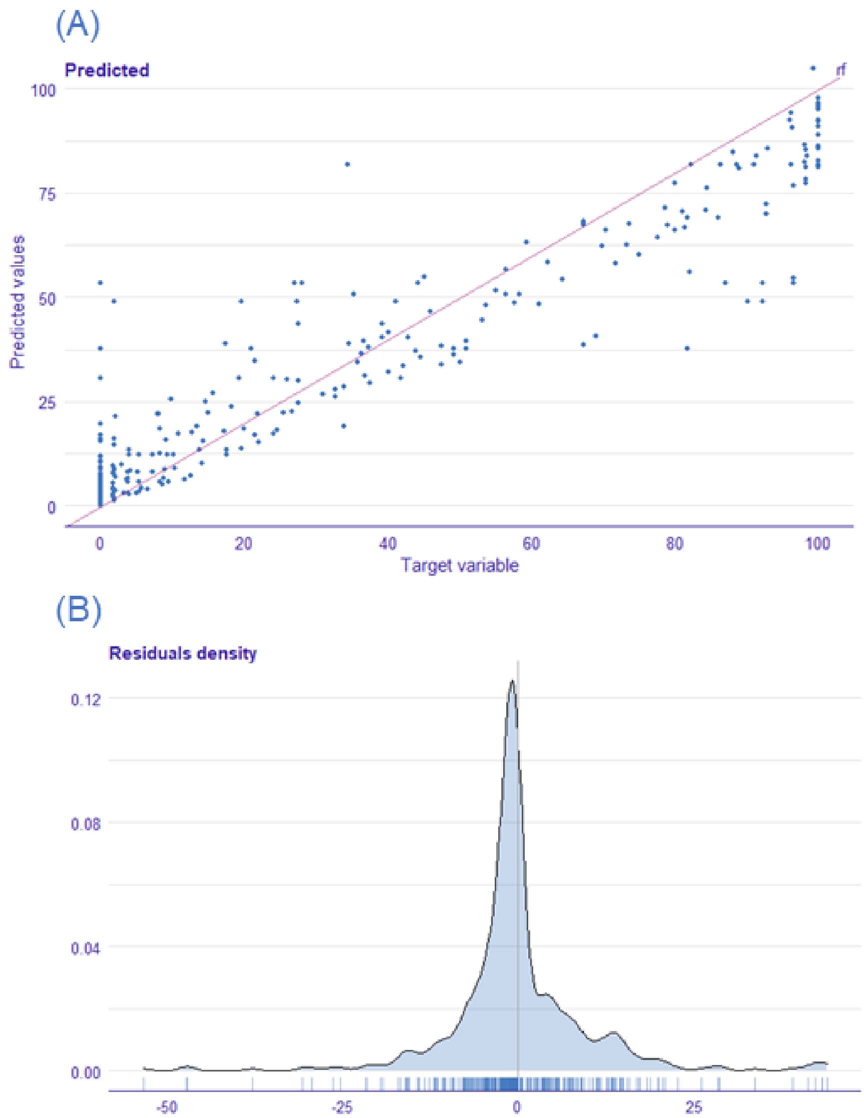


Figure 2 Diagnostic plots for the *rf* model generated with *auditor*: an R Package for model-agnostic visual validation and diagnostic (Gosiewska and Biecek, 2018), generally indicate a well fitted model. The plots show: (A) vegetation cover observations (%) (Target variable, x-axis) versus individual predictions (%) (Predicted values, y-axis). We see a uniform scattering around the diagonal and close dispersion of the predicted against the target variable at the center of the plot. The dispersion

around the diagonal line indicates over-prediction for small values of observed response and under-prediction for large values. (B) A common use of the residuals density plot is to analyze the variance of the error of the regression approach (here *rf*). The histogram indicate that our error is normally distributed close to zero.

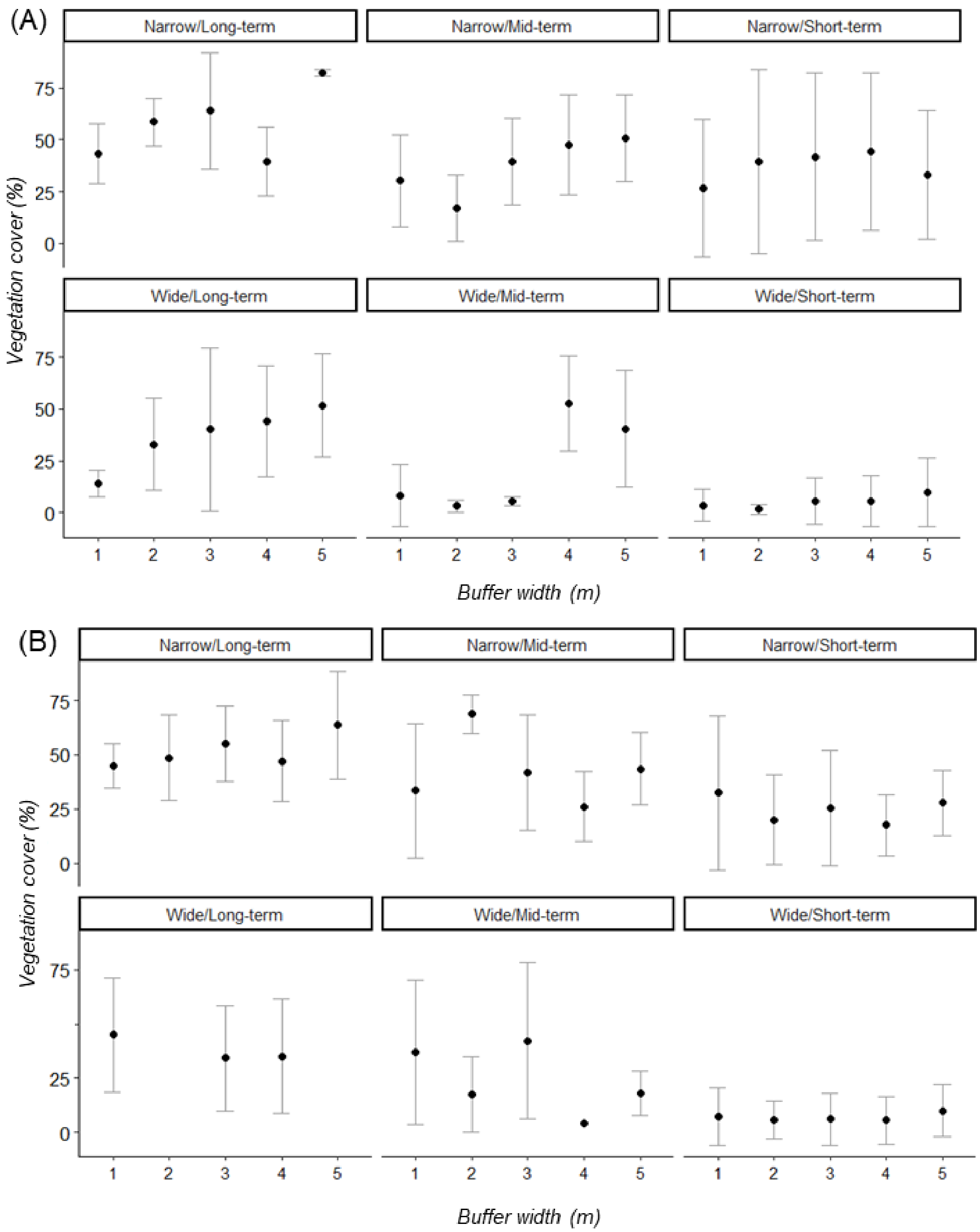


Figure 3 (A) Summary of vegetation cover predictions (means and means +/- standard deviation error bars) grouped by different forest road categories and timeframes, from

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.

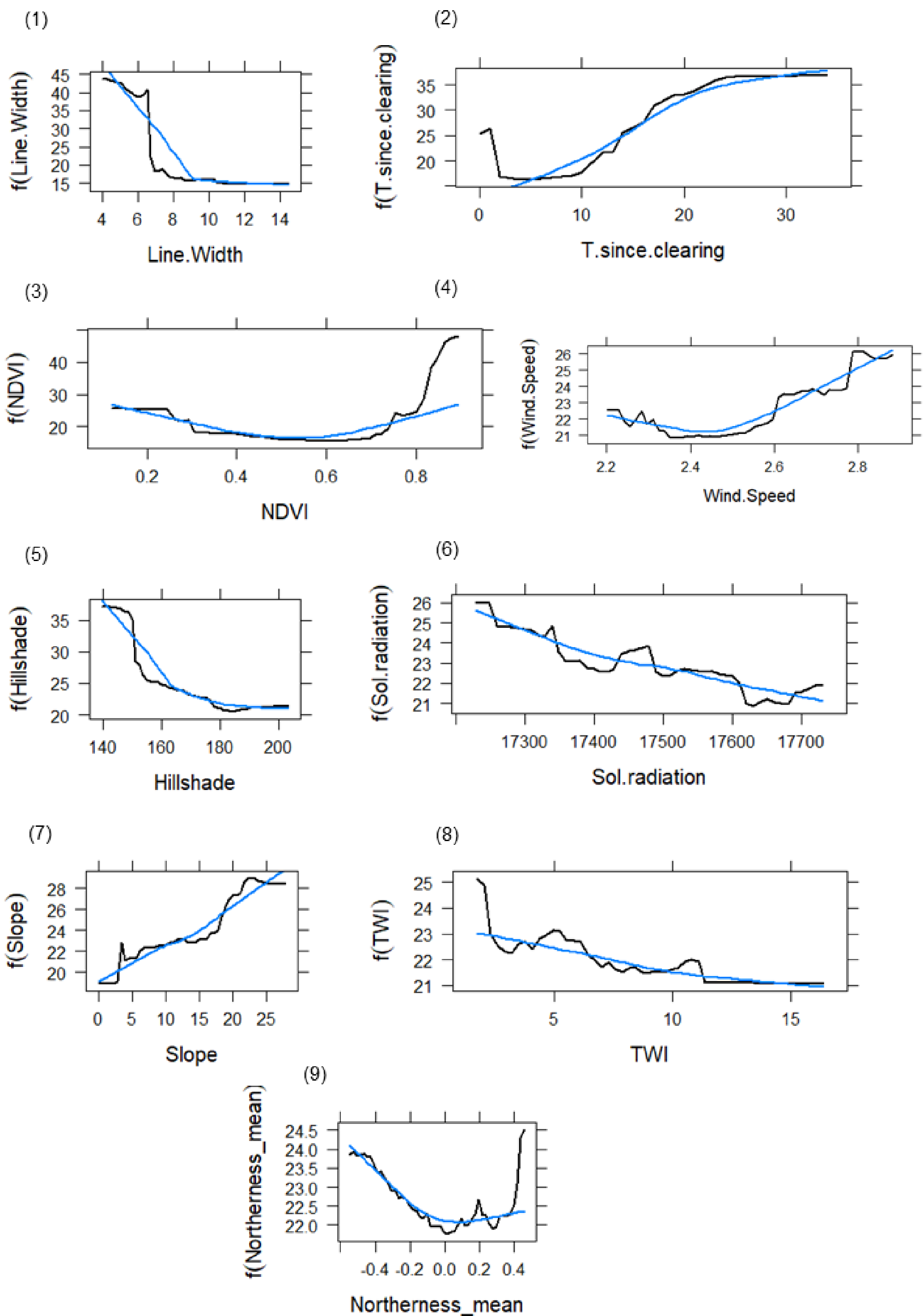

(1)
f(Line.Width)
Line.Width
(2)
f(T.since.clearing)
T.since.clearing
(3)
f(NDVI)
NDVI
(4)
f(Wind.Speed)
Wind.Speed
(5)
f(Hillshade)
Hillshade
(6)
f(Sol.radiation)
Sol.radiation
(7)
f(Slope)
Slope
(8)
f(TWI)
TWI
(9)
f(Northerness_mean)
Northerness_mean

Figure 4 *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. Abbreviations: (1) clearing width (Line.Width), (2) time (years) since last clearing (T.since.clearing), (3) NDVI, (4) Wind speed (Wind.Speed), (5) Hillshade, (6) Sunlight (Sol.radiation), (7) Slope, (8) TWI, and (9) Northerness (Northerness_mean).

BIBLIOGRAPHY

Abdi, A.M. (2020) Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing 57, 1-20.

Abib, T.H., (2018) Analysis of the vertical canopy structure in native forest fragments and Eucalyptus plantations to detect edge effects. Universidade de São Paulo.

Abib, T.H., Chasmer, L., Hopkinson, C., Mahoney, C., Rodriguez, L.C. (2019) Seismic line impacts on proximal boreal forest and wetland environments in Alberta. Science of the Total Environment 658, 1601-1613.

Ahmad, M.W., Mourshed, M., Rezgui, Y. (2018) Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. Energy 164, 465-474.

Ahmed, O.S., Franklin, S.E., Wulder, M.A., White, J.C. (2015) Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm. ISPRS Journal of Photogrammetry and Remote Sensing 101, 89-101.

Atkins, J.W., Bohrer, G., Fahey, R.T., Hardiman, B.S., Morin, T.H., Stovall, A.E., Zimmerman, N., Gough, C.M. (2018) Quantifying vegetation and canopy structural complexity from terrestrial Li DAR data using the forestr r package. Methods in Ecology and Evolution 9, 2057-2066.

Bagherzadeh, F., Mehrani, M.-J., Basirifard, M., Roostaei, J. (2021) Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. Journal of Water Process Engineering 41, 102033.

Barber, Q.E., Bater, C.W., Dabros, A., Pinzon, J., Nielsen, S.E., Parisien, M.-A. (2021) Persistent impact of conventional seismic lines on boreal vegetation structure following wildfire. Canadian Journal of Forest Research 51, 1581-1594.

Bartels, S.F., Chen, H.Y., Wulder, M.A., White, J.C. (2016) Trends in post-disturbance recovery rates of Canada's forests following wildfire and harvest. Forest Ecology and Management 361, 194-207.

Barton, K. (2009) MuMIn: multi-model inference. http://r-forge. r-project. org/projects/mumin/.

Bennett, V.J. (2017) Effects of road density and pattern on the conservation of species and biodiversity. Current Landscape Ecology Reports 2, 1-11.

Bera, D., Chatterjee, N.D., Bera, S. (2021) Comparative performance of linear regression, polynomial regression and generalized additive model for canopy cover estimation in the dry deciduous forest of West Bengal. Remote Sensing Applications: Society and Environment 22, 100502.

Blouin, J., Berger, J. (2001) Guide de reconnaissance des types écologiques de la région écologique 5b coteaux du réservoir Gouin. Québec, Canada: Ministère des Ressources Naturelles du Québec, Forêt Québec, Direction des inventaires forestiers, Division de la classification écologique et productivité des stations.

Boehmke, B., Greenwell, B. (2019a) Hands-on machine learning with R.

Boehmke, B., Greenwell, B.M. (2019b) Hands-on machine learning with R. CRC Press.

Boucher, D., Gauthier, S., Grandpré, L.D. (2006) Structural changes in coniferous stands along a chronosequence and a productivity gradient in the northeastern boreal forest of Québec. Ecoscience 13, 172-180.

Bour, B., Danneyrolles, V., Boucher, Y., Fournier, R.A., Guindon, L. (2021) Modeling post-logging height growth of black spruce-dominated boreal forests by combining airborne LiDAR and time since harvest maps. Forest Ecology and Management 502, 119697.

Bourgeois, L., Kneeshaw, D., Boisseau, G. (2005) Les routes forestières au Québec: Les impacts environnementaux, sociaux et économiques. VertigO-la revue électronique en sciences de l'environnement 6.

Breiman, L. (2001) Random forests. Machine Learning 45, 5-32.

Brubaker, K.M., Myers, W.L., Drohan, P.J., Miller, D.A., Boyer, E.W. (2013) The use of LiDAR terrain data in characterizing surface roughness and microtopography. Applied and Environmental Soil Science 2013.

Bruce, P., Bruce, A., Gedeck, P. (2020) Practical statistics for data scientists: 50+ essential concepts using R and Python. O'Reilly Media.

Burnham, K.P., Anderson, D.R. (2002) A practical information-theoretic approach. Model selection and multimodel inference 2.

Cambi, M., Certini, G., Neri, F., Marchi, E. (2015) The impact of heavy traffic on forest soils: A review. Forest Ecology and Management 338, 124-138.

Carlson, T.N., Ripley, D.A. (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sensing of Environment 62, 241-252.

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E.O., Waser, L.T., Travaglini, D., McRoberts, R.E. (2016) A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. Remote Sensing of Environment 176, 282-294.

Clawges, R., Vierling, K., Vierling, L., Rowell, E. (2008) The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. Remote Sensing of Environment 112, 2064-2073.

Cosenza, D.N., Korhonen, L., Maltamo, M., Packalen, P., Strunk, J.L., Næsset, E., Gobakken, T., Soares, P., Tomé, M. (2021) Comparison of linear regression, k-nearest neighbour and random forest methods in airborne laser-scanning-based prediction of growing stock. Forestry: An International Journal of Forest Research 94, 311-323.

Crookston, N.L., Finley, A.O. (2008) yaImpute: an R package for kNN imputation. Journal of Statistical Software. 23 (10). 16 p.

Dabros, A., Hammond, H.J., Pinzon, J., Pinno, B., Langor, D. (2017) Edge influence of low-impact seismic lines for oil exploration on upland forest vegetation in northern Alberta (Canada). Forest Ecology and Management 400, 278-288.

Dabros, A., Higgins, K.L., Santala, K., Aubin, I. (2022) Plant functional trait approach to assess the persistence of seismic line footprint in boreal peatlands of Alberta, Canada. Forest Ecology and Management 503, 119751.

Dabros, A., Pyper, M., Castilla, G. (2018) Seismic lines in the boreal and arctic ecosystems of North America: environmental impacts, challenges, and opportunities. Environmental Reviews 26, 214-229.

Dalla Corte, A.P., Souza, D.V., Rex, F.E., Sanquetta, C.R., Mohan, M., Silva, C.A., Zambrano, A.M.A., Prata, G., de Almeida, D.R.A., Trautenmüller, J.W. (2020) Forest inventory with high-density UAV-Lidar: Machine learning approaches for predicting individual tree attributes. Computers and Electronics in Agriculture 179, 105815.

Davidson, S.J., Goud, E.M., Malhotra, A., Estey, C.O., Korsah, P., Strack, M. (2021) Linear disturbances shift boreal peatland plant communities toward earlier peak greenness. Journal of Geophysical Research: Biogeosciences 126, e2021JG006403.

Eldegard, K., Totland, Ø., Moe, S.R. (2015) Edge effects on plant communities along power line clearings. Journal of Applied Ecology 52, 871-880.

Elith, J., Leathwick, J.R., Hastie, T. (2008) A working guide to boosted regression trees. Journal of Animal Ecology 77, 802-813.

Fahrig, L. (2003) Effects of habitat fragmentation on biodiversity. Annual review of ecology, evolution, and systematics 34, 487-515.

Filicetti, A.T. (2021) Fire and tree recovery on seismic lines.

Filicetti, A.T., Cody, M., Nielsen, S.E. (2019) Caribou conservation: restoring trees on seismic lines in Alberta, Canada. Forests 10, 185.

Filicetti, A.T., Nielsen, S.E. (2018) Fire and forest recovery on seismic lines in sandy upland jack pine (Pinus banksiana) forests. Forest Ecology and Management 421, 32-39.

Filicetti, A.T., Nielsen, S.E. (2020) Tree regeneration on industrial linear disturbances in treed peatlands is hastened by wildfire and delayed by loss of microtopography. Canadian Journal of Forest Research 50, 936-945.

Finley, A.O., McRoberts, R.E. (2008) Efficient k-nearest neighbor searches for multi-source forest attribute mapping. Remote Sensing of Environment 112, 2203-2211.

Finnegan, L., MacNearney, D., Pigeon, K.E. (2018a) Divergent patterns of understory forage growth after seismic line exploration: Implications for caribou habitat restoration. Forest Ecology and Management 409, 634-652.

Finnegan, L., Pigeon, K.E., Cranston, J., Hebblewhite, M., Musiani, M., Neufeld, L., Schmiegelow, F., Duval, J., Stenhouse, G.B. (2018b) Natural regeneration on seismic lines influences movement behaviour of wolves and grizzly bears. PLoS One 13, e0195480.

Finnegan, L., Pigeon, K.E., MacNearney, D. (2019) Predicting patterns of vegetation recovery on seismic lines: Informing restoration based on understory species composition and growth. Forest Ecology and Management 446, 175-192.

Fisher, J.T., Burton, A.C. (2018) Wildlife winners and losers in an oil sands landscape. Frontiers in Ecology and the Environment 16, 323-328.

Forkuor, G., Hounkpatin, O.K., Welp, G., Thiel, M. (2017) 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 12, e0170478.

Forman, R.T. (2000) Estimate of the area affected ecologically by the road system in the United States. Conservation Biology 14, 31-35.

Franco-Lopez, H., Ek, A.R., Bauer, M.E. (2001) Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbors method. Remote Sensing of Environment 77, 251-274.

Franklin, C.M., Filicetti, A.T., Nielsen, S.E. (2021) Seismic line width and orientation influence microclimatic forest edge gradients and tree regeneration. Forest Ecology and Management 492, 119216.

Friedman, J., Hastie, T., Tibshirani, R. (2001) The elements of statistical learning. Springer series in statistics New York.

Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Friedman, J.H. (2002) Stochastic gradient boosting. Computational Statistics & Data Analysis 38, 367-378.

Gangappa, M., Kiran, C., Sammulal, P. (2017) Techniques for machine learning based spatial data analysis: research directions. Int. J. Comput. Appl 170, 9-13.

Garbasevschi, O.M., Schmiedt, J.E., Verma, T., Lefter, I., Altes, W.K.K., Droin, A., Schiricke, B., Wurm, M. (2021) Spatial factors influencing building age prediction and implications for urban residential energy modelling. Computers, Environment and Urban Systems 88, 101637.

García-Gutiérrez, J., Martínez-Álvarez, F., Troncoso, A., Riquelme, J.C. (2015) A comparison of machine learning regression techniques for LiDAR-derived estimation of forest variables. Neurocomputing 167, 24-31.

Gartzia-Bengoetxea, N., de Arano, I.M., Arias-González, A. (2021) Forest productivity and associated soil ecosystem services remain altered 15years after mechanized site preparation for reforestation with Pinus radiata. Soil and Tillage Research 213, 105150.

Geiß, C., Pelizari, P.A., Schrade, H., Brenning, A., Taubenböck, H. (2017) On the effect of spatially non-disjoint training and test samples on estimated model generalization capabilities in supervised classification with spatial features. IEEE Geoscience and Remote Sensing Letters 14, 2008-2012.

Girardin, P., Valeria, O., Girard, F. (2022) Measuring Spatial and Temporal Gravelled Forest Road Degradation in the Boreal Forest. Remote Sensing 14, 457.

Gleason, C.J., Im, J. (2012) Forest biomass estimation from airborne LiDAR data using machine learning approaches. Remote Sensing of Environment 125, 80-91.

Gosiewska, A., Biecek, P. (2018) Auditor: An R package for model-agnostic visual validation and diagnostics. arXiv preprint arXiv:1809.07763.

Gosselin, J., Berger, J.-P. (2002) 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, Forêt Québec, Direction des inventaires ….

Government of Alberta (2016) Alberta Timber Harvest Planning and Operating Ground Rules Framework for Renewal. Institution: Calgary, AB, Canada,.

Greenwell, B.M. (2017) pdp: An R package for constructing partial dependence plots. R J. 9, 421.

Groot, A., Carlson, D.W., Fleming, R.L., Wood, J.E. (1997) Small openings in trembling aspen forest: microclimate and regeneration of white spruce and trembling aspen. NODA NFP Technical Report.

Guariguata, M.R., Dupuy, J.M. (1997) Forest Regeneration in Abandoned Logging Roads in Lowland Costa Rica 1. Biotropica 29, 15-28.

Guo, Q., Su, Y., Hu, T., Guan, H., Jin, S., Zhang, J., Zhao, X., Xu, K., Wei, D., Kelly, M. (2020) LiDAR boosts 3D ecological observations and modelings: A review and perspective. IEEE Geoscience and Remote Sensing Magazine 9, 232-257.

Guo, X., Coops, N.C., Tompalski, P., Nielsen, S.E., Bater, C.W., Stadt, J.J. (2017) Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. Ecological Informatics 38, 50-61.

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. (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. Science advances 1, e1500052.

Hajjem, A., Bellavance, F., Larocque, D. (2014) Mixed-effects random forest for clustered data. Journal of Statistical Computation and Simulation 84, 1313-1328.

Hansen, A.J., Phillips, L.B., Dubayah, R., Goetz, S., Hofton, M. (2014) Regional-scale application of lidar: Variation in forest canopy structure across the southeastern US. Forest Ecology and Management 329, 214-226.

Harper, K.A., Macdonald, S.E., Burton, P.J., Chen, J., Brosofske, K.D., Saunders, S.C., Euskirchen, E.S., Roberts, D., Jaiteh, M.S., Esseen, P.A. (2005) Edge influence on forest structure and composition in fragmented landscapes. Conservation Biology 19, 768-782.

Harper, K.A., Macdonald, S.E., Mayerhofer, M.S., Biswas, S.R., Esseen, P.A., Hylander, K., Stewart, K.J., Mallik, A.U., Drapeau, P., Jonsson, B.G. (2015) Edge influence on vegetation at natural and anthropogenic edges of boreal forests in C anada and F ennoscandia. Journal of Ecology 103, 550-562.

Hart, S.A., Chen, H.Y. (2006) Understory vegetation dynamics of North American boreal forests. Critical Reviews in Plant Sciences 25, 381-397.

Hassan, M.A., Khalil, A., Kaseb, S., Kassem, M. (2017) Exploring the potential of tree-based ensemble methods in solar radiation modeling. Applied Energy 203, 897-916.

Heidari, M.J., Najafi, A., Alavi, S. (2018) Pavement deterioration modeling for forest roads based on logistic regression and artificial neural networks. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 39, 271-287.

Hong, T., Lee, M., Koo, C., Jeong, K., Kim, J. (2017) Development of a method for estimating the rooftop solar photovoltaic (PV) potential by analyzing the available rooftop area using Hillshade analysis. Applied Energy 194, 320-332.

Hornseth, M., Pigeon, K., MacNearney, D., Larsen, T., Stenhouse, G., Cranston, J., Finnegan, L. (2018) Motorized activity on legacy seismic lines: a predictive modeling approach to prioritize restoration efforts. Environmental management 62, 595-607.

Hultquist, C., Chen, G., Zhao, K. (2014) A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests. Remote sensing letters 5, 723-732.

Ilisson, T., Chen, H.Y. (2009) The direct regeneration hypothesis in northern forests. Journal of Vegetation Science 20, 735-744.

Kaakkurivaara, T., Vuorimies, N., Kolisoja, P., Uusitalo, J. (2015) Applicability of portable tools in assessing the bearing capacity of forest roads.

Kalantar, B., Ueda, N., Saeidi, V., Ahmadi, K., Halin, A.A., Shabani, F. (2020) Landslide susceptibility mapping: Machine and ensemble learning based on remote sensing big data. Remote Sensing 12, 1737.

Kansas, J.L., Charlebois, M.L., Skatter, H.G. (2015) 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. Manage 4, 137-149.

Kleinschroth, F., Healey, J.R., Sist, P., Mortier, F., Gourlet‐Fleury, S. (2016) How persistent are the impacts of logging roads on Central African forest vegetation? Journal of Applied Ecology 53, 1127-1137.

Koma, Z., Seijmonsbergen, A.C., Kissling, W.D. (2021) Classifying wetland‐related land cover types and habitats using fine‐scale lidar metrics derived from country‐wide Airborne Laser Scanning. Remote Sensing in Ecology and Conservation 7, 80-96.

Kopecký, M., Macek, M., Wild, J. (2021) Topographic Wetness Index calculation guidelines based on measured soil moisture and plant species composition. Science of the Total Environment 757, 143785.

Kosicki, J.Z. (2020) Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. Environmental and Ecological Statistics 27, 273-292.

Kosicki, J.Z. (2020) Generalised Additive Models and Random Forest Approach as effective methods for predictive species density and functional species richness. Environmental and Ecological Statistics 27, 273-292.

Kuhn, M. (2015) Caret: classification and regression training. Astrophysics Source Code Library, ascl: 1505.1003.

Lapointe, A. (2019) Road Restoration in Caribou Habitat in a Northern Québec Wildlife Reserve and Possible Impacts on Users.

Laurance, W.F., Goosem, M., Laurance, S.G. (2009) Impacts of roads and linear clearings on tropical forests. Trends in ecology & evolution 24, 659-669.

Lay, U.S., Pradhan, B., Yusoff, Z.B.M., Abdallah, A.F.B., Aryal, J., Park, H.-J. (2019) Data mining and statistical approaches in debris-flow susceptibility modelling using airborne LiDAR data. Sensors 19, 3451.

Lázaro-Lobo, A., Ervin, G.N. (2019) A global examination on the differential impacts of roadsides on native vs. exotic and weedy plant species. Global ecology and conservation 17, e00555.

Leathwick, J., Elith, J., Hastie, T. (2006) Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modelling 199, 188-196.

Lee, P., Boutin, S. (2006) Persistence and developmental transition of wide seismic lines in the western Boreal Plains of Canada. Journal of Environmental Management 78, 240-250.

Leitold, V., Morton, D.C., Martinuzzi, S., Paynter, I., Uriarte, M., Keller, M., Ferraz, A., Cook, B.D., González, G. (2021) Tracking the rates and mechanisms of canopy damage and recovery following Hurricane Maria using multitemporal lidar data. bioRxiv.

Liaw, A., Wiener, M. (2002a) Classification and regression by randomForest. R news 2, 18-22.

Liaw, A., Wiener, M. (2002b) The randomforest package. R news 2, 18-22.

Liu, D., (2020) Work-in-Process Decision Support System with Predictive Modeling in the Food Manufacturing Industry. The George Washington University.

Liu, R., Yang, X., Xu, C., Wei, L., Zeng, X. (2022) Comparative Study of Convolutional Neural Network and Conventional Machine Learning Methods for Landslide Susceptibility Mapping. Remote Sensing 14, 321.

Liu, Z., Peng, C., Work, T., Candau, J.-N., DesRochers, A., Kneeshaw, D. (2018) Application of machine-learning methods in forest ecology: recent progress and future challenges. Environmental Reviews 26, 339-350.

Lovitt, J., Rahman, M.M., Saraswati, S., McDermid, G.J., Strack, M., Xu, B. (2018) UAV remote sensing can reveal the effects of low‐impact seismic lines on surface morphology, hydrology, and methane (CH4) release in a boreal treed bog. Journal of Geophysical Research: Biogeosciences 123, 1117-1129.

Mahon, C.L., Holloway, G.L., Bayne, E.M., Toms, J.D. (2019) Additive and interactive cumulative effects on boreal landbirds: winners and losers in a multi‐stressor landscape. Ecological Applications 29, e01895.

Malcolm, J.R., Ray, J.C. (2000) Influence of timber extraction routes on central African small‐mammal communities, forest structure, and tree diversity. Conservation Biology 14, 1623-1638.

Mansuy, N., Gauthier, S., Robitaille, A., Bergeron, Y. (2012) Regional patterns of postfire canopy recovery in the northern boreal forest of Quebec: interactions between surficial deposit, climate, and fire cycle. Canadian Journal of Forest Research 42, 1328-1343.

Martin, M., Cerrejón, C., Valeria, O. (2021) Complementary airborne LiDAR and satellite indices are reliable predictors of disturbance-induced structural diversity in mixed old-growth forest landscapes. Remote Sensing of Environment 267, 112746.

Martin, M., Orton, T., Lacarce, E., Meersmans, J., Saby, N., Paroissien, J., Jolivet, C., Boulonne, L., Arrouays, D. (2014) Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma 223, 97-107.

Martin, M., Valeria, O. (2022) “Old” is not precise enough: Airborne laser scanning reveals age-related structural diversity within old-growth forests. Remote Sensing of Environment 278, 113098.

Martínez‐Santos, P., Renard, P. (2020) Mapping groundwater potential through an ensemble of big data methods. Groundwater 58, 583-597.

Martinuzzi, S., Vierling, L.A., Gould, W.A., Falkowski, M.J., Evans, J.S., Hudak, A.T., Vierling, K.T. (2009) Mapping snags and understory shrubs for a LiDAR-based assessment of wildlife habitat suitability. Remote Sensing of Environment 113, 2533-2546.

Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Bolton, D.K., Tompalski, P., Bater, C.W. (2018a) Three decades of forest structural dynamics over Canada's forested ecosystems using Landsat time-series and lidar plots. Remote Sensing of Environment 216, 697-714.

Matasci, G., Hermosilla, T., Wulder, M.A., White, J.C., Coops, N.C., Hobart, G.W., Zald, H.S. (2018b) Large-area mapping of Canadian boreal forest cover, height, biomass and other structural attributes using Landsat composites and lidar plots. Remote Sensing of Environment 209, 90-106.

Maxwell, A.E., Warner, T.A., Fang, F. (2018) Implementation of machine-learning classification in remote sensing: An applied review. International Journal of Remote Sensing 39, 2784-2817.

McRoberts, R.E. (2012) Estimating forest attribute parameters for small areas using nearest neighbors techniques. Forest Ecology and Management 272, 3-12.

Mercier, P., Aas, G., Dengler, J. (2019) Effects of skid trails on understory vegetation in forests: a case study from Northern Bavaria (Germany). Forest Ecology and Management 453, 117579.

Meyer, H., Pebesma, E. (2021) Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution 12, 1620-1633.

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T. (2018) Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental Modelling & Software 101, 1-9.

Ministère de l'Environnement et de la Lutte contre les changements climatiques, (1999) Les provinces naturelles. Niveau I du cadre écologique de référence du Québec [Natural Province. Level I of the Quebec reference ecological framework]. Ministère des Ressources naturelles et des Forêts, Quebec City, QC.

Ministère des Ressources naturelles et des Forêts, (2020) Guide d’utilisation des produits dérivés du LiDAR, Québec, Canada. Ministère des Ressources naturelles et des Forêts.

Ministère des Ressources naturelles et des Forêts, (2021a) Guide D’application du Règlement sur L’aménagement Durable des Forêts du Domaine de L’état; Institution: Québec, QC, Canada. Ministère des Ressources naturelles et des Forêts.

Ministère des Ressources naturelles et des Forêts, (2021b) Guide D'application du Règlement sur L'aménagement Durable des Forêts du Domaine de L'état; Institution: Québec, QC, Canada

Annexe 4 – Caractéristiques des chemins selon leur classement. Ministère des Ressources naturelles et des Forêts.

Ministère des Ressources naturelles et des Forêts, (2022) Métadonnées des acquisitions LiDAR. Ministère des Ressources naturelles et des Forêts.

Moisen, G.G., Frescino, T.S. (2002) Comparing five modelling techniques for predicting forest characteristics. Ecological Modelling 157, 209-225.

Molnar, C., (2020) Interpretable machine learning. Lulu. com.

Moreau, G., Fortin, D., Couturier, S., Duchesne, T. (2012) Multi‐level functional responses for wildlife conservation: the case of threatened caribou in managed boreal forests. Journal of Applied Ecology 49, 611-620.

Morellos, A., Pantazi, X.-E., Moshou, D., Alexandridis, T., Whetton, R., Tziotzios, G., Wiebensohn, J., Bill, R., Mouazen, A.M. (2016) Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy. Biosystems Engineering 152, 104-116.

Moudrý, V., Moudrá, L., Barták, V., Bejček, V., Gdulová, K., Hendrychová, M., Moravec, D., Musil, P., Rocchini, D., Šťastný, K. (2021) The role of the vegetation structure, primary productivity and senescence derived from airborne LiDAR and hyperspectral data for birds diversity and rarity on a restored site. Landscape and urban planning 210, 104064.

Musetta-Lambert, J.L., Enanga, E.M., Teichert, S., Creed, I.F., Kidd, K.A., Kreutzweiser, D.P., Sibley, P.K. (2019) Industrial innovation and infrastructure as drivers of change in the Canadian boreal zone1. Environmental Reviews 27, 275-294.

Næsset, E. (2002) Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. Remote Sensing of Environment 80, 88-99.

Natekin, A., Knoll, A. (2013) Gradient boosting machines, a tutorial. Frontiers in neurorobotics 7, 21.

Nawar, S., Mouazen, A.M. (2017) 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 17, 2428.

Norden, N., Chazdon, R.L., Chao, A., Jiang, Y.H., Vílchez‐Alvarado, B. (2009) Resilience of tropical rain forests: tree community reassembly in secondary forests. Ecology Letters 12, 385-394.

O'mahony, M., Ueberschaer, A., Owende, P., Ward, S. (2000) Bearing capacity of forest access roads built on peat soils. Journal of terramechanics 37, 127-138.

Olander, L.P., Scatena, F., Silver, W.L. (1998) Impacts of disturbance initiated by road construction in a subtropical cloud forest in the Luquillo Experimental Forest, Puerto Rico. Forest Ecology and Management 109, 33-49.

Oliver, C.D., Larson, B.A. (1996) Forest stand dynamics. Formerly published by John Wiley & Sons; Copyright now held by CD Oliver ….

Pasher, J., Seed, E., Duffe, J. (2013) Development of boreal ecosystem anthropogenic disturbance layers for Canada based on 2008 to 2010 Landsat imagery. Canadian Journal of Remote Sensing 39, 42-58.

Pattison, C.A., Quinn, M.S., Dale, P., Catterall, C.P. (2016) The landscape impact of linear seismic clearings for oil and gas development in boreal forest. Northwest Science 90, 340-354.

Perez-Luque, A.J., Benito, B.M., Bonet-Garcia, F.J., Zamora, R. (2020) Ecological Diversity within Rear-Edge: A Case Study from Mediterranean Quercus pyrenaica Willd. Forests 12, 10.

Piedallu, C., Gégout, J.-C. (2007) Multiscale computation of solar radiation for predictive vegetation modelling. Annals of forest science 64, 899-909.

Pigeon, K.E., Anderson, M., MacNearney, D., Cranston, J., Stenhouse, G., Finnegan, L. (2016) Toward the restoration of caribou habitat: understanding factors associated with human motorized use of legacy seismic lines. Environmental management 58, 821-832.

Pinard, M., Barker, M., Tay, J. (2000) Soil disturbance and post-logging forest recovery on bulldozer paths in Sabah, Malaysia. Forest Ecology and Management 130, 213-225.

Poggio, L., Simonetti, E., Gimona, A. (2018) Enhancing the WorldClim data set for national and regional applications. Science of the Total Environment 625, 1628-1643.

Pohjankukka, J., Pahikkala, T., Nevalainen, P., Heikkonen, J. (2017) Estimating the prediction performance of spatial models via spatial k-fold cross validation. International Journal of Geographical Information Science 31, 2001-2019.

Pradhan, P., Setyawan, A.D. (2021) Filtering multi-collinear predictor variables from multi-resolution rasters of WorldClim 2.1 for Ecological Niche Modeling in Indonesian context. Asian Journal of Forestry 5. Gangappa

Prasad, A.M., Iverson, L.R., Liaw, A. (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9, 181-199.

R Core Team, (2020) R: A language and environment for statistical computing; 2020.

Ray, J.C. (2014) Defining habitat restoration for boreal caribou in the context of national recovery: a discussion paper. Prep. for Environment and Climate Change Canada. Prep. by Wildlife Conservation Society Canada, Toronto, ON. 51p.

Revel, R.D., Dougherty, T.D., Downing, D.J. (1984) Forest growth and revegetation along seismic lines.

Rioux, J. (2018) Nouvelle approche opérationnelle de récolte forestière et de déploiement du réseau routier à l'essai dans l'habitat du caribou forestier.

Roberts, D., Ciuti, S., Barber, Q.E., Willier, C., Nielsen, S.E. (2018) Accelerated seed dispersal along linear disturbances in the Canadian oil sands region. Scientific reports 8, 1-9.

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. (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography 40, 913-929.

Robitaille, A., Saucier, J. (1998) Paysages régionaux du Québec méridional. Direction de la gestion des stocks forestiers et Direction des relations publiques, Ministère des Ressources Naturelles du Québec. Publication du Québec, Québec.

Rossi, S., Cairo, E., Krause, C., Deslauriers, A. (2015) Growth and basic wood properties of black spruce along an alti-latitudinal gradient in Quebec, Canada. Annals of forest science 72, 77-87.

Rummer, B., Wear, D. (2002) Forest operations technology. Southern Forest Resource Assessment. General Technical Report. SRS-53, USDA-Forest Service, Southern Research Station, Asheville, NC, 341-353.

Schönauer, M. (2022) Prediction of forest soil trafficability by topography-based algorithms and in-situ test procedures.

Schratz, P., Muenchow, J., Iturritxa, E., Richter, J., Brenning, A. (2019) Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data. Ecological Modelling 406, 109-120.

Senf, C., Müller, J., Seidl, R. (2019) Post-disturbance recovery of forest cover and tree height differ with management in Central Europe. Landscape Ecology 34, 2837-2850.

Shin, J., Temesgen, H., Strunk, J.L., Hilker, T. (2016) Comparing modeling methods for predicting forest attributes using LiDAR metrics and ground measurements. Canadian Journal of Remote Sensing 42, 739-765.

St-Laurent, M.-H., Beauchesne, D., Lesmerises, F., (2014) Évaluation des impacts des vieux chemins forestiers et des modalités de fermeture dans un contexte de restauration de l'habitat du caribou forestier au Québec. Rapport scientifique présenté au ministère des Forêts, de la Faune et des ….

St-Pierre, F., Drapeau, P., St-Laurent, M.-H. (2021) Drivers of vegetation regrowth on logging roads in the boreal forest: implications for restoration of woodland caribou habitat. Forest Ecology and Management 482, 118846.

Startsev, A., McNabb, D. (2009) Effects of compaction on aeration and morphology of boreal forest soils in Alberta, Canada. Canadian Journal of Soil Science 89, 45-56.

Stern, E.R., Riva, F., Nielsen, S.E. (2018) Effects of narrow linear disturbances on light and wind patterns in fragmented boreal forests in northeastern Alberta. Forests 9, 486.

Stevenson, C.J., Filicetti, A.T., Nielsen, S.E. (2019) High precision altimeter demonstrates simplification and depression of microtopography on seismic lines in treed peatlands. Forests 10, 295.

Stojanova, D., Panov, P., Gjorgjioski, V., Kobler, A., Džeroski, S. (2010) Estimating vegetation height and canopy cover from remotely sensed data with machine learning. Ecological Informatics 5, 256-266.

Sun, C., Beirne, C., Burgar, J.M., Howey, T., Fisher, J.T., Burton, A.C. (2021) Simultaneous monitoring of vegetation dynamics and wildlife activity with camera traps to assess habitat change. Remote Sensing in Ecology and Conservation 7, 666-684.

Swanson, M.E., Franklin, J.F., Beschta, R.L., Crisafulli, C.M., DellaSala, D.A., Hutto, R.L., Lindenmayer, D.B., Swanson, F.J. (2011) The forgotten stage of forest succession: early - successional ecosystems on forest sites. Frontiers in Ecology and the Environment 9, 117-125.

Tarpley, J., Schneider, S., Money, R. (1984) Global vegetation indices from the NOAA-7 meteorological satellite. Journal of Climate and Applied Meteorology, 491-494.

Thompson, I., Mackey, B., McNulty, S., Mosseler, A., (2009) Forest resilience, biodiversity, and climate change, Secretariat of the Convention on Biological Diversity, Montreal. Technical Series no. 43. 1-67., pp. 1-67.

Toivio, J., Helmisaari, H.-S., Palviainen, M., Lindeman, H., Ala-Ilomäki, J., Sirén, M., Uusitalo, J. (2017) Impacts of timber forwarding on physical properties of forest soils in southern Finland. Forest Ecology and Management 405, 22-30.

Torre-Tojal, L., Bastarrika, A., Boyano, A., Lopez-Guede, J.M., Graña, M. (2022) Above-ground biomass estimation from LiDAR data using random forest algorithms. Journal of Computational Science 58, 101517.

van Rensen, C.K., Nielsen, S.E., White, B., Vinge, T., Lieffers, V.J. (2015) Natural regeneration of forest vegetation on legacy seismic lines in boreal habitats in Alberta's oil sands region. Biological Conservation 184, 127-135.

Venier, L., Thompson, I., Fleming, R., Malcolm, J., Aubin, I., Trofymow, J., Langor, D., Sturrock, R., Patry, C., Outerbridge, R. (2014) Effects of natural resource development on the terrestrial biodiversity of Canadian boreal forests. Environmental Reviews 22, 457-490.

Venier, L.A., Swystun, T., Mazerolle, M.J., Kreutzweiser, D.P., Wainio-Keizer, K.L., McIlwrick, K.A., Woods, M.E., Wang, X. (2019) Modelling vegetation understory cover using LiDAR metrics. PLoS One 14, e0220096.

Vepakomma, U., Kneeshaw, D.D., De Grandpré, L. (2018) Influence of natural and anthropogenic linear canopy openings on forest structural patterns investigated using lidar. Forests 9, 540.

Vepakomma, U., St-Onge, B., Kneeshaw, D. (2011) Response of a boreal forest to canopy opening: assessing vertical and lateral tree growth with multi-temporal lidar data. Ecological Applications 21, 99-121.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., Clawges, R.M. (2008) Lidar: shedding new light on habitat characterization and modeling. Frontiers in Ecology and the Environment 6, 90-98.

Vu, D.T., Tran, X.-L., Cao, M.-T., Tran, T.C., Hoang, N.-D. (2020) Machine learning based soil erosion susceptibility prediction using social spider algorithm optimized multivariate adaptive regression spline. Measurement 164, 108066.

Waga, K. (2021) Unpaved forest road quality assessment using airborne LiDAR data. methods 226, 354-364.

Waga, K., Tompalski, P., Coops, N.C., White, J.C., Wulder, M.A., Malinen, J., Tokola, T. (2020) Forest road status assessment using airborne laser scanning. Forest Science 66, 501-508.

Weiskittel, A.R., Crookston, N.L., Radtke, P.J. (2011) Linking climate, gross primary productivity, and site index across forests of the western United States. Canadian Journal of Forest Research 41, 1710-1721.

Weltz, M.A., Ritchie, J.C., Fox, H.D. (1994) Comparison of laser and field measurements of vegetation height and canopy cover. Water Resources Research 30, 1311-1319.

White, J.C., Saarinen, N., Kankare, V., Wulder, M.A., Hermosilla, T., Coops, N.C., Pickell, P.D., Holopainen, M., Hyyppä, J., Vastaranta, M. (2018) Confirmation of post-harvest spectral recovery from Landsat time series using measures of forest cover and height derived from airborne laser scanning data. Remote Sensing of Environment 216, 262-275.

Whitehurst, A.S., (2014) Assessing the relationships between vertical structure, biodiversity, and succession in a forest ecosystem using lidar remote sensing. University of Maryland, College Park.

WorldClim (2017) WorldClim version 2.

Wulder, M.A., White, J.C., Bater, C.W., Coops, N.C., Hopkinson, C., Chen, G. (2012) Lidar plots—A new large-area data collection option: Context, concepts, and case study. Canadian Journal of Remote Sensing 38, 600-618.

Yang, L., Liang, S., Zhang, Y. (2020) A new method for generating a global forest aboveground biomass map from multiple high-level satellite products and ancillary information. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 2587-2597.

Zang, R., Ding, Y. (2009) Forest recovery on abandoned logging roads in a tropical montane rain forest of Hainan Island, China. Acta Oecologica 35, 462-470.

Zenner, E.K., Fauskee, J.T., Berger, A.L., Puettmann, K.J. (2007) Impacts of skidding traffic intensity on soil disturbance, soil recovery, and aspen regeneration in north central Minnesota. Northern Journal of Applied Forestry 24, 177-183.

Zhang, Y., Haghani, A. (2015) A gradient boosting method to improve travel time prediction. Transportation Research Part C: Emerging Technologies 58, 308-324.

Zhang, Y., Ma, J., Liang, S., Li, X., Li, M. (2020) An evaluation of eight machine learning regression algorithms for forest aboveground biomass estimation from multiple satellite data products. Remote Sensing 12, 4015.

Zhao, K., Popescu, S., Meng, X., Pang, Y., Agca, M. (2011) Characterizing forest canopy structure with lidar composite metrics and machine learning. Remote Sensing of Environment 115, 1978-1996.

Zhou, T., Luo, X., Hou, Y., Xiang, Y., Peng, S. (2020) Quantifying the effects of road width on roadside vegetation and soil conditions in forests. Landscape Ecology 35, 69-81.

Zuur, A.F., Ieno, E.N., Elphick, C.S. (2010) A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1, 3-14.