

RAPORT

CARTOGRAPHIE NUMÉRIQUE DE PROPRIÉTÉS DE SOLS, CAS DE
L'ÉPAISSEUR DE LA COUCHE ORGANIQUE

CONTRAT DE SERVICES ABRÉGÉ

MINISTÈRE DES FORÊTS, DE LA FAUNE ET DES PARCS

No
2021-142332162-1

Par :

Oswaldo Valeria, Julien Beguin, Carlos Cerrejon et Jean Daniel Sylvain

Mars 2023

Avant propos

La Direction de la recherche forestière (DRF) poursuit des travaux de recherche portant sur le «Développement d'un système d'inférence et d'information géographique sur propriétés des sols forestiers au Québec» (projet 142332118). Le présent mandat vise à supporter les travaux de l'équipe de modélisation pour la cartographie des propriétés physiques et chimiques des sols et de l'incertitude associée dans le cadre d'une collaboration entre la Direction de la recherche forestière, l'Université du Québec en Abitibi-Témiscamingue et une équipe de recherche du centre de foresterie des Laurentides. Les travaux issus de ce projet permettront de compléter les approches utilisées dans le cadre du projet Développement d'un système d'information et d'inférence sur les sols forestiers (142332118). Les travaux seront réalisés par des étudiants gradués sous la supervision d'Oswaldo Valeria de l'institut de recherche sur les forêts et d'un chercheur de la DRF.

Table des matières

Avant propos.....	2
Introduction.....	4
Matériels et méthodes	7
2.1. Zone d'étude.....	7
2.2. Données de l'épaisseur de la couche organique.....	10
2.3. Covariables environnementales.....	10
2.4. Modélisation statistique.....	12
2.5. Validation.....	12
Résultats.....	14
Discussion.....	16
Mise en garde.....	17
Remerciements.....	17
Références.....	18

Introduction

Les sols fournissent des services écosystémiques (Greiner et al., 2017) essentiels pour le développement économique et social du Québec. En milieu forestier, ces services incluent la productivité et la croissance des forêts, la séquestration de carbone, la résistance des forêts à la sécheresse, la qualité de l'eau et de l'air, la disponibilité en matériel granulaire pour la construction de chemins forestiers, l'immobilisation de la pollution et autres contaminants, ainsi qu'un rôle majeur de réserve de biodiversité (Adhikari et Hartemink 2016). Cartographier avec précision les propriétés physico-chimiques des sols forestiers est donc une priorité à la fois économique et environnementale et une étape incontournable si on souhaite quantifier spatialement l'influence des sols sur certains services écosystémiques clefs comme la productivité et la croissance des forêts. Les méthodes traditionnelles de cartographie des sols n'ont toutefois pas été conçues pour assurer une représentation quantitative et matricielle des propriétés physico-chimiques du sol au niveau nationale et provinciale (CanSIS - <https://sis.agr.gc.ca/cansis/index.html>, <https://www.donneesquebec.ca/recherche/fr/dataset/depots-de-surface>) et de leur incertitude. Par ailleurs, la majorité des cartes des propriétés du sol actuellement disponibles ont été développées principalement en milieu agricole ce qui limite leur utilisation dans de nombreuses applications environnementales (Arrouays et al., 2014, Arrouays et al., 2017) et notamment dans le cadre de la planification des ressources naturelles du territoire. Ce manque d'information spatiale détaillée sur les sols forestiers est un frein qui limite actuellement le développement de stratégies d'aménagement forestier durable innovantes qui permettraient de prendre en compte l'effet des propriétés des substrats de croissance sur la productivité et la résilience des écosystèmes forestiers à travers le territoire.

La cartographie numérique des sols (CNS) (en anglais: *digital soil mapping*) consiste à générer des bases de données pédologiques géographiquement référencées, basées sur des relations quantitatives entre des données pédologiques ponctuelles, analytiques et géoréférencées (pedons) et des données environnementales matricielles (covariables) généralement extraites à partir de données de télédétection et de modèles numériques d'élévation. Des relations mathématiques (fonctions) sont ensuite appliquées afin de

générer un portrait matriciel en 3D de la variabilité spatiale des propriétés du sol à différentes profondeurs (McBratney et al., 2003). Les différentes étapes de la CNS sont détaillées dans (Bulmer et al., 2019).

Les développements en CNS connaissent un essor formidable depuis deux décennies grâce à l'émergence de méthodes statistiques innovantes et en particulier des algorithmes d'apprentissage automatisé. Par ailleurs, la disponibilité croissante des bases de données historiques géoréférencées et des données de télédétection contribue également à maximiser l'intérêt pour le développement de nouvelles approches en CNS [5]. De nombreux pays et continents (ex: Australie (Grundy et al., 2015), États-Unis (Ramcharan et al., 2018), Union européenne (Ballabio et al., 2016), Amérique latine (Guevara et al., 2018), (Vågen et al., 2016)) ont investi massivement dans des programmes d'observation et de suivi des propriétés des sols, ainsi que dans la modélisation spatiale de ces propriétés afin de rendre disponibles ces informations auprès des organisations gouvernementales et non-gouvernementales impliquées dans l'aménagement et l'utilisation des terres. Cependant, hors des zones agricoles, le développement de la CNS au Québec est resté modeste et fragmentaire jusqu'à ce jour. La plupart des études réalisées à l'échelle nationale ont été menées par le Service canadien des forêts (Mansuy et al., 2014, Beguin et al., 2017, Paré et al., 2021) et Agriculture Canada (Geng et al., 2010) alors qu'à l'échelle provinciale, seules quelques provinces ont investi dans un système moderne de CNS (e.g. Québec (Sylvain et al., 2021); New Brunswick (Furze et al., 2018); île du Prince Édouard (Nyiraneza et al., 2017)). Aux échelles régionales et locales, plusieurs initiatives ont démontré la faisabilité et l'importance de développer la cartographie quantitative des propriétés des sols forestiers (Akumu et al., 2015, Heung et al., 2014, Scarpone et al., 2016, Mansuy et al., 2018, Blackford et al., 2021). Ces études menées à différentes échelles spatiales ont notamment permis de mettre en lumière les écarts importants qui existent actuellement entre la qualité des cartes de propriétés des sols forestiers produites au Canada (Mansuy et al., 2014, Beguin et al., 2017, Sylvain et al., 2021) et les mêmes types de cartes prédictives générés par des modèles globaux (Hengl et al., 2017, Hengl et al., 2014). La faible performance, sur le territoire québécois, des produits issus des modèles globaux démontre la nécessité de développer une cartographie spécifique pour le Québec, et ce à

différentes échelles spatiales afin d'être en adéquation avec la nature des enjeux et des besoins.

Développer une foresterie durable, agile et résiliente aux changements environnementaux nécessite des données quantitatives spatialement explicites et à jour (non disponibles actuellement) qui décrivent précisément les conditions biophysiques (végétation + sols) à une échelle spatiale adaptée aux différents enjeux (échelle locale, régionale, provinciale et nationale). Ces dernières décennies, une emphase importante a été mise sur la compréhension et la quantification du rôle joué par les conditions climatiques et les régimes de perturbation sur le fonctionnement des écosystèmes forestiers (D'Orangeville et al., 2018, D'Orangeville et al., 2016). Toutefois, la réponse des arbres aux stress environnementaux dépend également des conditions locales des sites sur lesquels les arbres se trouvent (Marchand et al., 2019) et plusieurs études ont mis en évidence l'importance du rôle joué par les propriétés du sol sur la croissance forestière (Lévesque et al., 2016), Cienciala et al., 2016, Simard et al., 2007, Henneb et al., 2019).

L'objectif général de ce projet visait notamment à explorer le potentiel de différentes approches de modélisation issues du domaine de l'apprentissage automatisé pour la cartographie de la variabilité spatiale de l'épaisseur de la matière organique. La cartographie de l'épaisseur de la matière organique permettrait de bonifier les modèles de croissance utilisés pour simuler la croissance forestière et de mieux estimer le potentiel de séquestration de carbone dans les sols du Québec.

Ce rapport présente d'abord une description sommaire des données et des éléments méthodologiques utilisés pour cartographier l'épaisseur de la matière organique. Il présente ensuite les résultats obtenus en décrivant la performance des modèles ainsi que les cartes prédites et les incertitudes associées sur différents jeux de données. Finalement, les bénéfices et les limites de l'approche proposée, ainsi que les perspectives qu'offre la CNS à court terme sont abordés.

Matériels et méthodes

2.1. Zone d'étude

La zone d'étude d'environ 200 000 km² est située dans la région de la forêt boréale de l'ouest du Québec au Canada (Figure 1), dans l'écozone du bouclier boréal (Ecological Stratification Working Group, 1996). La partie nord est dominée par la pessière noire à mousse avec l'épinette noire (*Picea mariana* (Mill.) B.S.P.) comme principale espèce dominante et le pin gris (*Pinus banksiana* Lamb.) comme espèce associée. La partie sud de la région étudiée est principalement composée de peuplements mixtes dominés par le sapin baumier (*Abies balsamea* (L.) Mill), le bouleau jaune (*Betula alleghaniensis* Britt.) et le peuplier faux-tremble (*Populus tremuloides* Michx.). La paludification du sol dans cette région est principalement due à l'accumulation de sphaignes et d'éricacées (*Rhododendron groenlandicum*, *Kalmia angustifolia*, *Chamaedaphne calyculata* et *Vaccinium* spp.) (DeLuca et al., 2002, Thiffault et al., 2015). La température et les précipitations annuelles moyennes varient entre -2,4 C et 1,1C, et entre 727 mm et 930 mm (Environnement Canada, 2012: adapté de Mansuy et al., 2018). Les dépôts de surface sont dominés par des matériaux glaciaires, fluvio-glaciaires et marins (Soil Classification Working Group, 1998), dont la majorité est constituée d'argile compactée et de sable et gravier recouverts de tourbe (Veillete, 1994). Un ancien lac proglaciaire (le lac Barlow-Ojibway) a laissé un épais dépôt d'argile, formant l'unité physiographique connue aujourd'hui sous le nom de Ceinture d'argile (*Clay belt*), qui s'étend de part et d'autre de la frontière Québec-Ontario et couvre une superficie de 145 470 km² (Vincent et Hardy 1977). La topographie est plate, mais de petites collines rocheuses sont présentes.

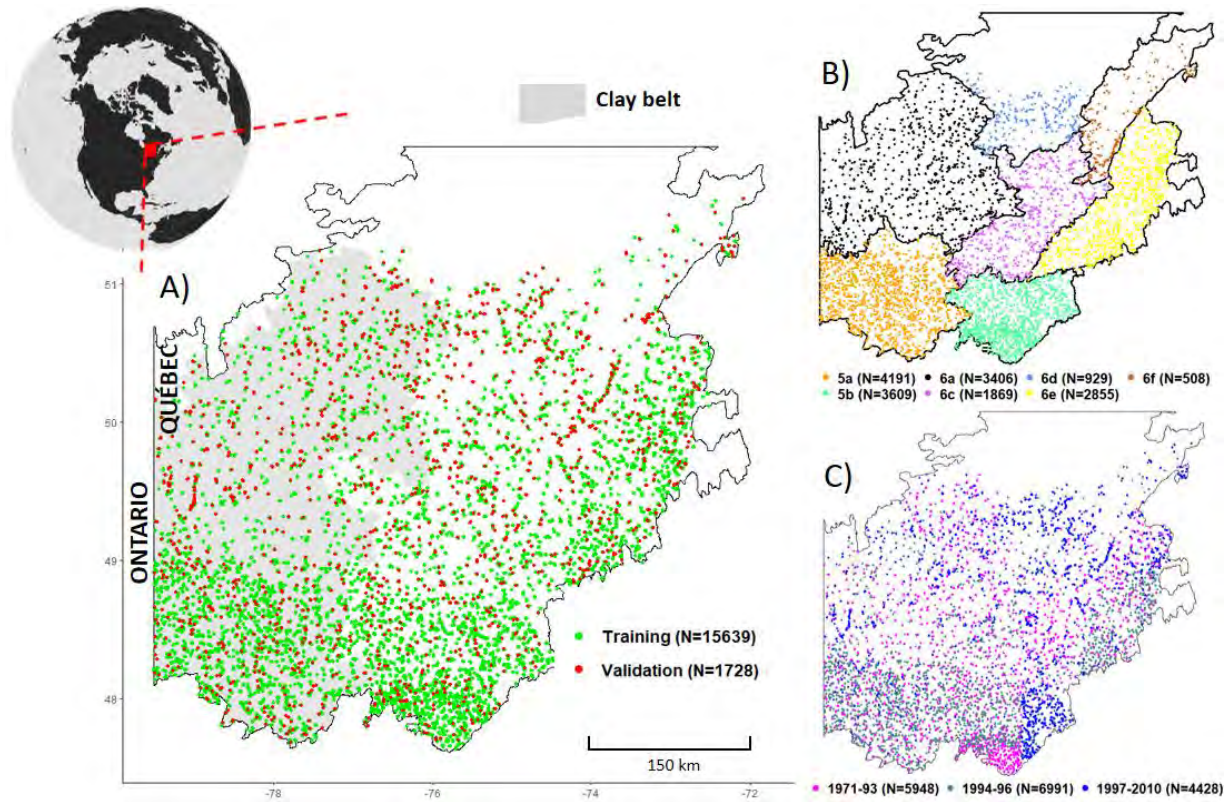


Figure 1. A) Aire d'étude et distribution spatiotemporelle des points d'échantillonnage selon B) les régions écologiques identifiées par le code alpha-numérique (blocs spatiaux) et C) l'année d'échantillonnage (regroupées en blocs temporels). N représente la taille de l'échantillon. Données d'entraînement (Training) et données de validation (validation).

Table 1. Les variables que nous avons utilisées dans nos modèles, les groupes auxquels elles appartiennent, les types de variables et les statistiques sommaires des variables continues. Le détail des variables est présenté dans le texte.

Variable	Variable groupe	Type de variable	Unité	Min.	Médiane	Moyenne	Max.	Std	Source	
<i>Variable réponse</i>										
OLT		Continue	centimètre	0.00	12.00	21.27	150.00	23.92	OLT échantillonnage	
<i>Variables explicatives</i>										
Latitude	Topo-Géographie	Continue	dégré	47.63			51.61		OLT Sample	
Longitude		Continue	dégré	-79.51			-72.08		OLT Sample	
Élévation		Continue	mètre	170	371.50	371.46	1156	72.42	DEM	
Pente		Continue	dégré	0	2.90	3.66	23.75	3.24	DEM	
Aspect		Continue	dégré	0	158.46	159.45	350.56	105.21	DEM	
S_WTNS	Indices de végétation	Continue		-145.29	-26.95	-35.84	103.3	37.84	Sentinel (2016-2018)	
Amp_HH		Continue		2327	6877	6763.07	10762	987.64	Radar	
NDVI		Continue		0,43	0.73	0.74	0.87	0.07	Landsat 7 (1985-2018)	
SAVI		Continue		1,62	2.75	2.78	3.27	0.26	Landsat 7 (1985-2018)	
MSAVI		Continue		0,6	0.85	0.85	0.93	0.05	Landsat 7 (1985-2018)	
Coupe_distance		Continue	mètre	7.11	598.86	1725.19	40199.10	5186.22	ForêtOuverte	
Feu_distance		Continue	mètre	0	3900.89	5287.12	27978.80	5527.66	ForêtOuverte	
Précipitation Mars	Climat	Continue	mm	21.54	38.39	38.59	47.33	3.42	BioSim	
Chute de neige Avril		Continue	mm	15.81	27.41	27.31	34.17	2.46	BioSim	
Dépôt surface	Condition écologique	Catégorique	classe (6)	1A, 2A, 2B, 4G, 7E and 7T						carte écoforestière
Drainage		Catégorique	classe (6)	1, 2, 3, 4, 5, 6						carte écoforestière
Type écologique		Catégorique	classe (7)	ME1, MS2, MS6, RE2, RE3, RS2 and RS3						carte écoforestière
Classe d'âge		Catégorique	classe (5)	Young, Mature, Old, Mixed, Other						carte écoforestière
Groupe d'espèces		Catégorique	classe (5)	Fi, Ft, Ri, Rt, SB						carte écoforestière

Les feux de forêt et les épidémies de tordeuse des bourgeons de l'épinette sont les deux principaux régimes de perturbation naturelle présents dans la région (Mansuy et al., 2012, Morin et al., 1993). L'exploitation forestière intensive, la principale perturbation anthropique actuelle, a débuté au cours des années 1970 (Belleau et Légaré, 2009) et ~80% de la zone d'étude est située dans la zone de forêt commerciale.

2.2. Données de l'épaisseur de la couche organique

Les données sur l'épaisseur de la couche organique (en cm) proviennent de plusieurs programmes d'inventaire forestier menés sur le terrain entre 1971 et 2010 et des projets de recherche expérimentaux menés en collaboration avec l'UQAT dans sept régions écologiques distinctes (5a, 5b, 6a, 6c, 6d, 6e et 6f) situées dans la zone d'étude (Figure 1). Ces programmes d'inventaire forestier comprennent 1) des parcelles d'inventaire forestier permanent; 2) des parcelles d'inventaire forestier temporaire; 3) des points d'observation écologique et 4) des données expérimentales pour un total de 17367 points de données géoréférencés situés dans des forêts commerciales. Pour permettre des comparaisons significatives dans le temps, nous avons divisé la période de l'étude en trois périodes distinctes de sorte que chaque période contienne un nombre équilibré de parcelles d'échantillonnage, ce qui conduit finalement aux périodes suivantes : 1) avant 1994 (N = 5948), 2) de 1994 à 1996 (N = 6991), et 3) de 1997 à 2010 (N = 4428) (Figure 1). Dans la base de données finale, les données de mesure des échantillons d'épaisseur de couche organique variaient entre zéro et 150 cm avec une moyenne de 21 cm (Tableau 1).

2.3. Covariables environnementales

Nous avons sélectionné sept classes de variables environnementales pour prédire et cartographier l'épaisseur de la couche organique dans la zone d'étude: (a) variables topographiques dérivées du modèle numérique d'élévation (MNE) à 30 m (2019 ; <https://www.foretouverte.gouv.qc.ca>), (b) variables climatiques à 30 m dérivées de BioSim (Régnière, et al., 2013), (c) variables de végétation dérivées de la télédétection (Landsat-7: *U.S. Geological Survey* <https://www.usgs.gov> et Radar ; <https://registry.opendata.aws/radarsat-1>), (d) variables de végétation dérivées de cartes forestières (2019 ; 4e

programme d'inventaire forestier ; <https://www.foretouverte.gouv.qc.ca>), (e) variables de site dérivées de cartes forestières (2019; 4e programme d'inventaire forestier; <https://www.foretouverte.gouv.qc.ca>), (f) dépôt de surface dérivé de cartes forestières (2019 ; 4e programme d'inventaire forestier ; <https://www.foretouverte.gouv.qc.ca>), et (g) variables géographiques utilisant des coordonnées géographiques obliques (N=6) (Møller et al. 2020) (Tableau 1). Les variables de classe décrivant les régimes de perturbation et les propriétés du sol forestier disponibles dans les cartes forestières ont également été testées mais n'ont pas été retenues car elles ne contribuaient pas à prédire l'épaisseur de la couche organique.

En ce qui concerne les données de télédétection, nous avons utilisé la moyenne annuelle sur 25 ans (1985-2010) de chaque bande spectrale de Landsat 7 pour calculer les indices de végétation suivants : indice de végétation par différence normalisée (NDVI), indice de végétation ajusté au sol (SAVI), indice de végétation ajusté au sol modifié (MSAVI), indice de surface foliaire (LAI), indice de couverture de la végétation forestière (FVCI), indice de végétation amélioré (EVI), taux de combustion normalisé (NBR), taux de combustion normalisé 2 (NBR2) et température de la surface terrestre (LST) (voir USGS <https://www.usgs.gov/> pour plus de détails). En outre, nous avons utilisé des données radar et calculé trois indices (Polaryzation_HH, Polaryzation_VV et Polaryzation_HVHH).

Pour les données climatiques, nous avons utilisé des données moyennes mensuelles (BioSim ; Régnière, et al., 2014) sur la période 1971-2010 pour vingt variables climatiques différentes, y compris : la température, les précipitations, le nombre de jours dans la saison de croissance, la quantité de neige tombée, et l'humidité relative. Nous avons simulé chaque variable climatique mensuelle pour chaque point d'échantillonnage avec une mesure de l'épaisseur de la couche organique.

Pour les données sur les sites et les dépôts de surface, nous avons extrait des cartes d'inventaire forestier (4ème inventaire ; <https://mffp.gouv.qc.ca/documents/forets/inventaire/norme-stratification.pdf>) les variables suivantes: dépôt de surface, classe de drainage, classe d'âge du peuplement, groupe d'espèces et type de couverture. Pour les variables catégorielles décrivant le dépôt de surface et la classe de drainage, nous avons

reclassé les classes originales en classes plus larges afin de garantir que suffisamment de points d'échantillonnage se trouvent dans chaque classe pour permettre une validation significative et des prédictions hors échantillon fiables.

Au total, un ensemble final de plus de 300 variables environnementales individuelles à une résolution spatiale de 30 m a été constitué. Conformément au principe de parcimonie, nous avons utilisé l'algorithme d'élimination récursive des variables (Granitto et al., 2006) pour supprimer une à une toutes les covariables environnementales qui ne contribuaient pas à améliorer les erreurs de prédiction finales sur la base d'une validation croisée à 10 plis répétée 20 fois. Cette procédure nous a également permis d'éviter tout problème potentiel de sur ajustement (Beguin et al. 2017). L'ensemble final de covariables environnementales contenait huit prédicteurs (Tableau 1).

2.4. Modélisation statistique

Nous avons comparé les performances croisées de six approches statistiques (*random forest*, *quantile random forest*, *weighted k-nearest neighbors*, *cubist*, *boosted regression trees* and *bayesian spatial modelling with stochastic partial differential equations* (SPDE)) avec trois formulations de modèles (covariables environnementales uniquement; fonctions spatiales uniquement avec coordonnées géographiques obliques; et covariables environnementales avec fonctions spatiales) sur la qualité des prédictions de l'épaisseur de la couche organique (voir Beguin et al. (2017) pour plus de détails). Pour évaluer la performance de chaque combinaison d'approche statistique-formulation de modèle, nous avons calculé trois indicateurs de performance : coefficient de détermination R^2 , l'erreur quadratique moyenne (RMSE) et le coefficient de corrélation de concordance de Lin (CCC de Lin) (voir Lin 1989 pour plus de détails sur cette indicateur).

2.5. Validation

Tout d'abord, nous avons divisé notre échantillon de points de données ($N = 17367$) en deux ensembles aléatoires distincts. Le premier ensemble contenant 90 % des données ($N= 15639$) a été utilisé pour les modèles d'entraînement et les 10 % restants ($N= 1728$) ont été mis de côté pour une validation indépendante. Ces données divisées étaient

approximativement proportionnelles entre les sept écorégions et les trois périodes sous analyse.

Pour chaque combinaison d'approche statistique et de formulation de modèle (Ncombinaison = 18), nous avons calculé et rapporté trois mesures différentes de la performance du modèle pour chaque indicateur, en fonction des données et de la méthodologie utilisées pour la validation du modèle. L'objectif de la comparaison des trois mesures suivantes est de permettre des évaluations explicites, transparentes et visuelles de la manière dont chaque combinaison d'approche statistique et de formulation de modèle se comporte en ce qui concerne le sur ajustement et la variabilité des erreurs de prédiction.

Dans la première mesure, une approche naïve mais néanmoins informative est utilisée : la performance du modèle est évaluée uniquement sur les données d'apprentissage (en vert sur la Figure 2) sans effectuer de validation externe ou de validation croisée (voir la légende "apprentissage" en violet sur la Figure 2). Dans la deuxième mesure, la performance du modèle est évaluée sur les mêmes données d'apprentissage que ci-dessus, mais une validation croisée de 10 plis répétée 100 fois est effectuée sur les données d'apprentissage (voir la légende "repeated 10-Folds CV" en blanc sur la Figure 2). Dans la dernière mesure, la performance du modèle est évaluée exclusivement sur des données indépendantes (en rouge sur la Figure 2) et une validation indépendante (hold-out) est effectuée (voir la légende "independent" en jaune sur la Figure 2).

Ces méthodes de validation sont désormais des procédures de routine dans la cartographie numérique de sols, mais aucune d'entre elles ne donne d'indications fiables sur le degré de confiance que l'on peut avoir dans la prévision de l'épaisseur de la couche organique en dehors des domaines spatiaux ou temporels utilisés pour l'étalonnage et le test des modèles. Pour combler cette lacune et évaluer spécifiquement la transférabilité de notre meilleur modèle ci-dessus, mais dans l'espace et dans le temps, nous avons effectué une validation croisée spatio-temporelle en utilisant des régions écologiques (5b, 6a, 6c, 6d, 6e et 6f) comme blocs spatiaux et des périodes d'échantillonnage ($t_1 = < 1994$, $t_2 = \text{de } 1994 \text{ à } 1996$, $t_3 = \geq 1997$) comme blocs temporels. Afin d'explorer quelle composante entre le temps et l'espace est la plus influente sur les résultats, nous avons comparé les résultats de la

validation croisée espace seulement, temps seulement et espace-temps. Pour la validation croisée temporelle, nous avons utilisé à la fois la validation prévisionnelle (prédiction dans le futur : $t_1 \rightarrow t_2$, $t_1 \rightarrow t_3$, $t_2 \rightarrow t_3$, $t_{1+2} \rightarrow t_3$) et la validation rétrospective (prédiction dans le passé : $t_3 \rightarrow t_1$, $t_3 \rightarrow t_2$, $t_2 \rightarrow t_1$, $t_{3+2} \rightarrow t_1$). Par conséquent, pour la validation croisée spatio-temporelle, nous avons utilisé et présenté à la fois les validations spatio-temporelles pour les prévisions et les prévisions rétrospectives (voir Figure 3).

Résultats

Voici les principaux constats (préliminaires) issus de résultats observés. L'ajout de fonctions spatiales n'a aucune influence sur les résultats observés (Figure 2). En effet, l'utilisation de régions écologiques ne semble pas avoir d'effet sur les résultats de la transférabilité du modèle. Nos résultats montrent par contre, un lien direct où les variables environnementales capturent à elles seules 100 % de l'information contenue dans les fonctions spatiales (100 % de chevauchement). La mesure de la performance du modèle est alors très élevée. Les méthodes statistiques utilisés dans la modélisation n'ont aucun impact sur la performance du modèle (Figure 2). Sans égard à la méthode statistique utilisée, la performance du modèle est cohérente entre les indicateurs et élevée pour tous les indicateurs, par exemple, avec un R^2 moyen avec validation croisée de 0,8 et un RMSE moyen de 9 cm. La variation entre les plis et les répétitions est faible, ce qui indique que la précision des prédictions du modèle est assez stable. La validation naïve sur les données d'entraînement surestime les performances du modèle de 5 à 20 % par rapport à la validation croisée et à la validation indépendante. Il est intéressant de noter que les résultats de la validation indépendante se situent dans la plage de variation décrite par le CV répété de 10 plis (Figure 4).

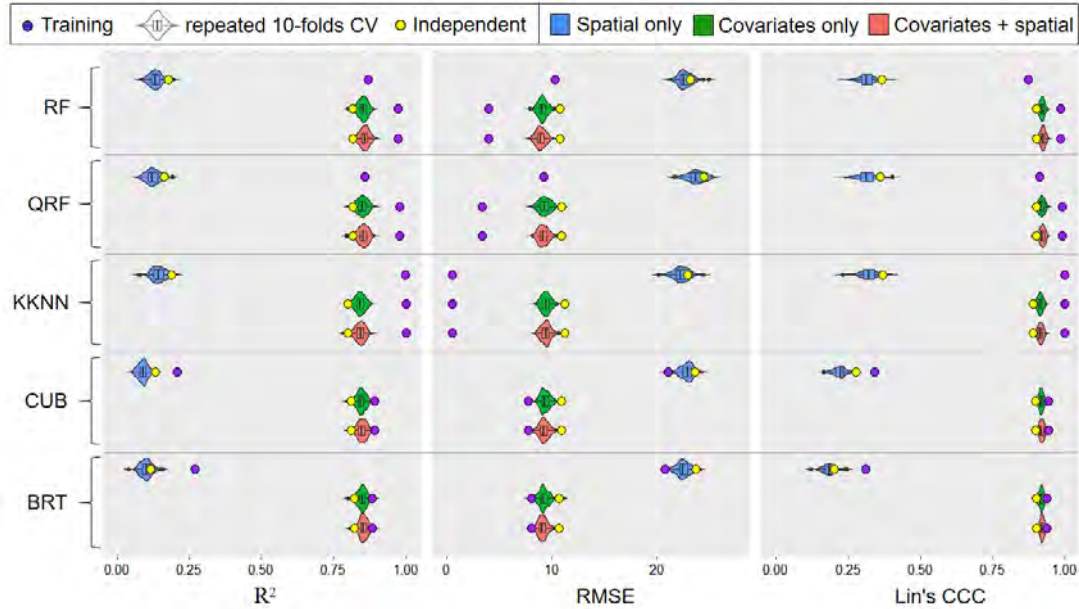


Figure 2. Performance des modèles selon A) l'algorithme utilisé (RF: Random Forest; QRF: Quantile Random Forest; KKNN: weighted K-nearest neighbors; CUB: Cubist; BRT: Boosted regression trees), B) l'utilisation de covariables environnementales uniquement (Covariates only), spatiales uniquement (Spatial only) ou des deux types de variables simultanément (Covariates + spatial), et C) du type de méthode de validation (pas de validation = training; validation croisée 10-folds répétée 20 fois = repeated 10-folds cross validation; indépendante = independent).

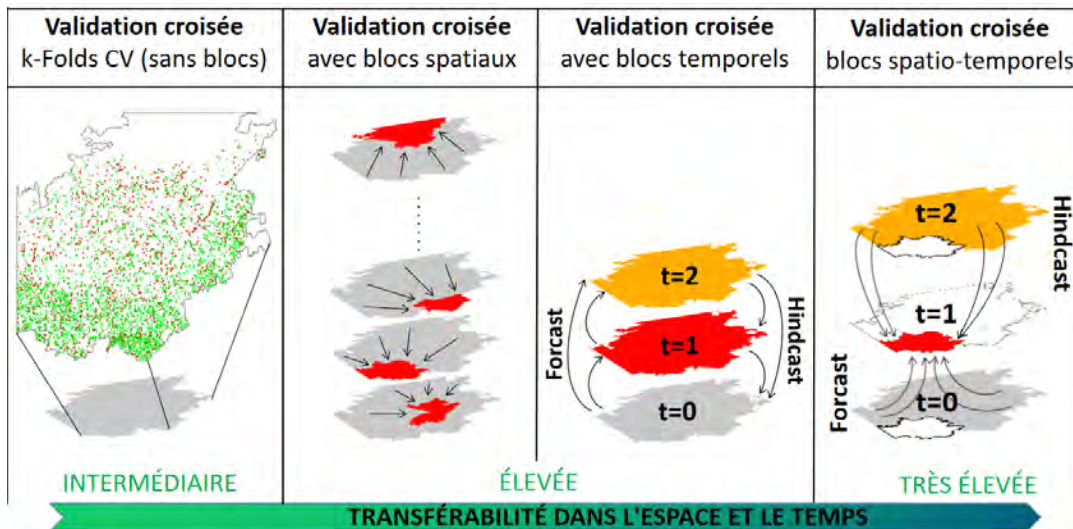


Figure 3. Illustration des méthodes de validation utilisées dans cette étude afin de quantifier la transférabilité des modèles prédictifs dans l'espace et le temps.

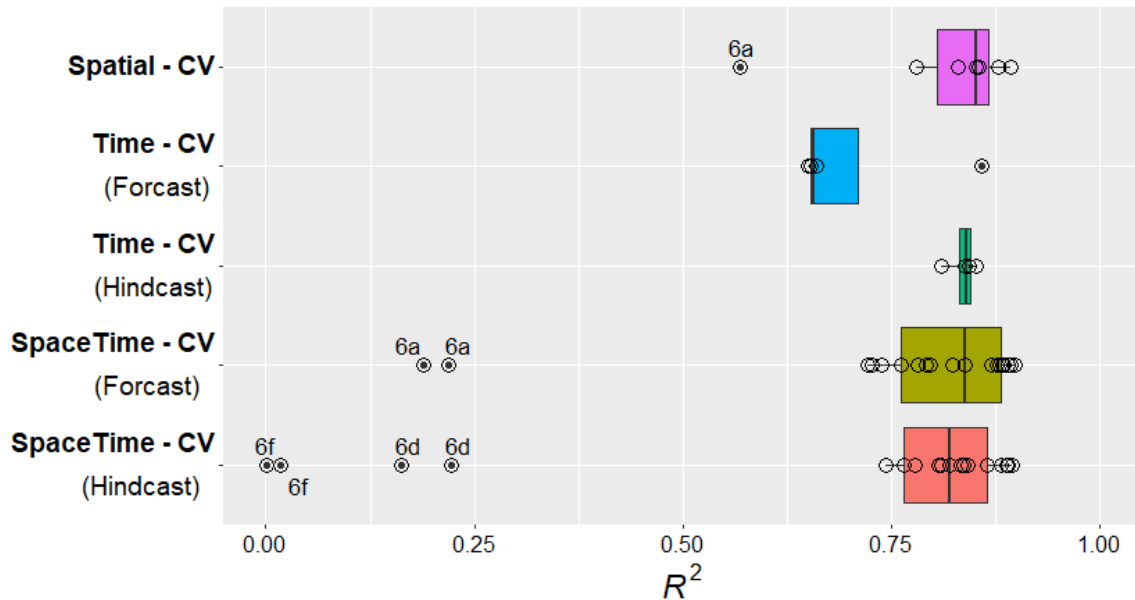


Figure 4. Performance des modèles avec l’algorithme *Random Forest* selon le type de validation croisée spatio-temporelle utilisée (spatiale seulement, temporelle seulement et spatio-temporelle combiné). Pour la validation temporelle, les résultats sont présentés pour la validation dans le futur (forecast) et dans le passé (hindcast). Les chiffres au-dessus des points correspondent à la région écologique.

Les prévisions sont robustes dans toutes les régions écologiques (Figure 4), tandis que l'écorégion "6a" semble se comporter différemment par rapport aux autres écorégions. Les prévisions sont également assez robustes sur l'ensemble des périodes, que ce soit en cas de forçage ou de prévision a posteriori, tandis que les régions écologiques et l'écorégion "6a" semblent se comporter différemment par rapport aux autres écorégions.

Discussion

Dans l'ensemble, les modèles montrent une performance très élevée avec un RMSE moyen de 9 cm et un R^2 élevés de 0,8, ce qui suggère que les modèles sont en mesure d’expliquer la variance observée. La performance peut être expliquée par la nature des données continues et catégoriques utilisées dans la description de l’épaisseur de la couche organique pour ce territoire où les variables environnementales capturent 100 % de l'information contenue dans les fonctions spatiales (chevauchement complet).

La faible performance temporelle (Figure 4) peut être interprétée comme une bonne nouvelle. En effet, la fenêtre du temps disponible n'est peut-être pas assez longue pour décrire un phénomène de la paludification qui peut prendre de siècles à s'installer.

La performance obtenue dans notre étude, démontre que l'estimation de l'épaisseur de la couche organique peut être estimée avec un faible erreur, alors cette cartographie peut s'avérer un outil puissant de précision de prédiction (Breiman 2001 ; Liaw et Wiener 2002). Les performances sont généralement équivalentes, entre les méthodes statistiques utilisées CNS, (Heung et al. 2016, Rudianto et al.2018).

Mise en garde

Malgré la haute performance de notre cartographie numérique de sols, nous nous sommes questionnées à la fin de la modélisation quant à la nature des données ayant servi à alimenter le modèle. En effet, une grande question quant à l'indépendance de l'estimation du dépôt et drainage obtenue dans les échantillons sur le terrain et celle disponible sur la carte écoforestière. Les deux approches permettent d'obtenir la classe dépôt drainage, mais ils sont issus de la même clé (<https://mffp.gouv.qc.ca/documents/forets/inventaire/norme-stratification.pdf>). Cette question nécessite être élucidé et il serait intéressant de réaliser une cartographie de l'épaisseur de la couche organique où la variable dépôt drainage est obtenue de la carte écoforestière et la comparer à celle issue des données échantillonnées sur le terrain. Le manque de temps et le retard dans le recrutement dans ce projet nous n'ont pas permis encore de les comparer.

Remerciements

Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce projet ainsi qu'aux discussions quant à la meilleure compréhension de la cartographie numérique de sols. Cette recherche continue avec un projet d'envergure (Projet Alliance : Améliorer les prédictions de la croissance forestière grâce à une nouvelle génération de cartes numériques des sols issues de l'intelligence artificielle) où nous ciblons améliorer notre compréhension de relations entre la cartographie et la productivité des forêts. Des méthodes d'intelligence artificielle

seront alors mises à contribution. L'équipe formée de Mickael Germain (U de S), Nicole Fenton (UQAT), Jean Daniel Sylvain et Guillaume Drolet (DRF), David Paré (Rcan) Mathieu Varin (Cerfo) Francis Perreault (PFR) Marie-Eve Sigouin (Greenfirst) et Alain Shink (Westfraser). Nous remercions également le soutien financier du MFFP, voire MRNF et la DRF pour la collaboration dans ce projet.

Références

- [1] Greiner, Lucie, Keller, A., Grêt-Regamey, A. & Papritz, A. (2017). Soil function assessment: review of methods for quantifying the contributions of soils to ecosystem services. *Land Use Policy*, 69: 224-237.
- [2] Adhikari, K., & Hartemink, A. E. (2016). Linking soils to ecosystem services - A global review. *Geoderma*, 262: 101-111.
- [3] Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B., Hong, S. Y. et al. (2014). GlobalSoilMap: toward a fine-resolution global grid of soil properties. In *Advances in agronomy* (Vol. 125, pp. 93-134). Academic Press.
- [4] Arrouays, D., Leenaars, J. G., Richer-de-Forges, A. C., Adhikari, K., Ballabio, C., Greve, M., et al. (2017). Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ*, 14: 1-19.
- [5] McBratney, A. B., Santos, M. M., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117: 3-52.
- [6] Bulmer, C., Paré, D., Domke, G.M. (2019). Cp. 14. A new era of digital soil mapping across forested landscapes.. In : *Global Change and Forest Soils*, Volume 36. 1st Edition. 538 p. Elsevier. Busse, M. Giardina, C., Morris, D. Page-Dumroese, D. (eds.).
- [7] Grundy, M. J., Rossel, R. V., Searle, R. D., Wilson, P. L., Chen, C., & Gregory, L. J. (2015). Soil and landscape grid of Australia. *Soil Research*, 53(8): 835-844.

- [8] Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., & Thompson, J. (2018). Soil property and class maps of the conterminous United States at 100-meter spatial resolution. *Soil Science Society of America Journal*, 82(1), 186-201.
- [9] Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, 261: 110-123.
- [10] Guevara, M., Olmedo, G. F., Stell, E., Yigini, Y., Duarte, Y. A., Hernández, C. A. et al. (2018). No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. *Soil*, 4(3): 173-193.
- [11] Vågen, T. G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbrecht, T. (2016). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma*, 263: 216-225.
- [12] Mansuy, N., Thiffault, E., Paré, D., Bernier, P., Guindon, L., Villemaire, P., ... & Beaudoin, A. (2014). Digital mapping of soil properties in Canadian managed forests at 250 m of resolution using the k-nearest neighbor method. *Geoderma*, 235, 59-73.
- [13] Beguin, J., Fuglstad, G. A., Mansuy, N., & Paré, D. (2017). Predicting soil properties in the Canadian boreal forest with limited data: Comparison of spatial and non-spatial statistical approaches. *Geoderma*, 306: 195-205.
- [14] Paré, D., Manka, F., Barrette, J., Augustin, F., Beguin, J. (2021). Indicators of site sensitivity to the removal of forest harvest residues at the sub-continental scale: mapping, comparisons and challenges. *Ecological indicators*, 125: 107516
- [15] Geng, X., Fraser, W., VandenBygaart, B., Smith, S., Waddell, A., Jiao, Y., & Patterson, G. (2010). Toward digital soil mapping in Canada: Existing soil survey data and related expert knowledge. In *Digital soil mapping* (pp. 325-335). Springer, Dordrecht.

- [16] Sylvain, J-D., Anctil, F. & Thiffault, E. (2021). Using bias correction and ensemble modelling for predictive mapping and related uncertainty: a case study in digital soil mapping. *Geoderma*, 403, 115153.
- [17] Furze, S., Arp, P., Li, S., Nocera, J., & Al-Tahir, R. (2018). A High-Resolution Digital Soil Mapping Framework for New Brunswick, Canada (Doctoral dissertation, PhD Thesis, Canada: University of New Brunswick).
- [18] Nyiraneza, J., Thompson, B., Geng, X., He, J., Jiang, Y., Fillmore, S., & Stiles, K. (2017). Changes in soil organic matter over 18 yr in Prince Edward Island, Canada. *Canadian Journal of Soil Science*, 97(4): 745-756.
- [19] Akumu, C. E., Johnson, J. A., Etheridge, D., Uhlig, P., Woods, M., Pitt, D. G., & McMurray, S. (2015). GIS-fuzzy logic based approach in modeling soil texture: using parts of the Clay Belt and Hornepayne region in Ontario Canada as a case study. *Geoderma*, 239: 13-24.
- [20] Heung, B., Bulmer, C. E., & Schmidt, M. G. (2014). Predictive soil parent material mapping at a regional-scale: a random forest approach. *Geoderma*, 214: 141-154.
- [21] Scarpone, C., Schmidt, M. G., Bulmer, C. E., & Knudby, A. (2016). Modelling soil thickness in the critical zone for Southern British Columbia. *Geoderma*, 282: 59-69.
- [22] Mansuy, N., Valeria, O., Laamrani, A., Fenton, N., Guindon, L., Bergeron, Y., & Légaré, S. (2018). Digital mapping of paludification in soils under black spruce forests of eastern Canada. *Geoderma Regional*, 15, e00194.
- [23] Blackford, C., Heung, B., Baldwin, K., Fleming, R. L., Hazlett, P. W., Morris, D. M., & Webster, K. L. (2021). Digital soil mapping workflow for forest resource applications: a case study in the Hearst Forest, Ontario. *Canadian Journal of Forest Research*, 51(1): 59-77.
- [24] Hengl, T., Jesus, J. M. De, Heuvelink, G. B. M., Ruiperez, M., Shangguan, W., Wright, M. N., Kilibarda, M., Blagoti, A., Geng, X., Bauer-marschallinger, B.,

- Guevara, M. A., Macmillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., Shangguan, W., ... Kempen, B. (2017). SoilGrids250m : Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2): 1–40.
- [25] Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE*, *9*(8), e105992.
- [26] D'Orangeville, L., Maxwell, J., Kneeshaw, D., Pederson, N., Duchesne, L., Logan, T., & Phillips, R. P. (2018). Drought timing and local climate determine the sensitivity of eastern temperate forests to drought. *Global Change Biology*, *24*(6): 2339-2351.
- [27] D'Orangeville, L., Duchesne, L., Houle, D., Kneeshaw, D., Côté, B., & Pederson, N. (2016). Northeastern North America as a potential refugium for boreal forests in a warming climate. *Science*, *352*(6292): 1452-1455.
- [28] Marchand, W., Girardin, M. P., Hartmann, H., Gauthier, S., & Bergeron, Y. (2019). Taxonomy, together with ontogeny and growing conditions, drives needleleaf species' sensitivity to climate in boreal North America. *Global change biology*, *25*(8): 2793-2809.
- [29] Lévesque, M., Walthert, L., & Weber, P. (2016). Soil nutrients influence growth response of temperate tree species to drought. *Journal of Ecology*, *104*(2): 377-387.
- [30] Cienciala, E., Russ, R., Šantrůčková, H., Altman, J., Kopáček, J., Hůnová, I., & Stáhl, G. (2016). Discerning environmental factors affecting current tree growth in Central Europe. *Science of the Total Environment*, *573*: 541-554.
- [31] Simard, M., Lecomte, N., Bergeron, Y., Bernier, P. Y., & Paré, D. (2007). Forest productivity decline caused by successional paludification of boreal soils. *Ecological Applications*, *17*: 1619-1637.

- [32] Henneb, M., Valeria, O., Thiffault, N., Fenton, N.J., & Bergeron, Y. (2019). Effects of Mechanical Site Preparation on Microsite Availability and Growth of Planted Black Spruce in Canadian Paludified Forests. *Forests*, 10(8).
- [33] Ecological Stratification Working Group. 1996. A national ecological framework for Canada. Agric. Agri-FoodCan., Res. Branch, Cent. Land Biol. Resour. Res.; Environ. Can., State Environ. Direct., Ecozone Anal.Branch, Ottawa, ON.
- [34] DeLuca, T.H., Nilsson, M.C., Zackrisson, O., 2002. Nitrogen mineralization and phenol accumulation along a fire chronosequence in northern Sweden. *Oecologia* 133, 206–214
- [35] Thiffault, N., P. Grondin, J. Noël and V. Poirier 2015. Ecological gradients driving the distribution of four Ericaceae in boreal Quebec, Canada. *Ecology and Evolution*. Volume 5, Issue 9 p. 1837-1853. <https://doi.org/10.1002/ece3.1476>
- [36] Veillette J.J. 1994. Evolution and paleohydrology of glacial lakes Barlow and Ojibway. *Quaternary Sciences Reviews*, 13 : 945-971.
- [37] Vincent JS, Hardy L 1977. L'évolution et l'extension des lacs glaciaires Barlow et Ojibway en territoire québécois. *Géographie Physique et Quaternaire* 31, 357–372. doi:10.7202/1000283AR
- [38] 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. *Can J For Res* 42, 1328–1343.
- [39] Morin, H., Danielle Laprise, and Yves Bergeron. 1993. Chronology of spruce budworm outbreaks near Lake Duparquet, Abitibi region, Quebec. *Canadian Journal of Forest Research*. 23(8): 1497-1506. <https://doi.org/10.1139/x93-189>
- [40] Régnière J. – Rémi Saint-Amant – Ariane Bécharde 2014. BioSIM 10 – GUIDE DE L'UTILISATEUR. Ressources naturelles Canada Service canadien des forêts Centre

de foresterie des Laurentides Rapport d'information LAU-X-137F.
<https://dl1ied5g1xfpx8.cloudfront.net/pdfs/34817.pdf>

- [41] Granitto P., C. Furlanello, F. Biasioli and F. Gasperi 2006. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometrics and Intelligent Laboratory Systems*}, 2006, 83,83-90.
- [42] Lin L (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 45: 255 - 268.
- [43] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [44] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [45] Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, 265, 62-77
- [46] Rudiyanto, Budiman Minasny, Budi Indra Setiawan, Satyanto Krido Saptomo, Alex B. McBratney, 2018. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands, *Geoderma*, Volume 313, 2018, Pages 25-40, <https://doi.org/10.1016/j.geoderma.2017.10.018>.