



Rapport scientifique

Évolution d'indicateurs spatialisés de la santé des sols sous l'effet des changements climatiques au Québec et en Ontario

Date : 26 septembre 2023

Responsable scientifique : Marc-Olivier Gasser, Ph. D.,

Ce rapport a été produit à l'attention du ministère de l'Agriculture, de l'Alimentation et des Affaires rurales de l'Ontario (MAAARO), du ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ) et du consortium Ouranos



À l'IRDA, on **collabore**, on se **questionne**, on **explore** et on **progressé** ensemble dans la même direction : celle d'une agriculture saine, dynamique et performante.

Nous sommes des **scientifiques**, mais aussi des **gens de terrain** qui **collaborent** avec l'ensemble du milieu agricole.

Notre mission consiste à innover en agroenvironnement pour créer ensemble la production agricole de demain. Consulter le www.irda.qc.ca pour en connaître davantage sur l'Institut et ses activités.

Question ou commentaire

Marc-Olivier Gasser, Ph. D.

Chercheur en conservation des sols et de l'eau

Responsable scientifique et auteur principal

T : 418 643-2380 p.650

marc-o.gasser@irda.qc.ca

Auteurs du rapport

Marc-Olivier Gasser, Ph. D., IRDA

Asim Biswas, Ph. D., U. of Guelph

Gaëtan Martinelli, itk

Tahmid Huq Easher, Ph. D., Olds College

Claude Bernard, Ph. D., retraité de l'IRDA

Antoine Zita, M. Sc. IRDA, INRS

Équipe de réalisation

Gaëtan Martinelli, itk

Tahmid Huq Easher, Ph. D., Olds College

Antoine Zita, M. Sc. IRDA, INRS

Farzaneh Heidari, M. Sc., UQAM

Marc-Olivier Gasser, Ph. D., IRDA

Asim Biswas, Ph. D., U. of Guelph

Claude Bernard, Ph. D., retraité de l'IRDA

Collaborateurs

Maxime Paré, Ph. D., UQAC

Maxime Leduc, Ph. D., U. McGill

Abdoulaye Baniré Diallo, Ph. D., UQAM

Pierre-Yves Gasser, M.B.A., AAC

Merci à nos partenaires financiers

Ce projet de recherche a été réalisé avec des aides financières accordées par une initiative conjointe entre le ministère de l'Agriculture, de l'Alimentation et des Affaires rurales de l'Ontario (MAAARO) et le ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ) dans le cadre de la Coopération Québec-Ontario pour la recherche agroalimentaire, ainsi que par le consortium Ouranos et le Fonds vert dans le cadre de la mise en œuvre du Plan d'action 2013-2020 sur les changements climatiques du Gouvernement du Québec.



Ce rapport peut être cité comme suit :

Gasser, Marc-Olivier, Asim Biswas, Gaëtan Martinelli, Tahmid Huq Easher, Claude Bernard, et Antoine Zita. 2023. Évolution d'indicateurs spatialisés de la santé des sols sous l'effet des changements climatiques au Québec et en Ontario. Rapport final. IRDA et partenaires. 118 pages.

© Institut de recherche et de développement en agroenvironnement inc. (IRDA)

INTRODUCTION

On évalue au Québec à 2,3 Mha (1,4 %) et en Ontario à 7,8 Mha (7,2 %), les superficies de sol qui ont un bon potentiel agricole. Les changements climatiques vont probablement conduire à une intensification de la production agricole sur les superficies actuelles et aussi à une expansion de l'agriculture dans des régions où le climat actuel ne la permet pas. Cette intensification de l'utilisation des terres conduira inévitablement à une minéralisation accrue de la matière organique des sols ainsi qu'à une dégradation de la structure et de la santé des sols. Le ruissellement et l'érosion des sols (sédiments, nutriments et pesticides) vers les cours d'eau ainsi que les émissions de gaz à effet de serre risquent de fortement augmenter, tandis que l'impact sur la productivité des cultures reste incertain. Il importe donc d'anticiper la nature et l'étendue de l'évolution de la santé des sols et de mettre en place les pratiques de conservation des sols qui permettront d'atténuer ces impacts et d'assurer la productivité des sols.

Objectifs et axes de recherche

Le projet visait à mieux comprendre l'évolution et la variabilité spatiale des indicateurs de santé des sols sous l'effet des CC et à valider les pratiques de conservation des sols qui permettront d'atténuer les impacts des CC et d'assurer la productivité des sols agricoles du Québec et de l'Ontario. Pour réaliser ces objectifs, la recherche a été menée en trois axes :

1. Développer un modèle prédictif de l'évolution de la matière organique en utilisant des bases de données d'analyse de sols, du climat et de l'affectation des terres et en utilisant différentes approches de modélisation par apprentissage machine à l'échelle des provinces;
2. Développer et spatialiser un index de santé des sols, basé sur des données historiques de propriétés des sols, à partir desquels les risques de dégradation des sols sous l'effet des changements climatiques seront quantifiés et cartographiés;
3. Mesurer avec des essais sous simulateur de pluie, l'efficacité de pratiques de conservation à contrôler l'érosion hydrique sous les conditions édaphiques et climatiques actuelles et futures.

Table des matières

Introduction	1
Objectifs et axes de recherche	1
Axe 1 : Projection de l'évolution de la matière organique et de la santé des sols sous climat futur à l'échelle du Québec et de l'Ontario.....	8
Résumé	8
Introduction	9
Objectif	11
Matériel et méthodes	12
1 Analyses de sols	12
1.1 Pré-traitement et spatialisation des échantillons de sols.....	12
1.2 Estimation de la texture des échantillons de sols à partir des autres analyses de sol.....	14
1.2.1 Réajustement des valeurs d'entraînement du modèle de prédiction des fractions granulométriques	15
1.2.2 Comparaison des modèles de prédiction des fractions granulométriques	15
1.3 Répartition de la matière organique des sols agricoles au Québec.....	16
2 Données climatiques.....	20
2.1 Source des données et des simulations en climat futur.....	20
2.2 Indicateurs climatiques	21
2.3 Découpage agroclimatique du territoire	23
2.4 Simulations du climat futur sur les zones agroclimatiques du Québec	25
3 Occupation des sols par les cultures	27
4 Variables dérivées du modèle numérique de terrain.....	29
Résultats	30
5 Modélisation de l'évolution de la matière organique du sol en climat futur.....	30
5.1 Corrélations simples entre la teneur en MOS et les autres variables incluses dans le modèle	30
5.2 Modélisation de la distribution spatiale de la matière organique du sol	32
5.3 Modélisation de la distribution spatiale du ratio de cultures pérennes	38
6 Projection de l'évolution des teneurs en MOS en climat futur	41
6.1 Application du modèle RF sur la MOS sans changement de cultures au Québec	41
6.2 Application du modèle XGBlinear pour simuler un changement de ratio de cultures pérennes en climat futur	43
6.3 Application du modèle RF avec changement de cultures	44
7 Application des modèles de simulation à l'Ontario	47
7.1 Simulations du climat futur sur les zones agroclimatiques de l'Ontario.....	50
7.2 Application du modèle RF sur la MOS sans changement de cultures en Ontario	52
Conclusion	55
Remerciements.....	57
Références	58
Annexes	62
Axe 2 : Spatialisation des indicateurs de santé des sols par la désagrégation de l'information sur les cartes de sols.....	67
Spatialization of soil health indicators through disaggregation of conventional soil maps	67
1 Soil Series Disaggregation of Agricultural Soils in Ontario.....	68
1.1 Background	68

1.2	Objective.....	68
1.3	Workflow and Results	69
2	Soil Organic Carbon Prediction of Ontario: A Probable Classifier Approach	89
2.1	Background.....	89
2.2	Scope and Objective	89
2.3	The specific objectives of the study were as follows:.....	90
2.4	Methodology	90
2.5	Results.....	91
3	Soil Organic Matter Disaggregation of Quebec Agricultural Soils.....	95
3.1	Background.....	95
3.2	Scope and Objectives.....	95
3.3	Methodology	95
3.4	Results.....	97
	Conclusion	103
	Annexe	104
	Axe 3 : Mesurer l'efficacité de pratiques de conservation à contrôler l'érosion hydrique sous les conditions édaphiques et climatiques actuelles et futures	111
	Résumé	111
	Annexe	114
	Analyse et discussion	117
	Conclusion et recommandations	118

Liste des tableaux

Tableau 1 :	Statistiques descriptives des propriétés de sol utilisées dans le modèle prédictif des fractions granulométriques.....	16
Tableau 2 :	Performances des modèles de prédictions des fractions granulométriques.....	16
Tableau 3 :	Définition des groupes texturaux des sols minéraux.	18
Tableau 4 :	Classification des simulations climatiques et origine des Earth System Models utilisés (ESM - CMIP5).....	21
Tableau 5 :	Valeurs propres et pourcentages de variance de l'ACP.....	24
Tableau 6 :	Normales climatiques de 30 ans (1981-2010) par zone agroclimatique du Québec.	25
Tableau 7 :	Définition des catégories de cultures et leur indice de culture.	28
Tableau 8 :	Normales climatiques par zone agroclimatique de l'Ontario.....	48
Tableau 9 :	Ratios de cultures pérennes actuels dans les six zones agroclimatiques du Québec et de l'Ontario.....	53
Table 1 :	Validation results of different modeling and sampling approaches of DSMART in predicting Medway Watershed soil great groups and soil series.	76
Table 2 :	Validation results of different modeling and sampling approaches of DSMART and Pure Polygon Disaggregation in predicting Middlesex County soil series.	80
Table 3 :	Entropy uncertainty of different modeling and sampling approaches of DSMART and Pure Polygon Disaggregation in predicting Middlesex County soil series.	80

Table 4 :	Model evaluation of SOC% prediction of Middlesex County using reassigned and weighted-probable models.	89
Table 5 :	Classification of 315,000 SOM% observations in agricultural soils of Quebec between 2009 and 2017, and 24 000 validation points collected from a private laboratory (LOGIAG) into categorical classes using uniform splitting interval.	97
Table 6 :	Confusion matrix of predicted soil organic matter classes using random forest in DSMART, compared to 24 000 validation points.	98

Liste des figures

Figure 1 :	Distribution des échantillons de sols AEL de la province de Québec géoréférencés au code postal.	12
Figure 2 :	Distribution de la teneur en matière organique du sol à l'échelle du Québec en fonction de trois tailles de polygones convexes (10, 30 et 50 km) regroupant les observations (n = 290 000) localisées par code postal.	13
Figure 3 :	Semi-variogrammes de la teneur en matière organique du sol en fonction de la taille des polygones convexe regroupant les observations (n = 290 000) géolocalisées par code postal. La grosseur des points est proportionnelle au nombre d'observations.	14
Figure 4 :	Distribution des fractions granulométriques initiales (a) et réajustées (b) des données du laboratoire AEL (8 364 échantillons) et distribution des sols de l'Étude sur l'état de santé des sols agricoles du Québec (c).	15
Figure 5 :	Dispersion des échantillons de sol d'AEL par codes postaux.	17
Figure 6 :	Distributions statistiques en violon et box-plot des teneurs en matière organique des sols d'AEL par zone climatique. Le x représente la moyenne. La ligne du bas, du centre et du haut de la boîte représentent le 25 ^e , 50 ^e (médiane) et 75 ^e percentiles des observations, respectivement.	18
Figure 7 :	Répartition de la teneur en matière organique des sols par zones et groupes texturaux.	19
Figure 8 :	Couverture spatiale des données climatiques avec le masque de la zone cultivée du Québec sur une grille de 10 km (exemple de la température maximale d'une journée).	20
Figure 9 :	Corrélation entre la MOS et le CDI calculé sur des saisons de croissance variant en termes de températures de base pour le calcul des degrés-jours (-10, -5, 0, 5, 10 °C) et du nombre d'années pour le calcul des normales climatiques (5, 10, 15, 20, 25, 30 ans).	23
Figure 10 :	Nombre optimal de clusters.	24
Figure 11 :	Découpage agroclimatique du Québec en 6 zones.	25
Figure 12 :	Projections en climat futur de quatre indicateurs climatiques dans six zones agroclimatiques en RCP 4.5 (a) et RCP 8.5 (b). Les lignes pleines et les parties ombragées représentent respectivement la moyenne et l'écart-type des simulations de 11 modèles climatiques.	26
Figure 13 :	Polygones des parcelles agricoles déclarées et des municipalités.	27
Figure 14 :	Pourcentage moyen d'occupation des grandes cultures par municipalité de 2003 à 2020.	28
Figure 15 :	Ratio moyen de cultures pérennes sur cultures annuelles (2003-2020) par municipalité.	29
Figure 16 :	Exemple de covariables dérivées du MNT : Élévation (haut gauche), pente (haut droite), aspect (bas gauche) et ombrage (bas droite).	30

Figure 17 :	Matrice de corrélation entre matière organique du sol (SOM), indicateur climatique CDI, bilan hydrique (<i>waterbalance</i>), température moyenne (tmoy) et ratio de cultures pérennes (<i>crop_ratio</i>), en moyenne (en gris) et par groupes texturaux (<i>GropTxt</i>) (en couleur).	31
Figure 18 :	Teneur moyenne en MOS (2007-2019) par municipalité.	32
Figure 19 :	Schéma des principales étapes du processus de modélisation utilisé.	33
Figure 20 :	Erreur quadratique moyenne (RMSE, %) et son intervalle de confiance à 95 % pour la prédiction de la matière organique du sol (MOS). Acronymes des méthodes statistiques : KNN = k le plus proche voisin (méthode pondérée) ; GLM = modèle linéaire généralisé ; GAM = modèle additif généralisé ; Cubist = Algorithme cubiste ; XGBtree = Extreme gradient boosting trees; XGBlinear = Extreme gradient boosting linear ; Ranger = forêts aléatoires.	34
Figure 21 :	Comparaison entre valeurs prédites et observées de la matière organique de sol (%) de six modèles testés sur le jeu de données indépendant (n = 61 167).	35
Figure 22 :	Importance des variables dans le modèle RF (forêts aléatoires) pour prédire la teneur en MOS.	36
Figure 23 :	Graphiques des dépendances partielles (<i>partial dependence plot</i>) entre la teneur en MOS et les variables prédictives les plus importantes, soit la teneur en sable (a), la teneur en argile (b), le ratio de cultures pérennes (c), le CDI (d), le bilan hydrique (e) et le modèle numérique de terrain (DEM) (f).	37
Figure 24 :	Erreur quadratique moyenne (RMSE en %) et son intervalle de confiance à 95 % de la prédiction du ratio de cultures pérennes.	38
Figure 25 :	Comparaison entre valeurs prédites et observées du ratio de cultures pérennes (%) pour quatre modèles testés sur le jeu de données indépendant (n = 154).	39
Figure 26 :	Importance des variables de quatre modèles développés pour prédire le ratio de cultures pérennes.	40
Figure 27 :	Projections de l'évolution des teneurs en MOS par zones climatiques et groupes texturaux représentées par la moyenne (ligne pleine) et l'écart-type (parties ombragées) des projections de 11 modèles climatiques en RCP 4.5 (a) et 8.5. (b).	42
Figure 28 :	Projections de l'évolution du ratio de cultures pérennes dans les six zones agroclimatiques avec le modèle XGBlinear en RCP 4.5 (a) et 8.5 (b), respectivement. La ligne pleine et les parties ombragées présentent la moyenne et l'écart-type des projections de 11 modèles climatiques.	43
Figure 29 :	Variations absolues des teneurs en MOS projetées de 2010 à 2100 résultant de la moyenne de 11 modèles climatiques évoluant sous l'effet de deux scénarios climatiques (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes, les groupes texturaux (a) ou les six zones agroclimatiques (b).	44
Figure 30 :	Variations relatives des teneurs en MOS projetées de 2010 à 2100 résultant de la moyenne de 11 modèles climatiques évoluant sous l'effet de deux scénarios climatiques (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes, les groupes texturaux (a) ou les six zones agroclimatiques (b).	44
Figure 31 :	Variations relatives des teneurs en MOS projetées de 2010 à 2100 en fonction de cinq groupes texturaux de sols, de six zones climatiques, de deux scénarios d'occupation du sol résultant de la moyenne de 11 modèles climatiques sous deux scénarios de réchauffement climatique (RCP 4.5 et 8.5).	46

Figure 32 :	Variations absolues des teneurs en MOS projetées à trois périodes (2011-2040, 2041-2070 et 2070-2100) sous l'effet de deux scénarios climatiques liés aux GES (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes dans les six zones agroclimatiques.	47
Figure 33 :	Découpage agroclimatique de l'Ontario en six zones.	48
Figure 34 :	Ratio de cultures pérennes moyen 2011-2020 présenté sur une grille de 250 m.	49
Figure 35 :	Polygones des municipalités (en bleu) et de la carte de sols (en beige) de l'Ontario.	50
Figure 36 :	Projections en climat futur de quatre indicateurs climatiques dans six zones agroclimatiques en RCP 4.5 (a) et RCP 8.5 (b). Les lignes pleines et les parties ombragées représentent respectivement la moyenne et l'écart-type des simulations de 11 modèles climatiques.	51
Figure 37 :	Projections de l'évolution des teneurs en MOS en Ontario par zones agroclimatiques et groupes texturaux représentées par la moyenne (ligne pleine) et l'écart-type (parties ombragées) des projections de 11 modèles climatiques en RCP 4.5 (a) et 8.5. (b).	53
Figure 38 :	Projections de l'évolution du ratio de cultures pérennes par zone climatique avec les modèles XGBtree (a et a'), XGBlinear (b et b') et RF (c et c') en RCP 4.5 () et 8.5 ('), respectivement. La ligne pleine et les parties ombragées présentent la moyenne et l'écart-type des projections de 11 modèles climatiques.....	62
Figure 39 :	Six Sub-Watersheds of Lake Erie.	69
Figure 40 :	Disaggregated Soil Series Maps of Garvey Glenn Creek Watershed at 30 m resolution.....	70
Figure 41 :	Disaggregated Soil Series Maps of Gully Creek Watershed at 30 m resolution.	71
Figure 42 :	Disaggregated Soil Series Maps of Jeanette's Creek Watershed at 30 m resolution.	72
Figure 43 :	Disaggregated Soil Series Maps of Wigle Creek Watershed at 30 m resolution.	73
Figure 44 :	Disaggregated Soil Series Maps of Kettle Creek Watershed at 30 m resolution.	74
Figure 45 :	Comparison of Soil Series Prediction between Different Models and Sampling Approaches in DSMART, Medway Watershed.	75
Figure 46 :	Study area, Middlesex County, shown by the red dot in the map of Canada (a) and by the red boundary area Ontario and in Southwestern Ontario (b). Soil series maps of Middlesex County and validation points represented by black dots from the 1992 soil survey report of Middlesex County (c).	77
Figure 47 :	Disaggregated Soil Series Map using DSMART and Pure Polygon Disaggregation of Middlesex County, ON at 50 m resolution (Area-Weighted Random Sampling Approach).....	78
Figure 48 :	Disaggregated Soil Series Map using DSMART and Pure Polygon Disaggregation of Middlesex County, ON at 50 m resolution (Conditional Latin Hypercube Sampling Approach).....	79
Figure 49 :	Disaggregated Soil Series Map of Lake Erie Basin at 250 m resolution ("by area" Sampling Approach in DSMART) *	82
Figure 50 :	Disaggregated Soil Series Map of Lake Erie Basin at 250 m resolution (Pure Polygon Approach) *	83
Figure 51 :	The study area located in North America (a) and the soils of Ontario represented by the red boundary area in Ontario (b). Validation points extracted from the National Pedon Database are represented by blue dots in the agricultural soil map of Ontario (c).....	85
Figure 52 :	Disaggregated Soil Series Map of Agricultural Soils of Ontario at 250 m resolution (Pure Polygon Approach) **	86
Figure 53 :	Entropy uncertainty of C5.0 in Pure Polygon Disaggregation using training data produced from area-weighted random approach at a 250m spatial resolution, Ontario.....	88
Figure 54 :	Mean SOC% prediction map at 0-30 cm depth of Middlesex County using reassigned and weighted-probable models.	92

Figure 55 :	Soil Organic Prediction Map of Agricultural Soils of Ontario using Probable-Classifer Approach.	94
Figure 56 :	The study area located in North America (a) and the agricultural soils of Quebec represented by a yellow boundary area in Quebec (b). Validation points extracted from a georeferenced private laboratory dataset are represented by green dots in the agricultural soil map of Quebec (c).	96
Figure 57 :	Disaggregated Soil Organic Matter (%) map of Quebec agricultural soils at 250m resolution using DSMART.....	99
Figure 58 :	Comparison between observed SOM (%) classes generated from the database compiled by IRDA and predicted SOM (%) classes from the disaggregation approach using DSMART.....	100
Figure 59 :	Entropy Uncertainty of Disaggregated Soil Organic Matter (%) Map of Quebec using DSMART.	101
Figure 60 :	Entropy Uncertainty of Disaggregated Soil Organic Matter (%) Map of Quebec using DSMART.	102

AXE 1 : PROJECTION DE L'ÉVOLUTION DE LA MATIÈRE ORGANIQUE ET DE LA SANTÉ DES SOLS SOUS CLIMAT FUTUR À L'ÉCHELLE DU QUÉBEC ET DE L'ONTARIO

Résumé

Dans le futur, les superficies agricoles vont probablement connaître une intensification de leur production sous l'effet des nouvelles normales climatiques et plusieurs propriétés des sols liées à la teneur en matière organique du sol (MOS) seront affectées. Le projet visait à mieux comprendre l'évolution et la variabilité spatiale des indicateurs de santé des sols sous l'effet des changements climatiques en analysant la structure spatiale et temporelle des propriétés du sol reliées à la matière organique, en utilisant des bases de données d'analyse de sols, du climat et de l'affectation des terres et en utilisant différentes approches de modélisation par apprentissage machine à l'échelle des provinces.

Après avoir filtré les analyses de sols du laboratoire privé et les avoir spatialisées sur l'ensemble du territoire agricole du Québec, les teneurs en MOS de la période de référence 2010-2018 présentaient des teneurs relativement élevées soient en moyenne de 4,5% dans la zone agroclimatique 1 la plus chaude en Montérégie et Centre-du-Québec, et de 8,7 % dans la zone agroclimatique 6 la plus froide en Gaspésie-Nord et Côte-Nord. En fonction de leur groupe textural prédit, les sols sableux présentaient les plus faibles teneurs en MOS et les sols argileux les teneurs les plus élevées. Différents modèles de régression développés en apprentissage machine ou linéaires ont prédit sur un jeu indépendant de données, la répartition de la MOS à l'échelle du Québec avec des R^2 variant de 0,42 à 0,49. Le modèle Random Forest retenu pour faire ces prédictions inclut par ordre d'importance, les teneurs en argile et en sable du sol, la prépondérance des cultures pérennes, des indicateurs climatiques reliés à la température et au bilan hydrique, et finalement des variables reliées au modèle numérique de terrain.

La prédiction du ratio de cultures pérennes à l'échelle du Québec a été modélisée avec la technique d'apprentissage machine XGBlineair donnant un R^2 de 0,81 et l'indicateur climatique *Climate Decomposition Index* (CDI) combinant des effets de la température et du bilan-hydrique a été la première variable d'importance suivi de covariables liées au modèle numérique de terrain. L'application du modèle de régression en conditions de climat futur (2010-2100) a projeté des baisses de teneurs en MOS dans les régions plus froides. Les changements climatiques et le type de texture de sols ont généré plus de variabilité sur les teneurs en MOS que le changement de ratio de cultures pérennes, malgré une baisse importante de ce ratio dans certaines zones plus froides en lien avec le réchauffement climatique dans ces zones. Les scénarios d'émissions de GES (RCP 4.5 et 8.5) n'ont pas non plus généré de différence aussi importante que les changements climatiques anticipés dans les zones plus froides. Des diminutions de 0,4 à 1,0 % de MOS sur 90 ans dans les sols sableux et argileux équivalent respectivement à des pertes annuelles de 0,0026 et 0,006 % C/an soit de l'ordre de grandeur du 0,008 % C/an rapporté par Warren et al. (2020) pour l'Ontario, mais plus élevées que les pertes de 0,02 à 0,055 Mg C/ha/an rapportées par Smith et al. (2009) et simulées avec le modèle Century dans des régions du sud de l'Ontario et du Québec.

L'application du modèle d'évolution de la MOS dans les conditions de l'Ontario a donné des résultats mitigés; c'est-à-dire des teneurs en MOS relativement élevées en climat actuel qui n'ont pas évolué à la baisse en climat futur. Les projections réalisées en dehors du domaine d'entraînement du modèle, en conditions plus chaudes en climat futur ainsi qu'en climat actuel dans le sud de l'Ontario pourrait expliquer ces résultats. Le ratio de

cultures pérennes très différents entre les deux provinces résultant soit d'une réalité agricole différente ou de la source de données différente utilisée pour calculer ces ratios pourrait aussi expliquer ces résultats.

L'approche utilisée dans les conditions du Québec a néanmoins permis de souligner l'importance de mettre en place des mesures pour protéger les sols qui seront éventuellement convertis en cultures annuelles dans les régions plus froides où les cultures pérennes occupent encore des superficies importantes. Selon les prévisions des modèles, les quantités perdues de MOS et de C émis dans l'atmosphère seraient plus importantes dans les sols argileux des régions des Laurentides, du Témiscamingue et du pourtour du Lac-Saint-Jean, ainsi que dans les sols sableux du nord du Lac-Saint-Jean et de l'Abitibi. Dans les régions plus chaudes de la Montérégie et du Centre-du-Québec, les sols semblent avoir atteint des niveaux plus faibles de MOS, mais à l'équilibre sous la présence de cultures annuelles plus prépondérante et n'évolueraient plus sous l'effet des changements climatiques.

Introduction

Les superficies de sols à bon potentiel agricole sont limitées dans le Québec et l'Ontario et les problèmes de dégradation de la santé des sols devraient augmenter sous l'intensification des cultures et des changements climatiques. Les superficies agricoles actuelles vont probablement connaître une intensification de leur production sous l'effet des nouvelles normales climatiques. Ces changements climatiques (CC) vont aussi favoriser l'expansion de l'agriculture dans des régions où le climat actuel ne la permet pas. Sans mesure de protection et de conservation adéquate des sols, la plupart des travaux de recherches sur le sujet prévoient un déclin de la matière organique des sols, une dégradation de leur structure, une érodabilité accrue, du ruissellement, de l'érosion et une pollution accrue des plans d'eau (sédiments, nutriments, pesticides), ainsi qu'une augmentation des émissions de CO₂ vers l'atmosphère. Ces effets sont d'autant plus vraisemblables que les sols, susceptibles de voir leur vocation agricole s'accroître, présentent une vulnérabilité élevée aux processus de dégradation. Il convient donc d'étudier l'évolution des sols sous les CC prévisibles et de développer des techniques de conservation des sols qui permettront de contrôler les impacts de ces changements et d'assurer la productivité des sols agricoles du Québec et de l'Ontario, particulièrement dans les régions où les cultures sont présentement marginales, mais appelées à s'intensifier.

Dans le futur plusieurs paramètres indicateurs de la qualité des sols, comme la stabilité structurale et l'érodabilité du sol seront affectés par l'évolution de la teneur en matière organique du sol (MOS) et de nombreux travaux ont démontré les liens entre la MOS et les propriétés physiques comme la masse volumique apparente, ou des indicateurs comme l'érodabilité de sols en fonction des types de sols souvent fonction aussi de la granulométrie du matériau ou de la texture du sol.

La matière organique du sol est un élément clé pour la santé et la conservation des sols. À l'interface entre l'atmosphère, la lithosphère, la biosphère et l'hydrosphère, le sol et la matière organique représentent un ensemble multifonctionnel qui jouent un rôle primordial dans la productivité agricole mondiale (Lal et al., 2004). Ils participent à réguler le climat via l'émission de gaz à effet de serre et la séquestration du carbone, ils jouent un rôle essentiel dans la résilience des écosystèmes et protègent les ressources en eau par leur capacité à filtrer ou à retenir les polluants (Lal et al., 2004; Hoffland et al., 2020). La teneur en matière organique dans les sols permet d'améliorer les propriétés physiques en améliorant la structure du sol et l'infiltration de l'eau, les propriétés biologiques en stimulant l'activité microbienne, et les propriétés chimiques en contribuant à la fertilité des sols agricoles. Elle résulte de l'équilibre entre l'accumulation et la dégradation des débris végétaux et est affectée par de multiples facteurs tels que le climat, la texture du sol, l'activité agricole, ou encore le pH des sols. C'est un indicateur essentiel au suivi de la santé des sols au Québec dans le contexte des grands

changements globaux du climat et par rapport aux enjeux spécifiques tel que l'amélioration de la gestion des matières fertilisantes, l'optimisation de la gestion de l'eau ou l'amélioration de la biodiversité.

Plusieurs façons d'évaluer ou de simuler l'évolution de la matière organique et autres propriétés des sols sont présentées dans la littérature. Mises à part de nouvelles méthodes de mesures indirectes de flux de CO₂ ou spectrales, le suivi d'essais de longue durée, les inventaires nationaux ou régionaux à répétition ainsi que les modèles mécanistiques sont généralement les sources et les méthodes de bases pour évaluer les changements des teneurs en C dans les sols (Smith et al., 2020). Des essais de longue durée permettent dans un premier temps de qualifier l'effet des certaines pratiques agricoles dans des contextes pédoclimatiques particuliers et de capturer la dynamique du carbone au cours du temps (Smith et al., 2020). Les essais de Rothamsted démarrés au 19^e siècle en sont un exemple éloquent qui ont permis de dresser les effets à long terme de plusieurs pratiques incluant les apports d'amendement organiques, l'inclusion de prairies dans la rotation et le retrait des résidus de cultures sur la capacité des sols à accumuler du carbone et rencontrer les objectifs de l'Initiative 4 pour mille (Poulton et al., 2018). Ces essais ont également permis de dériver un modèle mécanistique de l'évolution de C dans les sols en fonction des pratiques agricoles (RothC) qui fait partie du système recommandé par les experts de la FAO pour évaluer le potentiel de séquestration du C dans les sols à travers le monde (Parton, 1996; FAO, 2022).

Plusieurs modèles de simulation mécanistiques intégrant différentes composantes du système sol-plante-climat (par exemple, Century, ICBM, RothC, STICS, SOCRATES, etc.) ont été développés le plus souvent à partir de données d'essais de longue durée pour simuler le devenir de la matière organique ou du carbone dans les sols, avec différents pools de matière organique ou de carbone variant en nombre et en composition. Un de ces premiers modèles, le Bilan Humique, davantage utilisé au Québec à des fins de vulgarisation, reposait sur un pool unique de matière organique (CRAAQ, 2010). Le modèle ICBM est basé sur les travaux de Hénin et Dupuis (1945) et compartimente la matière organique en deux pools, une fraction stable et une fraction plus légère ou jeune (Andrén et Kätterer 2000). Il a été calibré sous les conditions de la Suède et intégré dans l'application HOLOS adaptée aux conditions canadiennes (Kroebel et al., 2013) ainsi qu'OGEMOS adapté plus particulièrement aux conditions du Québec (Michaud et al., 2020). Le modèle Century a été utilisé au Canada pour anticiper les effets des pratiques agricoles sous l'effet de scénarios de changements climatiques et sur 24 sites à travers le Canada dans les climats arides, sub-humides et humides (Smith et al., 2009). Des pertes de C plus importantes ont été simulées et notées sous l'effet de conditions plus sèches dans l'Ouest canadien, résultant d'un facteur de décomposition du C plus élevé. L'augmentation de la productivité des cultures liées aux hausses de températures n'a pas compensé ces pertes de C. La conversion des cultures annuelles en prairies a augmenté de 0,27 à 0,78 Mg C ha⁻¹ an⁻¹ la quantité de C accumulée dans les sols dans les régions humides de l'Est du Canada. Gottschalk et al. (2012) ont utilisé RothC à l'échelle planétaire pour anticiper l'effet des changements climatiques sur l'évolution du C dans les sols et les résultats pour le Canada ont été comparés à ceux obtenus de Smith et al. (2009). Leurs simulations prévoient des baisses plus importantes de C (-20 à -40 t C ha⁻¹) sur un horizon de 1971 à 2100 dans l'Est du Canada, sous l'effet de la déforestation ou conversion en terres arables accompagnées de plus faibles retours en biomasse des plantes, de hausses de la température moyenne annuelle de 4 à 6 °C et d'un bilan hydrique inférieur.

Des inventaires réalisés à l'échelle régionale à répétition dans le temps ont permis également d'anticiper l'effet des conditions pédoclimatiques et d'utilisation du territoire sur l'évolution de la matière organique et autres propriétés des sols (Nunes et al., 2020; Meersmans et al., 2011; Meersmans et al., 2013; Reijneveld et al., 2009; Dupla et al., 2021). Des modèles empiriques par exemple ont été développés pour spatialiser le teneur en C du sol en 1960 et 2006 en fonction du climat, de l'utilisation du sol, du type de sol et de sa gestion et comparer son évolution (Meersmans et al., 2011).

L'approche par substitution de l'espace pour le temps (space for time substitution) a été employée dans divers domaines pour modéliser des processus dynamiques et anticiper l'effet des changements climatiques. Par exemple, Blois et al. (2013) l'ont utilisé pour modéliser l'évolution de la biodiversité en Amérique du Nord et l'ont comparé avec succès à des mesures en séries temporelles historiques. Gray et al. (2016) a développé un tel modèle combinant des variables climatiques et des covariables utilisées en pédologie numérique tel que le relief, le matériau parental et des indices du degré d'altération de roche mère et de perturbation du sol lié à son usage pour prédire en climat futur l'évolution du C dans les sols en Australie à l'orée de 2070. À différentes profondeurs, la teneur en C était plus reliée à la température maximale, la teneur en silice et l'usage du sol qu'aux précipitations et aux variables liées au relief. Les auteurs soulignent l'importance de tenir compte de l'évolution de l'usage du sol qui était maintenu dans leur modèle, car son influence pourrait être considérable et devrait être testée. Smith et al. (2016) anticipent que le changement d'affectation des terres et de structure de végétation aura plus d'effet que le climat sur l'évolution du C dans les sols forestiers en Europe.

Objectif

Cet axe de recherche visait à mieux comprendre l'évolution et la variabilité spatiale des indicateurs de santé des sols sous l'effet des CC en analysant la structure spatiale et temporelle des propriétés du sol reliées à la matière organique, en utilisant des bases de données d'analyse de sols, du climat et de l'affectation des terres et en utilisant différentes approches de modélisation par apprentissage machine et déterministe à l'échelle des provinces.

Matériel et méthodes

1 Analyses de sols

Le laboratoire AgroEnvirolab (AEL), partenaire du projet, a fourni une base de données d'analyses courantes de la chimie-fertilité des sols pour environ 600 000 échantillons de sols analysés entre 2007 et 2019, prélevés à travers le Québec et identifiés avec le code postal du client (Figure 1).

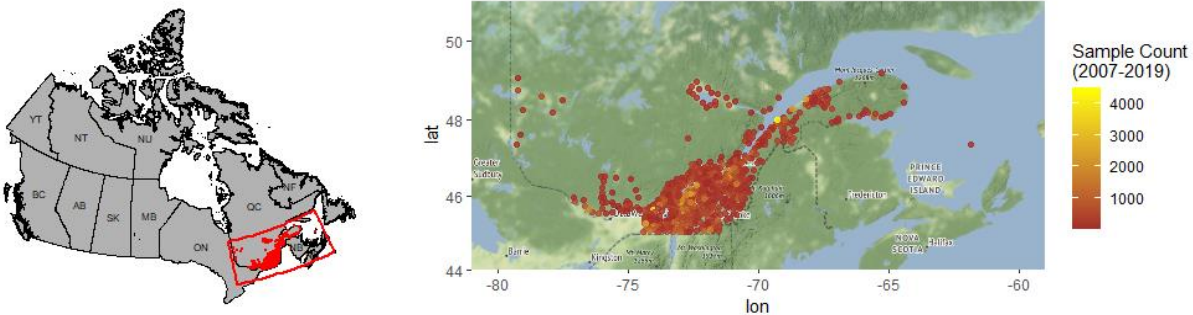


Figure 1 : Distribution des échantillons de sols AEL de la province de Québec géoréférencés au code postal.

1.1 Pré-traitement et spatialisation des échantillons de sols

La base de données des sols d'AEL contenait des sols qui n'étaient pas exclusivement des sols agricoles. Plusieurs filtres se basant sur une base de données indépendantes de sols agricoles ont été utilisés afin de faire un premier tri de la base de données et d'en conserver uniquement les sols jugés a priori agricoles (~290 000 échantillons conservés).

Les échantillons de sols du laboratoire étaient identifiés avec des codes postaux reliés aux clients, permettant ainsi de les géolocaliser à l'intérieur de polygones de municipalité couvrant les codes postaux. La plupart des centroïdes de polygones de municipalité associés à un code postal étaient situés à moins de 15 km des champs agricoles dans le sud du Québec, mais cette distance pouvait atteindre plus de 100 km en régions périphériques. Par ailleurs, les champs échantillonnés pouvaient se retrouver dans un polygone de municipalité associés à un code postal (et son centroïde) différent de celui de l'adresse administrative des producteurs ou des clubs conseils qui ont fait la demande d'analyses. Toutefois, une analyse préliminaire avec 340 observations géoréférencées du même ensemble de données de laboratoire a révélé que le centroïde du champ pouvait se situer en moyenne à 14 km (valeur min-max de 0,1 à 90 km) du centroïde du code postal déclaré par le demandeur de l'analyse. Une autre analyse préliminaire a été menée pour révéler comment la corrélation spatiale de la MOS évolue lorsque les observations géolocalisées au code postal sont regroupées à différentes tailles de grilles (10 à 100 km). La Figure 2 présente en exemple les données regroupées à des tailles de grilles de 10, 30 et 50 km.

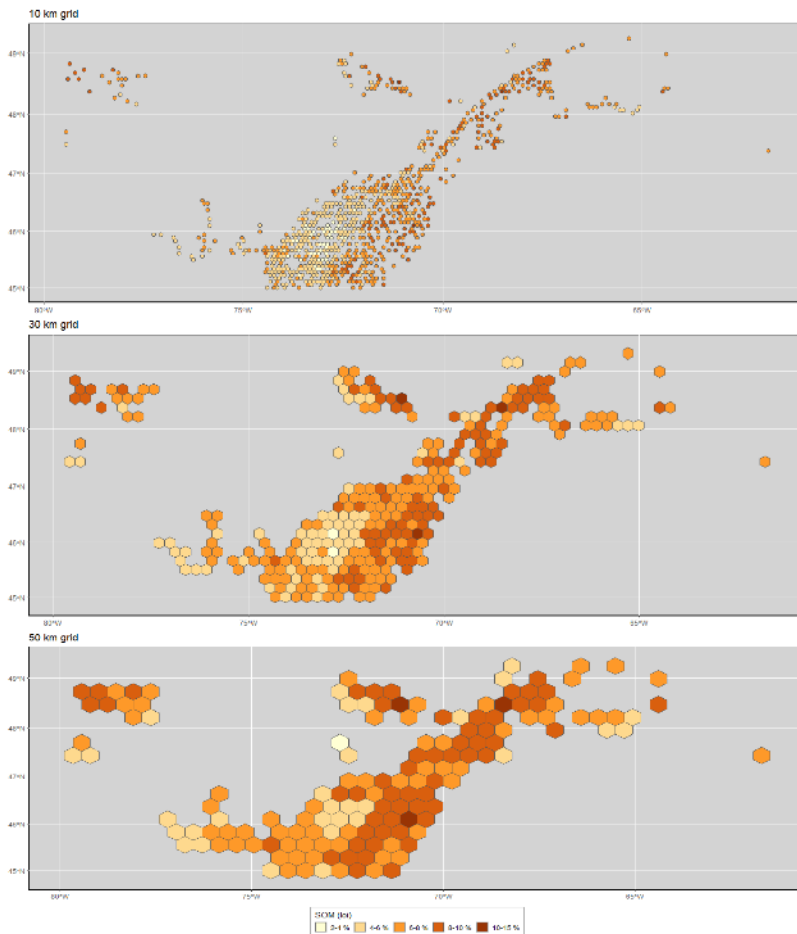


Figure 2 : Distribution de la teneur en matière organique du sol à l'échelle du Québec en fonction de trois tailles de polygones convexes (10, 30 et 50 km) regroupant les observations (n = 290 000) localisées par code postal.

En augmentant la taille des polygones convexes de 10 à 100 km, les effets de pépité et l'autocorrélation n'ont pas diminué de façon significative aux plus petites distances, indiquant qu'il n'y avait pas d'avantage apparent à agréger les données sur de plus grandes surfaces pour réduire l'incertitude sur la géoréférence des analyses de sols (Figure 3).

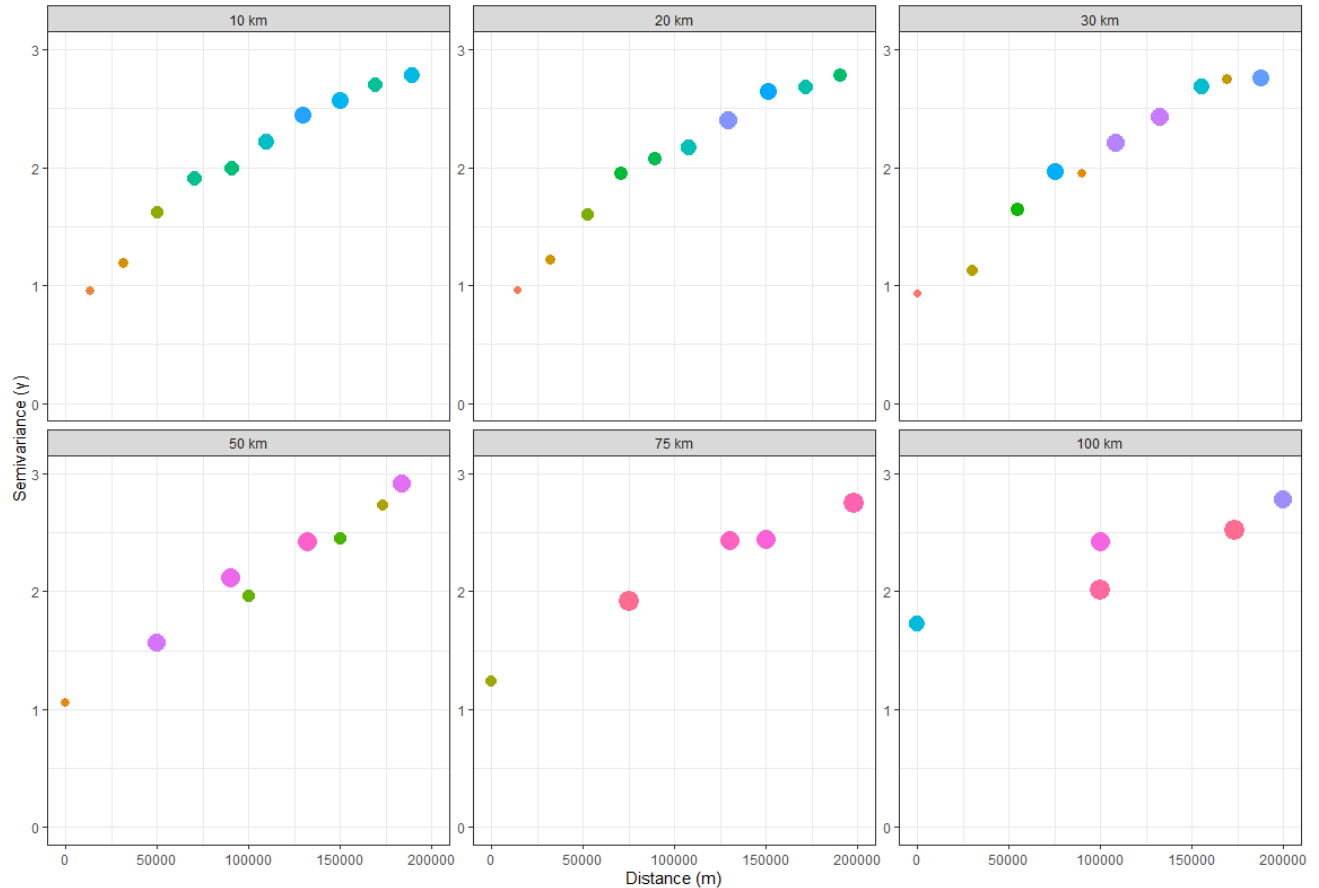


Figure 3 : Semi-variogrammes de la teneur en matière organique du sol en fonction de la taille des polygones convexe regroupant les observations (n = 290 000) géolocalisées par code postal. La grosseur des points est proportionnelle au nombre d'observations.

1.2 Estimation de la texture des échantillons de sols à partir des autres analyses de sol

([Machine learning models for predicting soil particles size fractions from chemical and physical analyses; Martinelli et Gasser, 2022; Article publié dans Soil Science Society of America Journal](#))

Les niveaux de matière organique et son évolution dans le sol sont conditionnés par la texture du sol sous l'effet de sa stabilisation dans les complexes argilo-humiques et à l'inverse, de sa minéralisation accélérée dans les sols sableux (Dexter et al. 2008; Johannes et al., 2017; Gray et al., 2015). D'après une méta-analyses récente, les teneurs en argile et limon fin sont les principaux facteurs définissant la capacité d'accumulation de C et N dans les sols (Matus, 2021). La texture est donc une variable importante à inclure dans un modèle prévisionnel de l'évolution de la matière organique du sol en climat futur. Sur les 290 000 échantillons de sols analysés et retenus, seulement 8 364 échantillons avaient une analyse de leur texture en trois fractions granulométriques (argile, limon, sable). Des modèles en apprentissage machine ont été développés et comparés, afin de prédire la texture des autres échantillons de sols à partir des autres propriétés physico-chimiques comme le pH à l'eau, les teneurs en éléments Mehlich-3 et la densité du sol mesurée dans une cuillère de 3 cc, dans le but d'augmenter l'information reliée aux analyses de MOS.

1.2.1 Réajustement des valeurs d'entraînement du modèle de prédiction des fractions granulométriques

Le laboratoire AEL utilise la méthode originale de Bouyoucos (1962) par mesure à l'hydromètre pour déterminer les fractions granulométriques de sable, limon et argile. Depuis le temps, cette méthode a été revue et modifiée pour considérer la présence d'agrégats et de matière organique dans l'échantillon analysé qui influencent la détermination de la granulométrie (Day, 1965 ; Gee & Bauder, 1986).

Le laboratoire de l'IRDA utilise une méthode de Bouyoucos modifiée qui traite la présence d'agrégats et de matière organique dans l'échantillon (lorsque MOS > 5%). Avec une base de données appariées entre les deux méthodes, un modèle de régression linéaire multiple (R^2 de 0.93 en validation) a permis de réajuster les fractions granulométriques des échantillons de sols d'AEL. Le modèle de réajustement a tendance à augmenter les fractions d'argile et de limon au détriment des fractions de sable, pour se rapprocher d'une distribution plus représentative des sols du Québec (Figure 3).

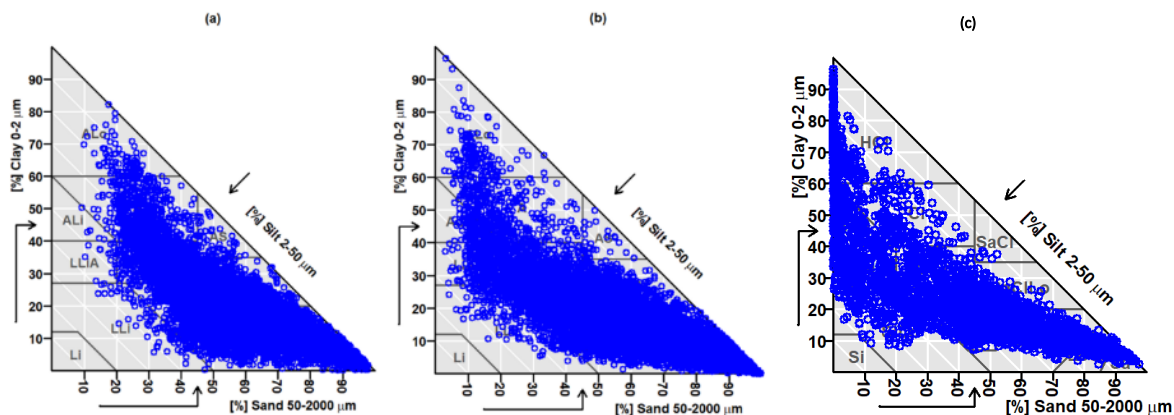


Figure 4 : Distribution des fractions granulométriques initiales (a) et réajustées (b) des données du laboratoire AEL (8 364 échantillons) et distribution des sols de l'Étude sur l'état de santé des sols agricoles du Québec (c).

1.2.2 Comparaison des modèles de prédiction des fractions granulométriques

Les statistiques des variables prédictives et des trois fractions granulométriques utilisées pour développer le modèle de prédiction des fractions granulométriques sont présentées au Tableau 1.

Le modèle retenu a été choisi en analysant ses performances prédictives, soit la robustesse et le degré de précision du modèle. Cinq modèles ont été comparés :-le modèle de régression multiple (LinearRegression),-la méthode des K plus proche voisin (K-NN),-le modèle de forêt aléatoire (RandomForest),-le modèle Optimized gradient boosting (XGBoost) et-le modèle de réseaux de neurones multicouches (NeuralNetwork). Les modèles RandomForest et XGBoost apparaissent les plus performants en observant les indicateurs (Tableau 2). En revanche, le modèle en RandomForest a tendance à lisser les valeurs extrêmes, ce qui provoquait une forme particulière du nuage de point des valeurs projetées dans le triangle des textures. Le modèle XGBoost a donc été choisi pour prédire les fractions granulométriques sur le reste des échantillons d'AEL.

Tableau 1 : Statistiques descriptives des propriétés de sol utilisées dans le modèle prédictif des fractions granulométriques.

Soil property	Moyenne^a	Écart type	Médiane	Minimum	Maximum
DSL ^b , g cm ⁻³	0,962	0,16	0,969	0,4	1,9
pH eau	6,3	0,72	6,3	5,0	8,5
CEC, cmol _c kg ⁻¹	17,9	5,7	17,2	0,51	65,4
P-M3, mg kg ⁻¹	57,9	86,3	42,8	0,7	1076
K-M3, mg kg ⁻¹	110,9	107,2	89,9	0,0	1966,5
Ca-M3, mg kg ⁻¹	1802,3	1260,9	1651,3	9,5	12169,6
Mg-M3, mg kg ⁻¹	183,8	162,6	121,3	0,9	1691,1
Al-M3, mg kg ⁻¹	1072,1	425,7	1046,0	26,4	2865
Mn-M3, mg kg ⁻¹	31,8	34,9	23,4	0,27	725,1
Cu-M3, mg kg ⁻¹	2,34	1,9	1,89	0,03	37,5
Zn-M3, mg kg ⁻¹	4,06	5,7	2,9	0,07	130,1
Fe-M3, mg kg ⁻¹	250,9	96,8	243,0	0,0	1420
Argile, %	18,7	13,5	15,7	0,01	96,4
Sable, %	55,8	24,3	56,8	2,85	99,7
Limon, %	25,6	15,5	24,6	0,1	74,8

^a Statistiques sur la base de données triées et utilisées pour calibrer le modèle (n = 8 364)

^b DSL, Densité du sol tamisé à 2 mm et évaluée avec une cuillère de 3 ml au laboratoire.

Tableau 2 : Performances des modèles de prédictions des fractions granulométriques.

Models	Sable		Limon		Argile	
	RMSE	R ²	RMSE	R ²	RMSE	R ²
Regression multiple linéaire	16,9	0,52	12,5	0,33	9,33	0,53
K-NN	12,7	0,72	10,5	0,52	7,24	0,72
Random Forest	12,5	0,73	10,2	0,55	7,16	0,73
Neural Network	14,0	0,67	11,4	0,45	7,59	0,68
XGBoost	12,4	0,73	10,1	0,57	6,57	0,77

Le modèle de prédiction final des fractions granulométriques a été transmis au laboratoire AEL sous forme d'une application *RShiny* afin de faciliter l'utilisation par les employés du laboratoire.

1.3 Répartition de la matière organique des sols agricoles au Québec

Tel que décrit précédemment, les analyses de sol provenaient du Laboratoire AgroEnviroLab et étaient géolocalisées à l'aide de codes postaux déclarés par le client dans la demande d'analyse et associés aux polygones de municipalités et à leur centroïde. Chaque polygone rassemblait de 10 à environ 3000 échantillons de sol répartis dans les différentes zones agroclimatiques du Québec (Figure 5).

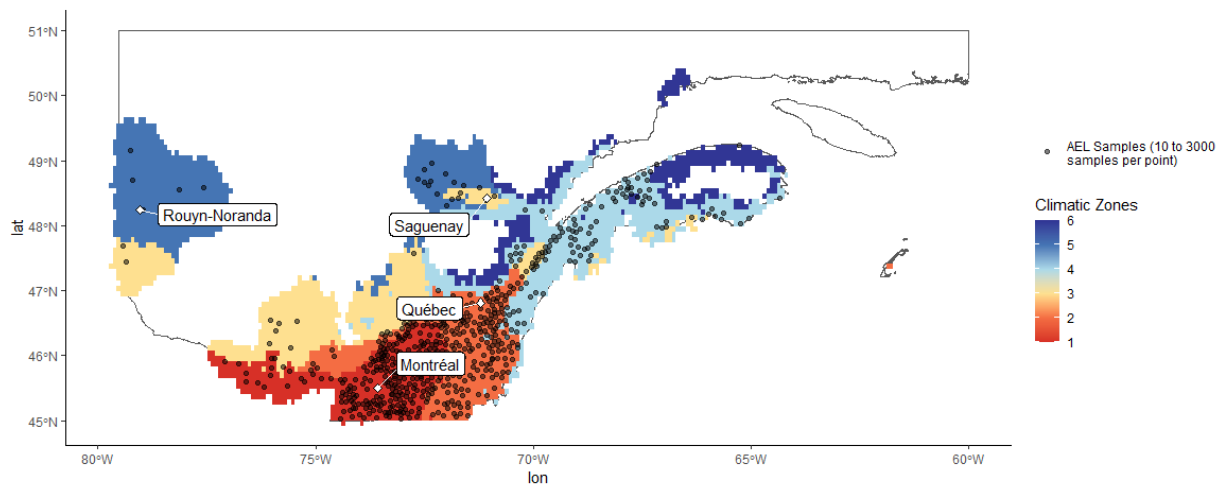


Figure 5 : Dispersion des échantillons de sol d'AEL par codes postaux.

Des essais ont été entrepris pour désagréger l'information sur les sols (MOS, texture) reliées aux polygones des municipalités pour l'associer à des pixels de plus petites tailles (250 m) par des techniques de pédologies numériques pour désagréger les unités cartographiques des cartes de sols en séries de sols uniques (Easher et al., 2022). L'information sur le modèle numérique de terrain et le climat étaient les principales co-variables utilisées pour désagréger les cartes de sols. Malgré des résultats prometteurs, cette approche n'a pas été retenues pour désagréger les différentes teneurs en MOS compilées par municipalité.

Une analyse par régression simple sur les 300 000 analyses de teneur en MOS réparties de 2010 à 2018 n'a pas révélé de tendance ni à la hausse ni à la baisse des teneurs en MOS sur cette période à l'échelle du Québec.

La Figure 6 présente les distributions statistiques des teneurs en MOS par zone agroclimatique. La région 1 (Montréal, Montérégie, Outaouais) considérée comme la plus chaude, présente la teneur moyenne en MOS la plus faible avec une moyenne pour la région de 4,5 %. À l'autre extrême, la zone 6 (Gaspésie, Côte nord) avec des températures moyennes les plus froides et un bilan hydrique (P-ETP) nettement plus élevé que dans la zone 5 (nord de l'Abitibi et du Lac-Saint-Jean) présente des teneurs moyennes en MOS nettement plus élevées soit de 8,7 %. L'étude en cours sur l'état de santé de sols agricoles du Québec révèlent des teneurs moyennes plus faibles en MOS de 3,5 % dans la zone 1 et de 5,4 % dans la zone 5, basé sur des analyses de C total par combustion sèche et un facteur d'équivalence de 58 %.

Sur la base des fractions granulométriques prédites grâce à notre modèle présenté à la section 1.2, les classes texturales de chaque échantillon ont été identifiées et nous ont permis de grouper les échantillons en 5 groupes texturaux (Tableau 3). Ce regroupement des classes texturales a été proposée pour réduire le nombre de type de sols et simplifier l'interprétation des résultats tout en conservant les caractères distinctifs de ces groupes en termes de comportement agroenvironnemental (Catherine Bossé et Lucie Grenon, comm. pers.).

Les analyses de sol déclarées dans la zone 6 (Gaspésie, Côte nord) ne semblent pas révéler de sols argileux, ni limoneux. Les sols limoneux ont le plus faible nombre d'échantillons en opposition aux sols loameux qui représentent plus de la moitié des échantillons de la base de données (Figure 7).

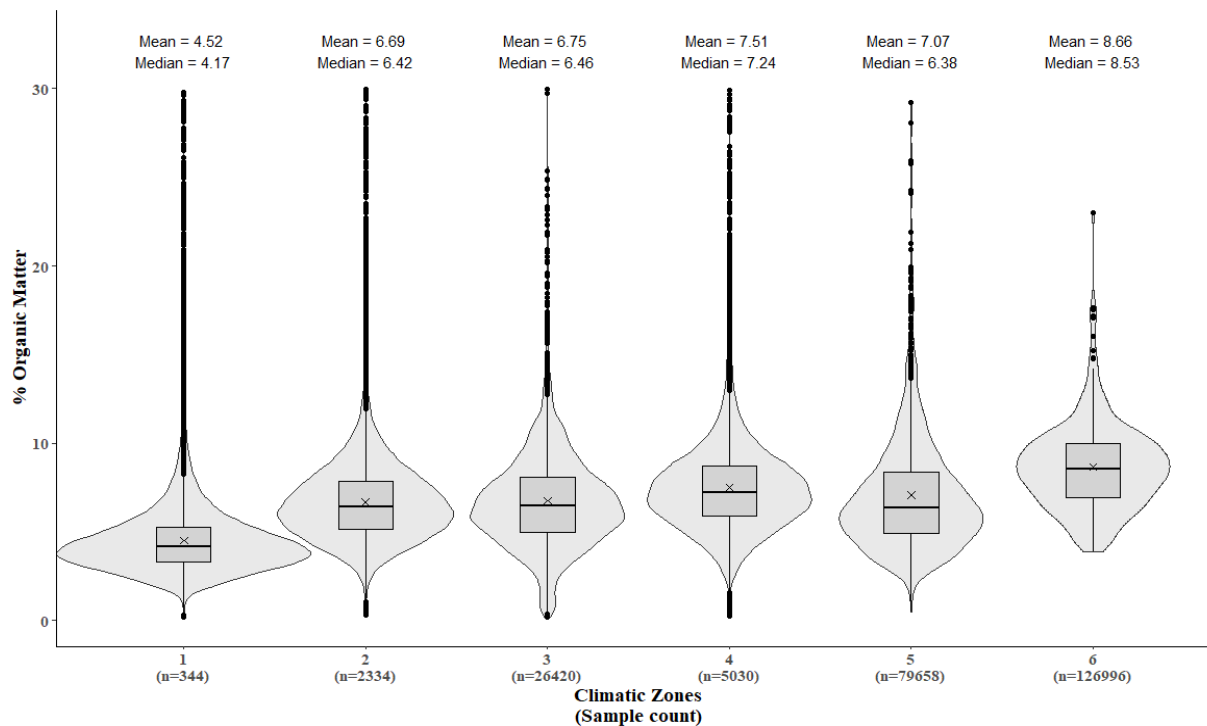


Figure 6 : Distributions statistiques en violon et box-plot des teneurs en matière organique des sols d’AEL par zone climatique. Le x représente la moyenne. La ligne du bas, du centre et du haut de la boîte représentent le 25^e, 50^e (médiane) et 75^e percentiles des observations, respectivement.

Tableau 3 : Définition des groupes texturaux des sols minéraux.

<i>Groupes texturaux</i>	<i>Abréviations</i>	<i>Classes texturales</i>	<i>Abréviations</i>
Argileux	A	Argile sableuse	AS
		Argile limoneuse	ALi
		Argile	A
		Argile lourde	AlO
Loameux argileux	LA	Loam argileux	LA
		Loam limono-argileux	LLiA
		Loam sablo-argileux	LSA
Limoneux	Li	Loam limoneux	LLi
		Limons	Li
Loameux	L	Loam	L
		Loam sableux	LS
Sableux	S	Sable loameux	SL
		Sable	S

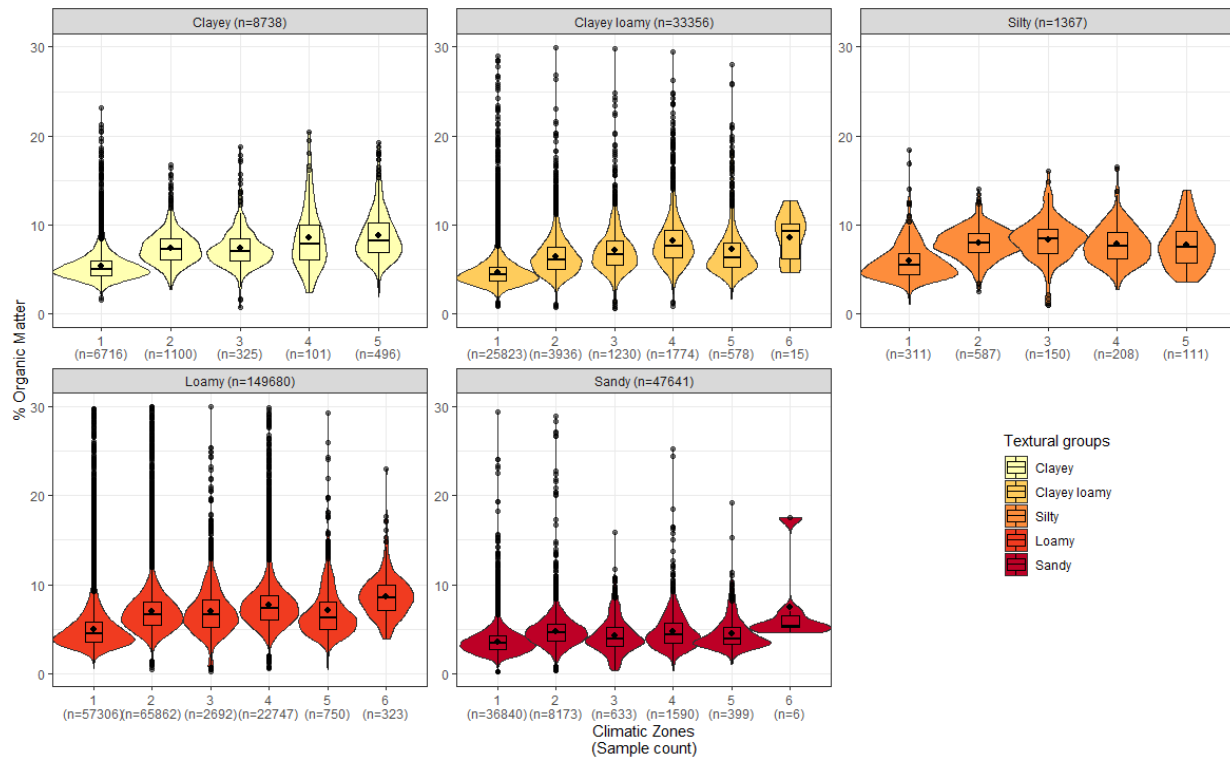


Figure 7 : Répartition de la teneur en matière organique des sols par zones et groupes texturaux.

2 Données climatiques

2.1 Source des données et des simulations en climat futur

Le consortium sur la climatologie Ouranos a fourni, pour les trois variables climatiques (température minimale, maximale et précipitation journalière), des données historiques (1950 – 2017) permettant de générer une période de référence de 30 ans (1981 – 2010), et des données modélisées provenant de différentes simulations climatiques pour des horizons futurs de 2011 – 2040, 2041 – 2070 et 2071 – 2100 à une résolution spatiale d'environ 10 km. Pour les données historiques et futures, les fichiers raster recouvraient la partie sud du 50^{ème} parallèle pour le Québec et l'Ontario (Figure 8). Pour produire des indicateurs climatiques sur la zone cultivée du Québec, nous avons appliqué le masque utilisé sur le site Agrométéo Québec (<https://www.agrometeo.org/>) et défini par Audet et al. (2012).

Au total, 22 simulations climatiques ont été fournies pour la période de référence et les trois horizons futurs : soit 11 modèles de simulation intégrant deux scénarios d'émission de gaz à effet de serre : le scénario d'émission modéré RCP 4.5, suppose une stabilisation des émissions de gaz à effet de serre d'ici la fin du siècle, tandis que le scénario d'émission élevé RCP 8.5, suppose une augmentation des émissions de GES.

Les trois sources d'incertitudes des projections climatiques (la variabilité naturelle du climat, les imprécisions des modèles climatiques et les émissions futures de GES) reflètent la complexité de prédire le climat futur et de sélectionner quelle simulation climatique ou quel scénario d'émission sera le plus près de la réalité dans un horizon proche ou plus lointain. Ainsi, il apparaît important de considérer un ensemble de simulations climatiques (22) afin d'avoir une représentativité des différentes réponses possibles des modèles et de couvrir l'incertitude associées aux simulations climatiques.

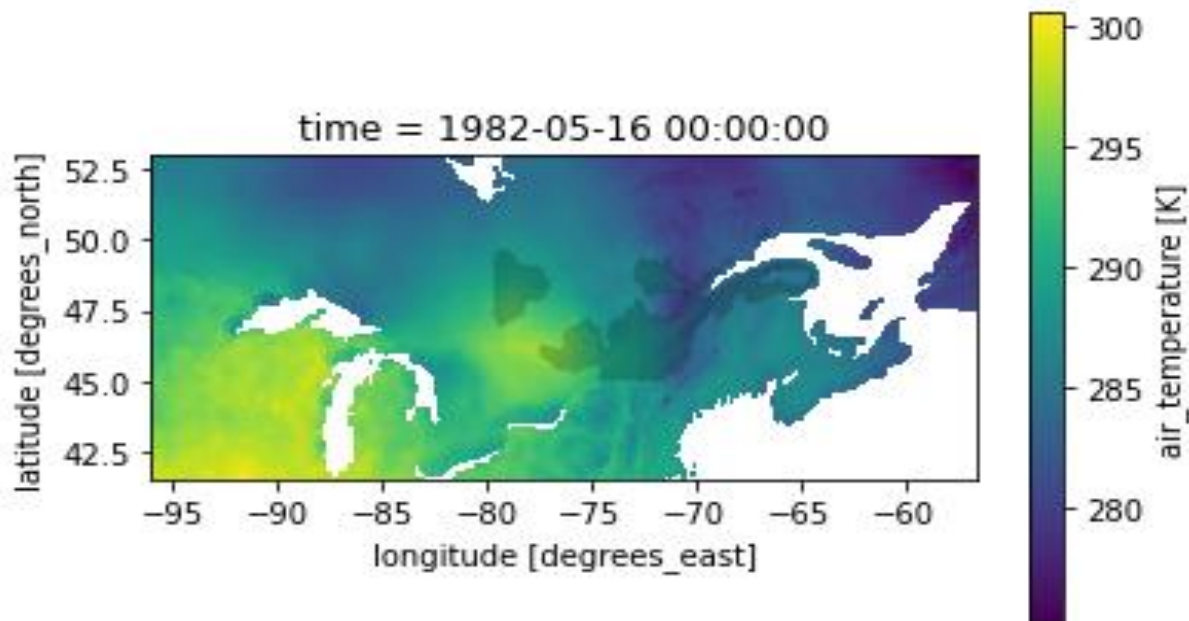


Figure 8 : Couverture spatiale des données climatiques avec le masque de la zone cultivée du Québec sur une grille de 10 km (exemple de la température maximale d'une journée).

L'analyse préliminaire et la classification faite chez OURANOS, a révélé que ces 22 simulations climatiques permettent une bonne représentation de la plage des multiples scénarios climatiques prédits dans le futur. La classification des 22 simulations (Tableau 4) se base sur 20 critères déterminés à priori. Les 20

critères/indicateurs utilisés incluent le changement (Δ) de la température moyenne et de la précipitation journalière, calculé sur une année et sur les 4 saisons, puis pour 2 horizons futurs (h50 = [2041,2070], h80 = [2071,2100]). Des détails supplémentaires sur la méthode utilisée pour la classification des simulations climatiques sont disponibles chez Ouranos.

Tableau 4 : Classification des simulations climatiques et origine des Earth System Models utilisés (ESM - CMIP5).

Rang	Simulations climatiques	Origine du modèle
1	CanESM2_RCP45	Canada
2	HadGEM2-CC_RCP85	UK
3	CMCC-CMS_RCP85	Italie
4	INM-CM4_RCP45	Russie
5	ACCESS1-3_RCP45	Australie
6	BNU-ESM_RCP85	Chine
7	INM-CM4_RCP85	Russie
8	IPSL-CM5A-LR_RCP45	France
9	IPSL-CM5A-LR_RCP85	France
10	CMCC-CMS_RCP45	Italie
11	MPI-ESM-LR_RCP85	Allemagne
12	GFDL-ESM2M_RCP45	USA
13	IPSL-CM5B-LR_RCP85	France
14	ACCESS1-3_RCP85	Australie
15	MPI-ESM-LR_RCP45	Allemagne
16	HadGEM2-CC_RCP45	UK
17	CanESM2_RCP85	Canada
18	NorESM1-M_RCP85	Norvège
19	NorESM1-M_RCP45	Norvège
20	IPSL-CM5B-LR_RCP45	France
21	BNU-ESM_RCP45	Chine
22	GFDL-ESM2M_RCP85	USA

2.2 Indicateurs climatiques

Onze indicateurs climatiques ont été dérivés et calculés sous forme de normale climatique sur 30 ans pour la période de référence et pour les trois horizons futurs : températures minimale, maximale et moyenne annuelle, précipitation journalière, ETP journalière calculée avec l'équation de Baier-Robertson recalibrée pour la vallée du Saint-Laurent (Rochette et Dubé, 1989) (Équation 1), le bilan hydrique (P-ETP), la longueur de la saison de croissance, les degrés jour de croissance, le nombre de jour de gel consécutif, le *Climate Decomposition Index* (CDI) (Adair et al., 2008) calculé sur une saison de croissance (Équation 2) et enfin le ratio de la température moyenne sur le bilan hydrique ($T_{\text{moy}}/\text{bilan hydrique}$), un ratio inverse à celui du CDI qui implique un effet multiplicatif de la température et du bilan hydrique.

Équation 1 :

$$ETp = -2,40 + 0,065 T_{\text{max}} + 0,083(T_{\text{max}} - T_{\text{min}}) + 0,00414R_A$$

ETp = évapotranspiration potentielle (mm/j)

T_{max} = température journalière maximale de l'air (°C)

T_{min} = température journalière minimale de l'air (°C)

R_A = radiation extra terrestre (cal/cm² j = Langleys)

Équation 2 :

$$CDI = F_T(T) \times F_W(PPT, PET)$$

$$F_T(T) = 0,5766 \times e^{308,56 \times \left(\frac{1}{56,02} - \frac{1}{(273+T)-227,13} \right)}$$

$$F_W(PPT, PET) = \frac{1}{1 + 30 \times e^{-8,5 \times \frac{PPT}{PET}}}$$

T = température moyenne mensuelle de l'air (°C)

PPT = précipitation mensuelle totale (mm)

PET = évapotranspiration potentielle mensuelle (mm)

La fonction du CDI décrit l'effet des variations de température et d'eau sur la décomposition du carbone. La fonction exponentielle décroissante utilisée pour qualifier l'effet de l'eau dans le sol sur la décomposition présuppose que cette dernière est principalement affectée par un manque d'eau. Ainsi à des niveaux très élevés de bilan hydrique (PPT/PET) reflétant un sol saturé en eau, le CDI approche sa valeur maximale et optimale de 1, alors que la décomposition pourrait théoriquement ralentir (Adair et al., 2008; Sierra et al. 2015).

Une analyse a été réalisée pour déterminer sur quelles bases temporelles les indicateurs climatiques devraient être calculés pour tenir compte de la longueur de la saison et le nombre d'années antérieures durant laquelle les processus biologiques affectent la dynamique de la matière organique du sol à l'échelle de la province. Pour faire cette analyse, le CDI a été calculé sur une base saisonnière en utilisant le concept des degrés-jours pour débuter et terminer la saison active durant laquelle ces processus opèrent. La température de base pour le calcul des degrés-jours a été modulée de -10 à +10°C par incrément de 5°C et le début et la fin de la saison ont été recalculés en tenant compte de la modulation de cette température de base. Pour évaluer, l'effet du nombre d'années antérieures sur les processus biologiques, le CDI a été calculé sur des périodes allant de 5 à 30 ans antérieures aux données de MOS par incrément de 5 ans.

L'analyse de corrélation entre le CDI calculé sur différentes bases saisonnières et de nombre d'années et la teneur en matière organique du sol (MOS) a démontré qu'à l'échelle de la province, le CDI est mieux corrélé (relation négative) à la MOS, lorsqu'il est calculé sur une saison avec un température de base de -5°C et des données climatiques sur 30 ans (Figure 9). Cela signifierait que les processus biologiques qui affectent l'évolution de la MOS sont toujours actifs lorsque la saison est définie par des températures allant sous le gel et qu'ils sont conditionnés à plus long terme par le climat.

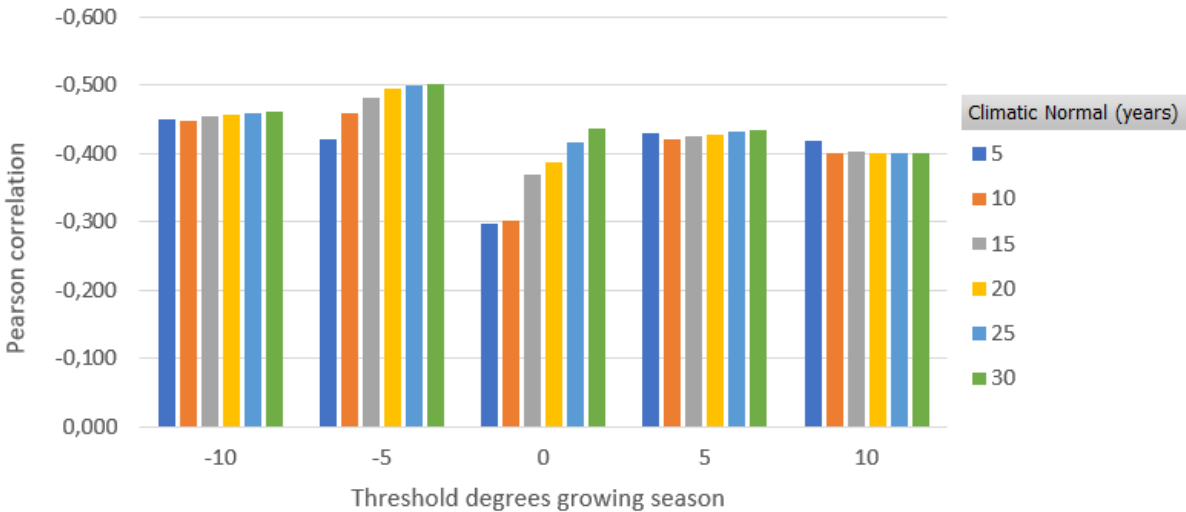


Figure 9 : Corrélation entre la MOS et le CDI calculé sur des saisons de croissance variant en termes de températures de base pour le calcul des degrés-jours (-10, -5, 0, 5, 10 °C) et du nombre d’années pour le calcul des normales climatiques (5, 10, 15, 20, 25, 30 ans).

2.3 Découpage agroclimatique du territoire

L’approche HCPC (*Hierarchical Clustering on Principal Components*) ou Classification hiérarchique de composantes principales a été utilisée pour découper la zone cultivée du Québec en zones agroclimatiques différenciées et homogènes. Cette approche combine trois méthodes utilisées en analyses multivariées (Husson et al., 2010) :

- Les méthodes en composantes principales (ACP, AFC, ACM, AFDM, AFM);
- La classification ascendante hiérarchique;
- Le partitionnement en k-moyennes.

Le jeu de données initial comprenait onze indicateurs climatiques répartis en données matricielles (grille de 10 km) à l’échelle de la zone cultivée du Québec. L’analyse en composantes principales (ACP) a été utilisée pour réduire le nombre de variables dans un nombre restreint de dimensions orthogonales à utiliser dans la classification hiérarchique. L’ACP a permis de réduire le bruit de fond dans les données et donc de produire une classification *a priori* plus stable. L’études des valeurs propres (*eigenvalues*) ou la variance des composantes principales a servi à sélectionner le nombre de composantes à conserver (Tableau 5).

La somme de toutes les valeurs propres (*eigenvalues*) étant égale à 10, les deux premières composantes expliquaient à elles seules 93 % de la variance totale. La première composante traduit l’effet de la température (temp. moy., degrés-jour, longueur de la saison de croissance) et la deuxième composante traduit l’effet de l’eau (bilan hydrique, précipitations). Ces deux premières composantes ont été retenues pour la classification hiérarchique.

La *classification ascendante hiérarchique (CAH)* a été utilisée pour identifier des groupes (i.e, clusters) d’observations similaires où le climat est plus homogène, c’est à dire où les corrélations entre les composantes principales formées des indices climatiques sont plus importantes permettant de former différents nombres de clusters.

Tableau 5 : Valeurs propres et pourcentages de variance de l'ACP.

<i>Composante</i>	<i>Valeur propre (eigenvalue)</i>	<i>% de variance expliquée</i>	<i>% de variance cumulée</i>
Comp 1	7.22	72.19	72.19
Comp 2	2.10	20.96	93.15
Comp 3	0.47	4.67	97.83
Comp 4	0.12	1.18	99.01
Comp 5	0.008	0.80	99.81
Comp 6	0.01	0.14	99.95
Comp 7	0.00	0.04	99.99
Comp 8	0.00	0.01	100.00
Comp 9	0.00	0.00	100.00
Comp 10	0.00	0.00	100.00

Le nombre optimal de cluster a été déterminé par la méthode « Elbow » qui selon l'allure de la courbe permet de déterminer le nombre optimal de cluster à conserver (Figure 10). L'allure de la courbe indiquait de conserver quatre groupes, mais afin d'avoir un découpage plus fin dans certaines zones (par exemple dans la région d'Abitibi-Témiscamingue), la carte a été découpée en six zones agroclimatiques (Figure 11). Le Tableau 6 présente les normales de cinq indicateurs climatiques pour les six zones.

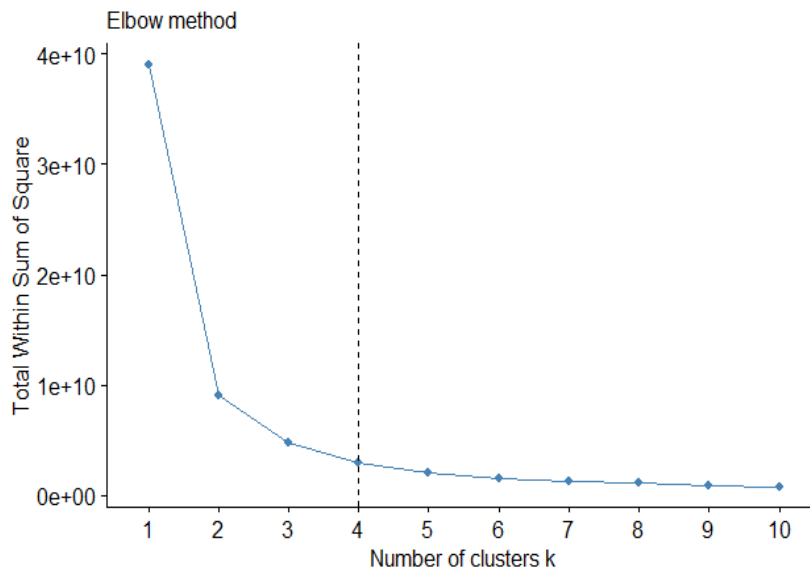


Figure 10 : Nombre optimal de clusters.

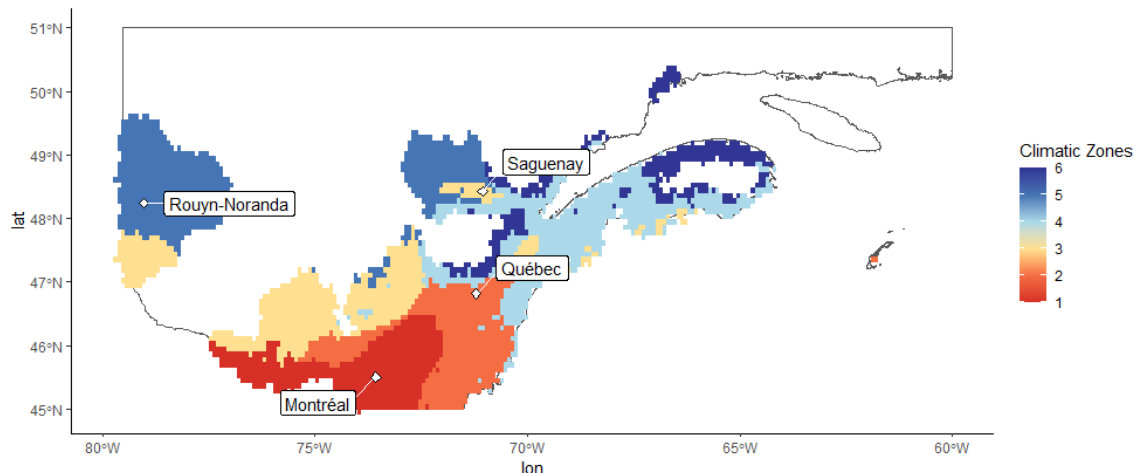


Figure 11 : Découpage agroclimatique du Québec en 6 zones.

Tableau 6 : Normales climatiques de 30 ans (1981-2010) par zone agroclimatique du Québec.

Zone Agroclimatique	T_{moy}^a (°C)	P_{cp}^a (mm)	ETP^a (mm)	DSC^a (jours)	$P-ETP^a$ (mm)
1	5.8	1044	639	199	404
2	4.7	1021	572	186	448
3	3.6	915	568	176	347
4	2.9	1026	533	169	493
5	2.2	877	565	170	311
6	1.8	964	462	158	501

^a, T_{moy} : température moyenne; P_{cp} : précipitation totale; ETP : évapotranspiration potentielle; DSC : durée de la saison de croissance; $P-ETP$: bilan hydrique résultant de précipitation totale - ETP .

- Zone 1 : Montréal – Montérégie – Outaouais
- Zone 2 : Laval – Centre du Québec – Estrie
- Zone 3 : Mauricie – Laurentides – Témiscamingue
- Zone 4 : Bas Saint-Laurent - Baie-des-chaleurs
- Zone 5 : Abitibi – Lac St-jean
- Zone 6 : Gaspésie nord– Côte-Nord

2.4 Simulations du climat futur sur les zones agroclimatiques du Québec

Les onze indicateurs climatiques ont été recalculés et projetés dans le futur sur les six zones agroclimatiques du Québec en utilisant les 11 modèles climatiques et les deux scénarios d'émissions de GES (RCP 4.5 et 8.5) fournis par Ouranos. La Figure 12 présente les projections de quatre indicateurs climatiques d'intérêt.

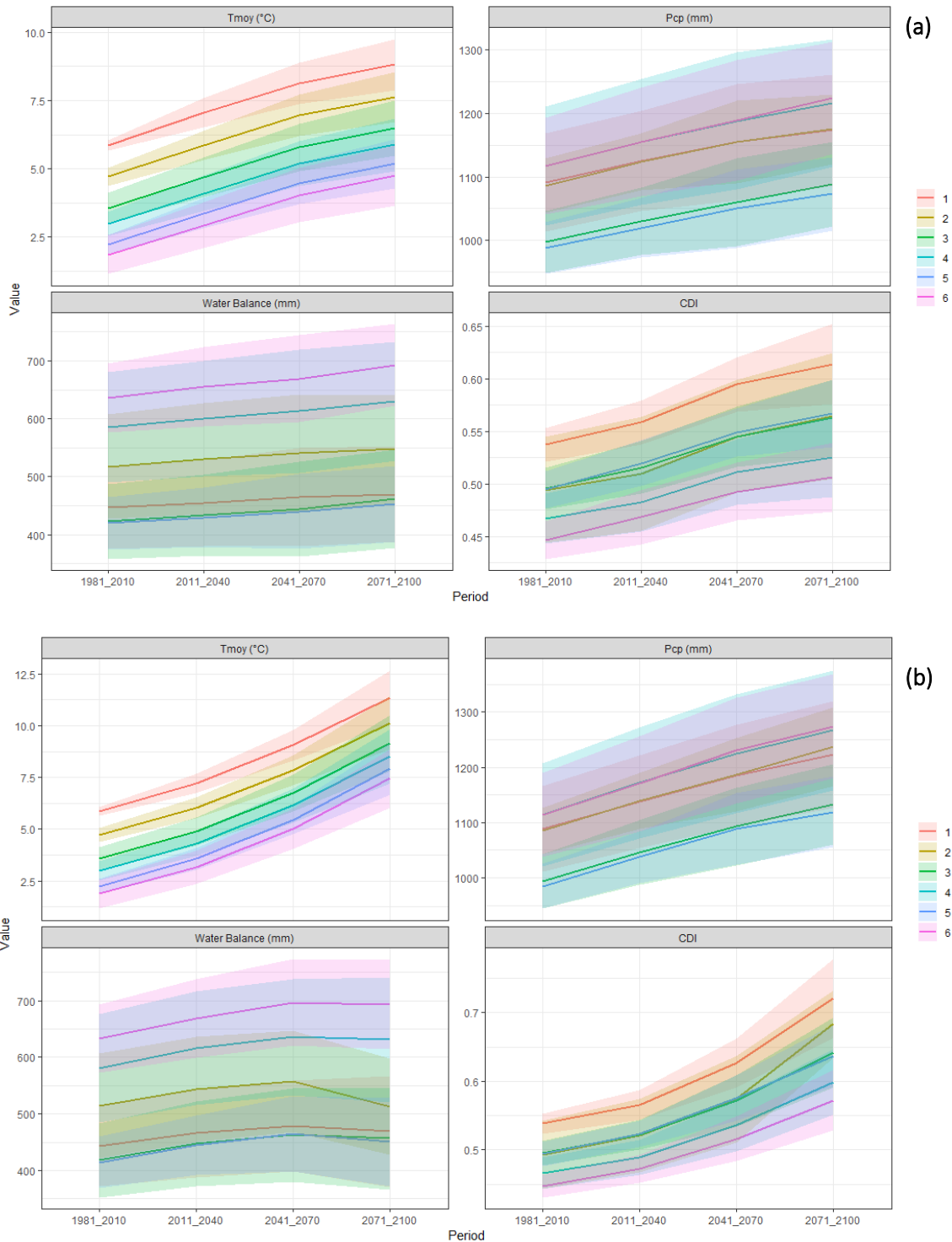


Figure 12 : Projections en climat futur de quatre indicateurs climatiques dans six zones agroclimatiques en RCP 4.5 (a) et RCP 8.5 (b). Les lignes pleines et les parties ombragées représentent respectivement la moyenne et l'écart-type des simulations de 11 modèles climatiques.

Sous ces scénarios du climat futur, la température moyenne annuelle devrait passer de 5,8°C à l'horizon 1980-2010 à près de 8,3°C à l'horizon de 2070-2100 en RCP 4.5 ou 11°C en RCP 8.5 dans la zone agroclimatique 1 la plus chaude, tandis qu'en région plus froide, cette variation serait de 1,8°C à près de 5°C en RCP 4.5 ou 7,5°C en RCP 8.5. Les variations de température sont importantes dans les deux cas, mais paraissent plus importantes

en termes relatifs en zone plus froide. En termes de précipitations et de bilan hydrique, les variations moyennes apparaissent moins importantes sur la période entre 1980-2010 et 2070-2100, mais les variations entre les 11 scénarios climatiques indiquées par les couleurs ombragées demeurent importantes. L'indicateur CDI qui intègre une composante température et bilan hydrique varie de 0,45 à 0,55 en climat actuel, de 0,5 à 0,6 en climat futur de 2070-2100 sous RCP 4.5 et de 5,7 à 7,2 pour la même période sous RCP 8.5. Il faut d'ores et déjà anticiper que les modèles calibrés sur des données actuelles de climat vont simuler des résultats en climat futur dans certaines zones (notamment la zone 1 en RCP 4.5 et les zones 1, 2, 3 et 4 en RCP 8.5) dans un domaine extérieur au domaine de calibration et donc plus prompt à générer des erreurs.

3 Occupation des sols par les cultures

L'occupation des sols a été définie à partir des cultures déclarées dans la base de données des parcelles et productions agricoles déclarées (BDPPAD) à la Financière Agricole du Québec (FADQ). Chaque année la FADQ relève les cultures déclarées par les producteurs sur chaque parcelle agricole assurée. Les parcelles agricoles sont géoréférencées sous format vectoriel (Figure 13) et les données des cultures et productions déclarées sont colligées dans des tables attributaires. L'information colligée est disponible à partir de 2003 sur le site internet de la FADQ.

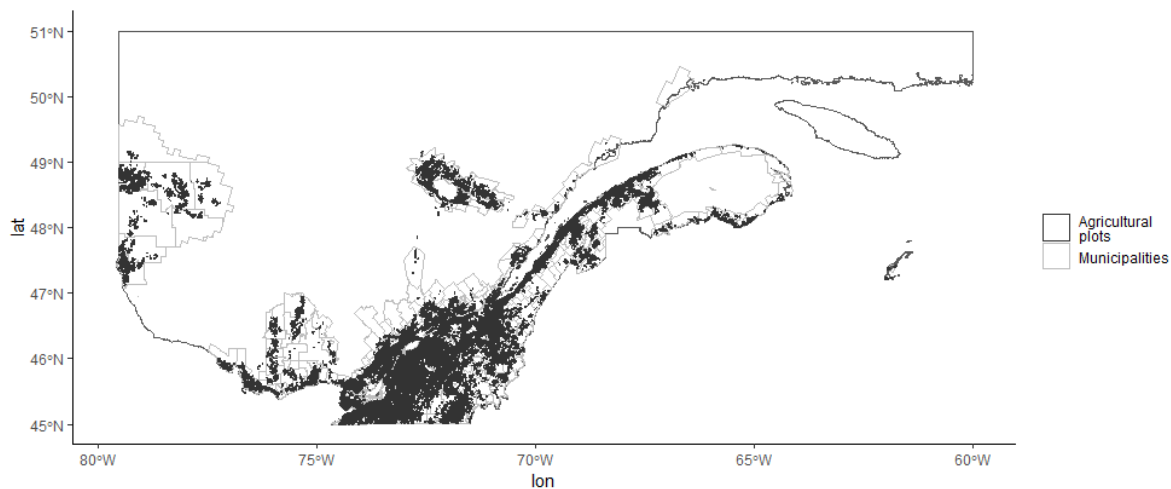


Figure 13 : Polygones des parcelles agricoles déclarées et des municipalités.

Plusieurs études ont tenté de relier certaines catégories de cultures à l'évolution des propriétés des sols dont la teneur en matière organique des sols. Par exemple, Gubler et al. (2019) ont utilisé les catégories prairies et pâturages, céréales, cultures sarclées (maïs, betterave, colza et pommes de terre) et autres cultures maraichères pour établir des relations avec l'évolution de la teneur en MOS. La présence de prairie dans la rotation était corrélée positivement à la teneur en MOS, tandis que les autres catégories de cultures corrélaient de façon négative. Meurer et al. (2018) ont attribué des indices de culture variant de 0 à 2 pour 11 catégories de cultures passant de la jachère sans culture à la prairie. Cet indice de culture est un concept simplifié pour estimer la contribution relative des biomasses aériennes et racinaires des différentes cultures aux apports de C

au sol (Bolinder et al., 2007 et Bolinder et al., 2015). Suivant la même logique, les cultures déclarées à la FADQ ont été regroupées dans les catégories de cultures présentées au Tableau 7.

Tableau 7 : Définition des catégories de cultures et leur indice de culture.

<i>Catégorie de cultures</i>	<i>Indice de culture</i>
Prairies et pâturages	2
Maïs-grain, et autres graminées en C4	1,5
Petites céréales (avoine, blé, orge, seigle), canola, lin, céréales fourragères, engrais vert, chanvre et autres cultures avec retour modéré de C au sol	1
Maïs-ensilage	0,75
Soya	0,75
Cultures de légumineuses	0,75
Cultures maraichères, pommes de terre, tabac et autres cultures avec peu de retour de C au sol	0,5

Les cultures déclarées à la FADQ ont été regroupées dans les différentes catégories de cultures présentées ci-haut pour calculer des ratios moyens d'occupation de sol pour chaque catégorie de grandes cultures par municipalité couvrant les années 2003 à 2020 (Figure 14).

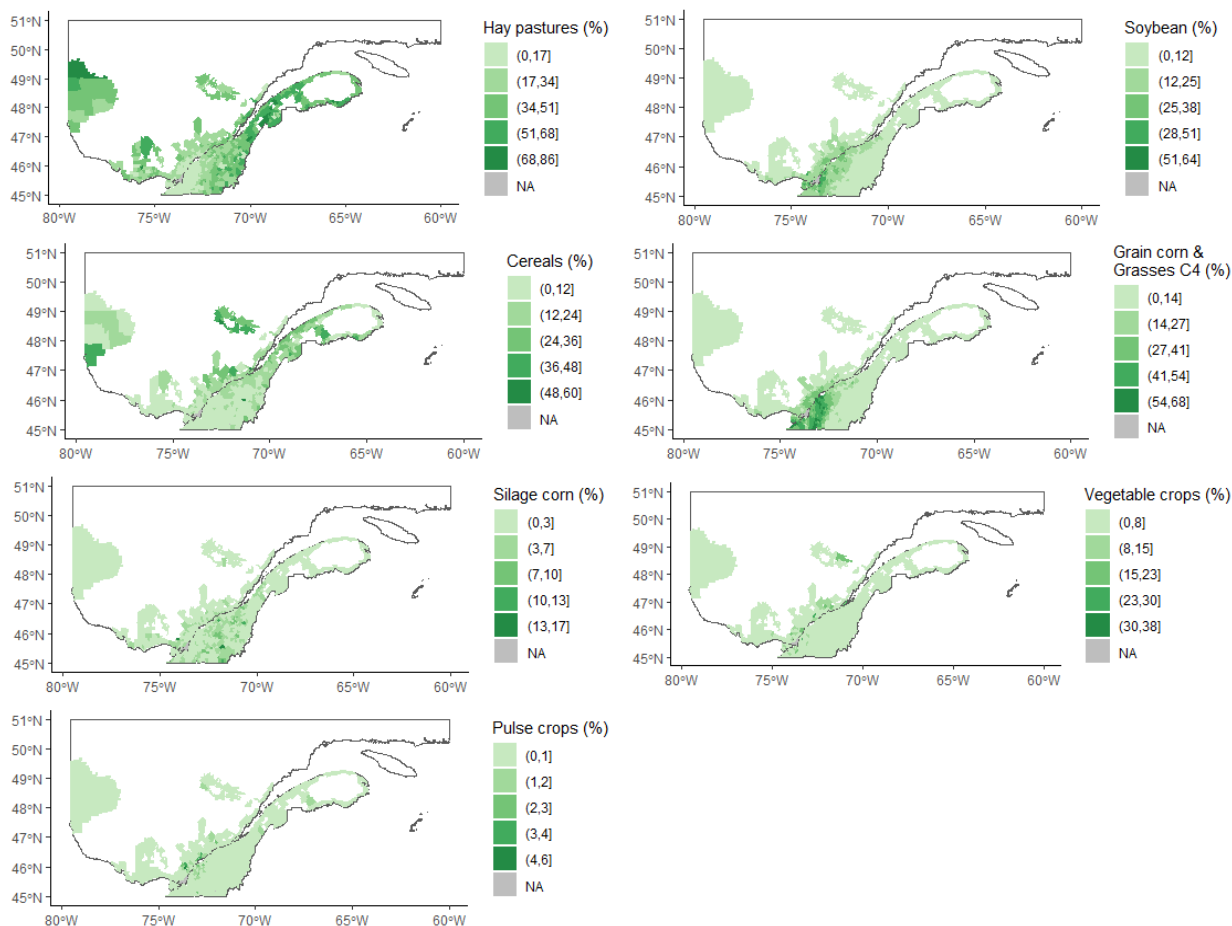


Figure 14 : Pourcentage moyen d'occupation des grandes cultures par municipalité de 2003 à 2020.

Sur la base de ces catégories de cultures, un ratio de cultures pérennes (prairies, pâturages) sur le total des superficies cultivées occupant le sol de 2003 à 2020 a aussi été calculé pour chaque municipalité. La présence de cultures annuelles est beaucoup plus élevée dans le sud de la province avec un climat plus chaud dans les régions autour de Montréal comme la Montérégie ou le Centre-du-Québec, alors que les cultures pérennes sont beaucoup plus présentes dans les régions au relief plus accidenté et plus froides comme en Estrie et Chaudière-Appalaches, et dans les régions plus froides plus au nord de la province (Figure 15).

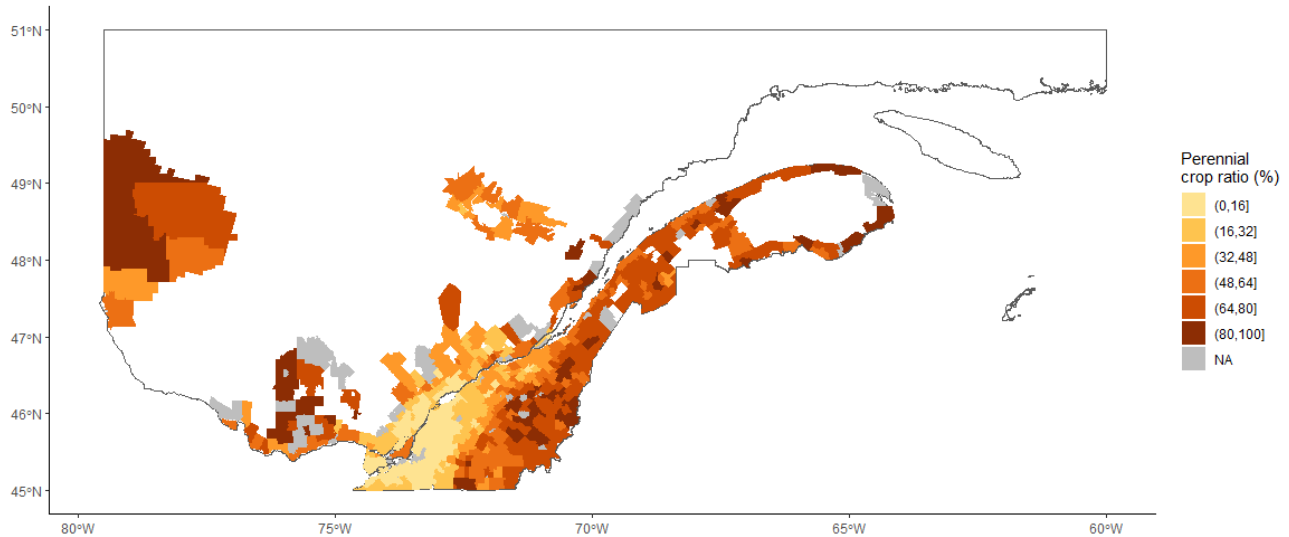


Figure 15 : Ratio moyen de cultures pérennes sur cultures annuelles (2003-2020) par municipalité.

4 Variables dérivées du modèle numérique de terrain

Le modèle numérique de terrain (MNT) (disponible à l'échelle 1/20 000 pour l'ensemble du Québec) est dérivé de relevés LIDAR et de photographies aériennes traités pour intégrer certaines données hypsométriques et hydrographiques. Des éléments de rupture, les lacs, les rivières, et autres infrastructures propres à modifier le paysage, ont été pris en compte afin d'affiner le produit final. Les données avaient une résolution spatiale initiale de 10 m et ont été agrégées au 250 m pour accélérer les traitements à l'échelle de la province. Le package RSAGA sur le logiciel R, faisant appel à des fonctionnalités de SAGA GIS et donnant accès aux fonctions de géotraitement et d'analyse de terrain (Conradt et al., 2015) ainsi que les outils de la suite WhiteboxTools (Lindsay, 2014) ont été utilisés pour dériver 61 covariables du MNT à être utilisées dans les modèles empiriques pour prédire l'évolution spatio-temporelle de la MOS. La Figure 16 présente quatre de ces covariables dérivées du MNT incluant l'élévation de départ.

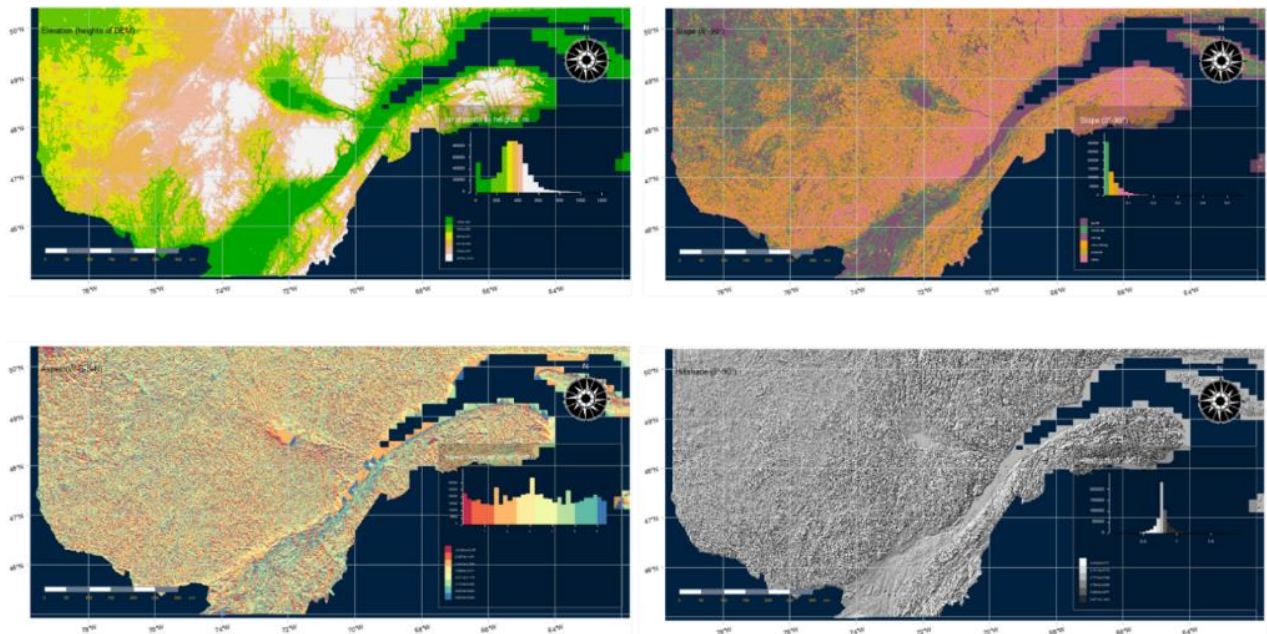


Figure 16 : Exemple de covariables dérivées du MNT : Élévation (haut gauche), pente (haut droite), aspect (bas gauche) et ombrage (bas droite).

Résultats

5 Modélisation de l'évolution de la matière organique du sol en climat futur

En rappel, les échantillons de sol d'AEL ont été géoréférencés par codes postaux municipaux. Pour ramener les autres variables (ratios de culture, climat et MNT) sur cette même base et en posant l'hypothèse que les échantillons de sol d'AEL proviennent majoritairement de zones cultivées, des polygones ont été créés par municipalité en regroupant les polygones des zones cultivées déclarées à la FADQ et en créant un polygone convexe autour de ces zones cultivées. Les 11 indicateurs climatiques calculés sur grille de 10 km, les covariables associées au modèle numérique de terrain (MNT) calculées sur une grille de 250 m et les ratios de cultures calculés sur une grille de 1 km ont été intersectés sur ces polygones convexes regroupant les zones cultivées par municipalité et leurs valeurs moyennes sont rapportées par polygone convexe de municipalité. L'information sur le climat, le MNT et les ratios de culture, pour par exemple, 1000 échantillons de sols contenus dans une municipalité est donc redondante. Les packages Recipe et Caret ont été utilisés dans le logiciel R pour normaliser et standardiser les variables prédictives avant de les inclure dans les modèles de régression (R Core Team 2022).

5.1 Corrélations simples entre la teneur en MOS et les autres variables incluses dans le modèle

Une première analyse de matrice de corrélations de Pearson entre les différentes variables a permis de mettre en relief les différentes corrélations entre la variable cible, la teneur en MOS et quelques variables les plus corrélées en fonction des cinq groupes texturaux (Figure 17).

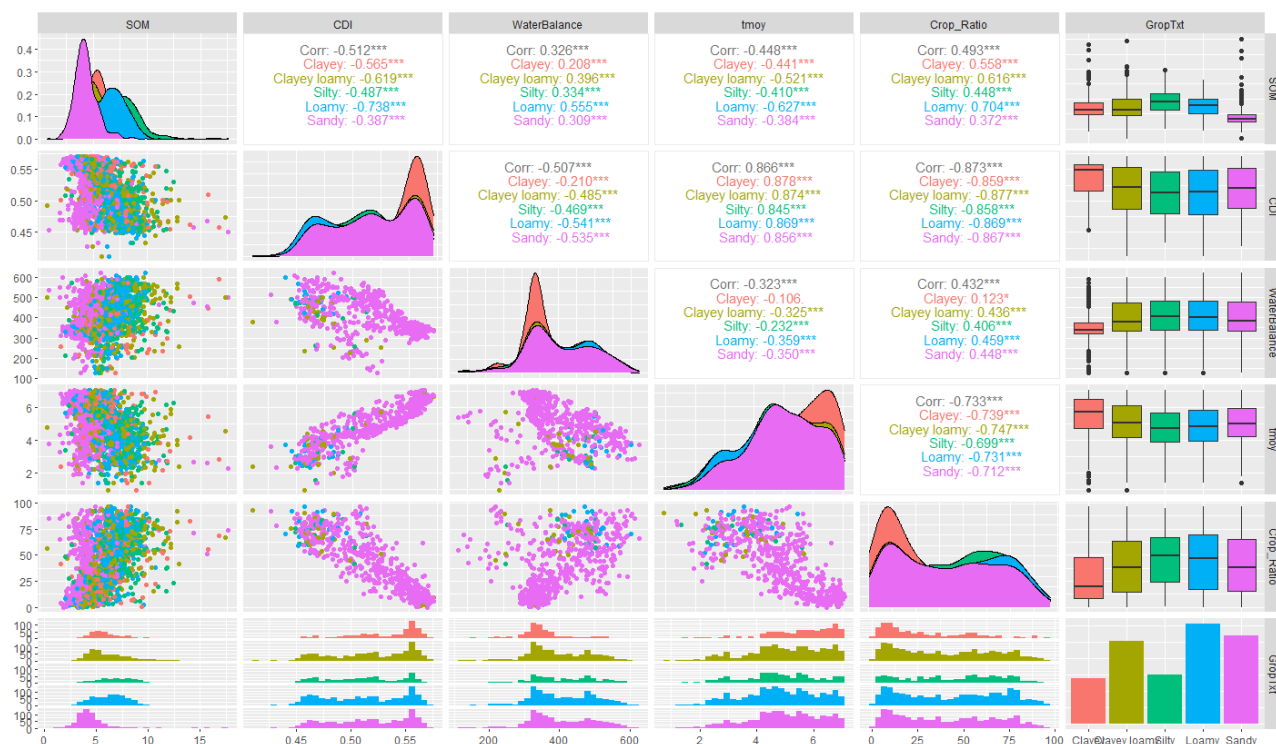


Figure 17 : Matrice de corrélation entre matière organique du sol (SOM), indicateur climatique CDI, bilan hydrique (*waterbalance*), température moyenne (*tmoy*) et ratio de cultures pérennes (*crop_ratio*), en moyenne (en gris) et par groupes texturaux (*GropTxt*) (en couleur).

Le *Climate Decomposition Index* (CDI) est l'indicateur le plus corrélé à la MOS avec une relation négative et un coefficient moyen de Pearson de -0,51 pour l'ensemble des sols et de -0,73 avec les sols loameux. Le CDI est formé de deux composantes modélisant la température et la teneur en eau dans le sol partagées dans plusieurs modèles simulant l'évolution du C dans les sols, dont Century et RothC (Sierra et al., 2015; Sierra et al., 2012; Adair et al., 2008).

Le ratio de cultures pérennes (*crop_ratio*) est la deuxième variable la plus corrélée avec la MOS, avec un coefficient moyen de 0,49. Cette corrélation entre le ratio de cultures pérennes et la MOS s'observe visuellement en comparant la Figure 18 ci-dessous et la carte des ratios de cultures pérennes présentée à la Figure 15.

Le CDI est aussi corrélé positivement à la température moyenne de l'air (0,87), mais négativement avec le ratio de cultures pérennes (-0,87). Les climats plus froids restreignent le choix des cultures annuelles et augmente la prépondérance des cultures pérennes.

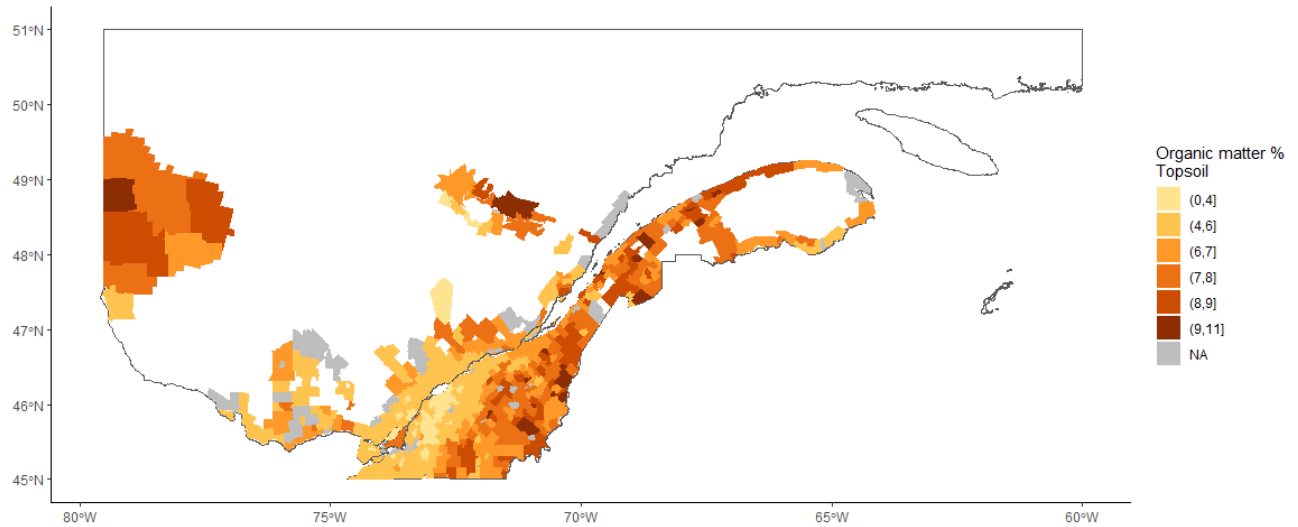


Figure 18 : Teneur moyenne en MOS (2007-2019) par municipalité.

5.2 Modélisation de la distribution spatiale de la matière organique du sol

Une approche de substitution de l'espace pour le temps (space for time substitution) est proposée pour prédire l'évolution de la matière organique du sol (MOS) en climat futur (Gray et al., 2016). Le modèle prédisant la distribution spatiale de la matière organique du sol (équation ci-dessous) intègre les ensembles de variables suivantes :

$$\% \text{ MOS} = f(\% \text{Sable} + \% \text{Argile} + \text{Indicateurs climatiques} + \text{covariables DEM} + \text{Ratio de culture})$$

Pour prédire la distribution spatiale de la matière organique du sol, les performances de sept modèles régressifs ont été analysées. Ces modèles comprenaient deux méthodes paramétriques *GLM* et *GAM* (modèles linéaires et additifs), et cinq méthodes non paramétriques en apprentissage machine (*XGBtrees*, arbres de régression boostés; *XGBlinear*, modèle multi-linéaire boosté; *Random Forest (Ranger)*, forêts d'arbres aléatoires; *Cubist*, modèle cubiste; et *KNN*, K plus proche voisin). Une méthode de validation reconnue en apprentissage automatique a été utilisée et des jeux de données indépendants ont été formés pour entraîner (calibration) et mesurer la performance des modèles (validation) (Figure 19). La sélection des covariables a été réalisée en s'inspirant des travaux de Lu et al., (2019) combinant l'analyse des corrélations de Pearson et la classification de l'importance des variables (*Feature importance*) dans les différents modèles.

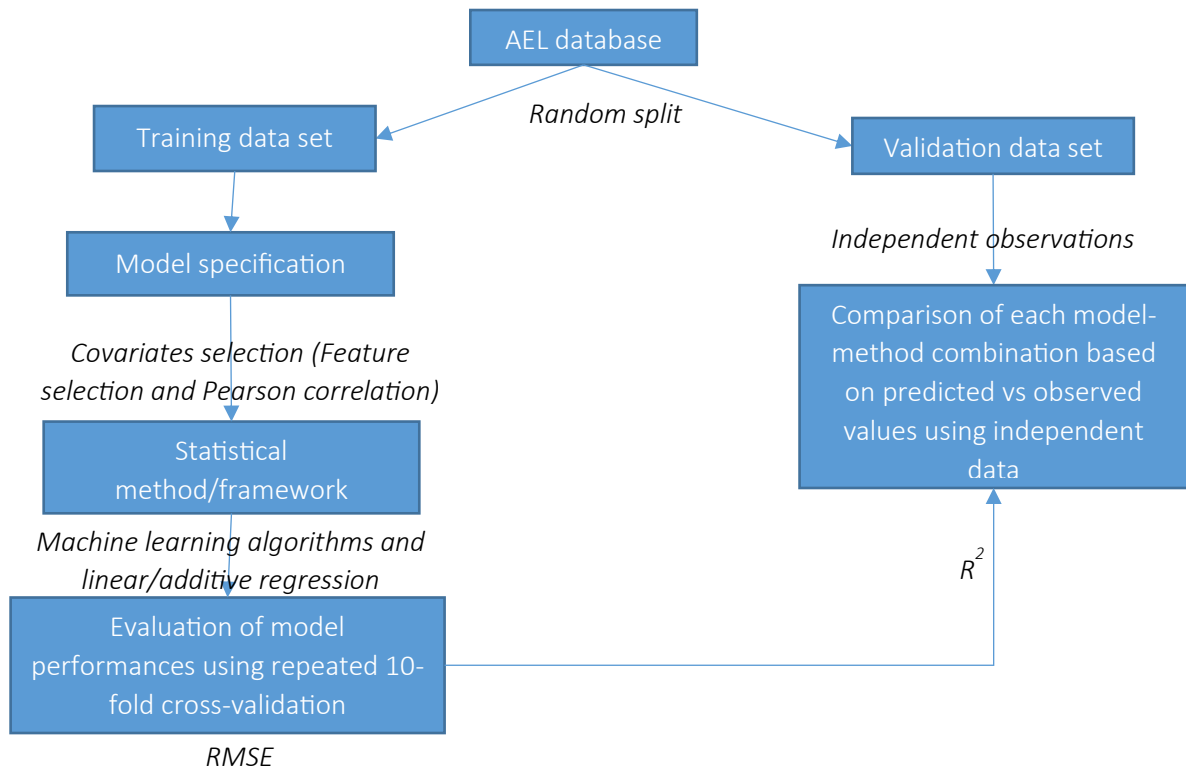


Figure 19 : Schéma des principales étapes du processus de modélisation utilisé.

Parmi les modèles non paramétriques en apprentissage automatique sélectionnés, deux techniques de modélisation d'ensemble ont retenu plus particulièrement notre attention : Les forêts aléatoires (RF) (*Random Forests*) (Breiman, 2001) et les arbres de régression boostés (*Extreme gradient boosting Trees* : XGBtrees) (Friedman, 2001, 2002). Ces deux techniques ont l'avantage de s'adapter aux valeurs aberrantes et de traiter les interactions entre les variables prédictives et les relations non-linéaires entre les variables prédictives et la variable prédite. Les prédictions du modèle RF ont été estimées sous forme de moyennes pondérées sur tous les arbres créés. Les poids des arbres sont calculés en fonction de la valeur prédite et la performance de chaque arbre, ce qui garantit que les meilleurs arbres contribuent le plus aux prédictions finales (Hastie et al., 2009). Deux hyperparamètres de RF ont été testés et calibrés avec différentes valeurs : le nombre maximum d'arbres utilisés, et le nombre de variables prédictives utilisées pour chaque arbre. Le modèle XGBtrees combine un grand nombre de modèles d'arbres de régression de manière adaptative afin d'optimiser les performances prédictives (Friedman, 2002). Il diffère d'autre technique d'arbre de décision car il utilise un algorithme de *boosting* permettant de minimiser l'erreur sur une fonction de perte moyennée sur un entraînement aléatoire. XGBtrees dispose de multiples paramètres calibrés par étapes hiérarchisées (taux d'apprentissage, profondeur maximale et poids minimale, indice Gamma, etc.). Les multiples étapes de paramétrisation des hyperparamètres du modèle XGBtrees ne sont pas détaillés ici, mais des informations détaillées peuvent être trouvées dans Elith et al. (2008) et Kuhn et Johnson (2013).

Tous les modèles ont été calibrés avec 75 % des observations de la base de données AEL par validation croisée et validés avec 25 % de la base de données restante (Figure 19). Pour des raisons de projections aberrantes avec le jeu de données de validation, le modèle GAM n'est pas présenté dans la figure 21. Pour tous les modèles, l'erreur quadratique moyenne (RMSE) et son intervalle de confiance à 95% ont été calculés avec la validation croisée *10-fold* réalisée à l'étape de la calibration du modèle (Figure 20). Le coefficient de détermination (R^2) a

été calculé entre les prédictions du modèle et les valeurs observées du jeu de données indépendant servant pour la validation. Tous les modèles ont été calibrés et validés à l'aide du package Caret (Kuhn, 2017) exécutés dans l'environnement R (R Core Team, 2022).

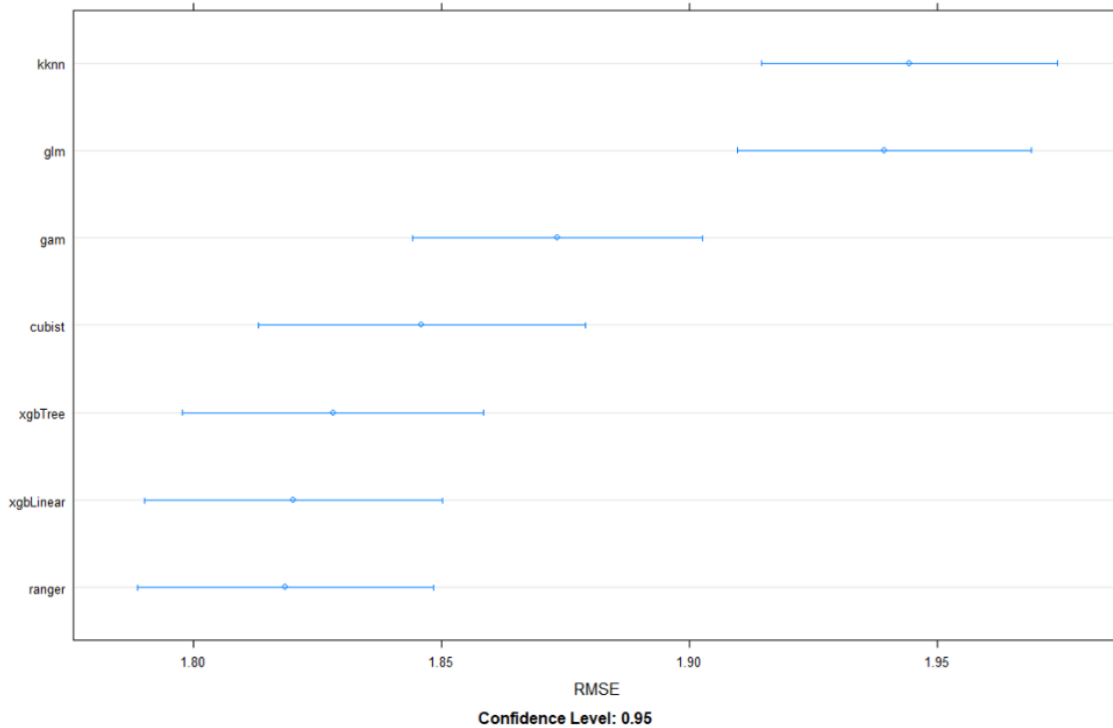


Figure 20 : Erreur quadratique moyenne (RMSE, %) et son intervalle de confiance à 95 % pour la prédiction de la matière organique du sol (MOS). Acronymes des méthodes statistiques : KKNN = k le plus proche voisin (méthode pondérée) ; GLM = modèle linéaire généralisé ; GAM = modèle additif généralisé ; Cubist = Algorithme cubiste ; XGBtree = Extreme gradient boosting trees ; XGBlinear = Extreme gradient boosting linear ; Ranger = forêts aléatoires.

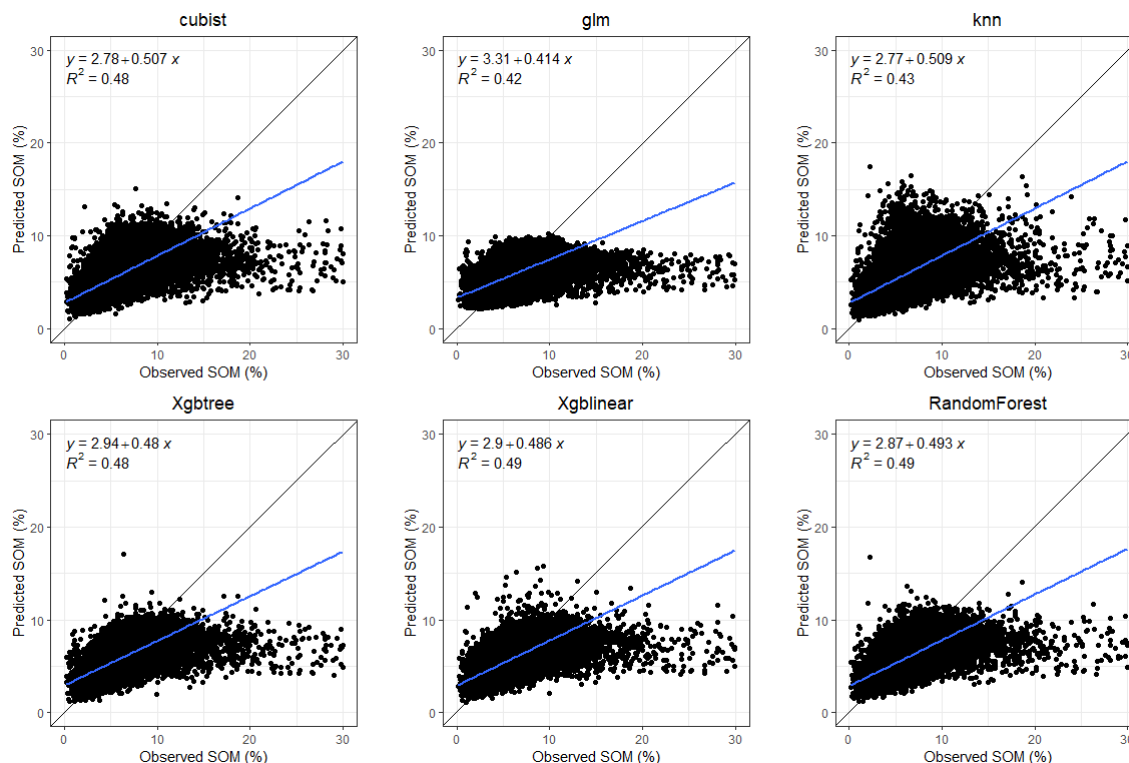


Figure 21 : Comparaison entre valeurs prédites et observées de la matière organique de sol (%) de six modèles testés sur le jeu de données indépendant (n = 61 167).

L'analyse comparative des performances des modèles a révélé que le choix de la technique d'apprentissage automatique ou du modèle de régression, n'influencent pas significativement les indicateurs de performance tels que le R² ou le RMSE. Les valeurs de R² sont légèrement plus faibles pour les modèles GLM et KNN (0,42 et 0,43) comparées aux autres modèles qui sont de l'ordre de 0,48. L'erreur quadratique (RMSE) est aussi plus élevée pour ces deux modèles (1,93 % pour les deux modèles). Le modèle de forêts aléatoires (RF) a produit les meilleures valeurs de R² (0,49) et de RMSE (1,82%), ce qui a conduit à son choix pour prédire l'évolution de la matière organique dans un climat futur. Toutefois, les modèles XGBlinear et XGBtree exhibaient des performances similaires et auraient aussi pu être utilisés.

La classification de l'importance des variables prédictives pour le modèle RF démontre une plus grande contribution des variables liées à la texture du sol (pourcentage de sable et d'argile), considérées plus stables dans le temps. Le ratio de cultures pérennes arrive en 3^{ème} position et enfin les deux indicateurs climatiques (CDI et Tmoy/Bilan hydrique) arrive en 4^{ème} et 5^{ème} position (Figure 22). Le CDI représente un produit des effets des composantes température et bilan hydrique dans le sol, tandis que la variable Tmoy/Bilan hydrique représente à l'inverse un ratio de température moyenne de l'air sur le bilan hydrique. Le modèle RF semble donner plus d'importance à ces formes d'interaction entre la température et le bilan hydrique pour expliquer la distribution de la MOS à l'échelle du Québec, que les autres variables liées à la température seulement (tmax, tmoy). Les covariables dérivées du DEM (swi, med3000, stanh, dem, etc.) et les autres indicateurs climatiques (GDD, tmax, tmoy) arrivent plus tard dans le classement de l'importance des variables. La variable swi est un indice d'humidité du sol (position basse sur le terrain) et med3000 est une différence d'élévation par rapport à une élévation médiane locale calculée sur 3000 x la largeur de la grille (Whitebox).

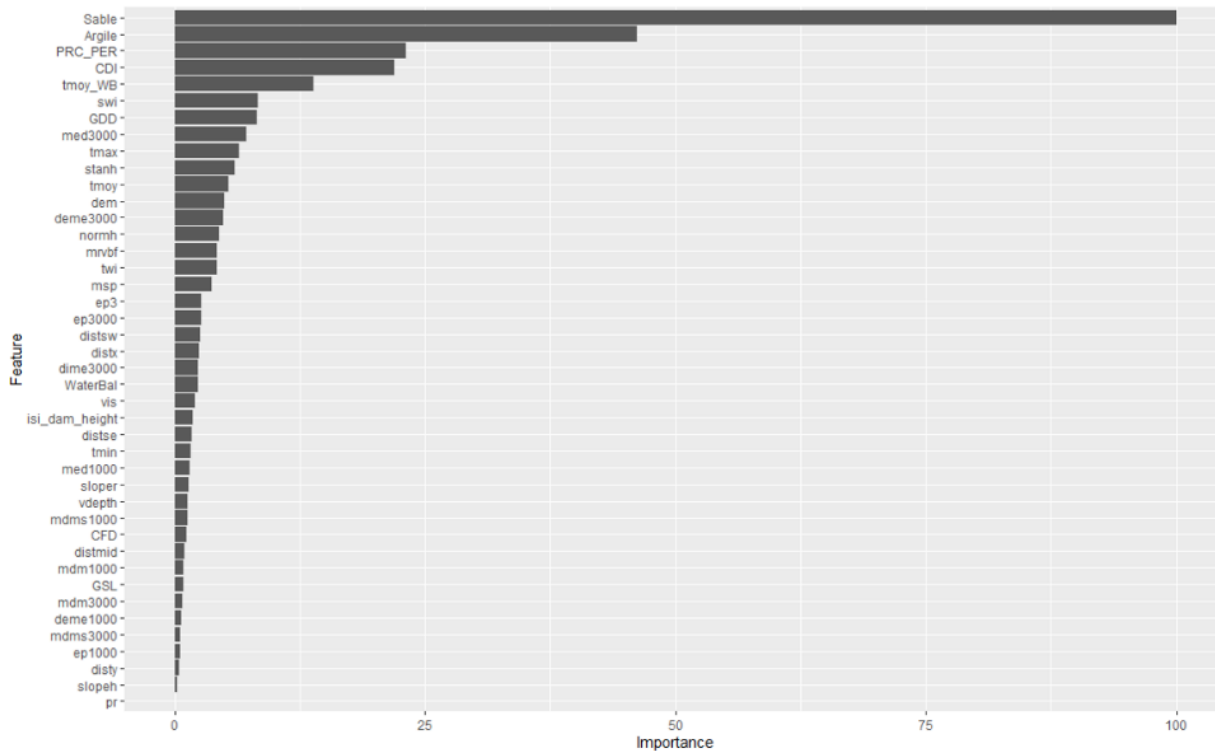


Figure 22 : Importance des variables dans le modèle RF (forêts aléatoires) pour prédire la teneur en MOS.

L'ordre d'importance des variables prédictives de la matière organique variait d'un modèle à l'autre (résultats non présentés). Cependant, les fractions granulométriques (Sable et Argile), l'indicateur climatique CDI et le ratio de cultures pérennes (PRC_PER) revenaient parmi les plus importantes variables prédictives de la MOS d'un modèle à l'autre.

La représentation des dépendances partielles (*partial dependence plots*) entre chaque variable prédictive et la MOS révèle des résultats similaires à l'analyse de corrélation de Pearson (Figure 23). Les dépendances partielles indiquent une relation positive entre la teneur en MOS et le ratio de cultures pérennes, le modèle numérique de terrain (DEM) ou le bilan hydrique, et à l'inverse une relation négative entre la teneur en MOS et la teneur en sable ou l'indicateur climatique CDI. La teneur en argile a quant à elle une influence non linéaire plus marquée sur la teneur en MOS, produisant de plus faibles teneurs entre 10 et 60 % d'argile. La teneur en argile dans les sols est généralement reconnue pour son effet de séquestration du C liée à la complexation de la MO autour et entre les particules plus fines d'argile et de limon dans le sol, tandis que la teneur en sable favorise l'aération et les réactions conduisant à la minéralisation de la MOS (Dexter et al., 2018; Johannes et al., 2017; Matus, 2021). Toutefois, on observe une augmentation sans raison a priori de la MOS à environ 5 % de teneur en argile (Figure 23). La prépondérance des cultures pérennes (PRC_PER ou Crop Ratio) favorise également une accumulation de MOS ou de C dans les sols. L'influence positive du modèle numérique de terrain (DEM) sur la teneur en MOS traduit fort probablement une prépondérance des cultures pérennes et de leur influence sur les sols à mesure qu'on s'éloigne de la plaine et des Basses-Terres du Saint-Laurent.

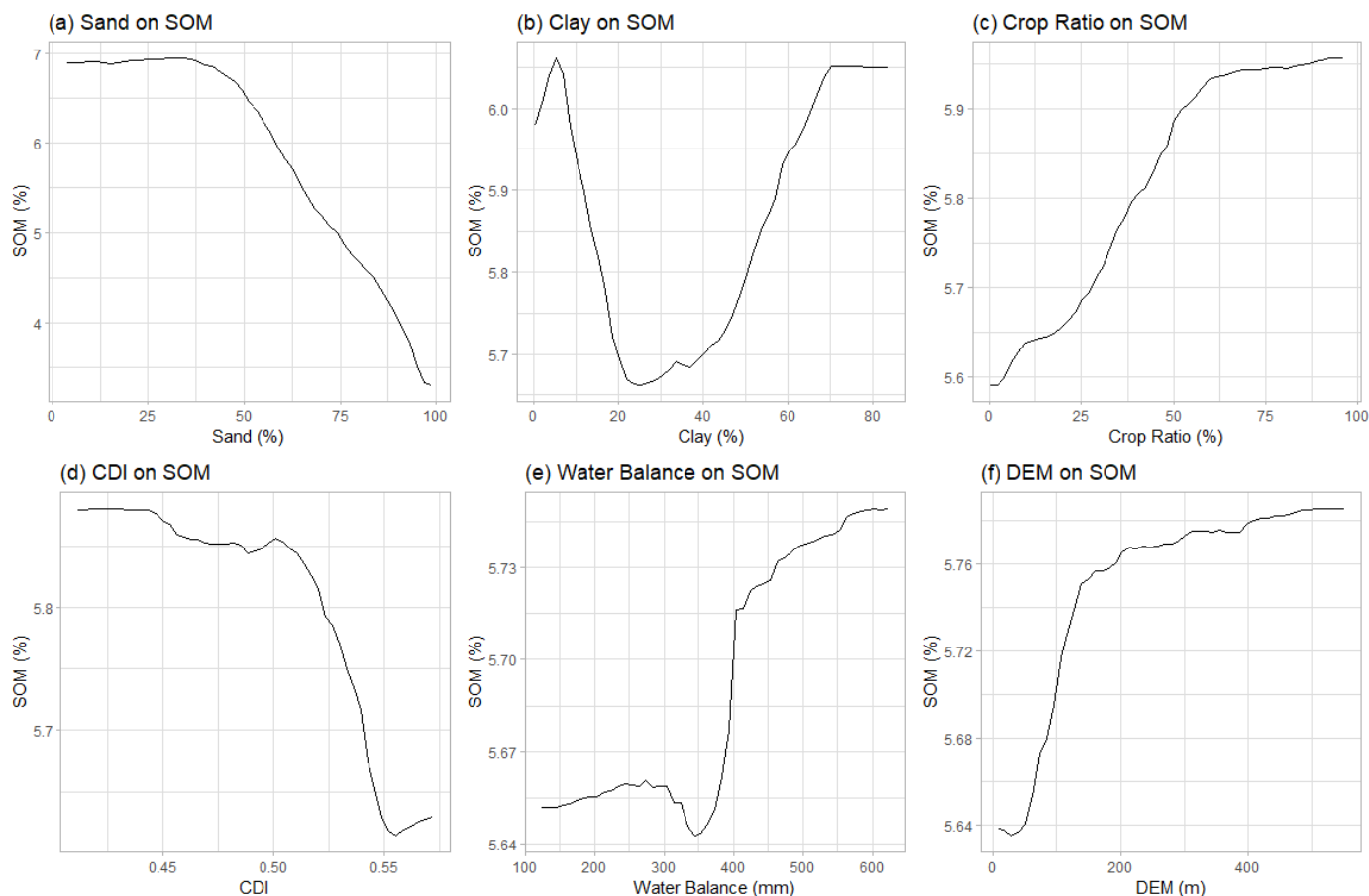


Figure 23 : Graphiques des dépendances partielles (*partial dependence plot*) entre la teneur en MOS et les variables prédictives les plus importantes, soit la teneur en sable (a), la teneur en argile (b), le ratio de cultures pérennes (c), le CDI (d), le bilan hydrique (e) et le modèle numérique de terrain (DEM) (f).

La performance du modèle RF pour prédire la distribution spatiale de la MOS (ou de son évolution en climat futur) n'est pas élevée avec un $R^2 = 0,49$, mais elle est comparable sinon supérieure à des études similaires relevées dans la littérature. Yigini et Panagos (2016) ont cartographié le stock de carbone dans les sols européens avec un modèle de régression et des performances de modèle de l'ordre de 0,40. De Brogniez et al. (2015) ont aussi cartographié la teneur en carbone organique de la couche arable de l'Europe à l'aide d'un modèle additif généralisé avec un R^2 de 0,29. Meersmans et al. (2011) ont mené une étude et construit un modèle pour évaluer la distribution spatiale du carbone organique du sol à l'échelle régionale en Belgique avec un R^2 ajusté de 0,42.

Le 50 % de variation inexpliquée par notre modèle est probablement dû en partie aux multiples facteurs influençant la teneur en matière organique de la couche arable des sols agricoles et qui ne sont pas intégrés dans notre modèle (par exemple, le travail de sol, l'érosion, la gestion des fumiers et des amendements organiques, etc.), mais aussi à une large part d'incertitude liée à la provenance des échantillons de sol et leur géoréférence associée à un code postal. Autour de 5 % des échantillons de sol avaient des teneurs égales ou supérieures à 15 % de MOS témoignant de sols relativement bien pourvus en MOS qui pourraient provenir de potager ou autres occupations non agricoles, qui étaient difficiles à identifier et éliminer de la base de données (Figure 21). La génération de moyennes pour les données climatiques, de MNT et d'occupation du sol pour pouvoir les associer aux analyses de sol (MOS, Argile, Sable) à l'échelle de la municipalité a aussi occasionné un

nivellement et une perte d'information que les modèles d'apprentissage machines ne captent pas dans la forme actuelle de la structure des données.

5.3 Modélisation de la distribution spatiale du ratio de cultures pérennes

La répartition des cultures et le climat au Québec sont deux variables étroitement liées (Figure 14). La prépondérance de cultures pérennes ayant un impact important sur la distribution spatiale de la MOS à l'échelle de la province, il devient important de considérer l'effet de l'évolution des cultures pérennes dans un horizon futur en fonction des changements climatiques anticipés afin d'intégrer ces changements à venir dans l'occupation du sol dans le modèle prédictif. Le modèle prédisant la distribution spatiale du ratio de cultures pérennes intègre les ensembles de variables suivantes :

$$\% \text{ Culture pérenne} = f(\text{Indicateurs climatiques} + \text{covariables DEM})$$

Quatre modèles de régression ont été testés en appliquant les méthodes d'apprentissage automatique présentées antérieurement, c'est-à-dire en séparant le jeu de données en jeu de calibration (75%) et de validation (25%), ainsi qu'en appliquant une validation croisée (*10-fold*) lors de la calibration du modèle. Le modèle paramétrique GLM a été comparé aux 3 modèles non paramétriques XGBtrees, XGBlinear et Random Forest (Ranger). La Figure 24 présente l'erreur quadratique moyenne (RMSE) du ratio de cultures pérennes prédit et son intervalle de confiance à 0,95 calculés avec la validation croisée (*10-fold*) de chaque modèle.

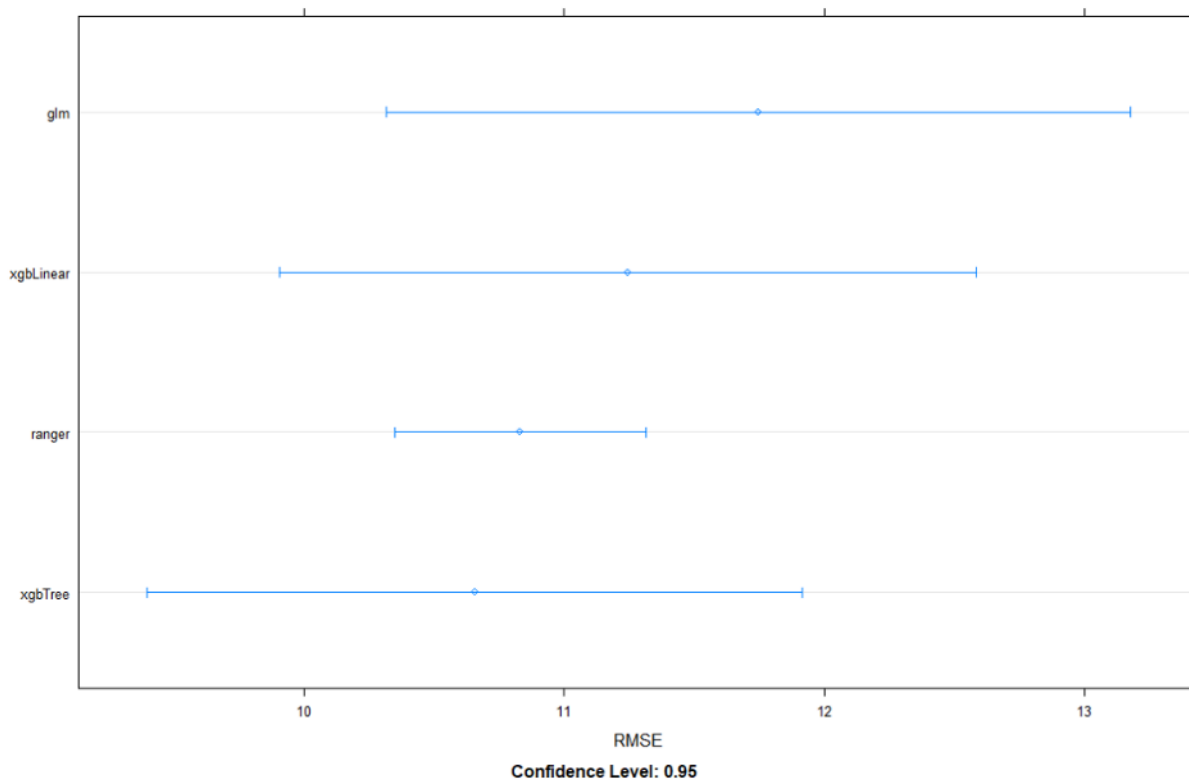


Figure 24 : Erreur quadratique moyenne (RMSE en %) et son intervalle de confiance à 95 % de la prédiction du ratio de cultures pérennes.

Les quatre modèles testés montrent des performances similaires. Les RMSE oscillent entre 10,5% et 11,7% respectivement pour XGBtrees (meilleur RMSE) et GLM (RMSE le plus élevé). Le modèle RF (Radom Forest-Ranger) présente toutefois un plus faible intervalle de confiance à 95 %. Comparant le ratio de cultures pérennes prédit à celui observé avec les données de validation, le modèle RF montre à nouveau les meilleures performances avec un $R^2 = 0,84$ (Figure 25). Le modèle GLM présente en revanche une pente plus près de 1 (0,89) que le modèle RF (0,83).

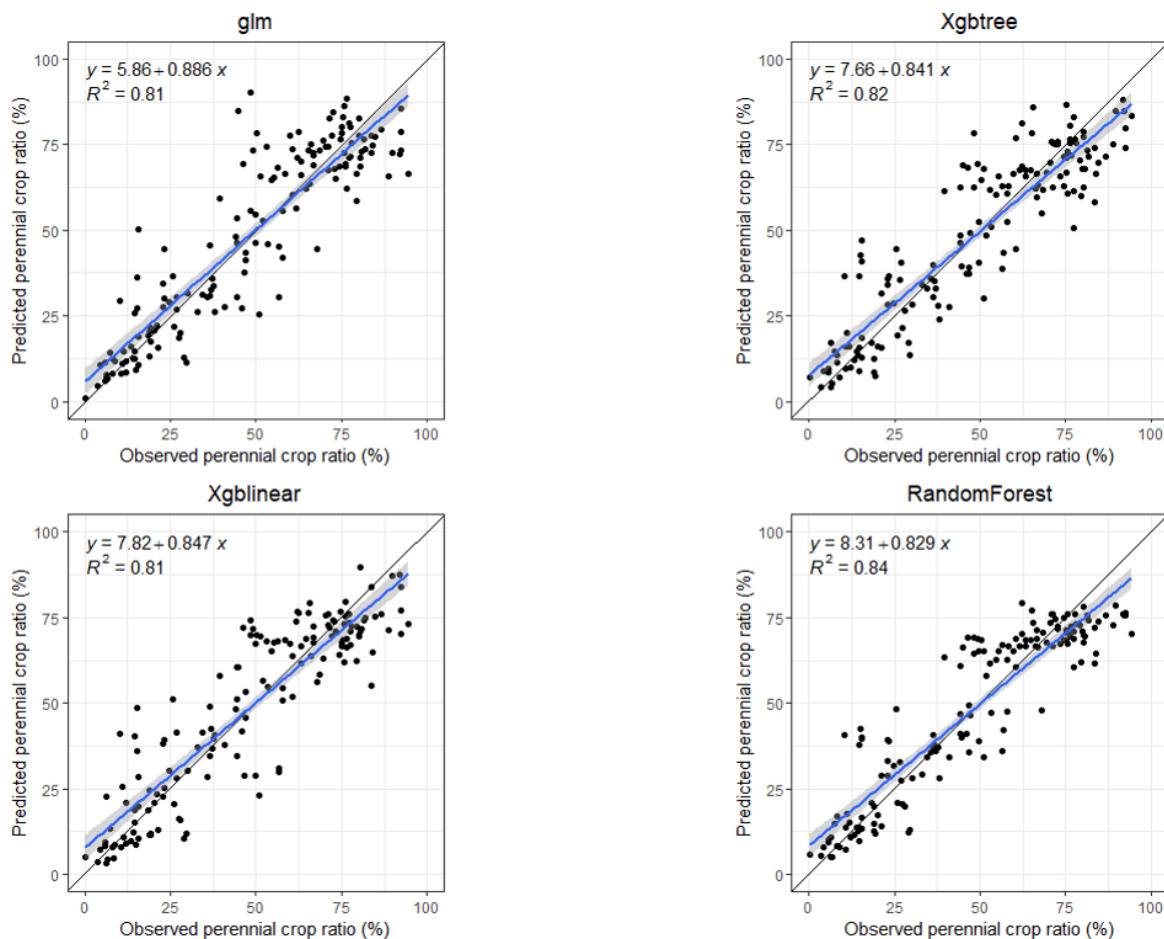


Figure 25 : Comparaison entre valeurs prédites et observées du ratio de cultures pérennes (%) pour quatre modèles testés sur le jeu de données indépendant (n = 154).

Hormis pour le modèle GLM, l'indicateur climatique CDI est systématiquement le meilleur prédicteur pour le ratio de culture (Figure 26). Intégrant deux facteurs liés à la température et au bilan hydrique, il est d'une part le mieux corrélé (corrélation inverse) au ratio de cultures pérennes moyenné par municipalité (Figure 17), mais aussi celui qui contribue le plus et qui a le plus de poids dans les modèles prédictifs de régression testés (Figure 26). Les autres variables climatiques telles que le GDD (growing degree days) et CFD (consecutive frost days) semblent avoir une plus grande importance aussi pour les modèles RF et XGBlinear. Le SAGA wetness index (swi), un indice topographique d'humidité apparaît comme la 3^e variable d'importance pour trois modèles RF, XGBlinear et XGBoost. Pour fin de projection en climat futur, le modèle XGBlinear a été choisi car il était plus sensible aux effets des changements climatiques que les modèles RF et XGBoost (voir Figure 38 à l'Annexe 1).

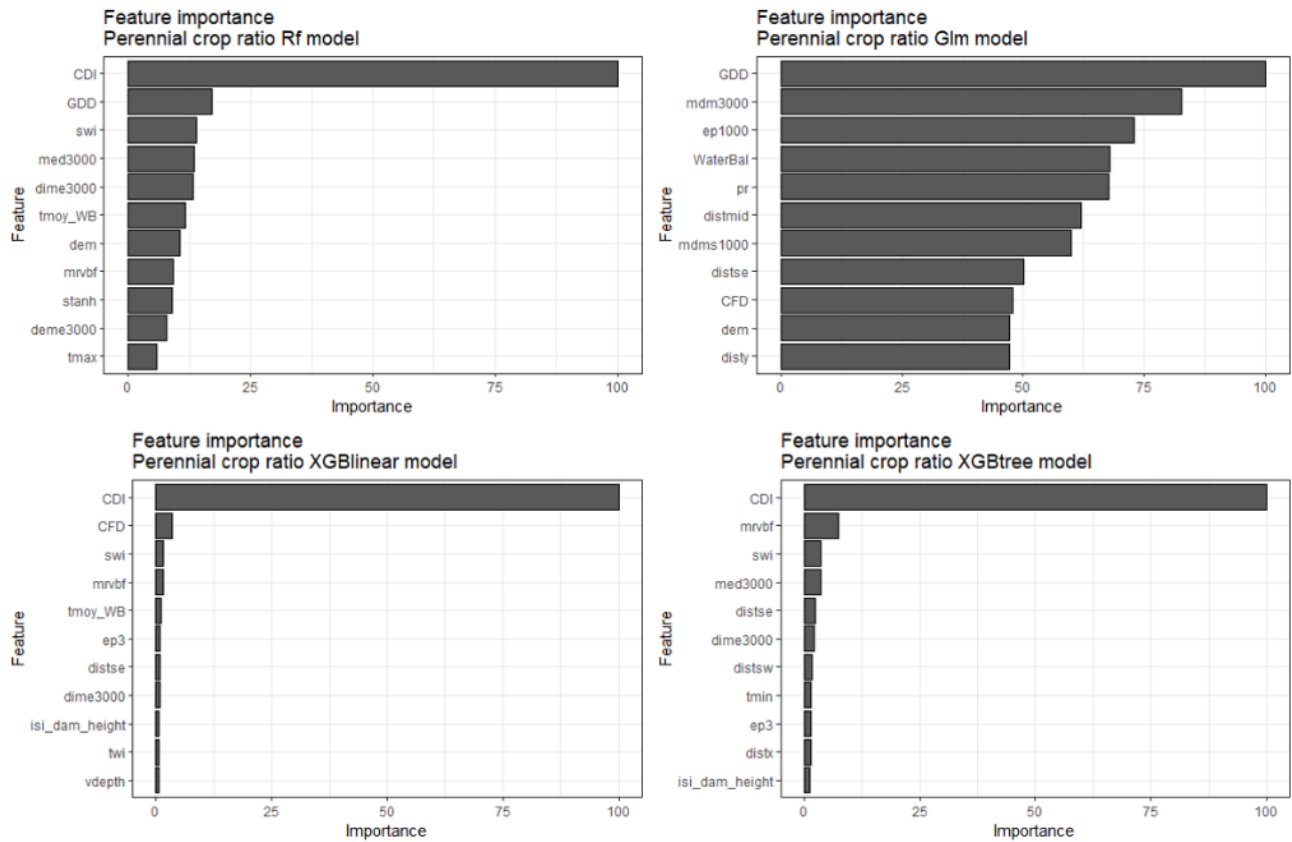


Figure 26 : Importance des variables de quatre modèles développés pour prédire le ratio de cultures pérennes.

Le développement et l'analyse de ces deux modèles de prédiction a permis de créer un modèle général permettant de faire des projections dans le futur de l'évolution de la MOS en intégrant ou non un ratio de cultures pérennes qui évolue dans le temps.

6 Projection de l'évolution des teneurs en MOS en climat futur

La projection de l'évolution des teneurs en MOS en climat futur (2040 à 2100) a été modélisée avec et sans changement de ratio de cultures pérennes dans le but d'estimer l'importance de cette variable en termes d'évolution des teneurs en MOS.

6.1 Application du modèle RF sur la MOS sans changement de cultures au Québec

Les projections de l'évolution des teneurs en MOS intégrant les 11 simulations climatiques en scénario d'émission modéré (RCP 4.5) et les 11 simulations climatiques en scénario d'émission élevé (RCP 8.5) sans changement de ratio de cultures pérennes démontrent globalement une baisse des taux de matière organique dans les sols agricoles du Québec, sous les deux RCP simulés et dans toutes les zones agroclimatiques sauf la zone 1 (Figure 27). Les moyennes et écarts-types des 11 modèles climatiques en RCP 4.5 et 8.5 présentés à la Figure 27 traduisent l'effet des multiples scénarios climatiques à l'horizon de 2100 sur les prédictions du modèle.

La teneur moyenne de MOS estimée en période initiale (1981-2010) dans la zone 1 (5 %) est légèrement plus élevée que la teneur initiale mesurée (4,5 %) dans le même zone (Figure 6), tandis que dans la zone 6, la teneur en période initiale (1981-2010) (7,8 %) tend à être plus faible que celle mesurée (8,6 %) dans la même zone. Les modèles ont souvent tendance à aplanir les extrêmes. Les teneurs initiales estimées sont aussi conditionnées par la texture révélée ici par les groupes texturaux, comme dans les observations de départ (Figure 7). Dans la zone climatique 1, les teneurs en MOS évolueraient très peu comparativement aux autres zones plus froides 3, 5 et 6 où le réchauffement climatique aurait plus d'impact.

Selon le modèle retenu, les teneurs initiales en matière organique sont plus faibles dans les sols plus sableux qu'argileux, là où les cultures annuelles sont prédominantes, dans les zones climatiques plus chaudes et présentant un bilan hydrique plus faible, et dans les zones de basses altitudes (Figure 22 et Figure 23). Les sols dans les basses terres de la plaine du Saint-Laurent recourent la plupart de ces facteurs sauf la texture qui est souvent plus argileuse. Ces facteurs expliqueraient les teneurs initiales en matière organique plus faibles dans les zones plus chaudes et tempérées du Québec. Le travail de sol plus profond réalisé par le passé et des apports réduits en engrais de ferme dans les exploitations dédiées aux cultures commerciales pourraient aussi expliquer ces plus faibles teneurs, mais ces facteurs ne sont pas inclus dans le modèle. Le maintien des niveaux de matière organique à de faibles niveaux et sans diminution supplémentaires en climat futur dans la zone 1, démontrerait l'atteinte d'un équilibre à l'intérieur du jeu de donnée sous un régime de température plus élevée.

Que ce soit en RCP 4.5 ou 8.5, les projections en zones 3, 5 et 6 démontrent des baisses plus importantes de teneurs en MOS de 2010 à 2100 (de l'ordre de -0,83 à -1,27 %), alors qu'en zones 2 et 4 les baisses sont plus modestes (de -0,31 à -0,61 %). Les simulations climatiques en RCP 8.5 induisent une baisse plus importante de MOS par rapport à celles en RCP 4.5 dans quelques conditions seulement, par exemple dans les sols loameux, loameux-argileux et sableux en zones 4 et 6, soit de -0,23 à -0,41 %.

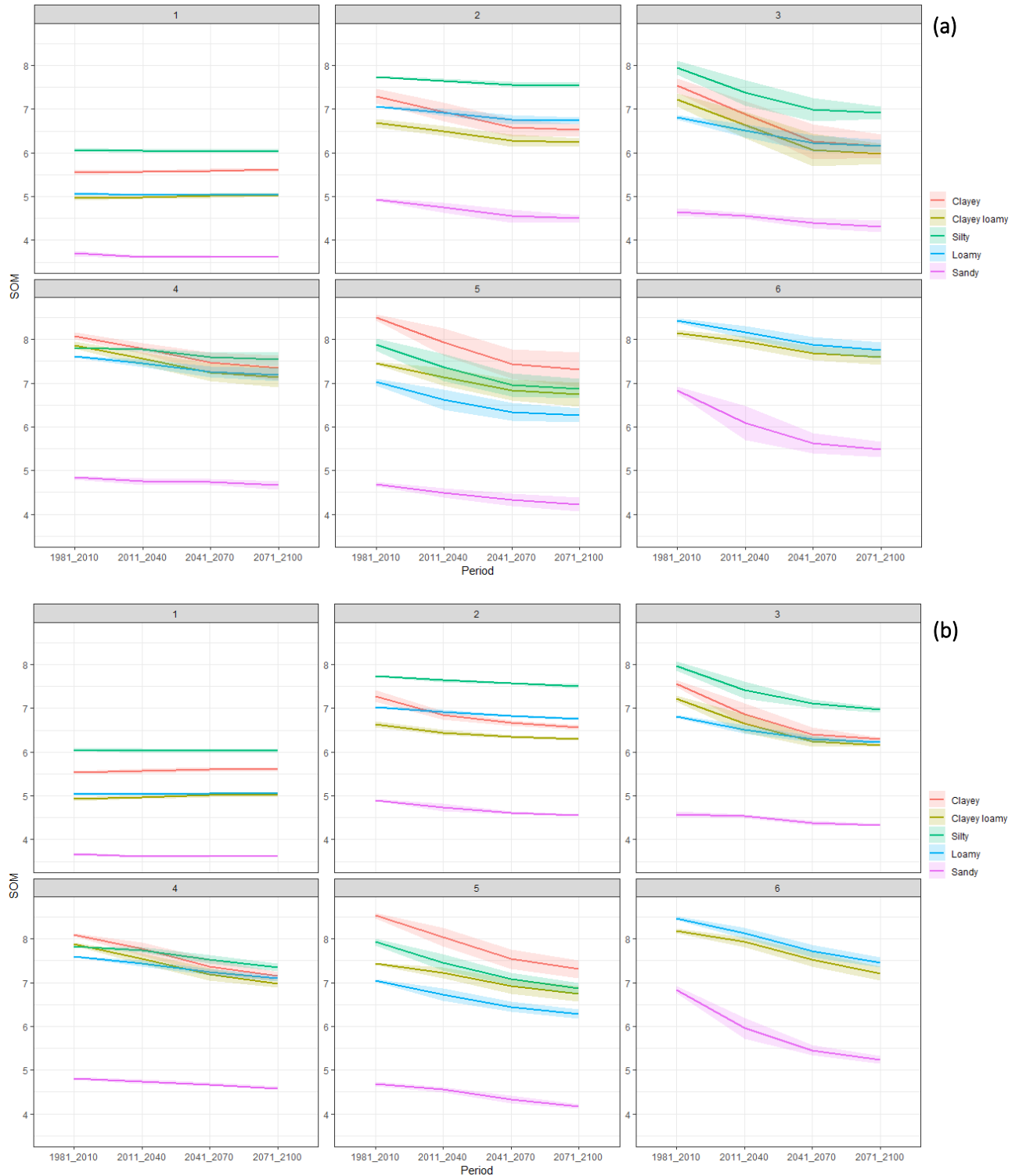
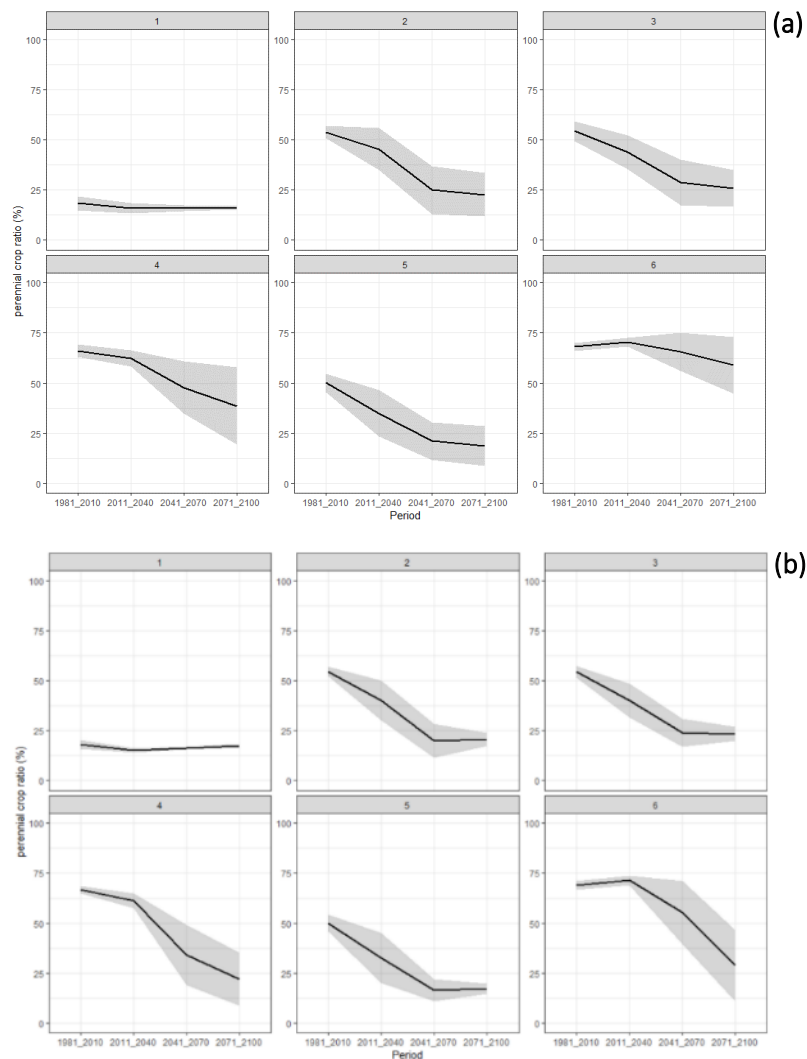


Figure 27 : Projections de l'évolution des teneurs en MOS par zones climatiques et groupes texturaux représentées par la moyenne (ligne pleine) et l'écart-type (parties ombragées) des projections de 11 modèles climatiques en RCP 4.5 (a) et 8.5. (b).

6.2 Application du modèle XGBlinear pour simuler un changement de ratio de cultures pérennes en climat futur

Le modèle XGBlinear retenu pour simuler un changement de ratio de cultures pérennes en climat futur démontre une baisse du ratio de cultures pérennes dans les différentes zones climatiques sauf la zone 1 (Figure 28). Les baisses sont plus importantes en RCP 8.5.

Le poids important du climat dans notre modèle de prédiction du ratio de cultures pérennes engendre des baisses significatives de ce ratio dans les zones plus froides (4 et 6) où il est initialement plus élevé. Le modèle XGBlinear a été retenu parce qu'il projette une évolution plus prononcée du ratio de cultures pérennes, alors que les modèles RF et XGBboost produisent des projections similaires et plus faibles à l'horizon 2100 (voir Figure 38 à l'Annexe 1).



6.3 Application du modèle RF avec changement de cultures

Le modèle RF pour projeter l'évolution de la MOS dans le temps a été appliqué en tenant compte ou non de l'évolution du ratio de cultures pérennes pour mesurer et pondérer l'impact des changements climatiques et du changement d'occupation des terres.

L'évolution absolue de la MOS en partant du niveau initial de MOS modélisé sur la période 1981-2010, jusqu'en 2100 est rapporté à la Figure 29 sous l'effet ou non du changement de ratio de cultures pérenne, des deux scénarios climatiques (RCP 4.5 et 8.5) et sous les effets du groupe textural (Figure 29a) et de la zone agroclimatique (Figure 29b). La même évolution de la MOS est rapportée sous forme relative à la Figure 30.

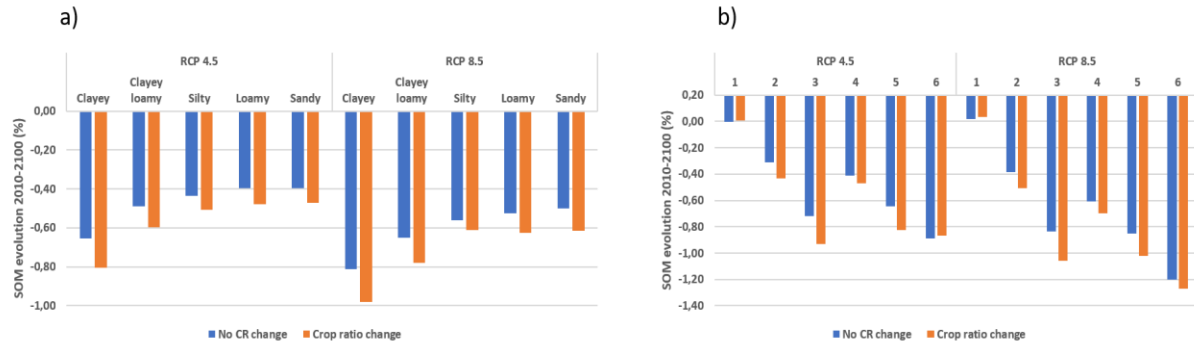


Figure 29 : Variations absolues des teneurs en MOS projetées de 2010 à 2100 résultant de la moyenne de 11 modèles climatiques évoluant sous l'effet de deux scénarios climatiques (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes, les groupes texturaux (a) ou les six zones agroclimatiques (b).

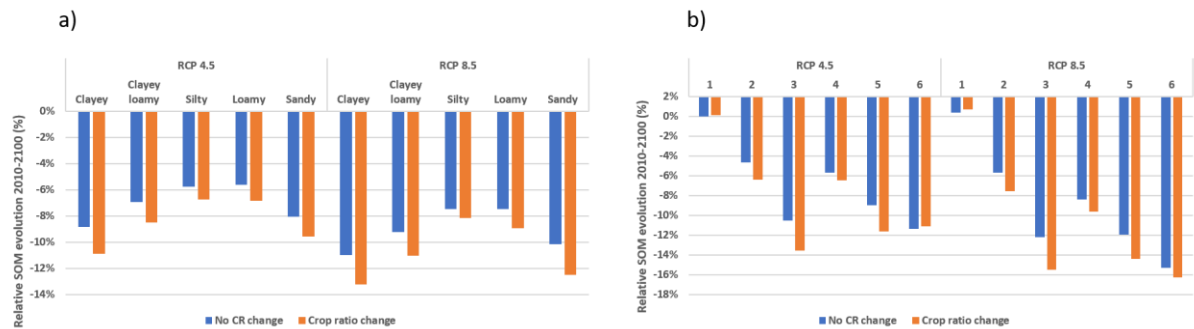


Figure 30 : Variations relatives des teneurs en MOS projetées de 2010 à 2100 résultant de la moyenne de 11 modèles climatiques évoluant sous l'effet de deux scénarios climatiques (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes, les groupes texturaux (a) ou les six zones agroclimatiques (b).

En terme absolu, le groupe textural fait diminuer les teneurs en MOS de 0,4 % à 1 % (Figure 29a). Les sols argileux pourraient perdre jusqu'à 1 % de MOS sous l'effet d'un changement de ratio de cultures pérennes et d'un scénario de GES RCP 8.5, tandis qu'à l'autre extrême les sols sableux pourraient ne perdre que 0,4 % de MOS sans changement de ratio de cultures pérennes et en RCP 4.5. Étant donné, leur niveau initialement plus faible en MOS, les sols plus sableux en auraient moins à perdre.

Rapporté sur une base annuelle (90 ans entre 2010 et 2100) et avec un facteur de conversion de 0,58 g C/g MOS, ces diminutions de 0,4 et 1,0 % de MOS équivalent respectivement à des pertes de 0,0026 et 0,006 % C/an soit de l'ordre de grandeur de 0,008 % C/an rapporté par Warren et al. (2020) pour décrire l'évolution des teneurs en C dans l'horizon de surface Ap des études pédologiques de l'Ontario entre 1950 et 2020. D'après leur analyse, les diminutions de teneurs en C observées l'Ap pourraient être en grande parties

imputées à une dilution de la MOS dans le profil de sol avec un approfondissement du travail de sol et de l'horizon Ap au cours de ces décennies.

Rapporté en termes de Mg C/ha en utilisant un horizon de sol de surface de 20 cm de profondeur et une masse volumique apparente de 1,3 g/cm³, ces baisses 0,4 et 1,0 % de MOS équivalent respectivement à des pertes de 6 et 15 Mg C/ha soit des baisses plus faibles mais dans l'ordre de grandeur de la perte de 20 à 40 Mg C/ha rapporté par Gottschalk et al. (2012) pour la période de 1970-2100 sous l'effet de la déforestation ou conversion en terres arables dans l'Est du Canada. Rapporté sur une base annuelle (90 ans entre 2010 et 2100), les pertes de 0,07 et 0,17 Mg C/ha/an sont plus élevées que les pertes de 0,02 à 0,055 Mg C/ha/an rapportées par Smith et al. (2009) et générées avec le modèle Century sur un horizon de 100 ans (2000-2099) sous différents scénarios de rotation de cultures dans des régions du sud l'Ontario et du Québec. De fait, notre modèle produit des pertes pratiquement nulle en moyenne dans la zone agroclimatique 1 la plus au sud du Québec (Figure 29b). Les baisses de MOS sont par ailleurs plus variables entre zones agroclimatiques, soit de près de 0 % en zone 1 à 1,2 % en zone 6 (Figure 29b). Par comparaison, le changement d'occupation des sols anticipé par une baisse du ratio de cultures pérennes en climat plus frais entraîne des baisses absolues plus faibles de l'ordre de 0,05 à 0,09 % selon les groupes texturaux (Figure 29a) et de 0 à 0,2% selon les zones climatiques (Figure 29b).

Sur une base relative, la MOS diminue de 6 à 12 % en fonction des groupes texturaux (Figure 30a) et de 0 à 16% en fonction des zones agroclimatiques (Figure 30b). Par comparaison, la diminution du ratio de cultures pérennes en climat plus frais entraîne des baisses relatives plus faibles de l'ordre de 0,5 à 2 % selon les groupes texturaux (Figure 30a) et de 0 à 3 % selon les zones climatiques (Figure 30b).

La Figure 31 présente les diminutions relatives plus spécifiquement sous l'interaction des groupes texturaux et des zones agroclimatiques. Les diminutions relatives sont plus importantes dans les sols argileux et loameux-argileux en zone agroclimatique 3 (environ 15 %), dans les sols limoneux et loameux en zones 3, 5 et 6 (environ 10 %) et dans les sols sableux en zone 2, 5 et 6 (de 5 % à 20 %).

On note également une légère augmentation relative (1%) dans les sols argileux et loameux-argileux en zone 1, mais une légère baisse dans les sols sableux. À l'inverse les sols plus argileux perdraient davantage de MOS (près de 20 % sur une base relative) comparativement aux sols plus sableux en zone 3 dans les régions des Laurentides, du Témiscamingue et du pourtour du Lac-Saint-Jean. De même les sols sableux en zone 6 (Abitibi et nord du Lac-Saint-Jean) perdraient plus de 20 % de leur teneur en MOS sur une base relative. Le changement d'occupation des sols en cultures annuelles aurait plus d'impact là où les mêmes interactions des groupes texturaux et zones climatiques engendrent des diminutions plus importantes, mais cette perte additionnelle n'est jamais supérieure à environ 5% de la teneur en MOS. Ces écarts sont cependant plus importants en zones agroclimatiques 2, 3 et 5 dans les sols argileux, loameux-argileux et sableux. La zone agroclimatique 2 inclut des régions de l'Estrie, de Chaudières-Appalaches et de Mauricie; la zone 3 inclut des régions des Laurentides, du Témiscamingue et du pourtour du Lac-Saint-Jean, et la zone 5, de l'Abitibi et du nord du Lac-Saint-Jean. Dans ces régions, l'expansion des cultures annuelles auraient le plus grand effet sur la diminution de la teneur en MOS. Dans les zones 1 (Montérégie) et 4 (Bellechasse, Bas-Saint-Laurent, Matapédia, Baie-des-Chaleurs, Portneuf, Saguenay) ces effets seraient moins importants.

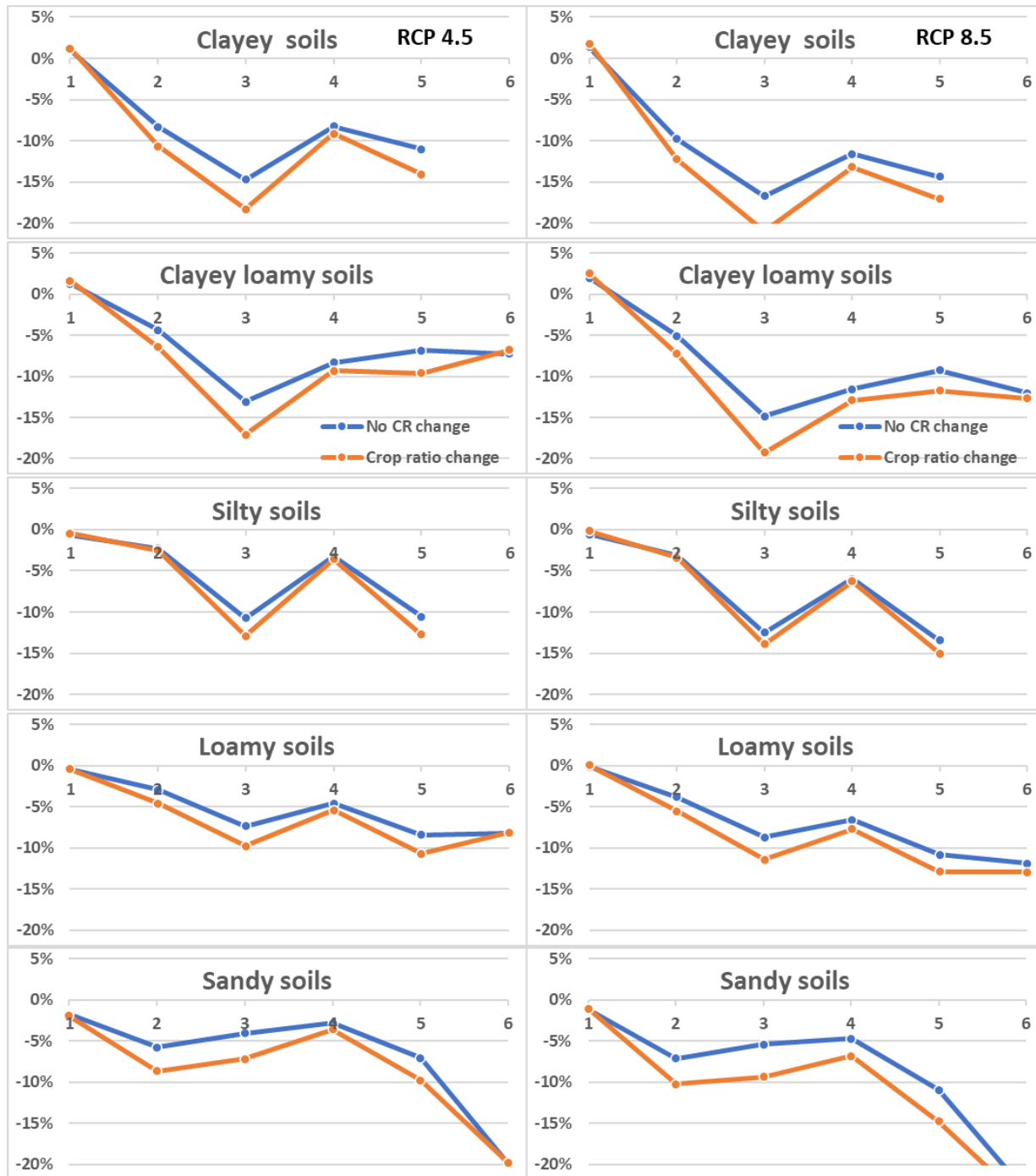


Figure 31 : Variations relatives des teneurs en MOS projetées de 2010 à 2100 en fonction de cinq groupes texturaux de sols, de six zones climatiques, de deux scénarios d'occupation du sol résultant de la moyenne de 11 modèles climatiques sous deux scénarios de réchauffement climatique (RCP 4.5 et 8.5).

Les simulations climatiques réalisées sous le scénario d'émissions élevées de GES en RCP 8.5 (peu de restrictions aux émissions ont été mises en place) augmentent légèrement les pertes de MOS (d'environ 5%) par rapport à un scénario d'émissions modérés (RCP 4.5), surtout pour les zones où le climat est plus froid (3, 4, 5 et 6). Outre ces effets de scénarios d'émissions (RCP 4.5 et 8.5), les variations dans les teneurs en MOS, sont surtout induites par les changements climatiques et la texture des sols, car la présence ou non d'un changement de ratio de

cultures pérennes dans le modèle induit tout au plus une variation relative d'environ 5 % de la teneur en MOS, alors que les changements climatiques peuvent induire des variations relatives de près de 20 %, dans les sols argileux en zone 3 et dans les sols sableux en zone climatique 6 par exemple.

Les restrictions sur les émissions de GES auront toutefois des effets sur les changements à long terme des teneurs en MOS. Plus les projections s'éloignent dans le temps, plus les écarts entre les émissions en RCP 4.5 et 8.5 sont importants.

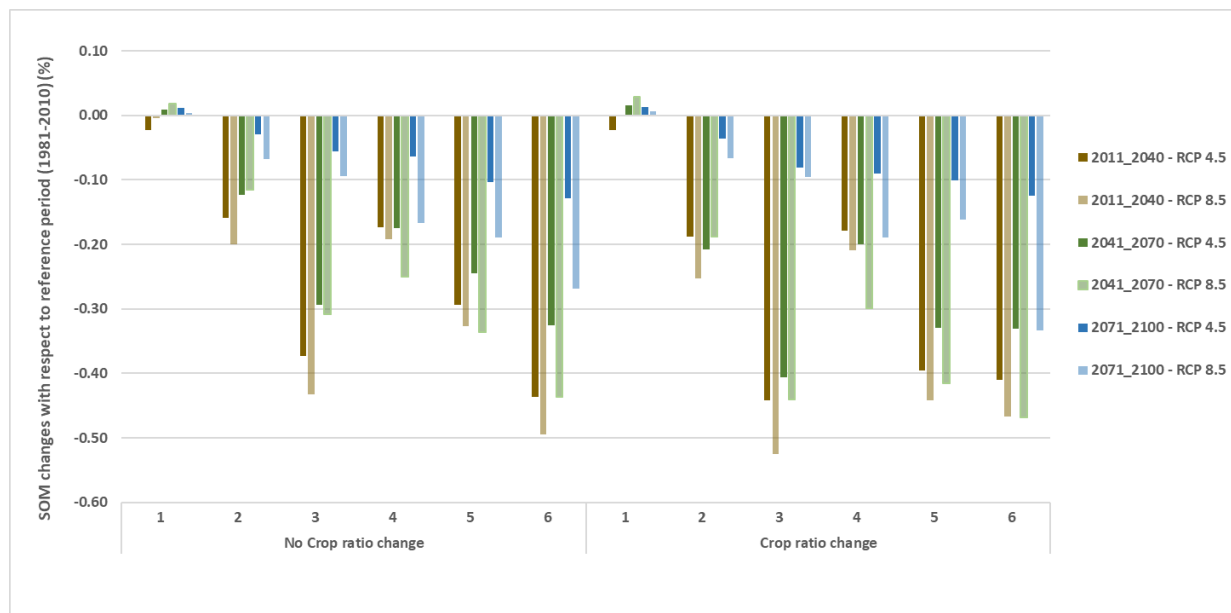


Figure 32 : Variations absolues des teneurs en MOS projetées à trois périodes (2011-2040, 2041-2070 et 2070-2100) sous l'effet de deux scénarios climatiques liés aux GES (RCP 4.5 et 8.5), le changement ou non de ratio de cultures pérennes dans les six zones agroclimatiques.

7 Application des modèles de simulation à l'Ontario

Les modèles calibrés et développés pour les conditions du Québec ont été appliqués à l'Ontario, dans le but d'observer leur comportement et les résultats produits dans un contexte en marge de celui du Québec. Les données sur le climat proviennent de la même source que pour le Québec soit les données historiques et simulées en climat futur par le consortium Ouranos. La zone agricole ontarienne a été découpé en six zones agroclimatiques en utilisant les mêmes onze indicateurs climatiques retenus pour le Québec. La Figure 33 illustre le découpage agroclimatique en six zones de l'Ontario, et le Tableau 8 présente par zone les normales climatiques sur 30 ans de six des onze indicateurs climatiques retenus.

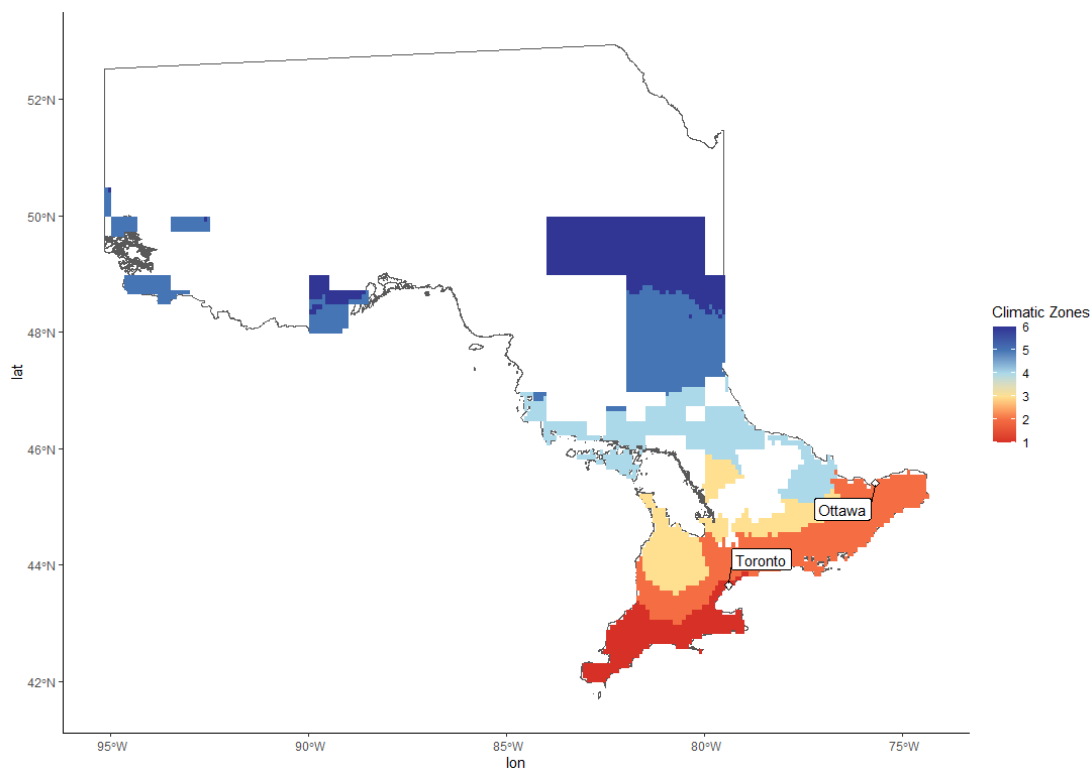


Figure 33 : Découpage agroclimatique de l'Ontario en six zones.

Tableau 8 : Normales climatiques par zone agroclimatique de l'Ontario.

Zone Agroclimatique	T_{moy}^a (°C)	P_{cp}^a (mm)	ETP^a (mm)	DSC^a (jours)	$P-ETP^a$ (mm)
1	8.6	800	668	225	132
2	6.8	845	661	211	184
3	6.1	907	647	201	260
4	5.2	753	586	193	167
5	2.8	680	545	176	135
6	2.0	671	532	168	139

^a T_{moy} : température moyenne; P_{cp} : précipitation totale; ETP : évapotranspiration potentielle; DSC : durée de la saison de croissance; $P-ETP$: bilan hydrique résultant de précipitation totale - ETP .

Les données reliées au MNT et ses covariables ont été générés sur une grille de 250 m. Les données sur la texture des sols proviennent de la carte de sols de l'Ontario (OMAFRA). Les données sur les cultures proviennent de [l'Inventaire des cultures du Canada](#) produit à chaque année par AAC sur une grille de 30 m. Les images produites depuis 2011 en Ontario ont été reclassifiées et rééchantillonnées sur une grille de 250 m afin de créer pour l'Ontario un ratio moyen de cultures pérennes similaire à celui généré pour le Québec à partir de la BDPPAD de la FADQ. La Figure 34 présente le ratio moyen de cultures pérennes de 2011 à 2020.

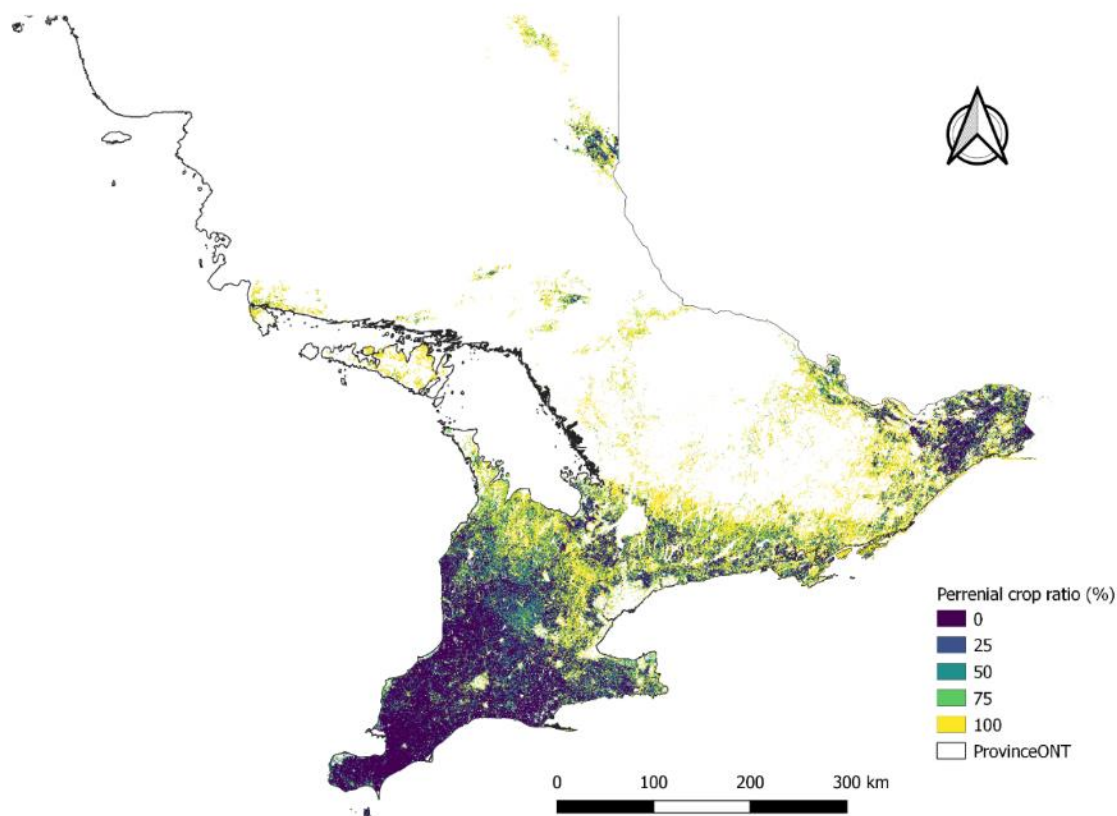


Figure 34 : Ratio de cultures pérennes moyen 2011-2020 présenté sur une grille de 250 m.

Les données reliées aux sols, aux MNT, au climat et aux cultures ont été intersectées et moyennées par municipalité ontarienne comme au Québec. Chaque polygone de la carte de sols de l'Ontario a été associé à une municipalité (Figure 35). Au total, environ 90 000 données de textures différentes ont été générées et associées aux municipalités ontariennes.

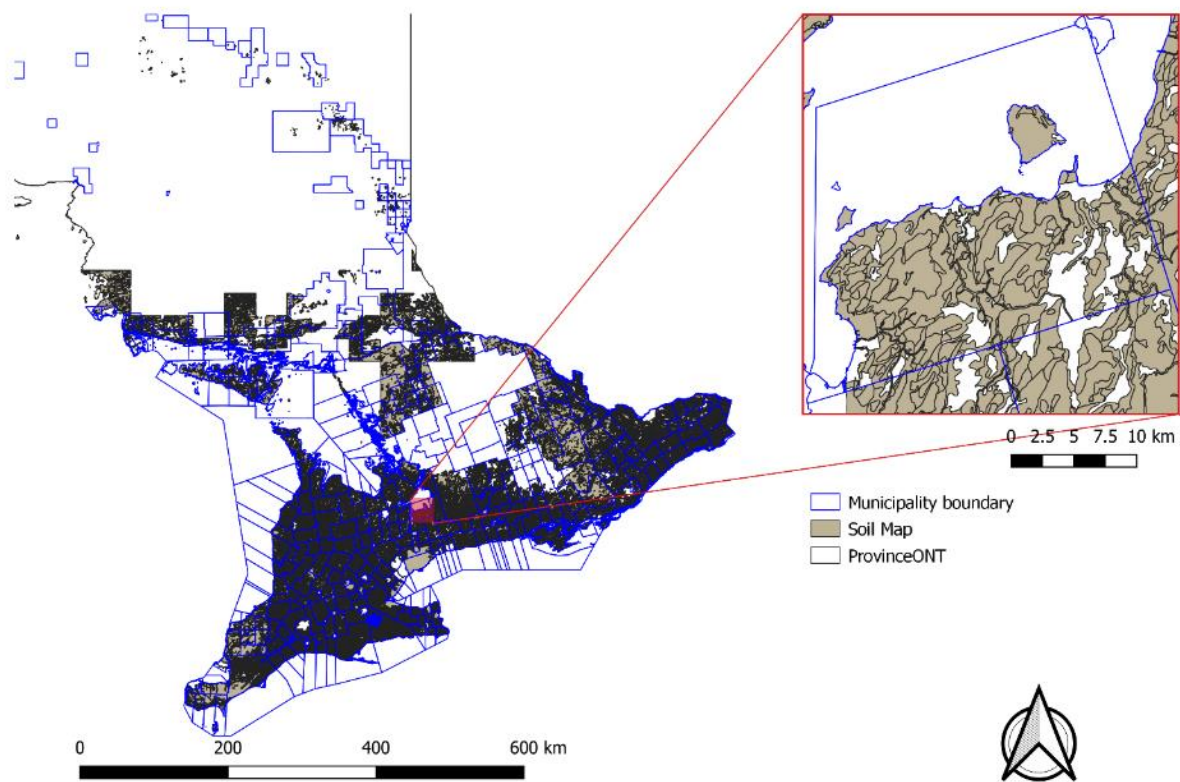


Figure 35 : Polygones des municipalités (en bleu) et de la carte de sols (en beige) de l’Ontario.

7.1 Simulations du climat futur sur les zones agroclimatiques de l’Ontario

Les onze indicateurs climatiques ont été recalculés et projetés dans le futur sur les six zones agroclimatiques de l’Ontario en utilisant les 11 modèles climatiques et les deux scénarios d’émissions de GES (RCP 4.5 et 8.5) fournis par Ouranos. La Figure 36 présente les projections de quatre indicateurs climatiques d’intérêt. Sous ces scénarios du climat futur, la température moyenne annuelle devrait passer d’environ 8,5°C à l’horizon 1980-2010 à près de 11,5°C à l’horizon de 2070-2100 en RCP 4.5 ou 14°C en RCP 8.5 dans la zone agroclimatique 1 la plus chaude, tandis qu’en région plus froide, cette variation serait d’environ 2°C à près de 5°C en RCP 4.5 ou 8°C en RCP 8.5. Les variations de température sont importantes dans les deux cas, mais paraissent plus importantes en termes relatifs en zone plus froide. Par comparaison avec le Québec, la zone 1 la plus chaude du Québec se compare aux zones 3 et 4 de l’Ontario. Comme pour le Québec, en termes de précipitations et de bilan hydrique, les variations moyennes apparaissent moins importantes sur la période entre 1980-2010 et 2070-2100, mais les variations obtenues entre les 11 scénarios climatiques, indiquées par les couleurs ombragées, demeurent importantes.

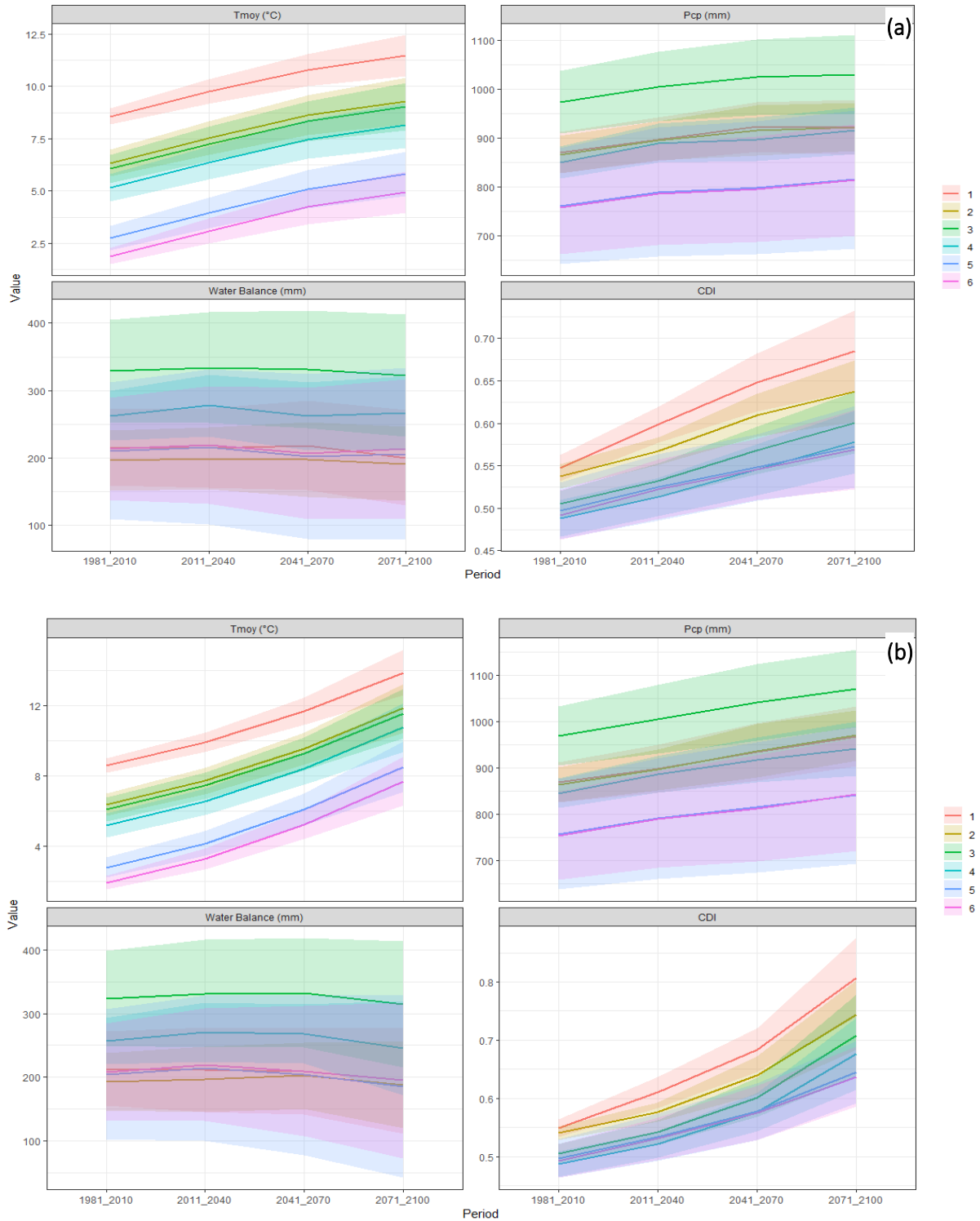


Figure 36 : Projections en climat futur de quatre indicateurs climatiques dans six zones agroclimatiques en RCP 4.5 (a) et RCP 8.5 (b). Les lignes pleines et les parties ombragées représentent respectivement la moyenne et l'écart-type des simulations de 11 modèles climatiques.

7.2 Application du modèle RF sur la MOS sans changement de cultures en Ontario

L'application du modèle de simulation de l'évolution de la MOS en climat futur dans les conditions de l'Ontario a donné des résultats mitigés et sans trop d'impact sur la MOS (Figure 37). Les niveaux de départ de MOS sont dans l'ordre de grandeur de ce qui a été observé au Québec, entre 4 et 8 % de MOS mais elles sont relativement élevées pour les zones agroclimatiques 1, 2 et 3 qui sont plus chaudes ou équivalentes à la zone 1 du Québec. Basés sur des études pédologiques de l'Ontario, Warren et al. (2020) ont calculé des teneurs moyennes en C évoluant de 2,85 % (ou 4,9 % de MOS) dans la période 1950-1980 à 2,34 % (ou 4,0 % de MOS) en 2000-2019, soit une baisse d'environ 0,014% de MOS par année. Nous notons au cours des trois prochaines périodes du futur une très faible évolution à la baisse des niveaux de MOS peu importe les zones agroclimatiques, soit de 0,0003 % par année sur un horizon de 2010 à 2100 dans la zone 1 et de 0,005 % dans la zone 6.

Plusieurs raisons peuvent expliquer le comportement inattendu du modèle, c'est-à-dire une très faible évolution à la baisse de la MOS, dans les conditions de l'Ontario, comparativement à ce qui est projeté au Québec. La plus importante, le modèle n'a pas été développé ou calibré à l'intérieur du même domaine de conditions initiales. Comme pour les simulations en climat futur au Québec, les températures moyennes des zones plus chaudes de l'Ontario sont plus chaudes que celles rencontrées au Québec dans la période de référence 1981-2010 qui a servi à la calibration (Tableau 6 et Tableau 8). Le CDI qui combine l'influence de la température et du bilan hydrique est une variable arrivant en 4^e place en termes d'importance après les teneurs en sable et argile, et le ratio de cultures pérennes (Figure 22). La relation inverse entre le CDI et la teneur en MOS devrait favoriser des baisses importantes de MOS en Ontario étant donnée son évolution à la hausse anticipée, mais ça ne semble pas être le cas. Par ailleurs, le bilan hydrique (P-ETP) est nettement plus sec en Ontario (entre 132 à 260 mm) qu'au Québec (entre 311 à 501 mm); ce qui devrait générer de plus faibles niveaux de MOS en climat actuel en Ontario selon la dépendance partielle entre le bilan hydrique et la MOS rapportée à la Figure 23.

Une autre raison pourrait être liée au ratio de cultures pérennes nettement plus élevé en Ontario qu'au Québec (Tableau 9). Ces ratios sont nettement plus élevés en Ontario mêmes dans ses zones agroclimatiques les plus méridionales où la présence de cultures annuelles seraient plus importantes. Dans la zone 1 plus méridionale du Québec, le ratio de cultures pérennes est de 17 % alors qu'il est de 65 et 71 % dans les zones 2 et 3 de l'Ontario respectivement où le climat et les cultures se ressemblent davantage. Ces différences de ratio sont importantes et pourraient être liées aux deux sources d'information différentes, soit la base de données des cultures déclarées à la FADQ pour le Québec et l'Inventaire des cultures du Canada pour l'Ontario. Ce constat amènerait à réévaluer la validité des observations et des calculs utilisés pour générer ces informations et d'ailleurs pousser l'analyse pour inclure d'autres types de cultures exploitables dans ces bases de données.

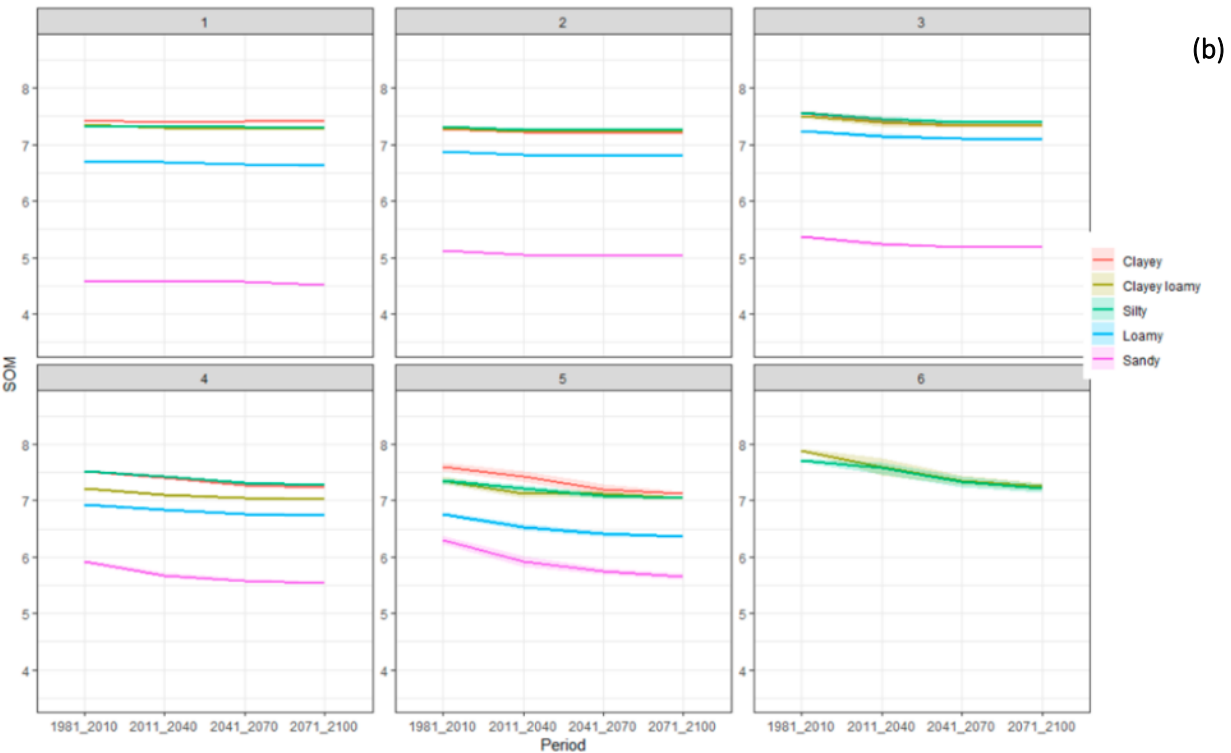
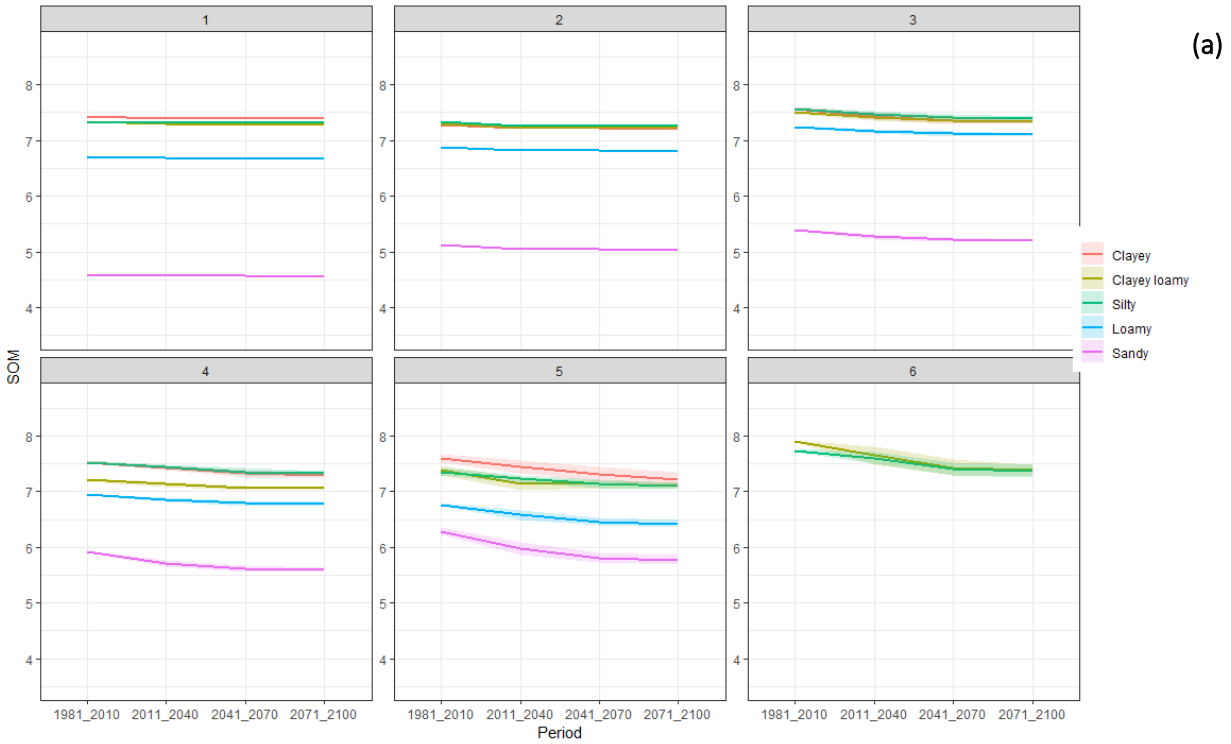


Figure 37 : Projections de l'évolution des teneurs en MOS en Ontario par zones agroclimatiques et groupes texturaux représentées par la moyenne (ligne pleine) et l'écart-type (parties ombragées) des projections de 11 modèles climatiques en RCP 4.5 (a) et 8.5. (b).

Tableau 9 : Ratios de cultures pérennes actuels dans les six zones agroclimatiques du Québec et de l'Ontario.

Zone agroclimatique	Québec (%)	Ontario (%)
1	16,9	38,0
2	52,5	65,2
3	57,3	71,1
4	66,8	93,9
5	60,6	73,6
6	74,0	95,9

Les résultats mitigés du modèle développé dans les conditions du Québec et appliqué dans les conditions de l'Ontario soulignent à quel point les données d'entraînement sont nécessaires et doivent à la fois être de mêmes sources et représenter le même domaine dans lequel les simulations sont réalisées. Par exemple, l'impossibilité d'obtenir des données sur la matière organique des sols en Ontario et le calcul du ratio de cultures pérennes avec des données de la FADQ pour le Québec et de AAC pour l'Ontario ont eu un impact important sur la capacité prédictive du modèle en Ontario. Pour ce qui est du domaine des simulations, la température moyenne actuelle de la zone agroclimatique 1 au Québec est de 5,8°C, alors que les modèles climatiques prévoient un réchauffement à plus de 11,5°C en climat futur (2071-2100 en RCP 4.5) dans la zone 1 de l'Ontario. Une calibration du modèle dans les conditions actuelles plus chaudes du sud de l'Ontario (8,6°C dans la zone agroclimatique 1) aurait été plus représentative pour des simulations en climat futur dans la zone 1 au Québec (8,3°C en 2071-2100 en RCP 4.5). Toutefois, le modèle d'apprentissage machine retenu Random Forest a été assez robuste pour ne pas prédire des évolutions incontrôlées de la MOS en climat futur contrairement à certains des autres modèles essayés.

Conclusion

L'objectif opérationnel de cette étude visait à utiliser une approche de substitution de l'espace avec le temps pour prédire les effets des changements climatiques et de l'occupation du sol sur l'évolution de la teneur en matière organique des sols agricoles (MOS). Certaines prémisses étaient toutefois sous-entendues. Par exemple, on suppose que les 290 000 observations fournies par le laboratoire commercial, filtrées et retenues proviennent essentiellement de champs agricoles dont la géoréférence est dans un rayon de 15 à 100 km du centroïde associé au code postal du client. Des analyses préliminaires avec des coordonnées connues ainsi que l'analyse de l'autocorrélation spatiale des données à différentes échelles de support d'échantillonnage n'ont pas démontré d'avantages à augmenter la taille du support au-delà du polygone de la municipalité associé au code postal. On suppose également que les teneurs en MOS observées chez les producteurs ont atteint un certain équilibre dans le temps qui est conditionné principalement par les conditions actuelles du climat, du choix des cultures, des pratiques agricoles, des types de sols et des conditions dans lesquelles ils se sont développés et révélés entre autres par le modèle numérique de terrain. L'analyse des données n'a pas révélé de tendance ni à la baisse ni à la hausse des teneurs en MOS sur la période de 2010 à 2018 à l'échelle du Québec.

Les fractions granulométriques des sols ont été dérivées à partir d'autres propriétés analysées et d'un jeu plus restreint d'analyses granulométriques réalisées par le laboratoire pour inclure ces paramètres dans la modélisation et l'analyse. Malgré la méthode simplifiée utilisée par le laboratoire pour déterminer la granulométrie, les modèles retenus pour corriger et prédire les fractions granulométriques ont été relativement efficaces pour prédire les fractions d'argile et de sable. Les fractions d'argile et de sable se sont d'ailleurs retrouvées parmi les premières variables d'importance intégrées dans les modèles développés pour spatialiser la teneur en MOS à l'échelle du Québec. Le ratio de cultures pérennes et l'indicateur climatique CDI sont arrivés en 3^e et 4^e position respectivement dans le modèle Random Forest présentant les meilleurs R^2 (0,49) et RMSE (1,82 %) parmi les modèles de régression développés.

La prédiction du ratio de cultures pérennes a été plus évidente à modéliser à l'échelle du Québec avec des modèles dont le R^2 passe de 0,81 en régression linéaire multiple à 0,84 en Random Forest. Le modèle XGBlineair a été retenu en raison de sa plus grande sensibilité aux changements climatiques. Les premières variables d'importance sont des indicateurs climatiques comme le CDI ou les degrés-jours de croissance, suivi d'une moindre importance par des covariables liées au modèle numérique de terrain.

L'application du modèle de régression en conditions de climat futur a produit des teneurs initiales de MOS par zone similaires aux observations de départ mais avec moins d'amplitude. Les changements climatiques et le type de texture de sols ont généré plus de variabilité sur les teneurs en MOS que le changement de ratio de cultures pérennes, malgré l'amplitude de ce changement dans certaines zones plus froides. Les scénarios d'émissions de GES (RCP 4.5 et 8.5) n'ont pas non plus généré de différence aussi importante que les changements climatiques anticipés dans les zones plus froides. Sur un horizon de 2070-2100, la teneur en MOS pourrait diminuer de façon plus importante dans certaines zones agroclimatiques plus froides (zones 3, 5 et 6), notamment dans les sols argileux en zone 3 dans les régions des Laurentides, du Témiscamingue et du pourtour du Lac-Saint-Jean et dans les sols sableux en zone 6 au nord du Lac-Saint-Jean.

Les diminutions projetées sur 90 ans (2010 à 2100), de 0,4 à 1,0 % de MOS dans les sols sableux et argileux équivalent respectivement à des pertes annuelles de 0,0026 et 0,006 % C/an soit de l'ordre de grandeur du 0,008 % C/an rapporté par Warren et al. (2020) pour l'Ontario.

Rapporté en termes de Mg C/ha en utilisant un horizon de sol de surface de 20 cm de profondeur et une masse volumique apparente de 1,3 g/cm³, ces baisses 0,4 et 1,0 % de MOS équivalent respectivement à des pertes de 6 et 15 Mg C/ha soit des baisses plus faibles mais dans l'ordre de grandeur de la perte de 20 à 40 Mg C/ha

rapporté par Gottschalk et al. (2012) pour la période de 1970-2100 sous l'effet de la déforestation ou conversion en terres arables dans l'Est du Canada.

Rapporté sur une base annuelle (90 ans entre 2010 et 2100), les pertes de 0,07 et 0,17 Mg C/ha/an sont plus élevées que les pertes de 0,02 à 0,055 Mg C/ha/an rapportées par Smith et al. (2009), et simulées avec le modèle Century dans des régions du sud de l'Ontario et du Québec. De fait, notre modèle produit aussi des pertes pratiquement nulles en moyenne dans la zone agroclimatique 1 la plus chaude du Québec.

L'application du modèle d'évolution de la MOS dans les conditions de l'Ontario a donné des résultats mitigés pour différentes raisons envisageables. Premièrement, certaines simulations ont été réalisées en dehors du domaine d'entraînement du modèle. Par exemple, les températures moyennes sont plus élevées en climat actuel dans les zones les plus chaudes en Ontario comme en climat futur pour les zones plus chaudes au Québec, par rapport au climat actuel qui a servi à la modélisation. Deuxièmement, le ratio de cultures pérennes très différent entre les deux provinces peut résulter soit d'une réalité agricole différente ou de la source de données différente utilisée pour calculer ces ratios. Cette deuxième raison mériterait d'être analysée.

D'autres travaux pourraient être réalisés pour améliorer la capacité prédictive de ces modèles dont l'inclusion d'autres ratios de cultures comme le soya reconnu pour laisser peu de résidus aux sols et qui a pris de l'expansion ces dernières années dans les régions plus froides à mesure que le climat se réchauffe. Les travaux pour désagréger les analyses de sols relevées à l'échelle des municipalités (travaux de T.H. Easher de l'équipe d'Asim Biswas) seraient à poursuivre pour associer les teneurs en MOS aux unités cartographiques des cartes de sols, ou pour les associer à des grilles plus fines (250 m) associées à des covariables du MNT ou du climat. Finalement, la calibration du modèle dans les conditions plus chaudes actuelles de l'Ontario permettrait de simuler des conditions plus chaudes rencontrées en climat futur au Québec avec plus d'assurance. Sur ce dernier point, l'incertitude des projections n'a pas été mesurée et rapportée. Seule la variabilité des 11 modèles de simulation du climat futur par RCP nous permet de mesurer l'incertitude générée en climat futur.

Un important travail a aussi été réalisé par Farzaneh Heidari sous la supervision du professeur Abdoulaye Baniré Diallo à l'UQAM qui consistait à développer des modèles de prédictions de l'évolution de certaines propriétés physiques des sols (masse volumique apparente, macroporosité, conductivité hydraulique, DMP) à partir des données de l'Inventaire des problèmes de dégradation des sols de 1990. Les modèles permettaient de prédire certaines de ces propriétés en profondeur à partir de propriétés de l'horizon de surface, dont la MOS qui aurait évolué sous l'effet des changements climatiques. Malgré les faibles performances de validation des modèles obtenus (R^2 entre 0,10 et 0,20), il aurait été intéressant de simuler la variation des propriétés physiques sous l'amplitude des changements de MOS observés en climat futur.

Remerciements

Nous sommes redevables à Michel Champagne du Laboratoire AgroEnvirolab d'avoir accepté de fournir le jeu de données d'analyses de sol essentiel à la réalisation de ces travaux. Le laboratoire a reçu en contrepartie une application lui permettant de prédire les fractions granulométriques du sol à partir des autres analyses de sols réalisées, mais nous voulons tout de même remercier Michel et le laboratoire pour l'importante contribution permettant de réaliser ces travaux.

Nous tenons à remercier Maxime Paré de l'UQAC pour son implication dans la demande de subvention et dans l'encadrement de l'étudiant inscrit aux études graduées pour la réalisation de ce projet à ses débuts. Nous tenons également à remercier Claude Bernard de l'IRDA qui était impliqué dès le départ dans l'écriture de la proposition de projet et la demande de subvention sans qui cette partie du projet n'aurait pu voir le jour. Nous voulons également remercier Farzaneh Heidari encadré par le professeur Abdoulaye Baniré Diallo à l'UQAM et son professionnel de recherche Maxime Leduc pour avoir réalisé des travaux de modélisation sur les propriétés physiques des sols. Nos remerciements vont également à Pierre-Yves Gasser pour sa contribution en expertise sur l'expansion des cultures en climat futur.

Nos remerciements vont également à Marco Braun du consortium Ouranos pour la fourniture des données climatiques en climat actuel et futur ainsi que le support pour le calcul d'indicateurs climatiques dérivés et à Anne Blondlot pour le suivi des échanges au sein du comité de suivi. À cet égard, nous voulons remercier et souligner l'implication des membres du comité de suivi : Denis Angers, retraité d'AAC, Martin Bolinder de SLU, Anne Blondlot d'Ouranos, Stéphane Martel du MAPAQ, Kevin McKague de OMAFRA, Michel Champagne retraité d'Agroenvirolab, Éric Thibeault de Pleineterre, et Gabe Ferguson du Rural Ontario Institute.

Références

- Adair, E. C., Parton, W. J., Del Grosso, S. J., Silver, W. L., Harmon, M. E., Hall, S. A., Burke, I. C. & Hart, S. C. (2008). Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates: Modeling long-term decomposition in diverse climates. *Global Change Biology*, 14(11), 2636–2660. <https://doi.org/10.1111/j.1365-2486.2008.01674.x>
- Andrén, O. & Kätterer, T. (1997). ICBM: The introductory carbon balance model for exploration of soil carbon balances. *Ecological Applications*, 7(4), 1226-1236. [https://doi.org/10.1890/1051-0761\(1997\)007\[1226:ITICBM\]2.0.CO;2](https://doi.org/10.1890/1051-0761(1997)007[1226:ITICBM]2.0.CO;2)
- Audet R., Côté, H., Bachand, D., & Mailhot, A. (2012). Atlas agroclimatique du Québec : évaluation des opportunités et des risques agroclimatiques dans un climat en évolution. Rapport final pour Ouranos. CRAAQ. Ouranos. Projet # 550001-XPI.
- Blois, J., Williams, J., Fitzpatrick, M., Jackson, S., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*. 110. <https://doi.org/10.1073/pnas.1220228110>
- Bolinder, M.A., Janzen, H.H., Gregorich, E.G., Angers, D.A., & VandenBygaart, A.J. (2007). An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. *Agriculture, Ecosystem & Environment*, 118, 29–42. <http://dx.doi.org/10.1016/j.agee.2006.05.013>
- Bolinder, M.A., Kätterer, T., Poeplau, C., Börjesson, G., & Parent, L.E. (2015). Net primary productivity and below-ground crop residue inputs for root crops: potato (*Solanum tuberosum* L.) and sugar beet (*Beta vulgaris* L.). *Canadian Journal of Soil Science*, 95, 87–93. <http://dx.doi.org/10.4141/CJSS-2014-091>
- Bouyoucos, G. J. (1962). Hydrometer Method Improved for Making Particle Size Analyses of Soils. *Agronomy Journal*, 54(5), 464–465. <https://doi.org/10.2134/agronj1962.00021962005400050028x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, 8(7), 1991–2007. <https://doi.org/10.5194/gmd-8-1991-2015>
- Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ). (2010). *Guide de référence en fertilisation*, 2^e édition, 473 p.
- Day, P. R., (1965). Particle fractionation and particle-size analysis. Chap 43. In *Methods of soil analysis: Part 1 Physical and mineralogical properties, including statistics of measurement and sampling*, C.A. Black (Ed.) <https://doi.org/10.2134/agronmonogr9.1.c43>
- de Brogniez, D., Ballabio, C., Stevens, A., Jones, R. J. A., Montanarella, L. & van Wesemael, B. (2015). A map of the topsoil organic carbon content of Europe generated by a generalized additive model: Soil organic carbon content at pan-European level. *European Journal of Soil Science*, 66(1), 121–134. <https://doi.org/10.1111/ejss.12193>
- Dexter, A.R., Richard, G., Arrouays, D., Czyz, E.A., Jolivet, C., & Duval, O. (2008). Complexed organic matter controls soil physical properties. *Geoderma*, 144, 620–627. <http://dx.doi.org/10.1016/j.geoderma.2008.01.022>
- Dupla, X., Gondret, K., Sauzet, O., Verrecchia, E. & Boivin, P. (2021). Changes in topsoil organic carbon content in the Swiss leman region cropland from 1993 to present. Insights from large scale on-farm study. *Geoderma*, 400. <https://doi.org/10.1016/j.geoderma.2021.115125>

- Easher, T.H., Saurette, D.D., Chappell, E., Lopez, F., Gasser, M.-O., Gillespie, A. Heck, R., Heung, B., Biswas, A. (2023). Sampling and classifier modification to DSMART for disaggregating soil polygon maps. *Geoderma*, 431. 116360. <https://doi.org/10.1016/j.geoderma.2023.116360> .
- Elith, J., Leathwick, J. R. & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- FAO. (2022). Global Soil Organic Carbon Sequestration Potential Map – SOCseq v.1.1. Technical report. Rome.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5). <https://doi.org/10.1214/aos/1013203451>
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367–378. [https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Martinelli, G., & Gasser, M.-O. (2022). Machine learning models for predicting soil particle size fractions from routine soil analyses in Quebec. *Soil Science Society of America Journal*, 00, 1– 14. <https://doi.org/10.1002/saj2.20469>
- Michaud, A.R., Comeau, S., Gasser, M.-O., & Ricard, S. 2022. OGEMOS. Outil de gestion de la matière organique des sols. <https://github.com/IRDA/OGEMOS>
- Gee, G. W. & Bauder, J.W. (1979). Particle Size Analysis by Hydrometer: A Simplified Method for Routine Textural Analysis and a Sensitivity Test of Measurement Parameters. *Soil Science Society of America Journal*, 43(5), 1004–1007. <https://doi.org/10.2136/sssaj1979.03615995004300050038x>
- Gottschalk, P., Smith, J. U., Wattenbach, M., Bellarby, J., Stehfest, E., Arnell, N., Osborn, T. J., Jones, C., & Smith, P. (2012). How will organic carbon stocks in mineral soils evolve under future climate? Global projections using RothC for a range of climate change scenarios. *Biogeosciences*, 9, 411-451. <https://doi.org/10.5194/bgd-9-411-2012>
- Gray, J.M., & Bishop, T.F.A. (2016). Change in Soil Organic Carbon Stocks under 12 Climate Change Projections over New South Wales, Australia. *Soil Science Society of America Journal*, 80: 1296–1307. <https://doi.org/10.2136/sssaj2016.02.0038>
- Gray, J.M., Bishop, T.F.A. & Wilson. B.R. (2015). Factors controlling soil organic carbon stocks with depth in eastern Australia. *Soil Science Society of America Journal*, 79:1741–1751. <https://doi.org/10.2136/sssaj2015.06.0224>
- Gubler, A., Wächter, D., Schwab, P., Müller, M. & Keller, A. (2019). Twenty-five years of observations of soil organic carbon in swiss croplands showing stability overall but with some divergent trends. *Environ. Monitoring and Assessment*, 191. 277. <https://doi.org/10.1007/s10661-019-7435-y>
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). Random Forests. In T. Hastie, R. Tibshirani & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, pp. 587–604. Springer New York. https://doi.org/10.1007/978-0-387-84858-7_15
- Hénin, S., & Dupuis, M. (1945). Essai de bilan de la matière organique des sols. *Annales Agronomiques*, 15 (1), 161-172.
- Hoffland, E., Kuyper, T.W., Comans, R.N.J., & Creamer, R.E. (2020). Eco-functionality of organic matter in soils. *Plant Soil*, 455, 1–22. <https://doi.org/10.1007/s11104-020-04651-9>
- Husson, F., Josse, J. & Pagès, J. (2010) Principal Component Methods Hierarchical Clustering Partitional Clustering: Why Would We Need to Choose for Visualizing Data? Technical Report – Agrocampus. Applied Mathematics Department. Data. http://www.sthda.com/english/upload/hcpc_husson_josse.pdf

- Johannes, A., Matter, A., Schulin, R., Weisskopf, P., Baveye, P., & Boivin, P. (2017). Optimal organic carbon values for soil structure quality of arable soils. Does clay content matter? *Geoderma*, 302. <https://doi.org/10.1016/j.geoderma.2017.04.021>
- Kröbel, R., Janzen, H. H., Beauchemin, K. A., Bonesmo, H., Little, S. M., & McAllister, T. A. (2013): A proposed approach to estimate and reduce the environmental impact from whole farms, *Acta Agriculturae Scandinavica, Section A – Animal Science*, 62(4), 225-232. <https://doi.org/10.1080/09064702.2013.770912>
- Kuhn M, (2017). Caret: classification and regression training. R package version 6.0-76. <https://CRAN.R-project.org/package=caret>
- Kuhn, M. & Johnson, K. (2013). Regression Trees and Rule-Based Models. In M. Kuhn & K. Johnson, *Applied Predictive Modeling*, pp. 173–220. Springer New York. https://doi.org/10.1007/978-1-4614-6849-3_8
- Lal Rattan, Griffin Michael, Apt Jay, Lave Lester & Morgan M. Granger. (2004). Managing Soil Carbon. *Science*, 304(5669), 393–393. <https://doi.org/10.1126/science.1093079>
- Lindsay, JB. (2014). The Whitebox Geospatial Analysis Tools project and open-access GIS. *Proceedings of the GIS Research UK 22nd Annual Conference*, The University of Glasgow, 16-18 April, <https://doi.org/10.13140/RG.2.1.1010.8962>
- Lu, Y., Liu, F., Zhao, Y., Song, X. & Zhang, G. (2019). An integrated method of selecting environmental covariates for predictive soil depth mapping. *Journal of Integrative Agriculture*, 18(2), 301–315. [https://doi.org/10.1016/S2095-3119\(18\)61936-7](https://doi.org/10.1016/S2095-3119(18)61936-7)
- Matus, F.J. (2021). Fine silt and clay content is the main factor defining maximal C and N accumulations in soils: a meta-analysis. *Science Report*, 11, 6438. <https://doi.org/10.1038/s41598-021-84821-6>
- Meersmans, J., De Ridder, F., Canters, F., De Baets, S. & Van Molle, M. (2008). A multiple regression approach to assess the spatial distribution of Soil Organic Carbon (SOC) at the regional scale (Flanders, Belgium). *Geoderma*, 143(1–2), 1–13. <https://doi.org/10.1016/j.geoderma.2007.08.025>
- Meersmans, J., van Wesemael, B. Goidts, E., Van Molle, M., De Baets, S. & De Ridder, F. (2010). Spatial Analysis of Soil Organic Carbon Evolution in Belgian Croplands and Grasslands, 1960-2006. *Global Change Biology*, 17, 466 - 479. <https://doi.org/10.1111/j.1365-2486.2010.02183.x>
- Meersmans, J., Martin, M., Lacarce, E., Orton, T.G., Baets, S., Gourat, M., Saby, N., Wetterlind, J., Bispo, A., Quine, T., & Arrouays, D. (2013). Estimation of Soil Carbon Input in France: An Inverse Modelling Approach. *Pedosphere*, 23, 422-436. [https://doi.org/10.1016/S1002-0160\(13\)60035-1](https://doi.org/10.1016/S1002-0160(13)60035-1)
- Meurer, K. H. E., Haddaway, N. R., Bolinder, M. A. & Kätterer, T. (2018). Tillage intensity affects total SOC stocks in boreo-temperate regions only in the topsoil: a systematic review using an ESM approach. *Earth-Science Reviews*, 177, 613–22. <https://doi.org/10.1016/j.earscirev.2017.12.015>
- Nunes, M.R.; van Es, H.M.; Veum, K.S.; Amsili, J.P.; Karlen, D.L. (2020). Anthropogenic and Inherent Effects on Soil Organic Carbon across the U.S. *Sustainability*, 12, 5695. <https://doi.org/10.3390/su12145695>
- Parton, W.J. (1996). The CENTURY model. In: Powlson, D.S., Smith, P., Smith, J.U. (eds) *Evaluation of Soil Organic Matter Models*. NATO ASI Series, vol 38. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-61094-3_23
- Poulton, P., Johnston, J., Macdonald, A., White, R. & Powlson, D. (2018). Major limitations to achieving “4 per 1000” increases in soil organic carbon stock in temperate regions: evidence from long-term

- experiments at Rothamsted Research, UK. *Global Change Biology*, 24(6), 2563-2584. <https://doi.org/10.1111/gcb.14066>
- R Core Team. 2022. R: A language and environment for statistical computing. *R Found. Statistics and Computing*, Vienna. <https://www.R-project.org>
- Reijneveld, A., vanWensem, J. & Oenema, O. (2009). Soil organic carbon contents of agricultural land in the Netherlands between 1984 and 2004. *Geoderma*, 152, 231-238. <https://doi.org/10.1016/j.geoderma.2009.06.007>
- Rochette, P. & Dubé, P.A. (1989). Calibration d'une équation simple pour l'estimation de l'évapotranspiration potentielle. *Naturaliste canadien*, 116 :193–203.
- Sierra, C. A., Müller, M., & Trumbore, S. E. 2012. Models of soil organic matter decomposition: the SoilR package, version 1.0. *Geoscientific Model Development*, 5, 1045–1060. <https://doi.org/10.5194/gmd-5-1045-2012>
- Sierra, C. A., Trumbore, S. E., Davidson, E. A., Vicca S. & I. Janssens (2015). Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture, *Journal of Advances Modeling Earth Systems*, 7, 335–356. <https://doi.org/10.1002/2014MS000358>
- Smith, P., Soussana, J. F., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., Batjes, N. H., van Egmond, F., McNeill, S., & Kuhnert, M. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology*, 26(1), 219-241. <https://doi.org/10.1111/gcb.14815>
- Smith, P., Smith, J., Wattenbach, M., Meyer, J., Lindner, M., Zaehle, S. et al. (2006). Projected changes in mineral soil carbon of European forests, 1990–2100. *Canadian Journal of Soil Science*, 86:159–169. <https://doi.org/10.4141/S05-078>
- Smith, W., Grant, B., Desjardins, R., Qian, B., Hutchinson, J., & Gameda, S. (2009). Potential impact of climate change on carbon in agricultural soils in Canada 2000–2099. *Climatic Change*, 93, 319-333. <https://doi.org/10.1007/s10584-008-9493-y>
- Warren, C.J., Saurette, D.D. & Gillespie, A.W. 2020. Soil organic carbon content: decreases partly attributed to dilution by increased depth of cultivation in southern Ontario. *Canadian Journal of Soil Science*, 101(2), 335-338. <https://doi.org/10.1139/cjss-2020-0092>
- Yigini, Y. & Panagos, P. (2016). Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Science of The Total Environment*, 557–558, 838–850. <https://doi.org/10.1016/j.scitotenv.2016.03.085>

ANNEXES

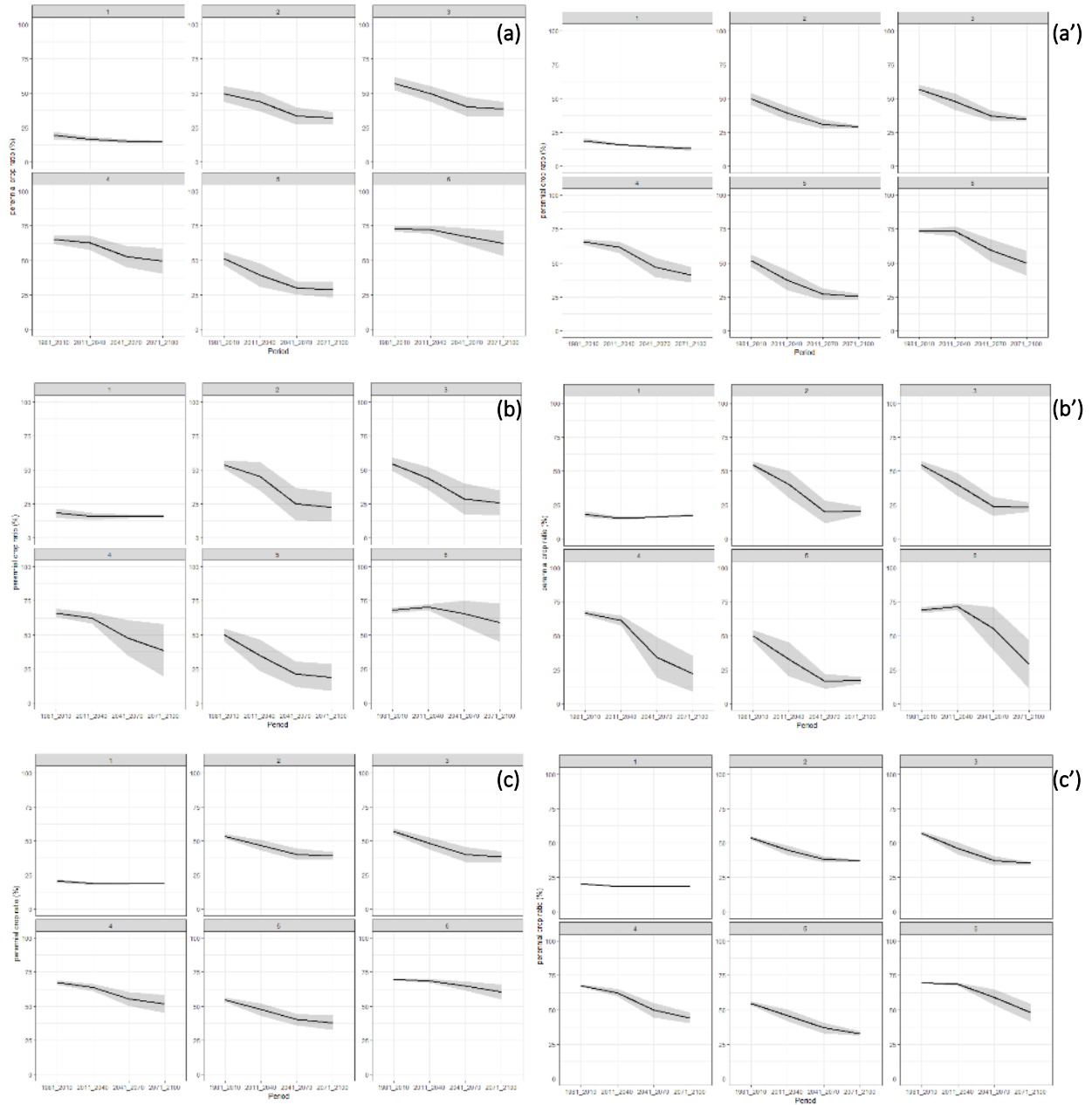


Figure 38 : Projections de l'évolution du ratio de cultures pérennes par zone climatique avec les modèles XGBtree (a et a'), XGBlinear (b et b') et RF (c et c') en RCP 4.5 () et 8.5 ('), respectivement. La ligne pleine et les parties ombragées présentent la moyenne et l'écart-type des projections de 11 modèles climatiques.

Évolution d'indicateurs spatialisés de la santé des sols sous l'effet des changements climatiques au Québec et en Ontario

Affiche présentée dans le cadre du Colloque Ouranos, Déc. 2022

Gasser, M.-O. Biswas, A., Bernard, C., A., Martinelli, G., Easher, T.H., Ondo Z.A.

Projet Ouranos 550028 Durée : 08/2018 – 10/2022

Les sols agricoles vont connaître une intensification de leur utilisation sous l'effet des changements climatiques et plusieurs propriétés des sols liées à la teneur en matière organique du sol (MOS) seront affectées. L'évolution de la MOS et du ratio de cultures pérennes a été modélisée en climat futur à l'échelle du Québec et de l'Ontario avec une approche de substitution de l'espace pour le temps et des techniques d'apprentissage machine. Des analyses de sols d'un laboratoire privé, des données climatiques actuelles et futures fournies par OURANOS, des données d'occupation de sols (zone cultivée et type de culture) provenant de la FADQ et de AAC, ainsi que des données de MNT ont été utilisées. Le modèle Random Forest a été retenu pour réaliser les simulations avec un R^2 de 0,49, incluant par ordre d'importance, la texture, le ratio de cultures pérennes, des indicateurs climatiques et finalement des variables liées au MNT. En 2100, les sols connaîtront une diminution plus importante de leur teneur en MOS dans les régions plus froides du Québec, ainsi qu'une diminution des superficies en cultures pérennes. Mais les diminutions de la MOS dans ces régions seront davantage imputées au climat qu'aux changements de cultures. La texture des sols jouera également un rôle important. La teneur en MOS étant généralement plus faible dans les sols sableux que dans les sols argileux, le climat futur les affectera différemment selon les régions. Les différents scénarios d'émissions de GES anticipés (RCP 4.5 et 8.5) auront moins d'effet et du même ordre de grandeur que le changement de cultures pérennes en annuelles. L'application du modèle climatique à l'échelle ontarienne n'a pas produit les résultats prévus en raison des domaines climatiques mal représentés et des sources de données différentes pour l'évaluation du ratio de cultures pérennes.

Les changements climatiques affecteront davantage la teneur en matière organique des sols dans les régions plus froides du Québec.

Affiche présentée dans le cadre du congrès de l'AQSSS, à Québec le 24 mai 2023.

Marc-Olivier Gasser¹, Gaëtan Martinelli², Tahmid Huq Easher³, Asim Biswas³

¹ Institut de recherche et de développement en agroenvironnement (IRDA), Québec, QC, G1P 3W8

² itk SAS, Montpellier, France

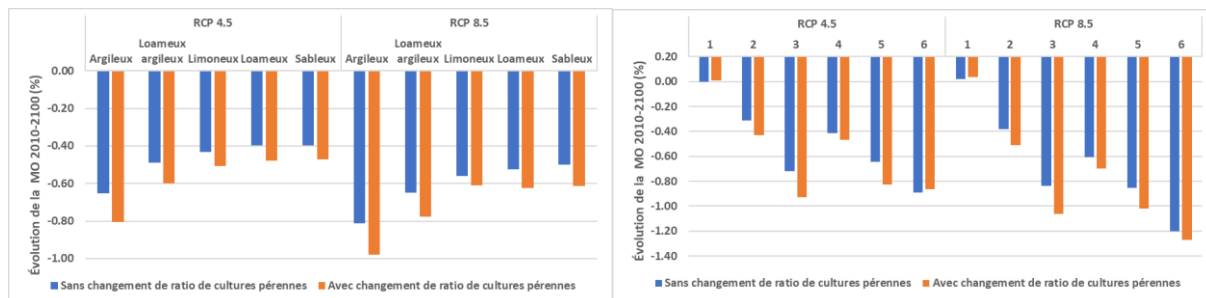
³ University of Guelph, Guelph, ON, N1G 2W1

marc-o.gasser@irda.qc.ca

Mots clés : matière organique, changements climatiques, santé des sols, conservation, dégradation des sols.

Les sols à bon potentiel agricole constituent une ressource limitée au Québec et ils vont connaître une intensification de leur utilisation sous l'effet des changements climatiques. Plusieurs des propriétés liées à la matière organique du sol (MOS) en seront affectées. Pour évaluer ces effets, l'évolution de la MOS et du ratio de cultures pérennes a été modélisé en climat futur à l'échelle du Québec et de l'Ontario avec une approche de substitution de l'espace pour le temps et des techniques d'apprentissage machine. Des analyses de sols d'un laboratoire privé, des données climatiques actuelles et futures fournies par OURANOS, des données d'occupation de sols (zone cultivée et type de culture), ainsi que des données issues du modèle numérique de terrain (MNT) ont été utilisées. Parmi les techniques d'apprentissage machine essayées, le modèle de régression généré avec des forêts d'arbres aléatoires (Random Forest) a été retenu pour réaliser les simulations avec un R² de 0,49. Selon les projections, les sols connaîtront en 2100 une diminution plus importante de leur teneur en MOS dans les régions plus froides du Québec. La teneur en MOS étant généralement plus faible dans les sols sableux que dans les sols argileux, le climat futur les affectera différemment selon les régions. La teneur en MOS pourrait diminuer de façon plus importante dans les sols plus argileux des Laurentides, du Témiscamingue et sur le pourtour du Lac-Saint-Jean et dans les sols sableux au nord du Lac-Saint-Jean. Les diminutions de la MOS dans ces régions seraient davantage affectées par le climat que par l'abandon des cultures pérennes. Les différents scénarios d'émissions de GES anticipés (RCP 4.5 et 8.5) auront aussi moins d'effet que le climat et du même ordre de grandeur que l'abandon des cultures pérennes. Les diminutions projetées sur 90 ans (2010 à 2100), de 0,4 à 1,0 % de MOS dans les sols sableux et argileux équivalent respectivement à des pertes annuelles de 0,0026 et 0,006 % C/an soit de l'ordre de grandeur du 0,008 % C/an rapporté par Warren et al. (2020) pour l'Ontario. Rapporté en termes de Mg C/ha en utilisant un horizon de sol de surface de 20 cm de profondeur et une masse volumique apparente de 1,3 g/cm³, ces baisses de 0,4 et 1,0 % de MOS équivalent respectivement à des pertes de 6 et 15 Mg C/ha soit des baisses plus faibles, mais dans l'ordre de grandeur des pertes de 20 à 40 Mg C/ha rapportées par Gottschalk et al. (2012) pour la période de 1970-2100 sous l'effet de la déforestation ou conversion en terres arables dans l'Est du Canada. Rapporté sur une base annuelle (90 ans entre 2010 et 2100), les pertes de 0,07 et 0,17 Mg C/ha/an sont plus élevées que les pertes de 0,02 à 0,055 Mg C/ha/an simulées avec le modèle Century dans des régions du sud de l'Ontario et du Québec (Smith et al., 2009). De fait, notre modèle produit aussi des pertes pratiquement nulles en moyenne dans la zone agroclimatique 1 la plus chaude du Québec.

Évolution moyenne des teneurs en MOS de 2010 à 2100 sous l'effet de 11 modèles climatiques, deux scénarios d'émissions de GES (RCP 4.5 et 8.5), du changement de ratio de cultures pérennes, selon les groupes texturaux (a) ou six zones agroclimatiques (b).



Références

- Gottschalk, P., Smith, J. U., Wattenbach, et al. (2012). How will organic carbon stocks in mineral soils evolve under future climate? Global projections using RothC for a range of climate change scenarios. *Biogeosciences*, 9, 411-451.
- Smith, W., Grant, B., Desjardins, R., Qian, B., Hutchinson, J., & Gameda, S. (2009). Potential impact of climate change on carbon in agricultural soils in Canada 2000–2099. *Climatic Change*, 93, 319-333.
- Warren, C.J., Saurette, D.D. & Gillespie, A.W. 2020. Soil organic carbon content: decreases partly attributed to dilution by increased depth of cultivation in southern Ontario. *Canadian Journal of Soil Science*, 101(2), 335-338.

Predicting soil organic matter evolution under climate change in Quebec.

Présentation donnée au congrès de la SCSS le 26 juin à Truro, Nouvelle-Écosse

Marc-Olivier Gasser¹, Gaëtan Martinelli¹, Tahmid Huq Easher², Asim Biswas²

¹ Institut de recherche et de développement en agroenvironnement (IRDA), Québec, QC, G1P 3W8

² University of Guelph, Guelph, ON, N1G 2W1

marc-o.gasser@irda.qc.ca

Keywords: organic matter, climate change, soil health, soil conservation, soil degradation.

Soils with good agricultural potential are limited in Québec and their use will intensify because of climate change. Soil organic matter (SOM) and related properties will be affected. To assess these effects, the evolution of the SOM and annual to perennial crop ratio was modeled under future climate using a space-for-time substitution approach and machine learning techniques. Soil analyses from a commercial laboratory, current and future climate data provided by OURANOS, land cover data (cultivated area and crop type), as well as data from a digital elevation model (DEM) were used. Among the machine learning techniques tested, the regression model generated with Random Forest and R^2 of 0.49 was selected to perform the simulations. In 2100, soils will experience a greater decrease in SOM content in the colder regions of Quebec. As the SOM content is generally lower in sandy soils than in clay soils, future climate will affect them differently in the different regions. The SOM content could decrease more significantly in the more clayey soils of the Laurentians, Témiscamingue and around Lac-Saint-Jean and in sandy soils north of Lac-Saint-Jean. The decreases in SOM in these regions will be affected more by climate than by the decline of perennial crops. The different scenarios of projected GHG emissions (RCPs 4.5 and 8.5) will also have less effect than climate and of the same order of magnitude as the decline of perennial crops. Projected over 90 years (2010 to 2100) decreases of 0.4 to 1.0% of SOM respectively in sandy and clay soils are equivalent to annual losses of 0.0026 and 0.006% C/year. Using a surface soil horizon of 20 cm depth and soil bulk density of 1.3 g/cm³, these decreases in SOM are equivalent to losses of 6 and 15 Mg C/ha, and on an annual basis to 0.07 and 0.17 Mg C/ha/year. Our model predicts lower losses of SOM in the warmest zone in Quebec and these losses are equivalent or in the range of what others have reported for the same region.

AXE 2 : SPATIALISATION DES INDICATEURS DE SANTÉ DES SOLS PAR LA DÉSAGRÉGATION DE L'INFORMATION SUR LES CARTES DE SOLS

Asim Biswas, U. of Guelph; Tahmid Huq Easher, U. of Guelph, Old College; Marc-Olivier Gasser, IRDA; Gaëtan Martinelli, IRDA

Spatialization of soil health indicators through disaggregation of conventional soil maps

Digital soil mapping (DSM) is a powerful tool with significant potential for agricultural and environmental management, providing accurate and up-to-date information on soil properties and distributions. It has the capacity to revolutionize our understanding of soils and their crucial role in sustaining life on Earth. With advancements in geospatial technology and data analysis, DSM is rapidly evolving, enhancing its accuracy and utility. Machine learning algorithms have enabled the automation and scalability of DSM, enabling the mapping of soil properties over larger areas and in finer detail than ever before.

The advantages of DSM are numerous. It allows for rapid and efficient coverage of large land areas through remote sensing technologies, providing detailed information on soil properties at a fine scale. This information supports informed decision-making in various domains, such as selecting appropriate crops, applying optimal amounts of fertilizer and water, and preventing soil degradation. Furthermore, DSM has significant applications in environmental management, aiding in the identification of soil degradation, erosion, and other environmental challenges. This knowledge helps develop targeted conservation and restoration strategies to maintain healthy soils and preserve ecosystem services.

DSM caters to diverse stakeholders, supporting evidence-based decision-making and sustainable land management practices. It assists farmers and land managers in making informed choices regarding land use and management practices. Environmental scientists and policymakers benefit from valuable information for the development of effective conservation and management strategies. Additionally, DSM contributes to the study of soil formation and evolution and aids land-use planners in evaluating the suitability of different areas for specific land uses. Overall, DSM plays a pivotal role in soil health and precision agriculture by generating detailed maps of soil variability and characteristics, facilitating informed decision-making and data-driven approaches to soil health management.

In Canada, particularly in Ontario, DSM has seen extensive development and utilization to enhance the understanding and management of soil resources. It has been employed for soil conservation, land use planning, agricultural management, and scientific research, including the study of climate change impacts on soil properties and the advancement of soil mapping methods. The Canadian government has provided funding and data access for DSM research and development, and the country actively participates in the International Union of Soil Sciences. Both federal and provincial governments have contributed significantly to the development and utilization of DSM in Canada by funding projects, supporting research and development, and providing access to data and tools.

DSM has enabled the production of high-resolution maps of soil properties and spatial variability in Ontario, Canada. These maps offer valuable information for diverse applications, such as agriculture, environmental management, and land use planning, aiding decision-making at the local and regional levels. They encompass mapping of soil organic matter/carbon stocks and fluxes, soil texture, and various soil properties across the boreal forest of Ontario. DSM techniques have been utilized in conjunction with topographic data, remote sensing data, and soil samples to generate these high-resolution maps. Furthermore, DSM has been employed

to map soil erosion risk, supporting land use planning and natural resource management. The methodology incorporates topographic data, climate data, and soil samples to produce detailed maps.

1 Soil Series Disaggregation of Agricultural Soils in Ontario

[\(Sampling and classifier modification to DSMART for disaggregating soil polygon maps. Article publié dans Geoderma, Volume 431, March 2023, 116360\)](#)

1.1 Background

Conventional soil maps (CSMs) serve as a valuable reference in digital soil mapping, providing information on soil type distribution and variability. CSMs play a crucial role in the development and validation of digital soil maps. They are used to calibrate and validate digital soil maps by comparing predicted soil properties and classifications with known data from conventional maps. CSMs also facilitate the updating of conventional data by combining information from conventional maps with new data from digital soil mapping, resulting in more detailed and up-to-date maps. Additionally, CSMs offer historical context and information on soil variability and characteristics, supporting the development of digital soil maps. They contribute to environmental monitoring and assessment by providing insights into historical trends in soil properties and facilitating comparisons with current data from digital soil mapping.

However, conventional maps do present certain challenges. They may be difficult to access due to ownership disparities and often lack precise geo-referencing. Inconsistent soil property measurement methods and classification systems across different conventional soil datasets pose additional issues. Temporal concerns arise due to the collection and compilation of soil data from various projects and different years, making integration with newly collected data in the DSM process complex, especially for non-static soil properties like soil organic carbon.

The spatial disaggregation of soil maps is a widely adopted method in DSM to enhance the spatial precision of polygon-based legacy soil maps, particularly when point data is unavailable. Spatial disaggregation involves dividing larger soil units into smaller, more detailed soil units, resulting in a more accurate representation of soil properties and distributions at a finer spatial scale. The process entails obtaining high-quality soil data, employing disaggregation algorithms, validating and refining the output, and integrating the disaggregated soil maps with other relevant data sources. Disaggregation techniques encompass rule-based models, geostatistical methods, decision-tree approaches, and machine learning algorithms.

1.2 Objective

The objective of this part of the study focuses on the spatial disaggregation of the soil series map of Ontario (ON) agricultural soils. The legacy soil series map of the agricultural soils of ON was extracted from the Soil Survey Complex compiled by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA). This polygon-based map provides information on soil classes, texture, pH, and drainage. However, the polygons in the map represent compound map units without a graphical representation of the distribution of the soil type units, only indicating proportions, environmental conditions, and soil characteristics.

To overcome the limitations of the legacy soil maps and enable digital modeling of the soil series, a spatial disaggregation approach was employed. The agricultural soils of Ontario (ON) were disaggregated by converting the polygon maps into point-based raster maps. This was achieved by redistributing multiple soil series within

the polygons, placing them in their correct spatial positions. The purpose of this disaggregation was to create more detailed and accurate representations of the soil series distribution.

1.3 Workflow and Results

For this research, six sub-watersheds within the Lake Erie Basin of Ontario were selected as study areas. These watersheds include Garvey Glenn, Gully Creek, Wigle Creek, Medway Creek, Kettle Creek, and Jeannette's Creek (Figure 39). The disaggregation algorithm used in this study is called The Disaggregation and Harmonisation of Soil Map Units Through Resampled Classification Trees (DSMART). DSMART generates virtual points from the polygon-based soil maps and assigns soil classes proportionately based on the composition of the map units using a machine learning algorithm (MLA) for soil class prediction.

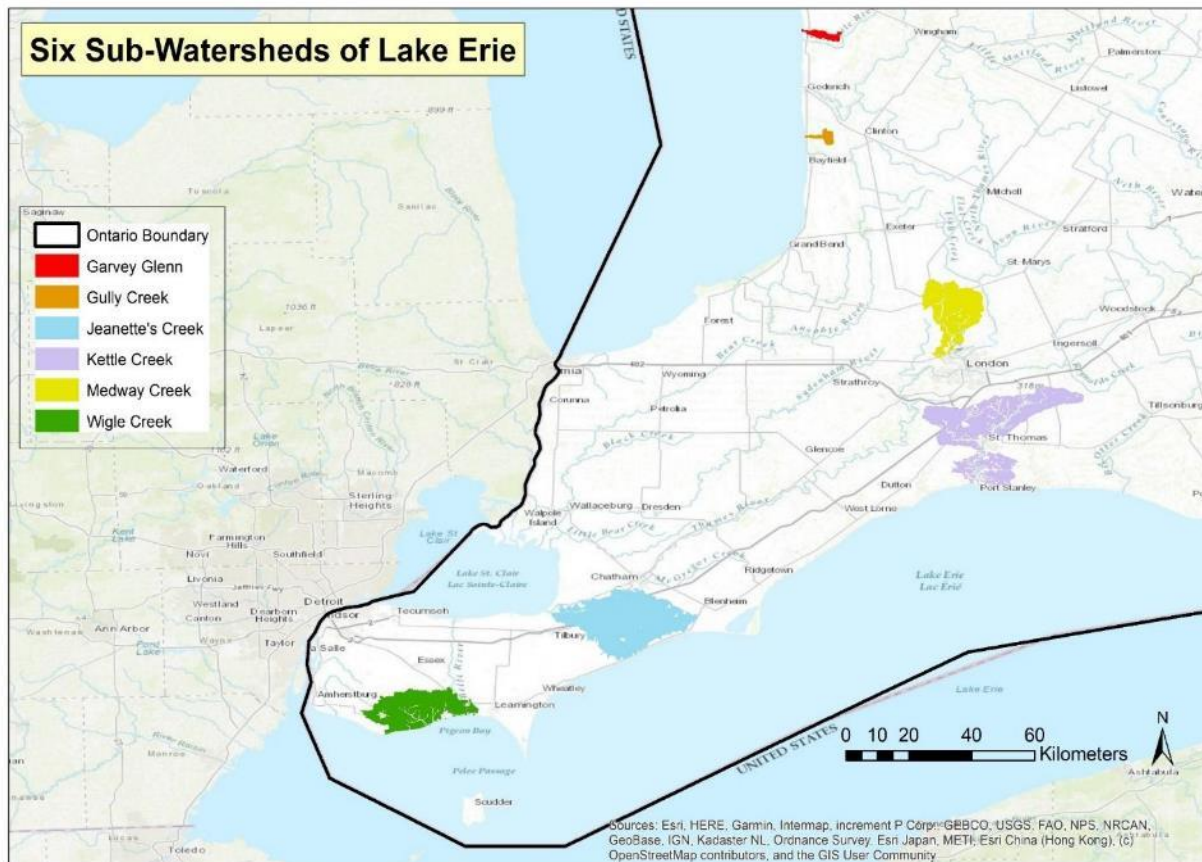


Figure 39 : Six Sub-Watersheds of Lake Erie.

Two sampling approaches of DSMART were considered: "by polygon" and "by area." The "by polygon" approach randomly assigns virtual sampling points within each polygon, while the "by area" approach randomly assigns virtual sampling points per square kilometer. The performance of these sampling approaches was evaluated for all the selected watersheds, and it was found that the "by area" sampling approach outperformed the "by polygon" approach in DSMART.

Figure 40 to Figure 44 illustrate the disaggregated maps of Garvey Glenn Creek, Gully Creek, Jeannette's Creek, Wigle Creek, and Kettle Creek watersheds, respectively, at a resolution of 30 meters. These maps provide a more detailed and accurate representation of the spatial distribution of soil series within the selected sub-watersheds, enabling better understanding and analysis of soil variability and characteristics in the Lake Erie Basin of Ontario.

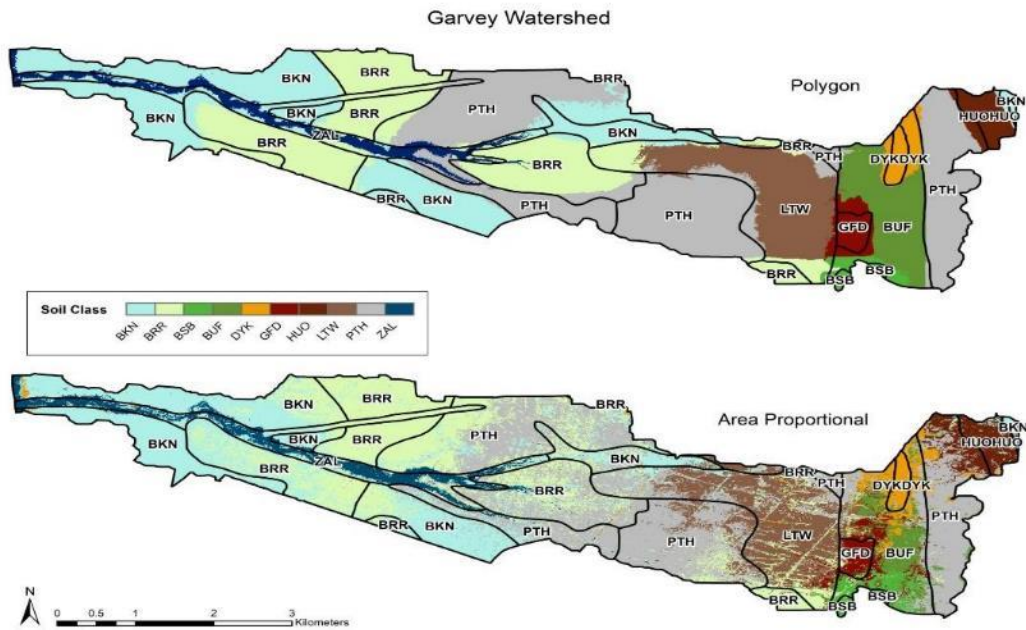


Figure 40 : Disaggregated Soil Series Maps of Garvey Glenn Creek Watershed at 30 m resolution.

In this part of the study, we explored different approaches and methodologies to disaggregate the soil series map of Ontario, specifically focusing on the Lake Erie Basin. Initially, we compared the default random sampling approach in DSMART with conditional latin hypercube sampling (cLHS) to optimize sample point selection based on background information or environmental covariates. Three classifier models or machine learning algorithms (C5.0, random forest, and k-nearest neighbor) were employed in DSMART for soil series disaggregation.

To evaluate the performance of these approaches, we conducted experiments in the Medway Watershed within the Lake Erie Basin. The DSMART package's original R-code was modified to incorporate cLHS sampling and change classifier algorithms. We compared the "by area" cLHS approach combined with the C5.0 machine learning algorithm (MLA) against the "by area" random sampling approach. The evaluation of these models, using 53 validation points, showed that the cLHS approach with C5.0 MLA achieved a Kappa score of 0.44, outperforming the random sampling approach with a Kappa score of 0.45. Figure 45 presents the prediction maps of the model combinations at a resolution of 30 meters, and the results are summarized in Table 1.

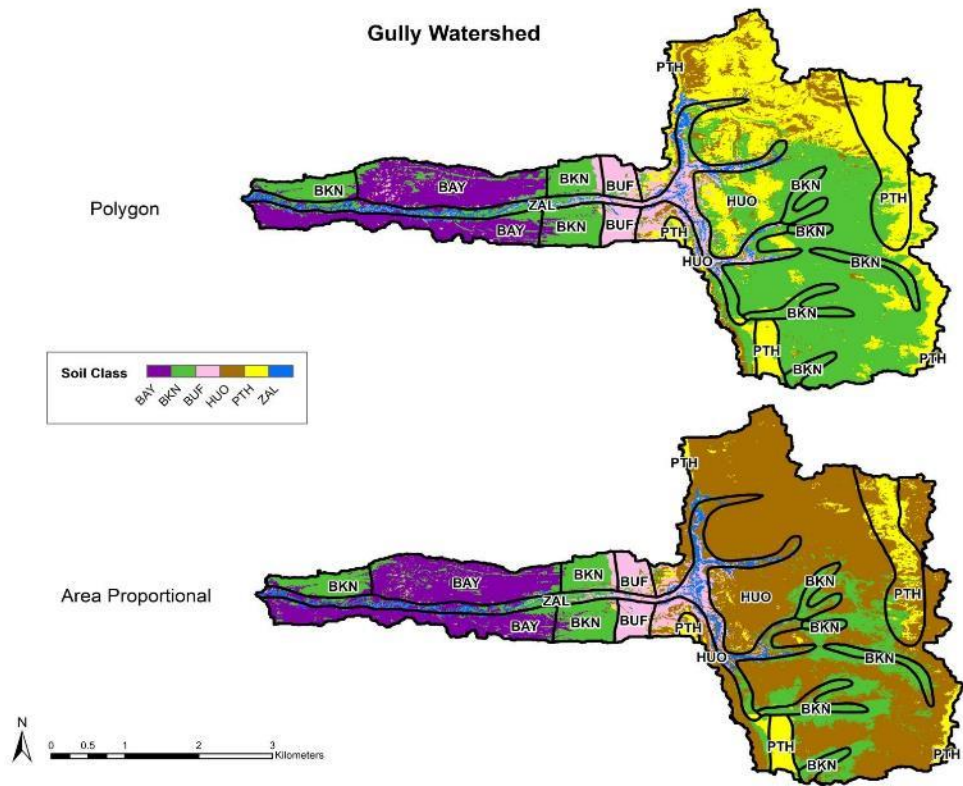


Figure 41 : Disaggregated Soil Series Maps of Gully Creek Watershed at 30 m resolution.

Jeanette's Creek

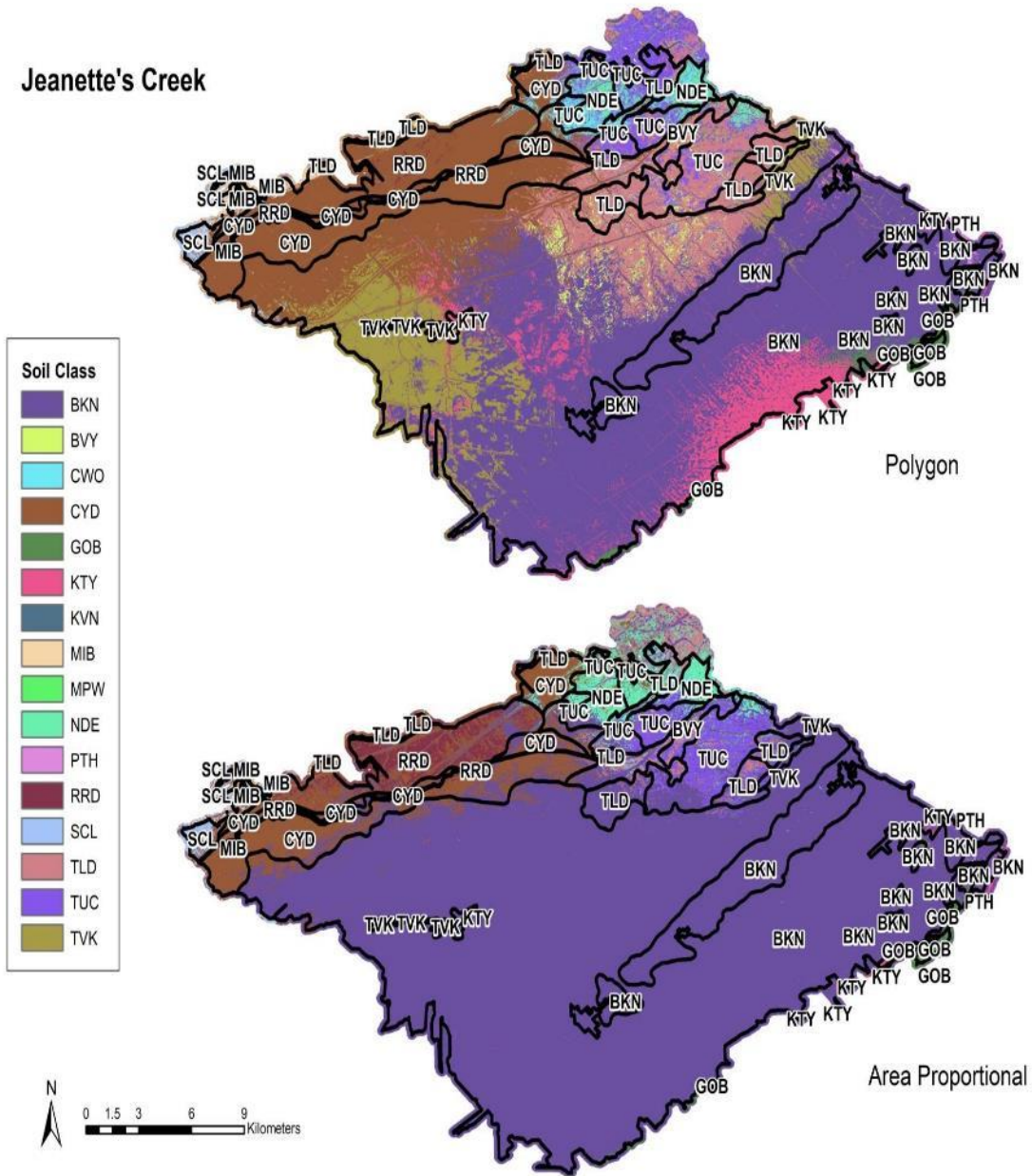


Figure 42 : Disaggregated Soil Series Maps of Jeanette's Creek Watershed at 30 m resolution.

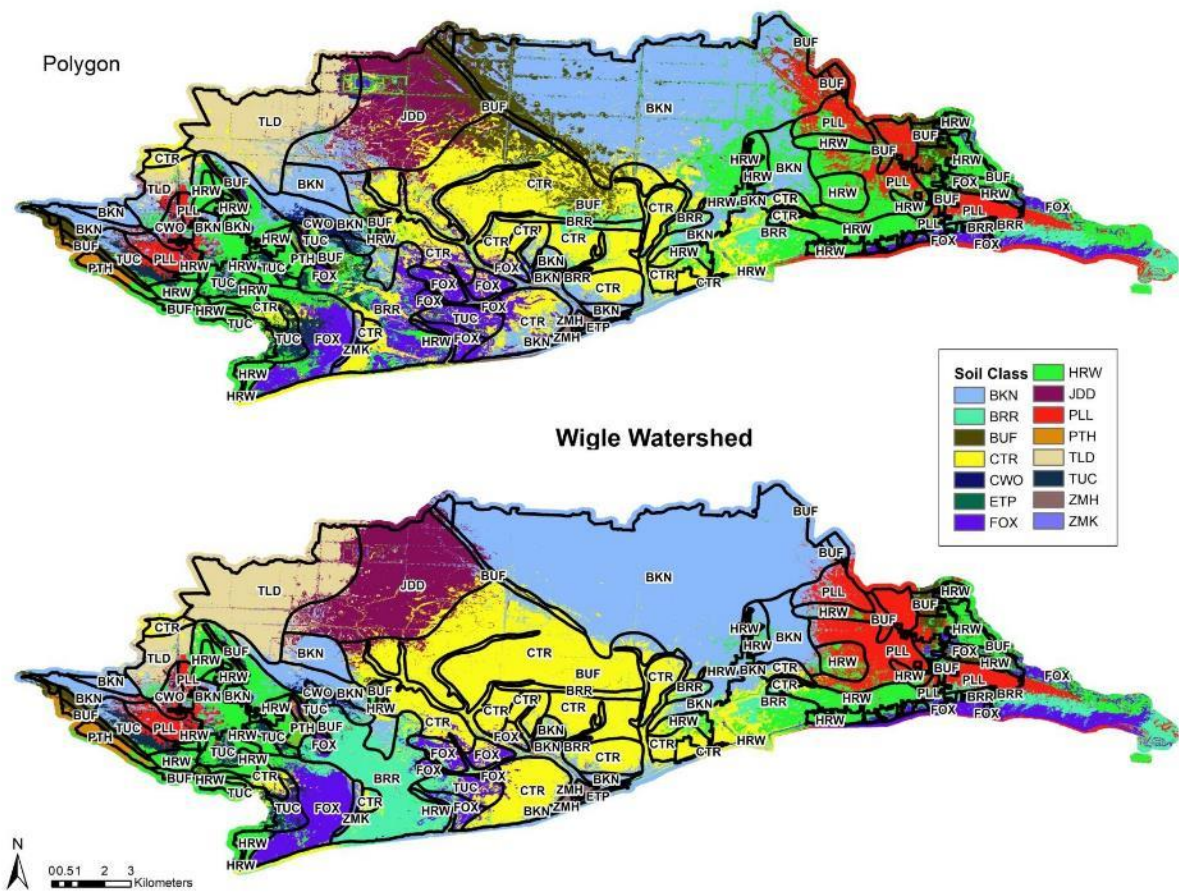


Figure 43 : Disaggregated Soil Series Maps of Wigle Creek Watershed at 30 m resolution.

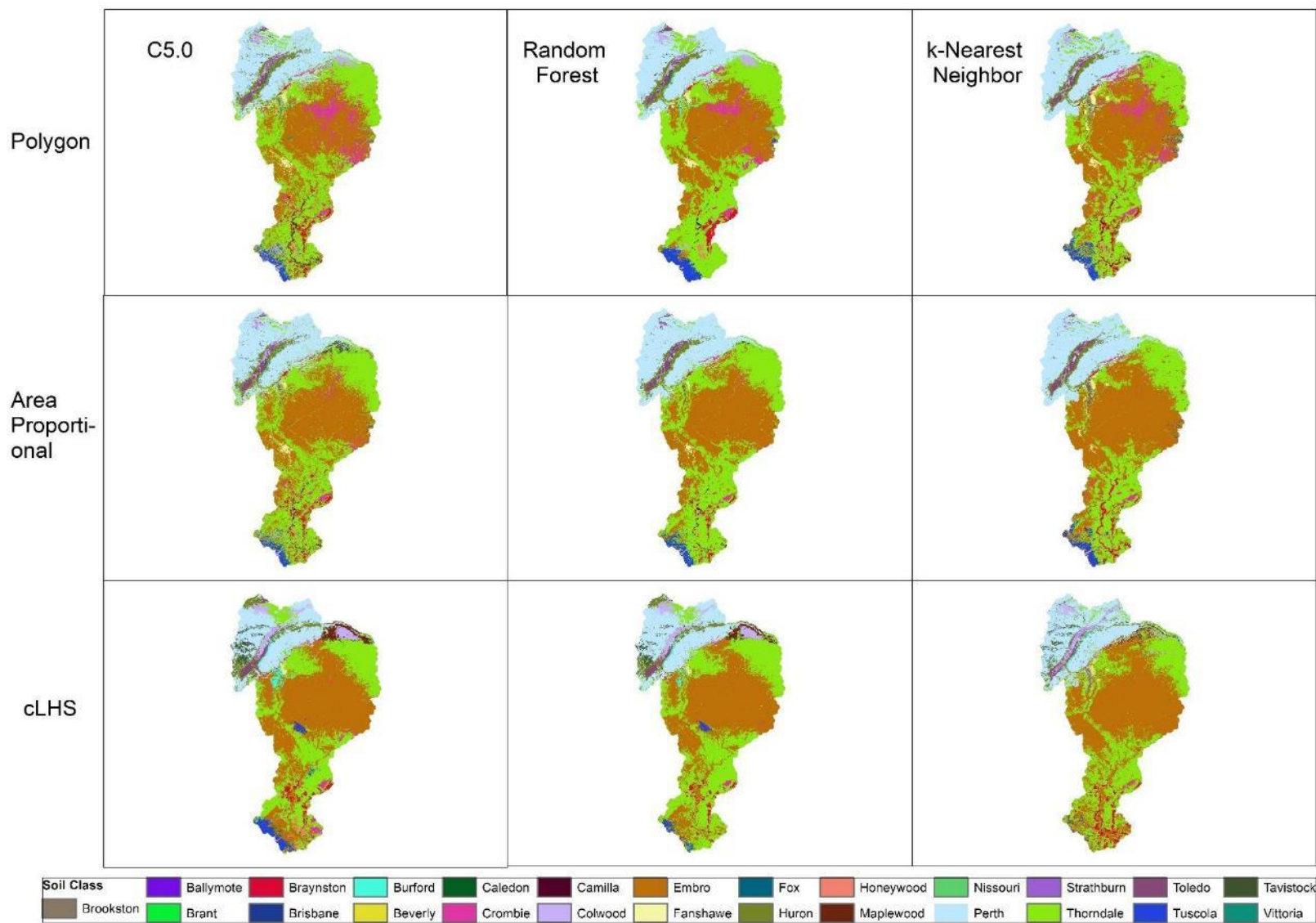


Figure 45 : Comparison of Soil Series Prediction between Different Models and Sampling Approaches in DSMART, Medway Watershed.

Table 1 : Validation results of different modeling and sampling approaches of DSMART in predicting Medway Watershed soil great groups and soil series.

Sampling Method		Random – “by polygon”			Random – “by area”			cLHS – “by polygon”			cLHS – “by area”		
		MLA	C5.0	RF	KNN	C5.0	RF	KNN	C5.0	RF	KNN	C5.0	RF
Soil Series	Overall Accuracy	44%	42%	40%	43%	43%	51%	45%	42%	29%	50%	38%	31%
	Kappa Score	0.38	0.34	0.33	0.37	0.37	0.45	0.39	0.35	0.2	0.44	0.3	0.22

Furthermore, we compared two disaggregation methods, DSMART and the Pure Polygon approach (PPD), for the soil series map of Middlesex County, Ontario (Figure 46). Both methods utilized the "by area" random sampling approach and C5.0 MLA. PPD, which allocated one soil class per polygon for training point development. The predicted maps were evaluated using 500 datapoints extracted from the Middlesex County Soil Survey report 1992. PPD demonstrated superior performance with a Kappa score of 0.66, surpassing the DSMART approach with a Kappa score of 0.57. Figure 47 and Figure 48 illustrate respectively the disaggregated soil series map of Middlesex County, ON at 50 m resolution using DSMART and PPD. The model evaluation results are presented in Table 2, and Table 3 showcases the uncertainty calculation outcomes.

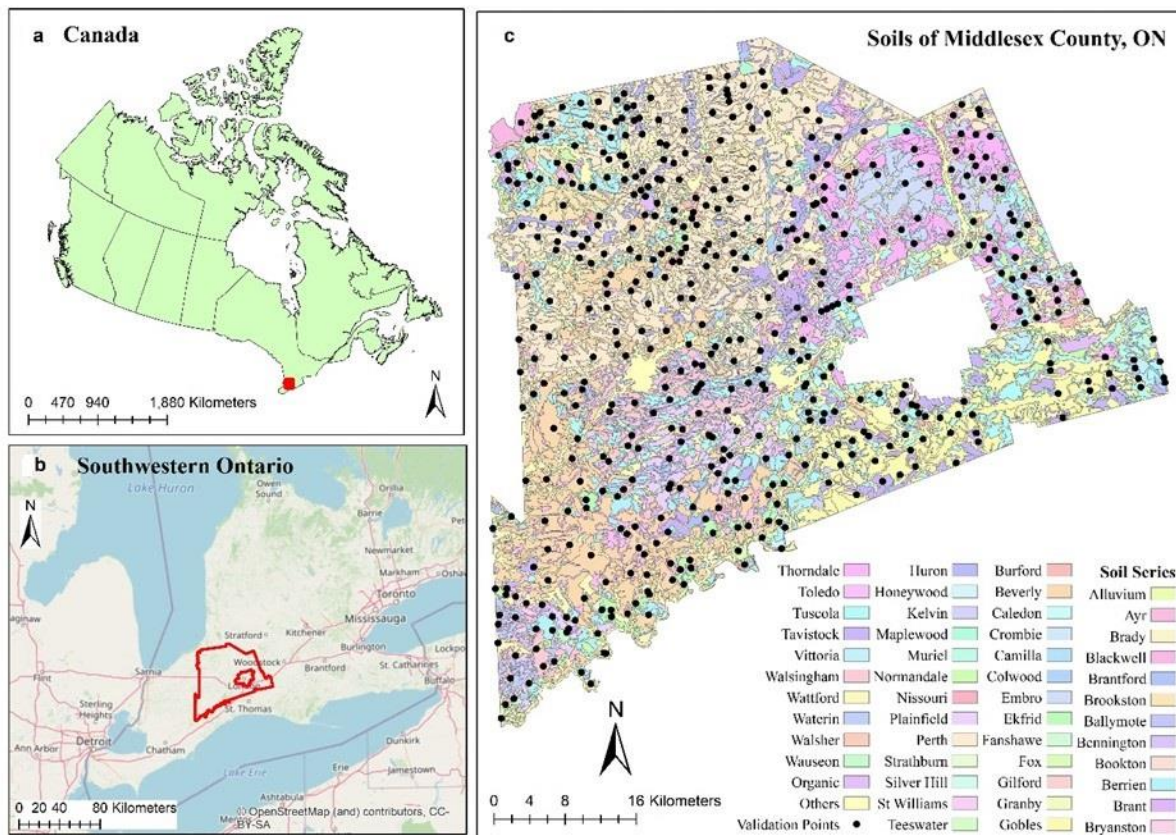


Figure 46 : Study area, Middlesex County, shown by the red dot in the map of Canada (a) and by the red boundary area Ontario and in Southwestern Ontario (b). Soil series maps of Middlesex County and validation points represented by black dots from the 1992 soil survey report of Middlesex County (c).

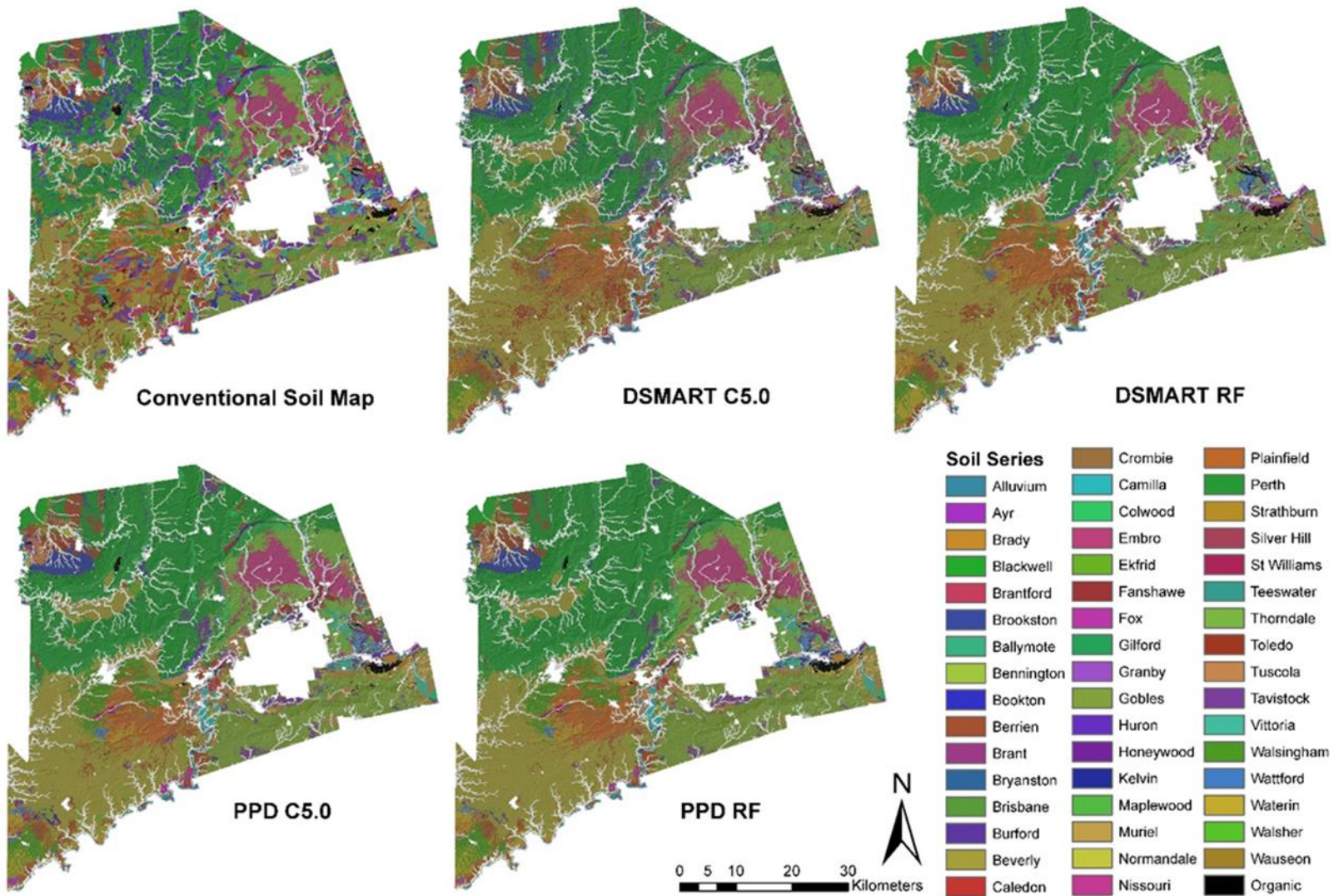


Figure 47 : Disaggregated Soil Series Map using DSMART and Pure Polygon Disaggregation of Middlesex County, ON at 50 m resolution (Area-Weighted Random Sampling Approach)

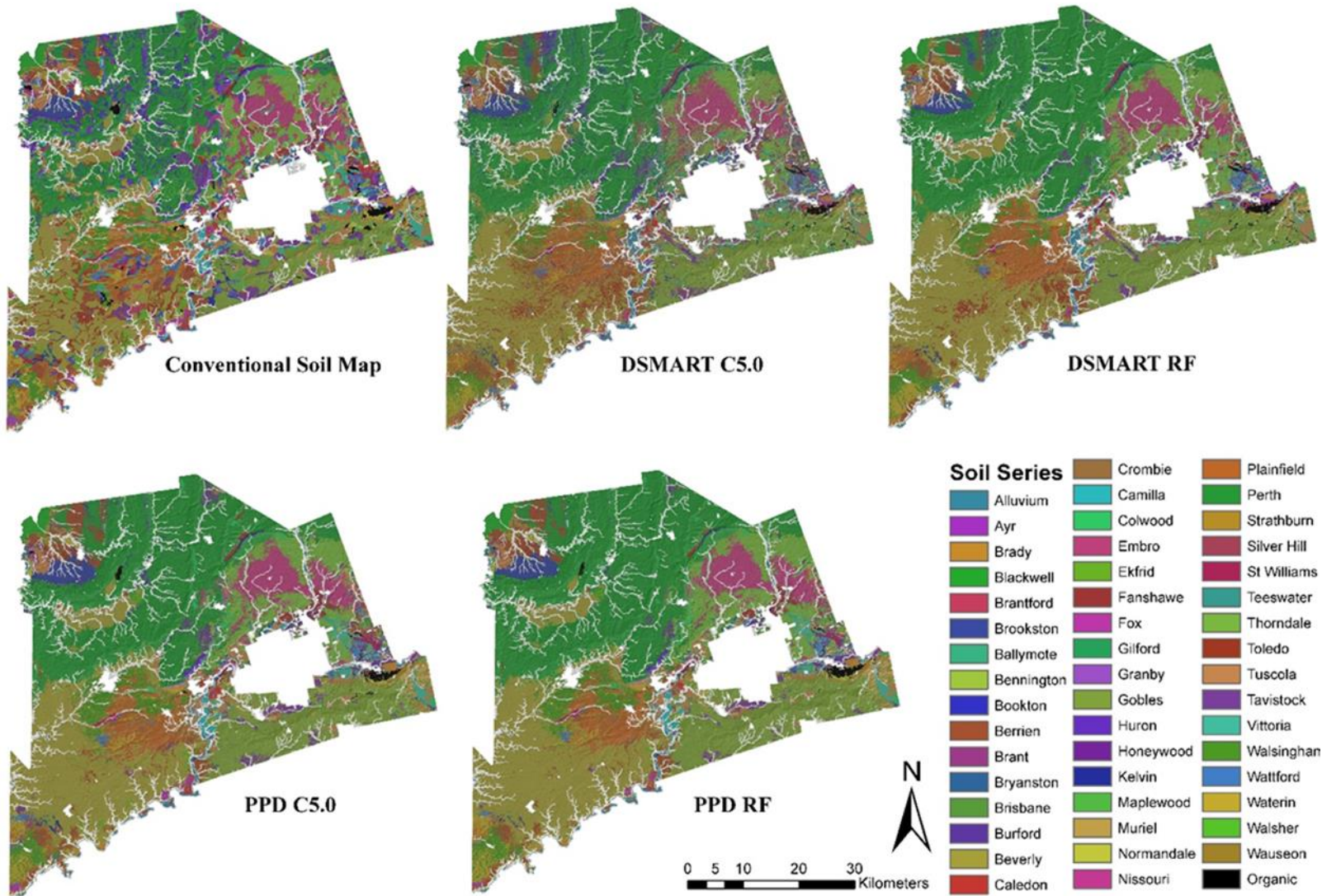


Figure 48 : Disaggregated Soil Series Map using DSMART and Pure Polygon Disaggregation of Middlesex County, ON at 50 m resolution (Conditional Latin Hypercube Sampling Approach)

Table 2 : Validation results of different modeling and sampling approaches of DSMART and Pure Polygon Disaggregation in predicting Middlesex County soil series.

Model		DSMART		PPD	
MLA	Sampling Method	RS	cLHS	RS	cLHS
C5.0	Overall Accuracy	61	61	66	61
	Kappa Index	0.58	0.58	0.63	0.58
RF	Overall Accuracy	57	58	59	58
	Kappa Index	0.53	0.54	0.55	0.54

Legend: MLA = Machine Learning Algorithm, RS = random sampling, cLHS = Conditional Latin Hypercube Sampling, RF = Random Forest.

Table 3 : Entropy uncertainty of different modeling and sampling approaches of DSMART and Pure Polygon Disaggregation in predicting Middlesex County soil series.

Model		DSMART		PPD	
MLA	Sampling Method	RS	cLHS	RS	cLHS
C5.0	Range	0-77%	0-77%	0-72%	0-70%
	Mean	33%	31%	13%	11%
	Standard Deviation	13%	13%	12%	12%
RF	Range	0-65%	0-59%	0-63%	0-61%
	Mean	12%	11%	9%	7%
	Standard Deviation	11%	10%	10%	10%

Legend: MLA = Machine Learning Algorithm, SRS = simple random sampling, cLHS = Conditional Latin Hypercube Sampling, RF = Random Forest.

Based on these findings, we proceeded to disaggregate the soil series map of the Lake Erie Watershed at a resolution of 250 meters (Figure 49 and Figure 50). Both the "by area" sampling approach in DSMART and PPD, employing C5.0 MLA, were utilized. While PPD successfully predicted all 137 soil series in the Lake Erie Basin, DSMART predicted 136 soil series. However, it is worth noting that DSMART required more computational resources and had a longer processing time compared to PPD. To enhance the performance of DSMART, we introduced principal component analysis to reduce the noise in the environmental covariates, which were used to spatially intersect the sampling points for training data generation. Overall, DSMART proved to be a valuable approach for disaggregating the multiple map components of polygon-based legacy soil maps, such as soil series, for modeling and prediction purposes. Nonetheless, if the original data primarily consists of more polygons with a single map unit, PPD is a more effective and computationally efficient alternative.

Based on the findings, we applied PPD, employing C5.0 MLA, to disaggregate the soil series information in the agricultural soil map of Ontario at a resolution of 250 m (Figure 51). In this process, we incorporated additional data such as climate data for the period of 1970-2000, vegetation data, and soil parent material data, which were not utilized during the model testing stages. Moreover, we employed the variance inflation factor algorithm to mitigate multicollinearity among the model predictors (environmental covariates). Figure 52 illustrates the disaggregated soil series map of Ontario agricultural soils at 250 m resolution. 1300 datapoints extracted from National Pedon Database were used to validate the predict soil series map. The kappa index of the model accuracy assessment for predicting soil series in Ontario using the PPD approach was 0.41, with an overall accuracy of 43%. An entropy map (Figure 53) was generated to calculate the model uncertainty, with the entropy uncertainty value (H) ranging from 0 to 0.55, a mean value of 0.15, and a standard deviation of 0.09. Lower values of H indicate lower uncertainty in the soil series classification model.

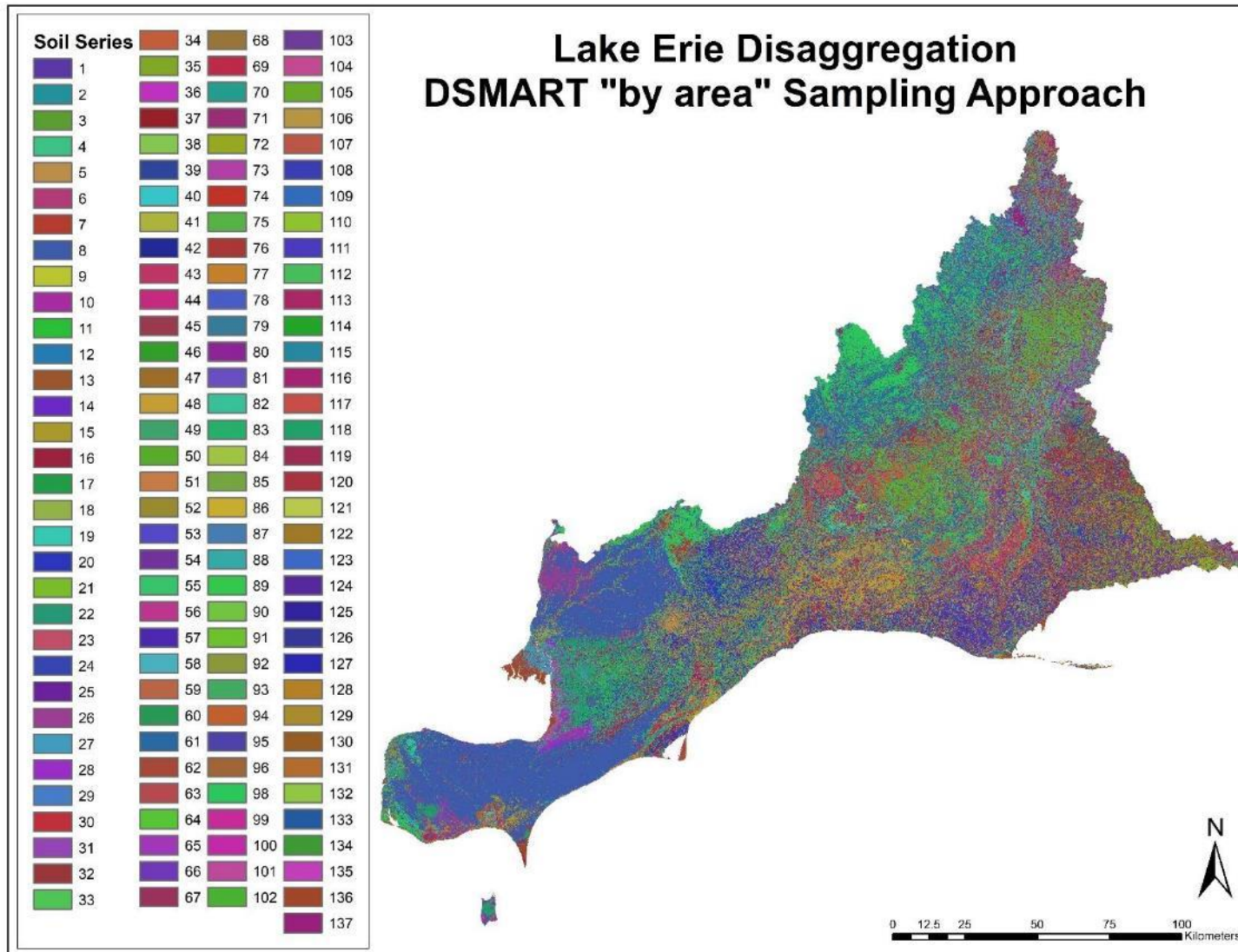


Figure 49 : Disaggregated Soil Series Map of Lake Erie Basin at 250 m resolution ("by area" Sampling Approach in DSMART) *

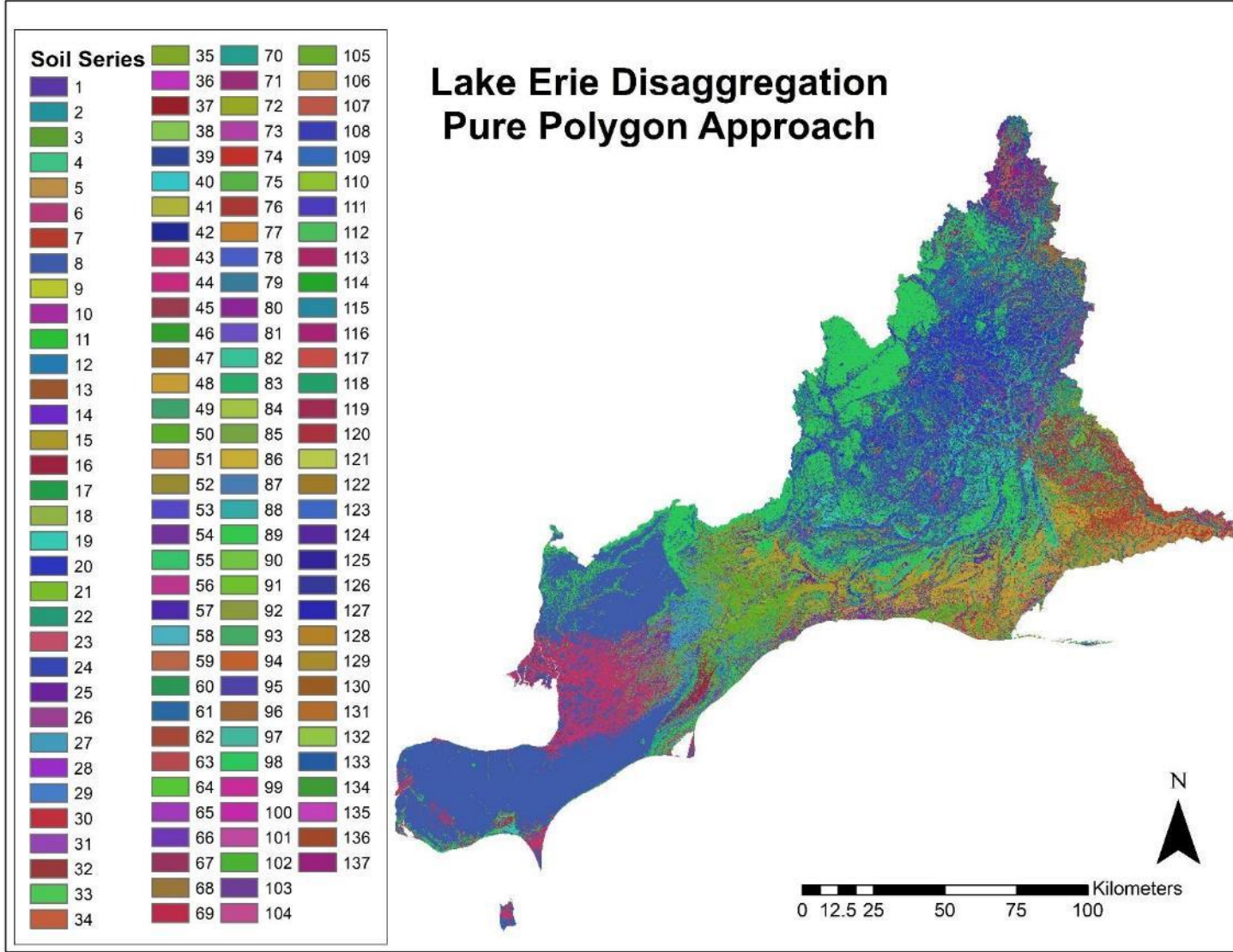


Figure 50 : Disaggregated Soil Series Map of Lake Erie Basin at 250 m resolution (Pure Polygon Approach) *

***Legends**

1	Alberton	35	Ekfrid	69	Kintyre	103	Shedden
2	Ancaster	36	Elmira	70	Kelvin	104	Springwater
3	Alluvium	37	Eastport	71	Lonsdale	105	Smithville
4	Ayr	38	Floradale	72	Lincoln	106	Silver Hill
5	Brady	39	Fanshawe	73	Lily	107	St Jacobs
6	Blackwell	40	Freeport	74	Lambton	108	Southwold
7	Brantford	41	Flamborough	75	London	109	St Williams
8	Brookston	42	Frome	76	Lowbanks	110	Seneca
9	Ballymote	43	Fox	77	Lisbon	111	Springvale
10	Bennington	44	Farmington	78	Listowel	112	Scotland
11	Binbrook	45	Gilford	79	Macton	113	Stayner
12	Brooke	46	Grimsby	80	Middlemarch	114	Styx
13	Boomer	47	Granby	81	Melbourne	115	Teeswater
14	Bookton	48	Gobles	82	Mitchell's Bay	116	Thorndale
15	Berrien	49	Grand	83	Muirkirk	117	Tioga
16	Brant	50	Guelph	84	Mannheim	118	Toledo
17	Bryanston	51	Harrisburg	85	Maplewood	119	Tuscola
18	Brisbane	52	Highgate	86	Martin	120	Tavistock
19	Burford	53	Heidelberg	87	Muriel	121	Vittoria
20	Beverly	54	Haldimand	88	Maryhill	122	Vineland
21	Caledon	55	Harkaway	89	Normandale	123	Vanessa
22	Churchville	56	Hillsburgh	90	Niagara	124	Walsingham
23	Crombie	57	Hampden	91	Nissouri	125	Wattford
24	Camilla	58	Harriston	92	Oakland	126	Wellesley
25	Conestogo	59	Harrow	93	Ontario	127	Wilsonville
26	Caistor	60	Hespeler	94	Oakview	128	Wiarion
27	Colwood	61	Huron	95	Plainfield	129	Welland
28	Clyde	62	Hawkesville	96	Parkhill	130	Wilmot
29	Dorking	63	Haysville	97	Preston	131	Woolwich
30	Donald	64	Honeywood	98	Perth	132	Waterin
31	Dumfries	65	Jeddo	99	Rivard	133	Walsher
32	Dundonald	66	Killean	100	Strathburn	134	Waterloo
33	Donnybrook	67	Kirkland	101	St Clair	135	Wauseon
34	Embro	68	Kossuth	102	St Clements	136	Marsh
						137	Organic

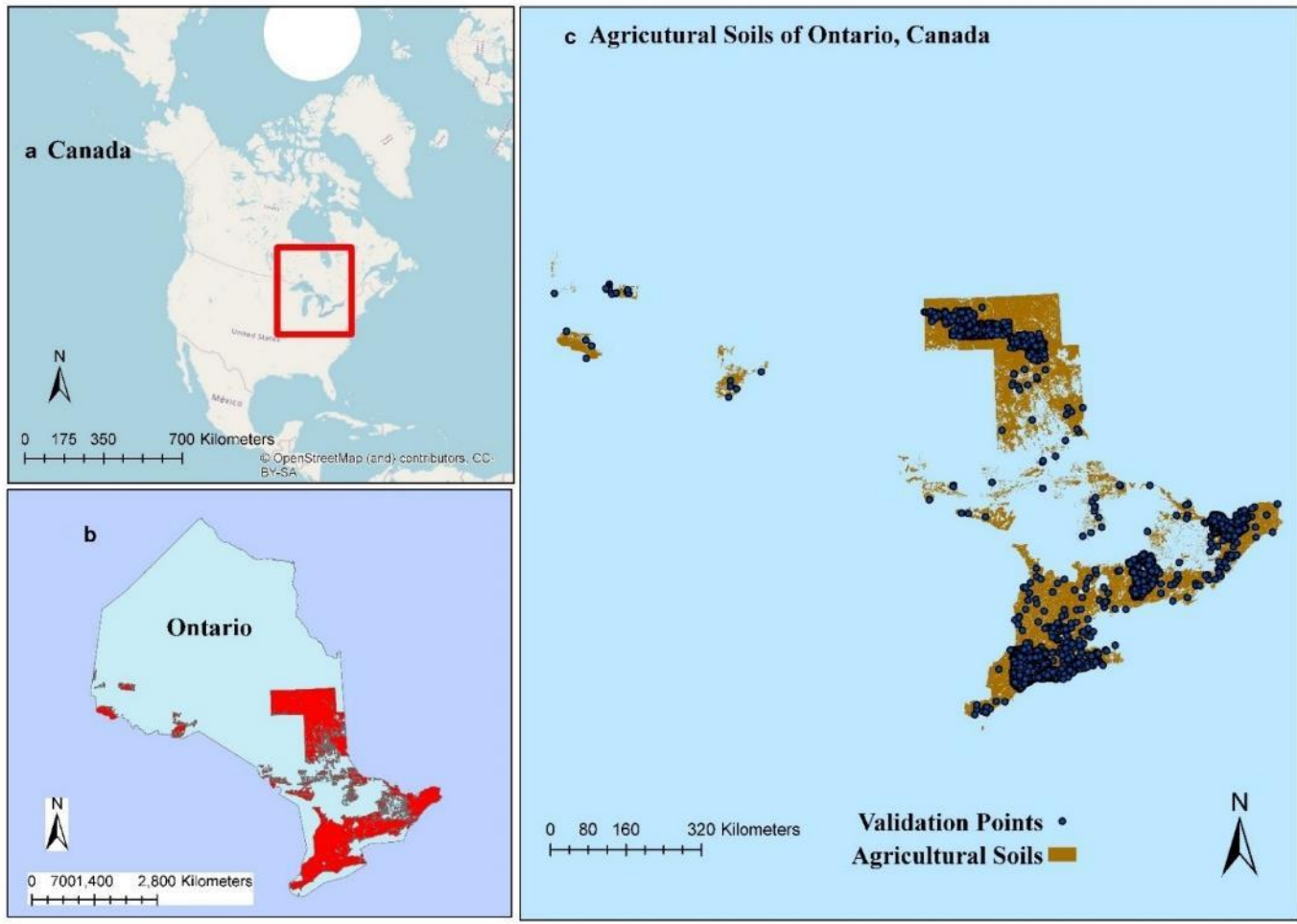


Figure 51 : The study area located in North America (a) and the soils of Ontario represented by the red boundary area in Ontario (b). Validation points extracted from the National Pedon Database are represented by blue dots in the agricultural soil map of Ontario (c).

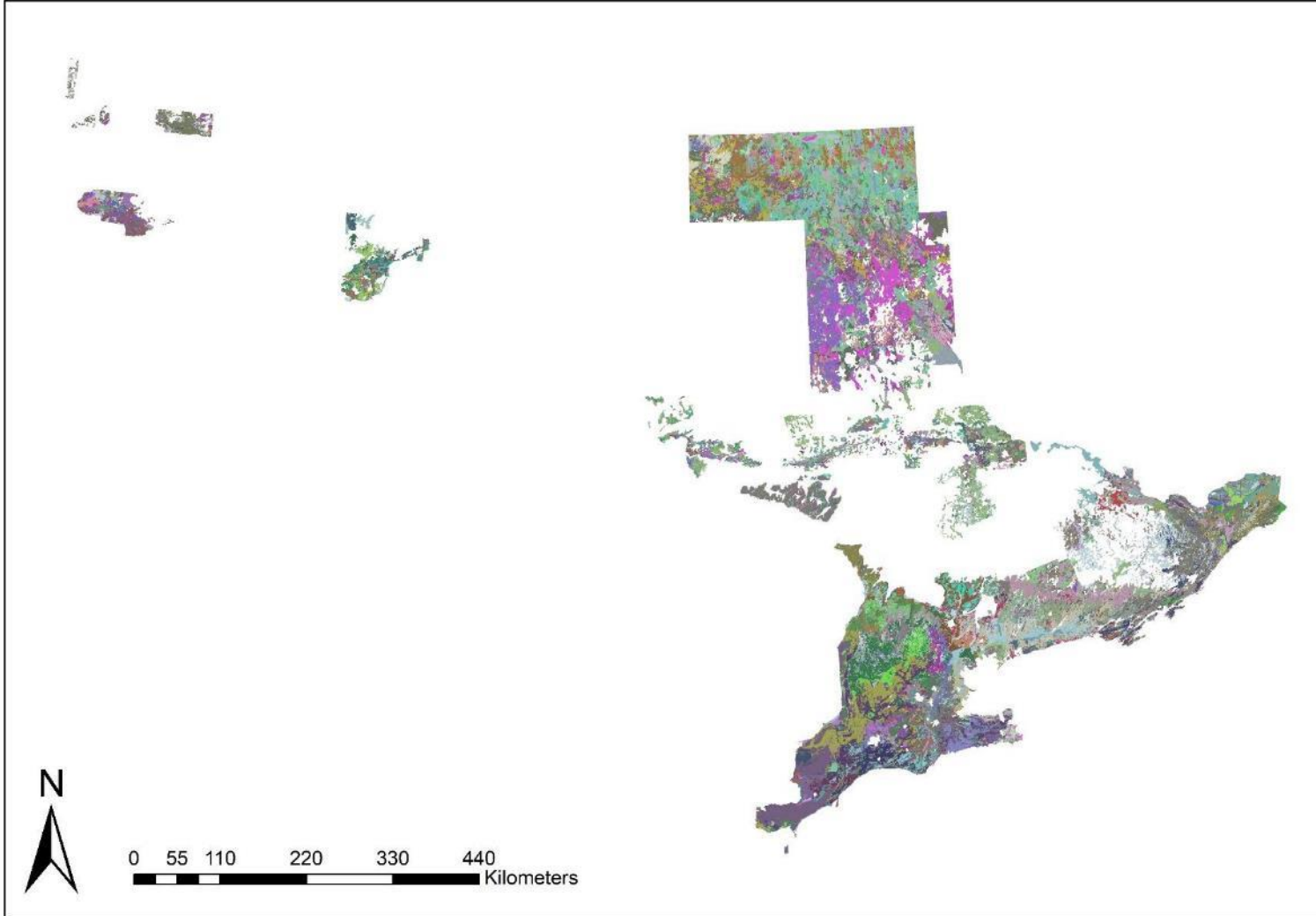


Figure 52 : Disaggregated Soil Series Map of Agricultural Soils of Ontario at 250 m resolution (Pure Polygon Approach) **

****Legends**

Ontario Soil Series	BID	BIV	CSM	DWS	FNT	HDD	JVS	LTM	MRY	OKD	RFW	SLI	TNY	WEY
ABM	BIN	BWI	CST	DWT	FOD	HDI	KBY	LTT	MSG	OKL	RHS	SLL	TPK	WFD
ABO	BIV	BWK	CSY	DYK	FOM	HEY	KEK	LTW	MSL	OPY	RKF	SMC	TRT	WFI
ABT	BIY	BYE	CTG	DYM	FOT	HGE	KEO	LUC	MSO	OTI	RKU	SMF	TSW	WFP
ABV	BIZ	BYK	CTR	DYN	FOX	HGR	KET	LVR	MSS	OTK	RMD	SMG	TTU	WGO
ACE	BKB	BZW	CTS	DYO	FRD	HGT	KGJ	LWH	MTD	OUL	RMI	SMK	TTY	WHG
ACK	BKC	CAC	CUA	EAK	FRM	HHG	KGW	LXX	MTI	OUM	ROS	SMV	TUC	WHK
AHG	BKG	CAD	CUL	EBR	FUQ	HHO	KHO	LYS	MTS	OVW	ROW	SNA	TUH	WHP
AHU	BKN	CAU	CUT	ECC	FWF	HIG	KIG	MAD	MTZ	PAI	RRD	SND	TUK	WID
ALL	BKS	CBE	CWK	EDO	FWL	HII	KIL	MAI	MUA	PAS	RSL	SOA	TUS	WIL
ALT	BKY	CBL	CWO	EDS	FXB	HIL	KIM	MAN	MUI	PAX	RTH	SOG	TVK	WIO
ALU	BLH	CBN	CKQ	EDV	FYX	HIM	KIV	MAT	MUR	PAY	RTL	SOY	TVL	WIT
AMO	BLK	CBT	CXT	EGH	GBO	HKY	KKD	MBG	MUY	PCH	RUB	SPD	TWE	WLF
ANG	BLL	CCY	CYD	EGL	GDN	HLG	KLL	MBT	MXB	PCY	RVC	SRD	TWI	WLK
ANW	BLU	CDR	DCK	EGV	GDT	HMH	KMB	MCC	MXG	PDW	RVE	SRI	TWN	WLL
ANY	BM B	CEG	DCP	EKF	GFD	HMP	KMW	MCT	MXJ	PEL	RYL	SRK	TYG	WLY
API	BM D	CER	DEK	EKP	GFF	HMW	KNG	MCW	MXP	PFD	RYV	SRN	UHF	WMH
APP	BM F	CEW	DGT	EKT	GFK	HNK	KPU	MDC	MYB	PFW	SAM	SRV	UPD	WMT
ASH	BM R	CEY	DHM	ELC	GIT	HNN	KRS	MDL	MYL	PHN	SAY	SSD	UPK	WMY
ASO	BM T	CFL	DHU	ELS	GIY	HOY	KSU	MDM	MYR	PIK	SBN	SSG	URT	WOD
ATH	BNB	COG	DHY	EMI	GLN	HRF	KTY	MDT	NDE	PIN	SCL	SSK	VAR	WOL
ATY	ENG	COH	DIK	EMO	GMD	HRH	KVN	MEL	NDM	PIO	SCM	SSM	VCO	WOO
AUG	BNH	CGT	DKG	EMR	GMY	HRR	KYH	MFD	NEG	PKL	SCT	SSW	VCT	WOW
AUH	BNK	CGU	DKI	EMY	GNY	HRW	LAV	MGL	NEP	PKT	SCX	STA	VIT	WOY
AUY	BNQ	CHD	DLM	ENG	GQB	HSP	LDE	MGW	NGG	PLL	SDD	STC	VLD	WPO
AVL	BNR	CHN	DLO	EOK	GOU	HST	LDW	MHS	NGR	PNK	SDK	STD	VLO	WRG
AVY	BOK	CHV	DMM	ESP	GOW	HTM	LEE	MHU	NGW	PNS	SEM	STN	VNY	WRN
AYR	BOL	CY	DNB	EST	GOU	HUU	LEG	MIB	NHB	PPP	SEN	STR	VRN	WRR
AZD	BOM	CLA	DNC	ESY	GRC	HWD	LEH	MIL	NHM	PPV	SES	STS	VRS	WSH
BAK	BON	CLB	DNM	ETO	GRD	HWO	LEI	MKB	NID	PQS	SET	STV	VSP	WSY
BAL	BOQ	CLS	DNR	ETP	GSE	HWV	LIC	MKG	NIS	PRL	SFD	SUS	VSS	WTC
BAY	BPE	CLV	DNT	EVU	GSS	HYU	LIL	MKK	NMG	PSN	SFL	SUX	VSY	WTF
BBH	EPN	CLY	DOD	FAD	GUP	HYW	LIY	MLA	NOU	PSS	SGE	SVN	VUD	WTK
BBK	BR A	CM B	DOE	FAL	GUR	IBT	LKO	MLE	NPE	PTH	SGH	SVR	VVI	WTO
BBO	BR G	CM H	DOI	FAN	GVI	IFD	LKP	MLP	NSG	PUM	SOR	SYB	WAF	WUS
BCF	BR R	CM L	DOO	FCK	GVR	IID	LMB	MLR	NVL	PWK	SGT	SYC	WAM	WVL
BCO	BRT	CNB	DPU	FCT	GWK	INK	LMR	MMN	NVN	PWO	SHN	SYR	WAS	WWN
BCW	BR Y	CNE	DPY	FEK	GWS	INV	LND	MNA	NWC	PWS	SHO	SYX	WAT	WYS
BDH	BSB	COB	DRR	FEL	GWY	ISH	LOD	MNE	NWF	PYF	SHV	SYZ	TAT	WAW
BDO	BSS	COR	DRY	FEP	GYK	JCY	LOW	MNM	NWG	PYN	SHY	TEW	WAY	ZMK
BDS	BTI	CPN	DTO	FFC	HAD	JDD	LPE	MNN	NWR	PYO	SIB	TFG	WBI	ZOR
BDT	ETR	GRE	DTY	FHH	HAE	JET	LPG	MNY	OAK	GRY	SH	THE	WBU	ZPT
BDY	BTW	CRI	DUD	FKW	HAR	JFY	LPP	MOB	OBE	QUI	SIR	THN	WBY	
BFO	BUB	CRO	DUF	FLK	HAS	JKV	LRK	MOG	OBK	QWY	SV	THT	WCK	
BGD	BUF	CRP	DUL	FLO	HAU	JMT	LSR	MOK	OBY	RBU	SIY	THW	WDG	
BGH	BUK	CRV	DUN	FMB	HBC	JND	LSB	MOS	OGO	RCL	SJB	TIG	WDK	
BGM	BUJ	CRY	DVN	FME	HBG	JNI	LSY	MOY	OID	RDU	SJO	TLD	WDV	
BHT	BVA	CRZ	DVT	FMI	HBS	JOR	LTC	MPL	OIK	RDV	SKD	TLO	WDY	
BHU	BVE	CSH	DWH	FNG	HBV	JUB	LTH	MPW	OKA	RFK	SLD	TMO	WEK	

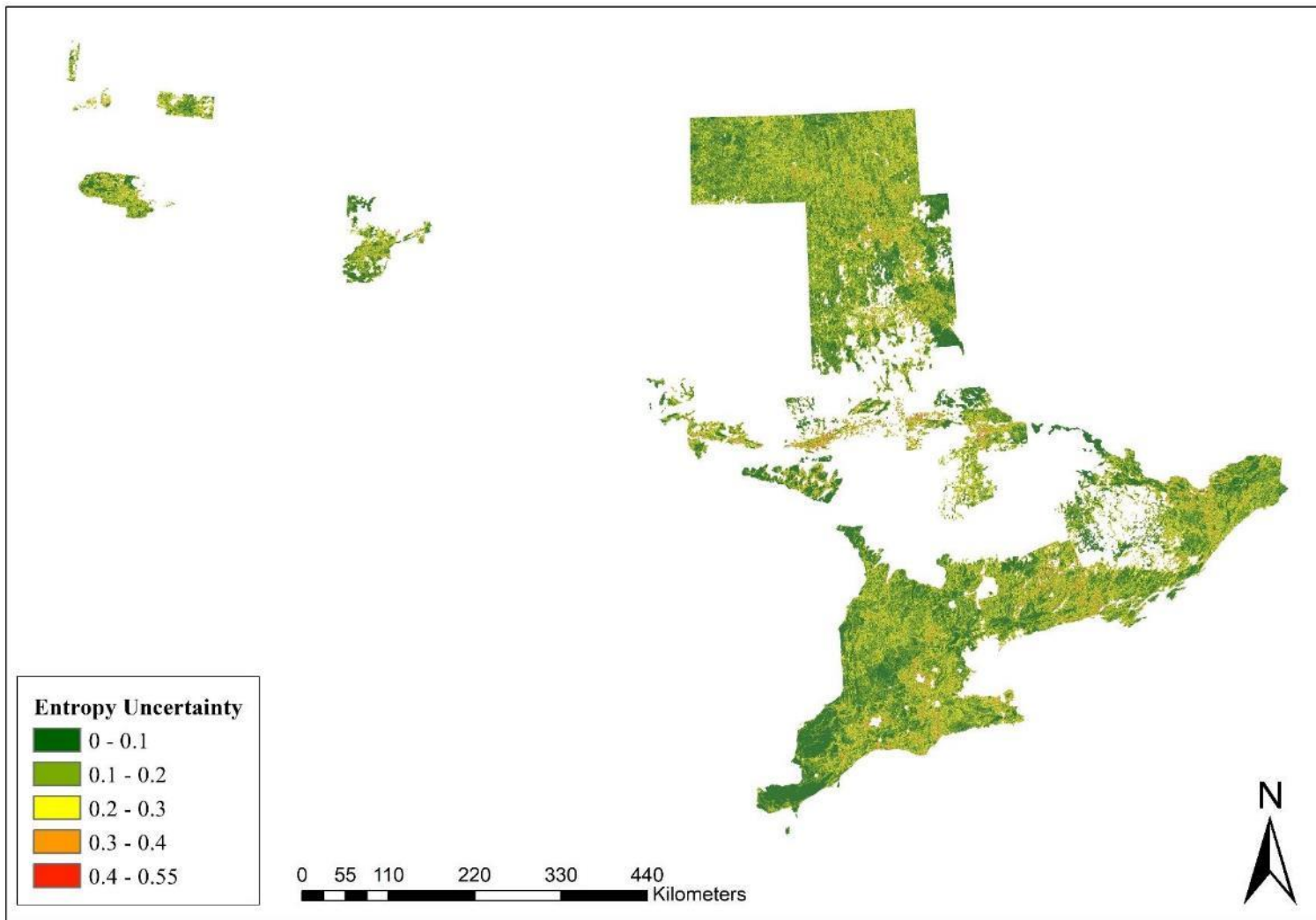


Figure 53 : Entropy uncertainty of C5.0 in Pure Polygon Disaggregation using training data produced from area-weighted random approach at a 250m spatial resolution, Ontario.

Table 4 : Model evaluation of SOC% prediction of Middlesex County using reassigned and weighted-probable models.

<i>Approach</i>	<i>Model</i>	<i>MLA</i>	<i>R²</i>	<i>CCC</i>	<i>RMSE</i>
Reassigned	DSMART	C5.0	0.15	0.29	0.98
		RF	0.16	0.30	0.96
	PPD	C5.0	0.16	0.30	0.91
		RF	0.17	0.32	0.88
Weighted-Probable	DSMART	C5.0	0.21	0.38	0.71
		RF	0.19	0.34	0.85
	PPD	C5.0	0.21	0.38	0.77
		RF	0.20	0.37	0.79

2 Soil Organic Carbon Prediction of Ontario: A Probable Classifier Approach

2.1 Background

DSM offers significant potential for predicting soil properties and generating high-resolution soil property maps. Through the integration of various data sources and the application of statistical techniques or machine learning algorithms, DSM enables accurate predictions of soil properties. The validation of DSM models using field observations or laboratory data ensures their reliability. The resulting high-resolution maps of soil properties have diverse applications in land use planning, soil conservation, and agricultural practices, providing valuable information for decision-making processes in these domains.

DSM is capable of predicting a range of soil properties, including but not limited to soil organic carbon (SOC) content, soil texture, pH, drainage, and nutrient availability. Geographically Weighted Regression (GWR) is a commonly used technique for mapping SOC content and storage, as well as other soil properties. The Classification and Regression Tree (CART) model, such as Random Forest (RF), has also shown success in mapping regional soil types, soil drainage, SOC stocks, and soil texture. These models provide improved error measurements and are less prone to overfitting. Additionally, other approaches like equal-area spline functions and regression kriging have been utilized to predict lateral and three-dimensional variations in SOC content.

2.2 Scope and Objective

The scope of this part of study was to explore a technique for predicting soil organic carbon (SOC) in regions with sparse and scarce data. The objective was to derive and spatialize soil property information, specifically SOC, by utilizing disaggregated soil series maps. By leveraging available soil profile information obtained from legacy soil maps (LSM) or expert knowledge, it becomes possible to predict soil properties using soil class probability rasters generated from disaggregation applications.

In this study, we focused on assessing the viability of utilizing soil series information obtained from a comprehensive soil map (CSM) and soil series maps that have been disaggregated through digital soil

mapping (DSM) techniques to predict soil property maps, with a specific emphasis on SOC%. The primary goal was to develop a technique that can accurately predict SOC% in regions where data is either absent or limited.

2.3 The specific objectives of the study were as follows:

- To employ a reassigned approach to assign SOC% values obtained from the conventional soil survey report to the disaggregated soil series values, thereby generating SOC raster maps.
- To utilize a weighted-probable approach to predict SOC% maps by assigning SOC% values to the soil series probability maps generated from disaggregation techniques.

By achieving these objectives, we aimed to demonstrate the potential of utilizing disaggregated soil series maps and existing soil profile information to predict SOC% and produce high-resolution maps of this critical soil property. This method offers a promising alternative for generating reliable soil property information in specific areas of interest until new and updated soil data become available through standard sampling and examination approaches.

2.4 Methodology

To evaluate the effectiveness of the proposed methods in generating soil organic carbon (SOC) maps, a study area was selected in Middlesex County. The study area consisted of a disaggregated soil series map created using the DSMART and Pure Polygon Disaggregation (PPD) approaches. Two distinct approaches, namely the reassigned approach and the weighted-probable approach, were implemented to generate SOC maps for the study area.

In the reassigned approach, mean SOC% values obtained from soil survey reports were utilized. Specifically, the mean SOC% values at the 0-30 cm depth were reassigned to the disaggregated soil series map. This involved replacing the existing values in the disaggregated map with their corresponding mean SOC% values derived from the soil survey reports. By applying this approach, a SOC% map was created, providing insights into the spatial distribution of SOC in Middlesex County.

Furthermore, the weighted-probable approach was employed to enhance the accuracy and detail of the SOC maps. Probability maps were generated for each soil series using both DSMART and PPD techniques. These probability maps were then stacked together. Next, the stacked probability map was multiplied by the mean SOC% values at the 0-30 cm depth, which were obtained from the soil survey reports and associated with each soil series. The resulting raster files were subsequently summed to generate a weighted mean SOC% map for the study area. The weighted-probable approach leveraged the probability maps to capture the nuanced distribution of SOC across Middlesex County, providing a more comprehensive representation of SOC variability.

By employing these two approaches, SOC maps were generated for the study area in Middlesex County. These maps provided valuable insights into the spatial distribution of soil organic carbon (SOC) and presented alternative approaches for predicting SOC% at a higher resolution. These maps served as an effective tool for assessing SOC content and exploring potential strategies for improving SOC prediction in regions with limited data availability. Through the integration of the reassigned and weighted-probable

approaches, this methodology facilitated the production of accurate SOC maps for the selected study area, aiding in the assessment of SOC content and variability.

In the next phase of this component of the study, the weighted-probable approach was chosen based on the validation results to generate the soil organic carbon percentage (SOC%) map for agricultural soils in Ontario. To obtain the final SOC map at a resolution of 250 m, the probability maps derived from the Pure Polygon Disaggregation (PPD) approach were stacked. These stacked probability maps were then multiplied by the mean SOC% values obtained from the National Soil Database (NSDB) for the corresponding soil series. Subsequently, all the raster files were aggregated to produce the mean SOC% map at a depth of 0-30 cm for the study area. The resulting map, depicting the spatial distribution of SOC in the agricultural soils of Ontario. This methodology effectively combined the reassignment of mean SOC% values from soil survey reports with the utilization of probability maps and weighted calculations, resulting in accurate and high-resolution SOC maps. A validation dataset comprising 500 points was extracted from the Middlesex County Soil Survey report of 1992. This dataset was utilized to evaluate the accuracy of the predicted soil organic carbon percentage (SOC%) maps generated from the reassigned and weighted-probable approaches, which were based on the disaggregated soil series maps. Furthermore, for the assessment of the disaggregated soil series map accuracy for Ontario, a separate validation dataset consisting of 1300 soil series observations was obtained from the National Pedon Database. To measure the goodness of fit, statistical analysis was performed using the R software, calculating R-squared (R^2), Concordance Correlation Coefficient (CCC), and Root Mean Square Error (RMSE).

2.5 Results

Figure 54 showcases the predicted soil organic carbon percentage (SOC%) maps generated using the reassigned and weighted-probable approaches with disaggregated soil series maps of Middlesex County. The evaluation of the model performance was conducted using Goodness of Fit statistics, and the results are presented in Table 4.

The accuracy assessment revealed that the weighted-probable approach outperformed the reassigned approach. The combination of machine learning algorithms (MLAs) with the weighted-probable approach resulted in more accurate SOC% maps compared to the reassigned approach. Specifically, the PPD-RF combination within the reassigned approach exhibited the highest performance among the reclassify approaches. However, in terms of accuracy results, the DSMART-C5.0 combination within the weighted-probable approach achieved the best outcomes, with a R^2 value of 0.21, a CCC of 0.38, and a RMSE of 0.71.

Figure 55 illustrates the average soil organic carbon percentage (SOC%) map for agricultural soils in Ontario, specifically at a depth of 0-30 cm. This map was generated using the weighted-probable approach, which relied on the disaggregated soil series map produced from the PPD approach.

To assess the model's performance, goodness of fit statistics were employed using an external dataset of 1300 validation points obtained from the National Pedon Dataset. The application of the weighted-probable approach on the disaggregated soil series maps within the PPD method yielded a R^2 value of 0.15, a RMSE of 9.4, and a CCC of 0.32.

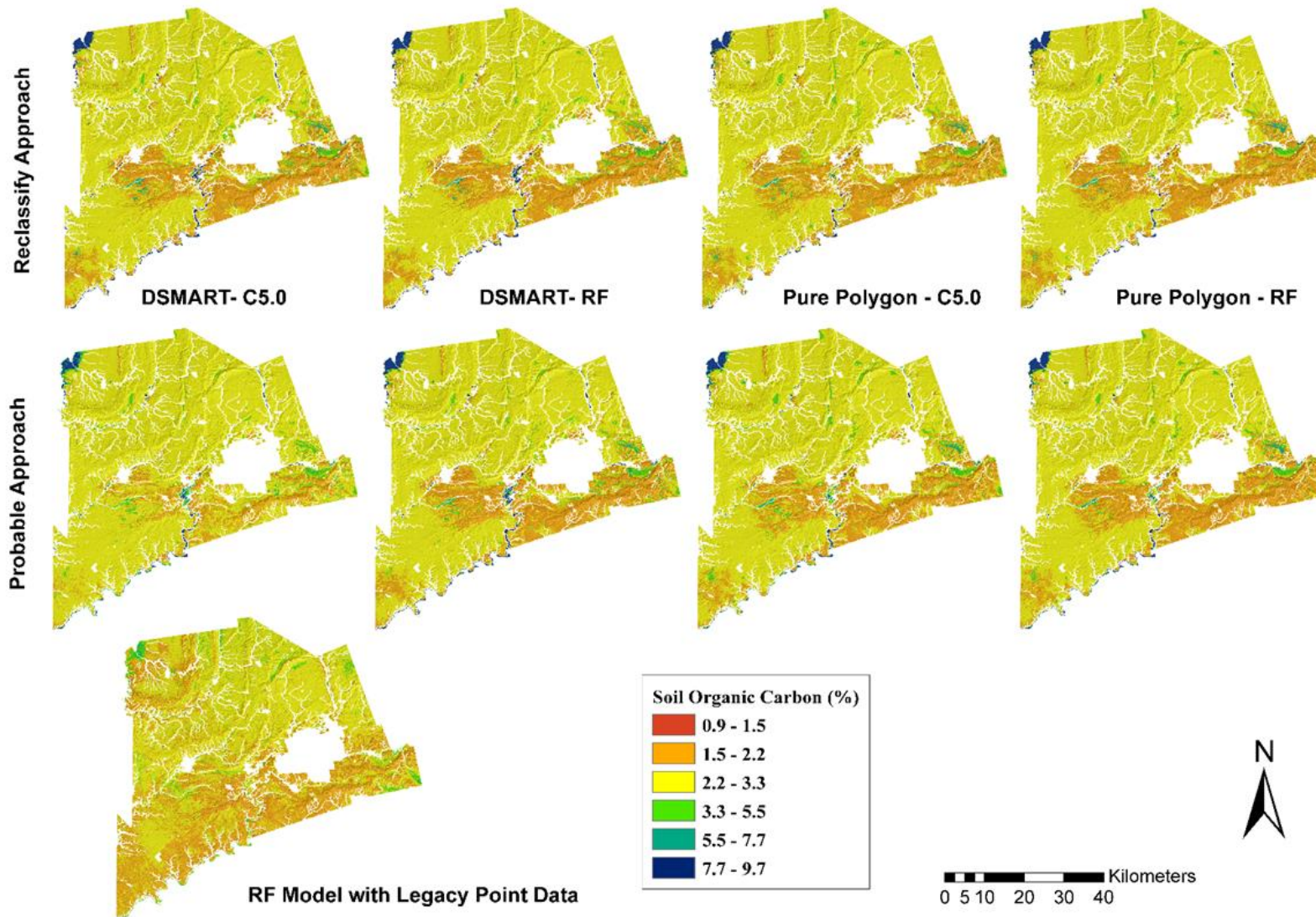


Figure 54 : Mean SOC% prediction map at 0-30 cm depth of Middlesex County using reassigned and weighted-probable models.

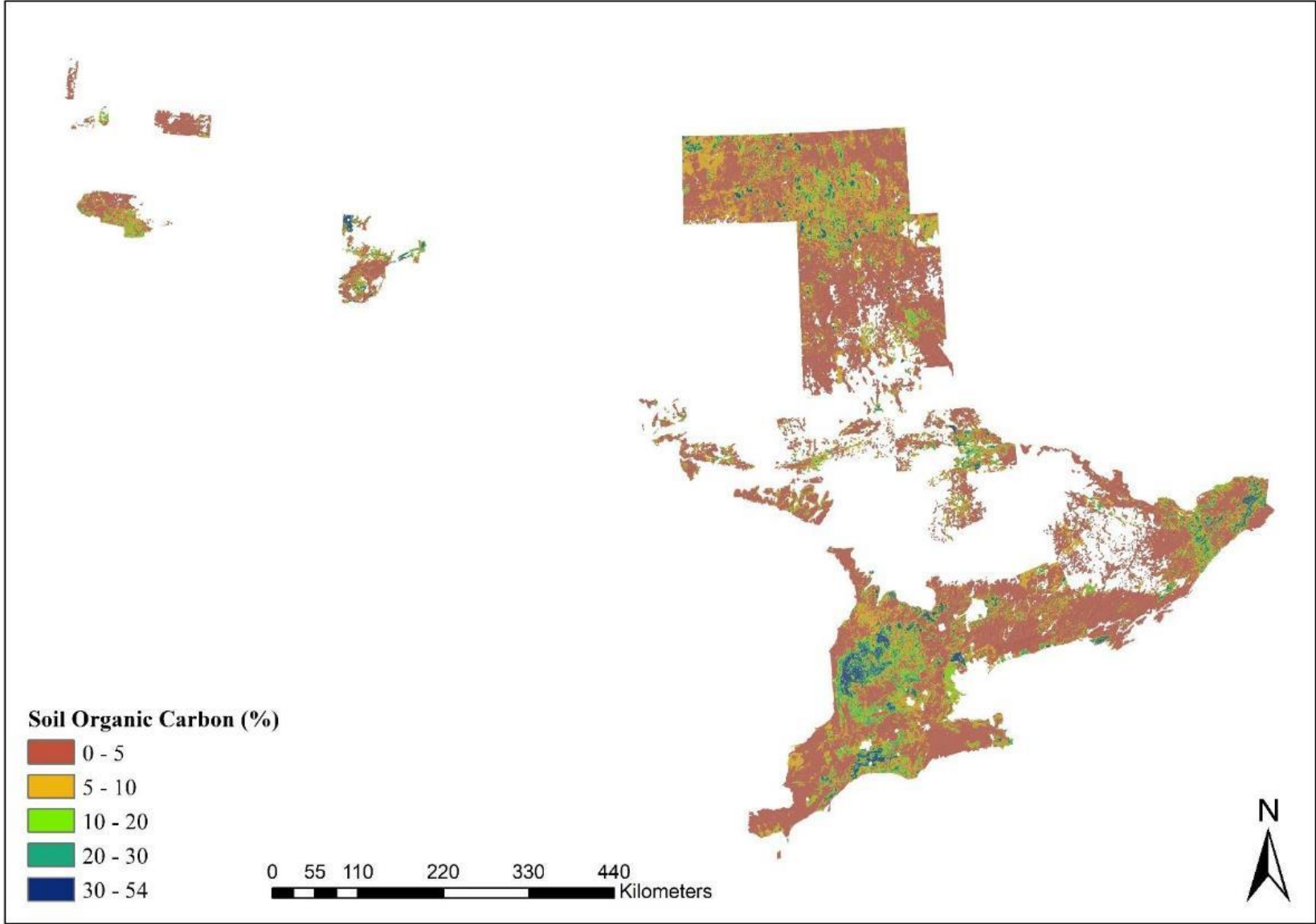


Figure 55 : Soil Organic Prediction Map of Agricultural Soils of Ontario using Probable-Classifer Approach.

3 Soil Organic Matter Disaggregation of Quebec Agricultural Soils

3.1 Background

DSM is a valuable approach for estimating soil properties, particularly in situations where precise georeferencing is lacking or conventional soil maps are unavailable. By establishing statistical relationships between limited soil data and environmental variables, DSM can predict soil properties in unsampled areas. The GlobalSoilMap project, initiated in 2010, defined specifications for soil properties, including soil carbon, to facilitate global-scale DSM at a spatial resolution of 100 meters. A more advanced version of SoilGrids was developed, incorporating machine learning techniques, higher resolution covariate layers, and additional soil profiles to estimate soil organic carbon (SOC) and other properties across standard soil depths ranging from 0 to 200 cm.

Various methods have been employed to predict SOC, including linear statistical models, geostatistical and hybrid approaches, and machine learning algorithms (MLAs) such as Random Forest, Boosted Regression Tree, Cubist, Geographically Weighted Regression, Regression Kriging, Support Vector Machine, and Artificial Neural Network. MLAs have demonstrated promising results in SOC prediction. Recent DSM studies have utilized these models to map changes in SOC% levels under different land use/land cover and climate change scenarios. These maps provide visual representations of potential scenarios and aid in proactive planning for climate change mitigation and carbon sequestration initiatives. However, it is essential to calibrate the models using appropriate datasets and algorithms and validate them with external datasets to ensure reliable performance.

3.2 Scope and Objectives

The DSM models mentioned in previous studies have relied on point databases that provide geospatially referenced SOC/SOM information and associated environmental covariates. However, in practical scenarios, privacy concerns often lead to the aggregation of data at the postal code or polygon level, without disclosing precise geospatial locations. For instance, the Institut de recherche et de développement en agroenvironnement (IRDA) in the Province of Québec compiled a dataset of 315,000 SOM observations from agricultural soils, which were analyzed by private labs between 2007 and 2018. In 2019, this dataset was aggregated at the postal code level to safeguard landowner privacy. However, the challenge arises when attempting to apply DSM techniques using this aggregated dataset, as multiple postal code units may contain multiple sampling points without precise georeferenced information due to potential disparities between producers' administrative addresses and the sampled fields. Therefore, the objective of this study is to disaggregate the aggregated SOM observations of agricultural soils in Quebec, which were compiled at the postal code level, using DSMART techniques.

3.3 Methodology

In 2019, the Institut de recherche et de développement en agroenvironnement (IRDA) assembled a dataset of 315,000 soil organic matter (SOM) observations obtained from private labs in Québec. These observations were collected at the postal code level between 2007 and 2018. The Québec postal code map was obtained from *Adresses Québec* (Figure 56).

The SOM values of each postal code were categorized into ten groups based on an equivalent number of samples found in each class at the provincial scale. The classification involved creating classes with approximately 1.5% intervals of SOM, determined by a fixed increment of SOM observations.

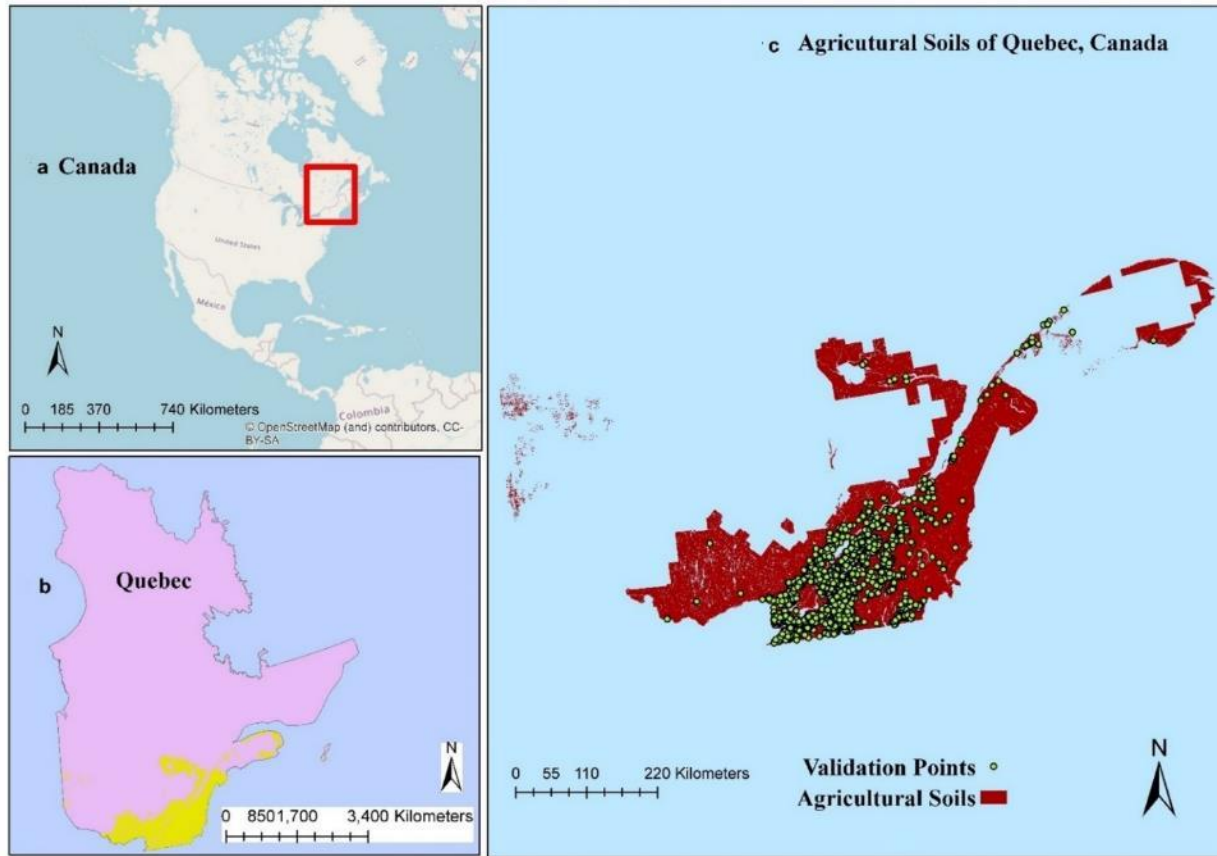


Figure 56 : The study area located in North America (a) and the agricultural soils of Quebec represented by a yellow boundary area in Quebec (b). Validation points extracted from a georeferenced private laboratory dataset are represented by green dots in the agricultural soil map of Quebec (c).

A shapefile was developed, representing the proportion of each SOM group, resulting in 663 polygons that depicted the sample proportions of each SOM class. Table 5 provides the classification details of the SOM values into these ten groups.

The DSMART algorithm, implemented using the `rdsmart` package in R (Odgers et al., 2021), was employed for disaggregating the SOM classes. The "by area" sampling procedure of DSMART was utilized to generate the training dataset, selecting 10 synthetic samples per square kilometer of the study area. A total of 25 unique synthetic datasets were created using this sample design, resulting in 1,451,110 synthetic sampling points. The environmental covariate values from the 21 retained predictor layers were extracted from these locations to build the model.

The Random Forest (RF) machine learning algorithm was applied within the DSMART framework using the 25 unique sampling plans to disaggregate the SOM classes. DSMART produced 25 realizations of SOM% prediction rasters, which were stacked and summarized using the `summarize` function in DSMART. This process computed the probabilities of occurrence for the SOM classes. The predicted SOM% classes were ranked based on occurrence, resulting in three "most probable" maps. The first most probable map was selected to showcase the results in this study.

Table 5 : Classification of 315,000 SOM% observations in agricultural soils of Quebec between 2009 and 2017, and 24 000 validation points collected from a private laboratory (LOGIAG) into categorical classes using uniform splitting interval.

<i>SOM (%)</i>		
Classes	Range (%)	Validation Points
Class 1	0 - 1.66	40
Class 2	1.66 - 3.14	6500
Class 3	3.14 - 4.62	8500
Class 4	4.62 - 6.1	5500
Class 5	6.1 - 7.58	2100
Class 6	7.58 - 9.07	1000
Class 7	9.07 - 10.5	300
Class 8	10.5 - 12	30
Class 9	12 - 13.5	20
Class 10	13.5 - 15	10

A total of 24 000 observations of soil organic matter (SOM) percentage were collected from Logiag, a private laboratory in Quebec. This dataset was utilized as a validation dataset to evaluate the performance of the SOM prediction model. The validation dataset encompassed all SOM class intervals as presented in Table 5. To assess the accuracy of SOM prediction using the DSMART algorithm, overall accuracy and Kappa index were calculated. These metrics provide measures of agreement between the predicted SOM classes and the observed values in the validation dataset. The following equations were employed to calculate the overall accuracy and Kappa index:

- Overall Accuracy = (Number of correctly classified observations) / (Total number of observations)
- Kappa Index = (Overall Accuracy- Expected Accuracy) / (1- Expected Accuracy)

Entropy uncertainty (H) was computed based on the probability maps generated from the 25 model realizations per sampling plan. The entropy uncertainty provides insights into the variability and uncertainty associated with the SOM predictions.

3.4 Results

Figure 57 depict the disaggregated map of SOM% for agricultural soils in the province of Quebec at a spatial resolution of 250 m. These maps provide valuable insights into the spatial distribution of SOM% and the level of uncertainty associated with the predictions.

The accuracy of the DSMART model in predicting SOM% disaggregation was assessed, resulting in an overall accuracy of 48% and a kappa index statistic of 0.23. The model demonstrated satisfactory performance in predicting SOM% values within classes 2 to 6, ranging from 1.67% to 9.07%, as confirmed by the confusion matrix comparing the predicted SOM% to the validation dataset, illustrated in Table 6. However, due to the low representation of classes 1, 8, 9, and 10 in the validation dataset, the model could not validate any points from

these classes. The close proximity of SOM class values and the minimal differences between each class may have contributed to challenges in accuracy assessment and validation of the prediction map.

Table 6 : Confusion matrix of predicted soil organic matter classes using random forest in DSMART, compared to 24 000 validation points.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Accuracy
Class 1	0	0	0	0	0	0	0	0	0	0	-
Class 2	13	150	55	15	1	0	0	0	1	0	64%
Class 3	15	275	623	267	83	29	16	3	6	3	48%
Class 4	0	24	169	230	111	28	14	9	3	1	40%
Class 5	0	5	30	40	65	37	12	3	3	1	34%
Class 6	0	0	1	4	12	15	4	1	0	1	40%
Class 7	0	0	0	0	0	2	1	0	0	0	34%
Class 8	0	0	0	0	0	0	0	0	0	0	-
Class 9	0	0	0	0	0	0	0	0	0	0	-
Class 10	0	0	0	0	0	0	0	0	0	0	-

Legend: Class1 = 0-1.66%, Class2 = 1.66-3.14%, Class3 = 3.14-4.62%, Class4 = 4.62-6.1%, Class5 = 6.1-7.58%, Class6 = 7.58-9.07%, Class7 = 9.07-10.5%, Class8 = 10.5-12%, Class9 = 12-13.5%, Class10 = 13.5-15%.

An analysis of the total cumulative area for each SOM% class was conducted for both the observed and predicted datasets. The results revealed that DSMART performed poorly in predicting lower SOM% values falling within class 1 (0-1.66% SOM) and higher values in classes 8 (10.35-12%), 9 (12-13.5%), and 10 (13.5-15%). The model tended to overpredict SOM% values for classes 4 and 5 while significantly underpredicting values for class 2. Although there is a consistent trend between both datasets, they exhibit a left-skewed distribution (Figure 58).

The distribution of entropy uncertainty values (H) for SOM% prediction is represented in Figure 59. The H values range from 0 to 0.73, with a mean of 0.26 and a standard deviation of 0.15. This indicates the variability and uncertainty present in the SOM% predictions. Figure 60 illustrates the overlay of the entropy map and elevation map in the study area. The visualization indicates a clear association between higher entropy values (indicating higher uncertainty) in SOM% prediction and areas of elevated terrain.

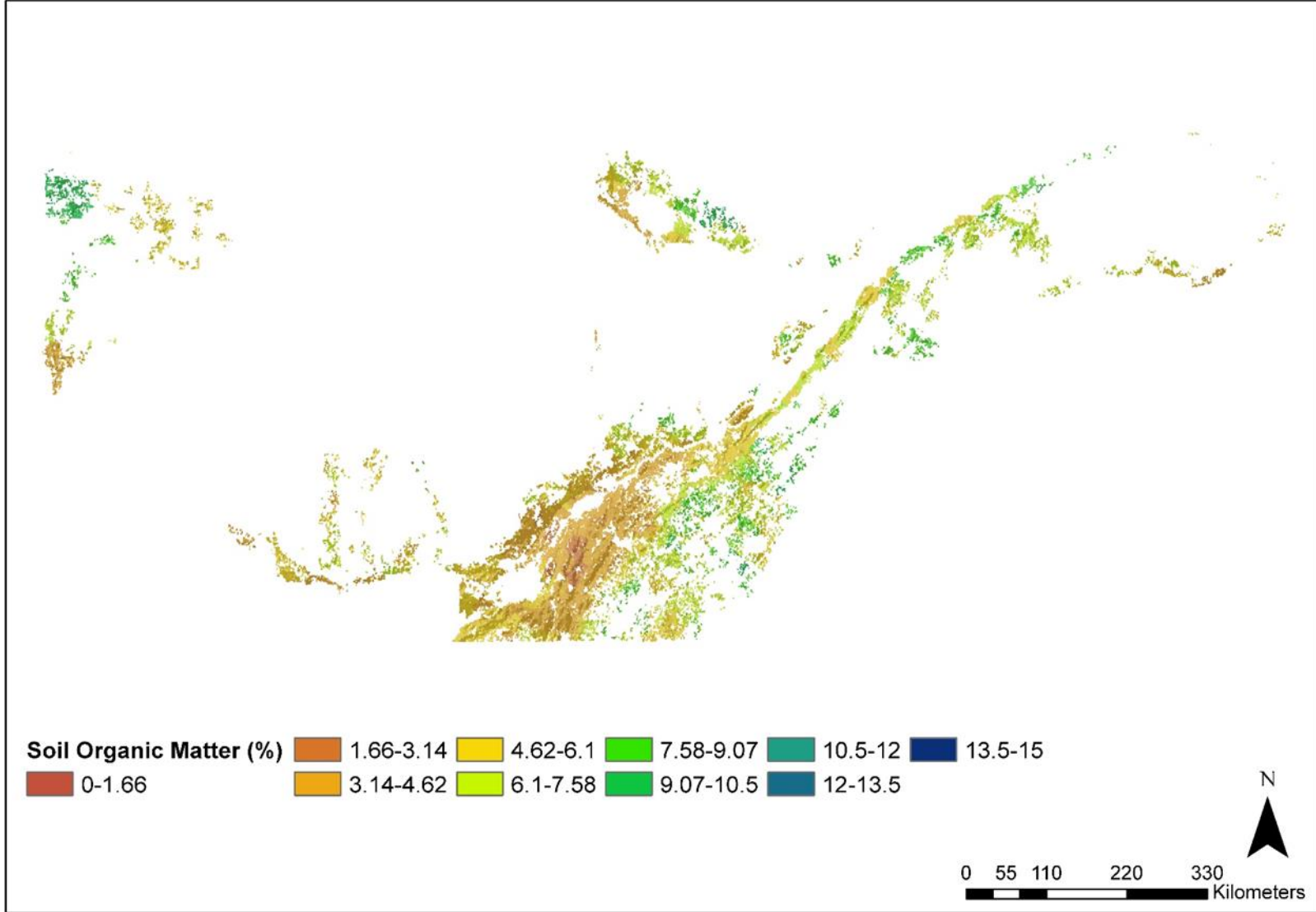


Figure 57 : Disaggregated Soil Organic Matter (%) map of Quebec agricultural soils at 250m resolution using DSMART.

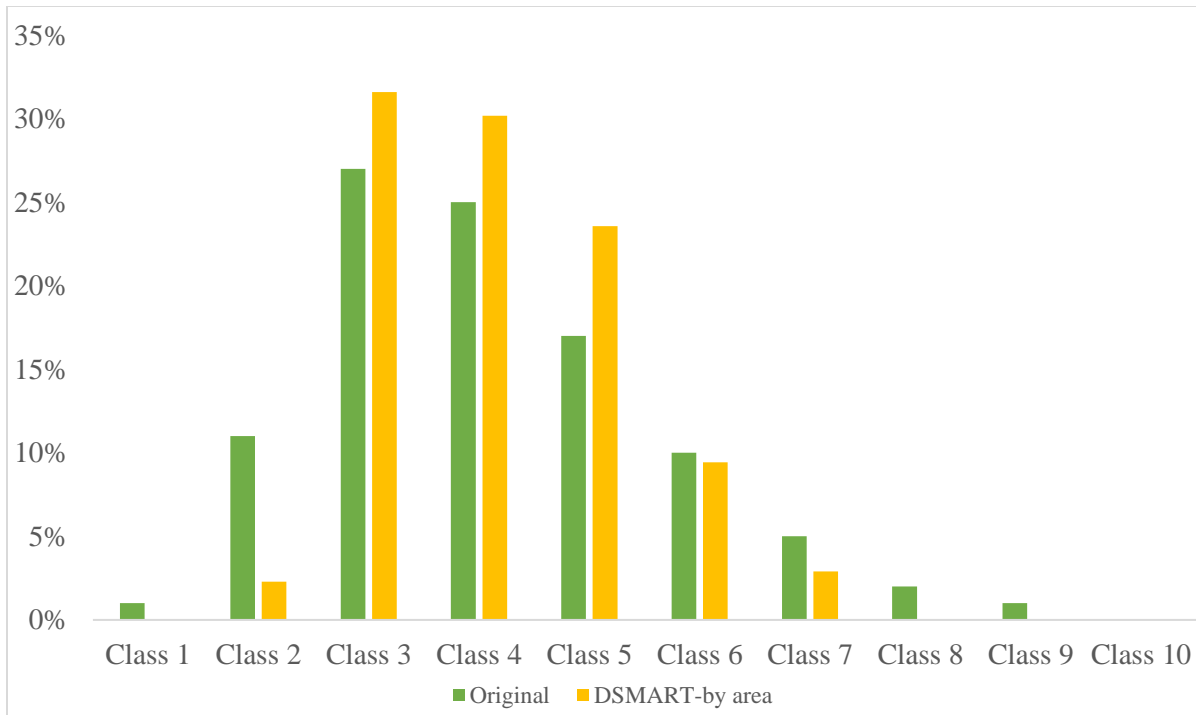


Figure 58 : Comparison between observed SOM (%) classes generated from the database compiled by IRDA and predicted SOM (%) classes from the disaggregation approach using DSMART.

Legend: Class 1 = 0-1.66%, Class 2 = 1.66-3.14%, Class 3 = 3.14-4.62%, Class 4 = 4.62-6.1%, Class 5 = 6.1-7.58%, Class 6 = 7.58-9.07, Class 7 = 9.07-10.5%, Class 8 = 10.5-12%, Class 9 = 12-13.5%, Class 10 = 13.5-15%

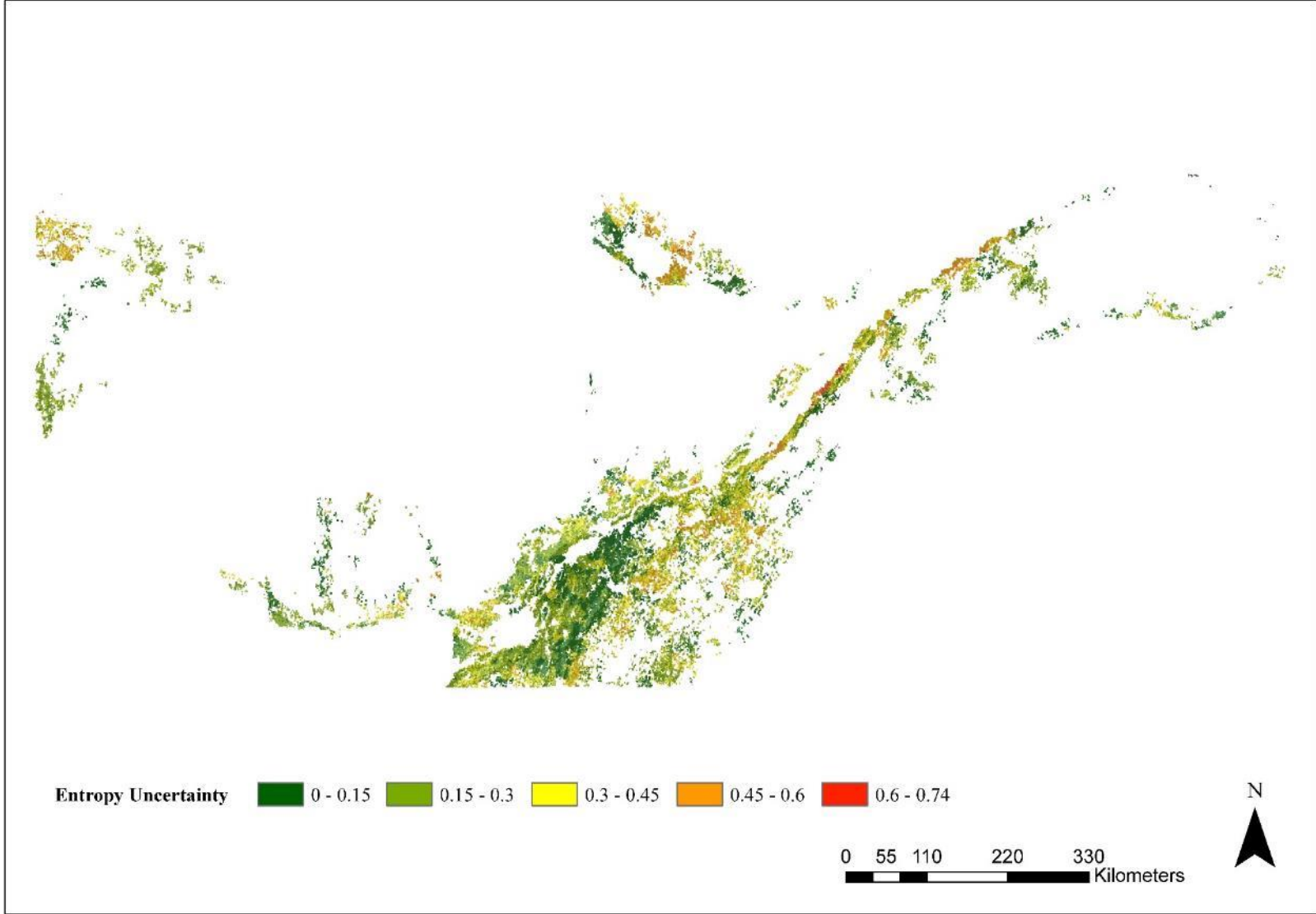


Figure 59 : Entropy Uncertainty of Disaggregated Soil Organic Matter (%) Map of Quebec using DSMART.

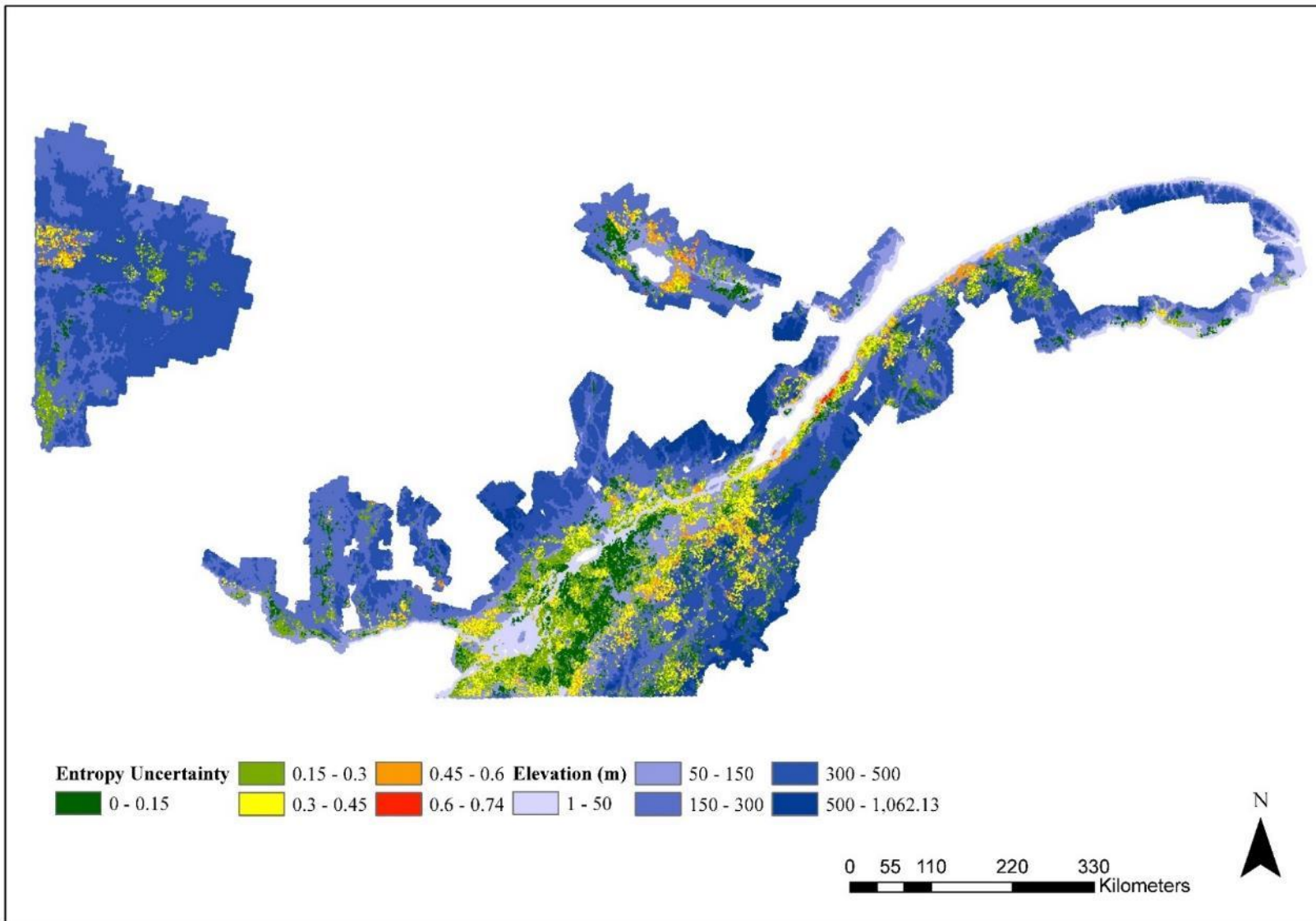


Figure 60 : Entropy Uncertainty of Disaggregated Soil Organic Matter (%) Map of Quebec using DSMART.

Conclusion

This part of the study is centered with the incorporation of conventional soil maps (CSMs) into the digital soil mapping (DSM) framework. CSMs and survey reports are a rich source of information on the distribution and properties of soils. Nonetheless, they often lack the necessary spatial resolution and detail required for contemporary applications and may fail to fully capture the complexity of soil variability within a specific area. Nevertheless, CSMs have potential for identifying general patterns and trends in soil properties and can serve as a point of reference for comparison with more detailed and accurate soil maps produced through DSM techniques.

The study examined and revised various modeling strategies for forecasting soil organic carbon stocks, establishing that digital soil mapping techniques can convert disaggregated maps into soil property maps utilizing conventional soil data. The analysis of uncertainty and accuracy attestation corroborated the maps produced via different digital soil mapping techniques.

The research conducted in this study is expected to have numerous practical applications in utilizing conventional soil maps in Canada to their fullest potential. Additionally, this research has opened new avenues for future investigation into the use of CSMs and DSM techniques for the spatial prediction of soil properties. Soil properties, particularly soil organic carbon (SOC), are crucial indicators of soil health. The DSM techniques used in this study to predict SOC and soil organic matter (SOM) maps at the 0-30cm depth of agricultural soils in Ontario provide an objective reference value for evaluating the effectiveness of initiatives aimed at increasing SOC levels at the provincial level.

The disaggregated soil series maps of Ontario are a valuable source of spatial soil information. While this data layer is useful on its own, it can also serve as a covariate for predicting other soil properties, such as SOC. The work carried out in this study can be applied on a national scale and can serve as a reference for similar projects in other parts of the world.

This research has modified two disaggregation techniques to predict not only soil series but also to generate soil property maps at the provincial level. The soil profile data stored at the NSDB platform can be associated with the disaggregated soil series maps to derive soil property information for a specific location. These techniques can potentially predict soil property maps in areas with limited availability of data.

The Quebec part of this study focused on a unique scenario where SOM% data were aggregated into postal codes without precise georeferencing, prioritizing the privacy of landowners and users. In light of this situation, the application of the DSMART disaggregation approach holds promise for predicting SOM% and other continuous soil attributes, particularly when observations are aggregated. The results of this study offer valuable insights into the disaggregation process of SOM% and its associated uncertainty, the accuracy of the DSMART model, and the challenges faced when predicting specific SOM% classes. These findings contribute to our understanding of the potential and limitations of using DSMART for disaggregating aggregated soil data.

Overall, this study highlights the need for further research to improve the predictive mapping of soil series and soil properties using DSM. By providing guidance on appropriate techniques and highlighting areas for future investigation, this research can contribute to the development of more accurate soil maps, which have significant implications for improving soil management practices and addressing global challenges related to soil health and climate change.

ANNEXE

Presentations and Paper

Conference 1: Comparison of two Digital Soil Mapping techniques to disaggregate soil great groups of six sub-watersheds at Lake Erie Basin using existing soil polygon maps at the Graduate Students' Association 2020 Virtual Conference, University of Guelph 2020

Authors: Tahmid Huq Easher¹, Daniel Saurette¹, Fernando de Jesus Montano Lopez¹, Emma Chappell¹, Brandon Heung², Asim Biswas¹

1. School of Environmental Sciences, University of Guelph
2. Department of Plant, Food, and Environmental Sciences, Dalhousie University

Abstract:

Soil health, climate change, water quality, and food security are all linked together in achieving a sustainable and low-carbon economy. To measure soil health, the first step is to understand the soil types of Ontario and its functionality. The existing soil survey maps and data are the key source to get this information. Although these maps and data rely heavily on broad soil measurement and subjective judgment of surveyors, thus lack efficiency, accuracy, predictive ability, and applicability for practical land management, as the data are polygon-based. Digital Soil Mapping (DSM) has the capability to disaggregate the polygon-based maps to develop predictive digital soil maps with soil information in 3D with pedological descriptions for specific geographical locations, using the existing maps. This presentation will be focused on using two different disaggregation techniques to disaggregate soil class maps of six sub-watersheds at the Lake Erie Basin. DSMART is a well-known and user-friendly approach and can be performed using existing soil class maps and environmental covariates. It disaggregates the polygons by taking existing polygon-based soil maps and correlating them with environmental covariates. DSMART uses multiple resampling techniques to select the same number of sample points randomly from each polygon. On the other hand, Pure Polygon Disaggregation is relatively a recent approach to disaggregate maps, which is developed through random sampling of single-component soil survey map units based on an area-weighted approach. In this approach, each soil observation is spatially intersected with the suite of soil-environmental covariates to develop training points. Then single-model and ensemble-model learners are applied for the prediction of soil great groups and order for point data. These two techniques will be applied to disaggregate the maps, they will be validated using existing data points and then will be compared using statistical methods, to find out which approach works better for Ontario.

Conference 2: A modified approach to disaggregate soil polygon maps using DSMART by incorporating sampling and classifying modifiers at Annual Meeting: Soil Science Society of America 2020.

Authors: Tahmid Huq Easher¹, Daniel Saurette¹, Fernando de Jesus Montano Lopez¹, Emma Chappell¹, Brandon Heung², Asim Biswas¹

3. School of Environmental Sciences, University of Guelph

4. Department of Plant, Food, and Environmental Sciences, Dalhousie University

Abstract:

The recent development in Digital Soil Mapping has provided stimulating options for using existing maps to predict soil classes, enabling researchers to model soil properties. Although data-driven approaches provide better results, pedologists are using statistical approaches to overcome insufficient data and sampling infrastructures. DSMART is a popular method to disaggregate soil series maps which is easier to perform and doesn't require a wide range of data. It trains a single decision tree model from virtual samples by selecting the same number of sample points from each polygon randomly and uses the C5.0 model of the caret package as a default method. It generates virtual samples randomly, which might not represent the polygon or the area properly, and may impact the results of DSMART. Recently Moller et al., 2019 modified the original DSMART approach and included area proportional sampling to improve the DSMART prediction. Also, Flynn et al., 2019 modified the classifier algorithm in DSMART and used 10 machine learning algorithms in the CARET package to produce disaggregated soil maps. Thus, there is an opportunity to modify the sampling technique and classifier modification in DSMART by incorporating different sampling techniques with polygon and area proportional approaches with different machine learning algorithms. Hence, the purpose of this paper is to compare different sampling (multiple resampling technique and conditioned Latin hypercube) and classifying modifiers (random forest, k-nearest neighbor, and multinomial ridge regression) with the default approach of DSMART to disaggregate the soil class maps (soil great groups and soil series) of Medway Creek Watershed at Lake Erie Basin in Ontario, Canada. The results are demonstrating that by modifying both default sampling and classifier algorithms, DSMART will be able to produce improved outputs and make it better. This paper will not only contribute to improving the DSMART approach to generate credible disaggregated maps but also will produce important soil data like soil class.

Conference 3: Comparison of two disaggregation techniques (DSMART & Pure Polygon Disaggregation) using existing soil class polygon maps of Lake Erie Basin, Ontario at Canadian Soil Science Society Meeting 2020.

Authors: Tahmid Huq Easher¹, Emma Chappell¹, Daniel Saurette¹, Brandon Heung², Asim Biswas¹

5. School of Environmental Sciences, University of Guelph

6. Department of Plant, Food, and Environmental Sciences, Dalhousie University

Abstract:

Traditional soil maps have one or multiple soil types in a polygon which restricts the machine learning applications to predict soils using those existing pedon data. Digital Soil Mapping (DSM) can disaggregate polygon-based maps to develop predictive digital soil maps with soil information in 3D with pedological descriptions for site-specific geographical locations, using the legacy maps. This presentation will be focused on comparing two different disaggregation techniques to disaggregate soil great group maps of six sub-watersheds at Lake Erie Basin. DSMART is a well-known user-friendly approach and can be performed using existing soil great group maps and environmental covariates. It disaggregates the polygons by taking existing polygon-based soil maps and correlating them with environmental covariates. DSMART uses a stratified random sampling technique to select the same number of sample points randomly from each polygon. On the other hand, Pure Polygon Disaggregation is relatively a recent approach to disaggregate maps, which is developed through random sampling of single-component soil survey map units based on an area-weighted approach. In this approach, each soil observation is spatially intersected with the suite of soil-environmental covariates to develop training points. Then single-model and ensemble-model learners are applied for the prediction of soil great groups and order for point data. These two techniques will be applied to disaggregate the soil great group maps of six sub-watersheds at Lake Erie Basin, Ontario. Then they will be validated using existing data points and will be compared using statistical methods, to find out which approach suits better for the study areas. Also, this presentation will focus on the benefits and drawbacks of both disaggregation approaches and the potential of these model outputs to be used for site-specific soil properties modeling.

Conference 4: Developing Soil Property (Soil Organic Carbon) Map from the Disaggregated Map of the Legacy Soil Maps of Middlesex County, ON at Canadian Soil Science Society Meeting 2021.

Authors: Tahmid Huq Easher¹, Emma Chappell¹, Daniel Saurette¹, Asim Biswas¹

1. School of Environmental Sciences, University of Guelph

Abstract:

Soil Property Maps, especially Soil Organic Carbon (SOC) data, play an integral role in forming and implementing policies at the regional and national levels, and in making the appropriate and reliable decisions at the farm level to mitigate greenhouse gases from agricultural activities. In Ontario, soil surveys and county reports (also known as legacy soil maps) dated back to the 1980s and 90s are the main sources of this soil data which need to be sampled, tested, and updated to know the current state of the soil. In response to that OMAFRA took the initiative “New Horizon” in “Sustaining Ontario Agricultural Soils” and one of the goals is to generate reliable soil data and tools to allow for informed decision-making and analysis. It is a time-demanding and resource-intensive activity to generate updated data, which is still under process. In the meantime, the authority requires an alternative way to gather soil information to take actions not only to mitigate climate change but also to manage agricultural land uses to meet the rapidly growing food demand. In these regards, this study is focused on developing a soil property map (in this case only soil organic carbon) using the legacy maps and data of Middlesex County, ON. The soil survey data and county maps of Middlesex are polygon-based and each polygon contains one or multiple soil-type units. They contain the proportions, environmental conditions, and soil characteristics (soil type, texture, properties like pH, SOC, etc.) in a detailed database, but don’t contain any graphical representation of the distribution of the soil type units. In this study, Digital Soil Mapping (DSM) techniques are used to disaggregate these compound polygon-based maps into point-based raster digital soil maps to predict the soil classes of each geographical point. Then based on soil survey data and county reports, the average and range of values of SOC for each soil class are assigned to the disaggregated soil class maps to develop a predictive SOC map. Then the maps are validated using existing data points extracted from the county reports and will be compared using statistical methods, to explore the acceptance of this method. Thus, this presentation is focused on showcasing the result of this method in predicting SOC from disaggregated soil maps, the benefits, and drawbacks of this approach, and the potential of this method to be used in predicting other soil properties.

Conference 5: Disaggregating Organic Matter Map of Agricultural Soils, Québec, Canada at at Canadian Soil Science Society Meeting 2022.

Authors: Tahmid Huq Easher^{1*}, Gaetan Martinelli², Marc-Olivier Gasser², Asim Biswas¹

¹ School of Environmental Sciences, University of Guelph, Guelph, ON, Canada

² Institut de recherche et de développement en agroenvironnement (IRDA), Québec City, Québec, Canada

Soil Organic Matter (SOM) has been identified as a major soil health indicator as it improves soil fertility, minimizes soil degradation, and enhances food production. Climate change and intensive farming practices may result in the accelerated decline of SOM, which will hinder achieving a sustainable and low-carbon agricultural economy. Hence, it is important to anticipate the nature and extent of SOM in the agricultural field. In response to that, IRDA (agricultural research and development institute of the Province of Québec) has accumulated 315 000 SOM observations, sampled between 2007-2018, and was assembled in 2019, at the postal code level from the Québec private labs. Most postal code centroids in Québec rural regions are less than 15 km apart. However, producers' administrative addresses could be more distant from related sampled fields and soil analysis reported by a private lab. It creates difficulties to perform digital soil mapping techniques using this dataset as multiple units at the postal code level have more than one sampling point without precise geo-referenced information. The objective of this study is to disaggregate the SOM observations at the postal code level into a point-based raster to produce a predictive SOM map of Québec agricultural soils. A reformed version of DSMART (cLHS as sampling approach and random forest as modeling algorithm) was used to disaggregate this dataset. Typically, DSMART is applied to categorical data (e.g., soil class). To fit this SOM dataset into DSMART, the observations were divided into 10 classes. These classes were built on an equivalent number of samples found in each class at the provincial scale. 1km Digital Terrain Model and the last 30 years of climate and cropping data were used as environmental covariates in the model. The final disaggregated map has been produced and a separate dataset is now being processed to validate the outputs. At the conference, this presentation will focus on showcasing the validation results and the disaggregated SOM map of Québec agricultural soils. Also, the drawback of this method and its potential will be further discussed through this oral presentation.

Keywords: Disaggregation, Soil Organic Matter, Digital Soil Mapping, DSMART

Paper 1: Sampling and classifier modified DSMART to disaggregate soil polygon maps (published at Geoderma)

Tahmid Huq Easher^a, Daniel Saurette^a, Emma Chappell^a, Fernando de Jesus Montano Lopez^a, Brandon Heung^b, Asim Biswas^{a*}

^a School of Environmental Sciences, University of Guelph

^b Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University

Highlights

- The C5.0 algorithm performed better as a model in predicting soil class using DSMART
- Area proportional sampling approach performed better in disaggregating maps
- Runtime using Conditioned Latin Hypercube Sampling and k-Nearest Neighbor was faster

Abstract

DSMART has been widely used to disaggregate soil polygon maps into point-based raster maps. It uses a stratified random sampling technique to select the same number of sample points from each polygon and uses the C5.0 model of the caret package as a default model method. Few published studies have modified the original DSMART approach to include an area proportional sampling approach and used different machine learning algorithms in the caret package, to produce disaggregated soil maps. Thus, there is an opportunity to modify the sampling technique of the DSMART and incorporate different sampling techniques with polygon and area proportional approaches with different machine learning algorithms of the caret package. Hence this paper compared different sampling approaches (stratified random sampling and conditioned Latin Hypercube Sampling) and applied three classifying modifiers (C5.0, random forest, and k-nearest neighbor) in DSMART to disaggregate the soil great group and soil class maps of a sub-watershed in Southern Ontario, Canada. We found that for soil great group prediction, the default approach of DSMART performed better compared to other sampling approaches and classifier models, compared with the validation dataset we extracted from the county soil report. For soil class prediction, the combination of the area proportional sampling approach and the C5.0 algorithm of DSMART performed better, compared with the validation dataset; whereas the default approach of DSMART and the combination of conditioned Latin Hypercube Sampling and the C5.0 algorithm produced a similar performance. The processing time of cLHS approaches was significantly less than the default approaches of DSMART.

Keywords: digital soil mapping, disaggregation, sampling techniques, classification and regression model, machine learning algorithms

Paper 2: Comparing DSMART and Pure Polygon Disaggregation Approach in Disaggregating Soil Series Maps. (Accepted at Pedosphere with moderate revision)

Tahmid Huq Easher¹, Daniel Saurette¹, Brandon Heung², Richard J. Heck¹, Adam Gillespie¹, Asim Biswas¹
1 School of Environmental Sciences, University of Guelph, 50 Stone Road East, Guelph, ON, Canada, N1G 2W1

2 Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, PO Box 550, Truro, Nova Scotia, B2N 5E3

Abstract

Conventional Soil Maps (CSMs) often have multiple soil types within a single polygon, which hinders the ability of machine learning to accurately predict soils. Soil disaggregation techniques are commonly used to improve the spatial and attribute precision of these maps. DSMART is a popular but computationally intensive method, as it generates and assigns synthetic samples to soil series based on the CSMs' areal coverage information. Alternatively, Pure Polygon Disaggregation (PPD) is a less computationally demanding technique and assigns soil series based solely on the proportion of soil series in pure polygons (polygons with a single soil series) across the CSM area. This study compared these two techniques by applying them to a CSM of Middlesex County, Canada. Four different sampling methods were used: two sampling designs (simple random sampling, SRS and conditional Latin hypercube sampling, cLHS) with two sample sizes (83,100 and 19,420) per sampling plan, both based on an area-weighted approach. The C5.0 and random forest (RF) were applied to both techniques to compare the result accuracy. Five hundred validation points were used for the accuracy assessment. C5.0 showed better performance (tau 0.6 to 0.65) than RF (0.55-0.57) with a larger sample size, and PPD with C5.0 using a larger sample size was the best performing (tau 0.65) approach. With a smaller sample size, both cLHS (tau 0.45 to 0.51) and SRS (0.44-0.50) produced similar accuracy results. PPD was found to be significantly less computationally intensive and time-consuming than DSMART. The maps generated through PPD had lower uncertainty values.

Keywords: Disaggregation, machine learning algorithm, conditioned latin hypercube sampling, soil classification prediction, prediction uncertainty.

AXE 3 : MESURER L'EFFICACITÉ DE PRATIQUES DE CONSERVATION À CONTRÔLER L'ÉROSION HYDRIQUE SOUS LES CONDITIONS ÉDAPHIQUES ET CLIMATIQUES ACTUELLES ET FUTURES

Claude Bernard, IRDA; Zita Ondo, IRDA, INRS;

Résumé

Les changements climatiques prévus à l'horizon 2050 en Ontario et au Québec se traduiront par une intensification des cultures. Sans mesure de conservation, cette intensification devrait favoriser une perte de carbone du sol et une fragilisation de sa structure, contribuant à accroître sa vulnérabilité à la compaction et son érodabilité. Les sols ainsi affectés verront leur potentiel d'érosion s'accroître bien au-delà des seuils agronomiquement acceptables, mettant en péril leur productivité à long terme.

Les pratiques de conservation, telles que le travail réduit du sol ou l'implantation de cultures intercalaires dans les cultures annuelles comme le maïs et le soya, favorisent le maintien des propriétés physico-chimiques (matière organique, stabilité structurale) du sol et la réduction des risques d'érosion et d'exportation de polluants qui y sont associés. La capacité de ces pratiques à maintenir les pertes de sol à l'intérieur de limites acceptables sous les conditions futures de climat et de qualité des sols demande cependant à être évaluée. Un dispositif expérimental en split-plot a été implanté à Saint-Lambert-de-Lauzon pour comparer sous pluies simulées en parcelles principales, l'effet des précipitations du futur (intensité, durée et fréquence plus élevées) et des conditions futures de sol dégradé (taux de matière organique et stabilité structurale diminuées) sur l'efficacité des cultures intercalaires et du travail réduit à protéger les sols de l'érosion hydrique par rapport aux conditions actuelles de sol et de précipitations.

Trois blocs de répétitions sur six ont été travaillé intensivement en 2020 avant l'implantation des pratiques de conservation pour simuler la dégradation de la structure et la diminution du taux de matière organique en climat futur. Sur chaque bloc de répétition, six traitements ont été implantés en sous-parcelles à partir de l'automne 2020 pour comparer l'effet de : 1) sol nu; 2) culture de soya sous travail conventionnel et sans plante couverture; 3) culture de soya sous travail conventionnel et avec plante couverture; 4) culture de soya sous travail réduit (chisel) et sans plante couverture; 5) culture de soya sous travail réduit et avec plante couverture; 6) prairie. En septembre 2021, deux pluies simulées de 30 minutes d'une récurrence de 1/20 ans représentatives du climat actuel ($61 \text{ mm}\cdot\text{hr}^{-1}$) et du climat futur, horizon 2050 ($71 \text{ mm}\cdot\text{hr}^{-1}$) ont été appliquées en trois répétitions (par blocs) sur les six systèmes culturaux en sous-parcelles. L'analyse des échantillons de sols prélevés et du ruissellement, les observations de terrain, l'ANOVA et l'ANCOVA sur les moyennes obtenues, et l'application de la USLE ont permis d'évaluer l'effet des différents systèmes culturaux sur l'érosion du sol et la qualité des eaux dans une perspective d'intensification des pluies dans le sud du Québec.

Sous régime de précipitations en climat futur, le ruissellement et l'érosion des sols a augmenté en moyenne de 48 %, tandis que les pertes d'orthophosphates, de P biodisponible et de N total a augmenté de 42 %, 53 % et 86 % respectivement. Par rapport au semis direct, le travail du sol au chisel a augmenté les pertes de sol et le P biodisponible par N total, de 126 %, 136 % et 140 %, respectivement. En moyenne, le sol laissé à nu a augmenté le coefficient de ruissellement de 5 % et les pertes totales de N jusqu'à 57 %. Sous régime de précipitations actuel, les pertes de sol et de N total étaient similaires avec ou sans couverture du sol, mais elles ont augmenté de 59 % et 87 % respectivement sous régime climatique du futur. L'effet du travail du sol a presque doublé les pertes de sol et de N total en climat futur par rapport à son effet en climat actuel. Les facteurs K et C de l'équation universelle de perte de sol (USLE) ont été calculés à partir des résultats des tests. Les valeurs de K variaient de 0,0028 à 0,0109 $\text{t ha}^{-1} \text{ h}^{-1} (\text{ha MJ mm})^{-1}$, tandis que les valeurs de C variaient de 0,141 sous prairie à 2,37 sous soya en travail réduit du sol et culture intercalaire, un résultat inattendu car sous culture ce facteur ne devrait pas être supérieur à 1,0, la référence sur sol nu. Un sol travaillé au burin aurait rendu le sol plus vulnérable à l'érosion qu'un sol laissé nu sans travail du sol, qu'une culture de couverture ou un

remplissage soit utilisé ou non, du moins en fin de saison dans les cultures espacées qui laissent peu de résidus sur le sol comme le soja.

Mots-clés : intensification des pluies, simulateur de pluie, érosion hydrique, systèmes culturaux, pertes de sol et des nutriments, USLE, érodabilité du sol, indice cultural C.

De plus ample détails sur la revue de littérature, le matériel et méthodes, les résultats et leur interprétation sont disponibles dans le mémoire de maîtrise de Zita Antoine Ondo :

Zita Antoine Ondo, 2022. Intensification des pluies et impacts sur l'érosion hydrique des sols et la qualité des eaux sous six systèmes culturaux à Saint-Lambert-de-Lauzon. Mémoire de Maîtrise en sciences de l'eau. INRS. 125 p. +Annexes. <https://espace.inrs.ca/id/eprint/13179/>

Deux présentations des résultats ont été faites par l'étudiant à la maîtrise en 2021 lors du Colloque de l'AQSSS (15-17 juin) et une lors de la réunion de l'ACRH (26-28 mai). L'étudiant a également présenté un séminaire dans le cadre de son programme de Maitrise M. Sc. au printemps 2022.

	RUNOFF.VOL	RUNOFF.COEFF	EROSION	o-PO4	Pbiotot	Nt
	(L)	(%)	(kg/ha)	(mg/m2)	(mg/m2)	(mg/m2)
Future	226	63	733	0,968	37	20644
Present	153	49	495	0,684	24	11077
Fut/Pre	1,48****	1,29**	1,48*	1,42	1,53	1,86*
Direct	181	52	426	0,665	20	10570
Chisel	201	60	963	0,755	48	25408
Ch/Dir	1,11	1,15	2,26***	1,14	2,36****	2,40**
W.Cover	177	55	605	0,652	31	14016
Wo.Cover	205	58	783	0,769	38	21962
Wo/W	1,16	1,05	1,29	1,18	1,24	1,57
Dir Fut	243	67	531	0,890	26	13200
Dir Pre	120	38	319	0,441	15	7940
Fut/Pre	2,03****	1,76****	1,66*	2,02**	1,72	1,66
Ch Fut	244	67	1344	0,931	66	39338
Ch Pre	158	53	582	0,580	30	11479
Fut/Pre	1,54*	1,26	2,31**	1,61	2,20**	3,43***
W.Cov Fut	223	65	724	0,814	35	18282
W.Cov Pre	132	44	485	0,489	26	9750
Fut/Pre	1,69*	1,48	1,49	1,66	1,36	1,88*
Wo.Cov Fut	264	68	1152	1,010	56	34256
Wo.Cov Pre	146	47	416	0,532	19	9669
Fut/Pre	1,81****	1,48**	2,77**	1,90**	2,97**	3,54*
W.Cov Pre	132	44	490	0,489	26	9750
Wo.Cov Pre	146	47	420	0,532	19	9669
Wo/W	1,11	1,07	0,86	1,09	0,73	0,99
W.Cov Fut	223	65	724	0,814	35	18282
Wo.Cov Fut	264	68	1151	1,010	56	34256
Wo/W	1,18	1,05	1,59	1,24	1,61	1,87*
Dir Pre	120	38	319	0,441	15	7940
Ch Pre	158	53	582	0,580	30	11479
Ch/Dir	1,32	1,39	1,82*	1,32	2,01*	1,45
Dir Fut	243	67	531	0,890	26	13200
Ch Fut	244	70	1344	0,931	66	39338
Ch/Dir	1,00	1,04	2,53**	1,05	2,56**	2,98**

Statistical significance: * 0,1 ** 0,05 *** 0,01 **** <0,01

ANNEXE

Estimation des facteurs K et C sur un site agricole dans le Bassin Versant de la rivière Etchemin

Zita Antoine Ondo - Antoine_Zita.Ondo@ete.inrs.ca
 Claude Bernard, Ph.D. - claud.bernard@irda.qc.ca
 Alain Rousseau, Ph.D. Ing. - alain.rousseau@ete.inrs.ca

Contexte : l'intensification de l'agriculture suite aux changements climatiques pressentis au Québec à l'horizon 2050 pourrait davantage :

- fragiliser la structure des sols (Ketcheson, 1980) et accroître son érodabilité (Bernard, 1997) et l'indice C des systèmes culturaux
- conduisant à terme à l'augmentation de l'érosion malgré l'adoption de pratiques de conservation (Bootsma et al., 2001).

Objectif : estimer l'érodabilité (K) ainsi que l'indice cultural (C) du sol de 36 parcelles regroupées dans 6 blocs disposés sur un sol Le Bras à la ferme de l'IRDA à Saint-Lambert-de-Lauzon : (Lat.: 46°36'20", Long.:71°10'17") en sept 2020

Méthodologie :

Dispositif expérimental = plan en parcelles partagées avec

1- Deux Intensités de pluie (mm.h⁻¹) de 30 min d'une récurrence de 1/20 ans au niveau des parcelles principales ou Blocs à savoir :

- 61 mm.h⁻¹ → 3 Blocs (condition du climat actuel)
- 71 mm.h⁻¹ → 3 Blocs (condition du climat futur)

2- 6 systèmes culturaux (par Bloc) installés de façon aléatoire dans les parcelles secondaires, à raison de 6 parcelles secondaires par Bloc soit :

- (i) sol nu, S-Nu; (ii) semis direct sans plante de couverture, SD-SPLC; (iii) semis direct avec plante de couverture, SD-PLC; (iv) labour au chisel sans plante couverture, Ch-SPLC; (v) labour au chisel avec plante couverture, Ch-PLC; et (vi) prairie, P;

sur un site de pente moyenne 4.5% (2.58°).

NB : parcelles secondaire (6 m x 3 m)
 parcelle principale-Bloc-(18 mx 6 m)

3- collecte des échantillons : de sol et de ruissellement respectivement avant et après chaque simulation de pluie.

4. Calcul du facteur K par d'une part :

• le Nomogramme (Wischmeier) → K_{nomog}
 $K_{nomog} = 2.77 \cdot 10^{-7} \cdot M^{1.4} \cdot (12 - a) + 0.0043 \cdot (b - 2) + 0.0033 \cdot (c - 3)$
 (Wischmeier et Smith, 1971)

Méthodologie (suite): et d'autre part par :

• l'équation $A = R \cdot K \cdot LS \cdot C \cdot P$ (USLE)
 → (parcelle S-Nu → référence par Bloc)

$$K_{mesurée.Bloc(i)} = \left(\frac{A(i)(S - Nu)}{(E \cdot 1.30 \max(S - Nu) \times L \cdot S(S - Nu))} \right)$$

5- Calcul du facteur [L.S] dans 36 parcelles :

$$[L \cdot S] = [(X/22.13)^{0.7}] \cdot [(65.41 \cdot \sin^2\theta + 4.65 \cdot \sin\theta + 0.065)]$$

Enfin ,

6- Calcul de l'indice C pour 6 syst. culturaux

$$C(Trt(i)) = \left(\frac{A(i)(Trt(i))}{(E \cdot 1.30 \max(i)(Trt(i)) \times K_{mesurée.Bloc(i)} \times L \cdot S(Trt(i)))} \right)$$

NB : C*P = 1 ; Cdt. = conditions; Fu. = futur; Ac. = actuel

I.2- Érodabilité : K (analyse-discussion)

Blocs.Cdt.Fu.	K _{nomo.W(moy.)}	K _{mesurée(Bloc)}	ratio = $\frac{K(cdt.Fu.)}{K(cdt.Ac.)}$
B.#4	0.014	0.0041	ratio = 1,11 Cdt.Fu. : ↑ 11% K (n.s) K _{pluie.SJM} < 2-5 fois K _{nomog.} Salehi et al.(1993) rapportent des différences similaires à Lennoxville (Québec)
B.#5	0.028	0.0028	
B.#1	0.018	0.0046	
Moy.	0.020 (a)	0.0038 (**)	
Blocs.Cdt.Ac.	K _{nomo.W(moy.)}	K _{mesurée(Bloc)}	
B.#2	0.010	0.0074	
B.#3	0.014	0.0073	
B.#6	0.031	0.0109	
Moy.	0.018 (a)	0.0085 (**)	
(a) test-F p-value 0.839 (n.s.)			
(**) test-F p-value 0.0224 (sig.)			

II- Indice cultural : C

Syt.culturaux	Cdts.	C _{nomo.Par}	C _{actuel}	C _{actuel} /C _{nomo.Par}	Sig.
P (n=5)	Futur	0.141(a)		0.73	p-value (0.120)(n.s)
	Actuelle	0.192(a)			
Ch-PLC (n=6)	Futur	1.397(a)		1.70	p-value (0.242)(n.s)
	Actuelle	0.823(a)			
Ch-SPLC (n=6)	Futur	2.370		5.35	p-value (0.023)**
	Actuelle	0.443			
SD-PLC (n=6)	Futur	1.112		4.76	p-value (0.046)**
	Actuelle	0.234			
SD-SPLC (n=6)	Futur	0.643(a)		1.45	p-value (0.609)(n.s)
	Actuelle	0.443(a)			
S-Nu (n=6)	Futur	1.00		1.00	Référence
	Actuel	1.00			

n.s = non significatif; ** = sig. à 0.05
 Sous pluie actuelle (C : 0.19 (P) – 1.00 (S-Nu))
 Sous pluie future (C : 0.14 (P) – 2.37 (Ch-SPLC))
 • sous pluies futures, l'indice C a significativement ▲ | sous le labour au Ch (2.370) comparativement au SD (1.112)

III- Facteurs explicatifs des variations du C

► parcelles Ch-SPLC plus perturbées → t.ha⁻¹:

DMP (mm)	Ch (1.36 - 1.71)(e)	SD (2.46 - 3.31)(b)
Syst.cult.(****)		P (2.41 - 2.77)(b) S-Nu(2.10 - 2.41)(b)
%ASE (> 1.00 mm)	Ch (45%)(A)	SD (69% - 79%)(B) P (73%)(B) S-Nu (63%)(B)
Covul.col. (%) (X) Y = Érosion spécifique par érosivité (g.m ⁻² .E _{150mm} ⁻²) (X < 30%) : Y = 0.1341.e ^{0.0019(X-1)} (X > 30%) : Y = 0.5248.e^{0.0003(X-1)}}}		
Influence du labour au Ch sur le D.M.P., le %ASE et l'érosion spécifique		
► Croûte de battance sur S-Nu (Coef. Ruiss. : 85% - 65%)		



Résultats

I.1- Érodabilité : K (ensemble du site)

Statistiques	K [t.ha.h.(ha.MJ.mm) ⁻¹]
Moy.	0.019
Méd.	0.017
Max.	0.035
Min.	0.007
Mode	0.015
E.T	0.0081
Erreur.Std.moy.	0.0014
N	36
I.C 95%	0.016 ≤ moy. ≤ 0.022

Perspectives

Estimer K et C au printemps

Intensification des pluies en climat futur et érosion des sols agricoles à Saint-Lambert-de-Lauzon

ZITA ANTOINE ONDOA¹, CLAUDE BERNARD², ALAIN N. ROUSSEAU¹

¹ Centre Eau Terre Environnement, INRS

² Institut de recherche et de développement en agroenvironnement (IRDA)

Antoine_Zita.Ondoa@ete.inrs.ca

Mots clés : pluies, systèmes culturaux, érosion, charges

Suite aux changements climatiques, l'intensification des pluies printanières et automnales, l'exposition précoce et prolongée du sol aux périodes critiques de gel et de fonte printanière, l'augmentation des fréquences de cycles de dégel et de gel, l'augmentation des superficies en cultures intensives contribueront, entre autres, à accroître la susceptibilité des sols à l'érosion hydrique (Ouranos, 2015). L'objet de la présente étude était d'évaluer l'efficacité des pratiques de conservation du sol sur le ruissellement, les pertes de sols, de matière organique (M.O.), de phosphore et d'azote.

Deux pluies simulées de 30 minutes d'une récurrence de 1/20 ans de 61 mm h⁻¹ et 70 mm h⁻¹ ont été retenues, pour représenter le climat actuel et le climat futur. Six (6) systèmes culturaux ont été étudiés sous les deux (2) intensités de pluie : sol nu, S-Nu; semis direct sans plante de couverture, SD-SPLC; semis direct avec plante de couverture, SD-PLC; labour au chisel sans plante de couverture, Ch-SPLC; labour au Chisel avec plante de couverture, Ch-PLC; Prairie, P.

Un dispositif en parcelles divisées a été retenu, avec les pluies en parcelles principales et les traitements culturaux en parcelles secondaires. Lors des simulations de pluie, des échantillons de sol et de ruissellement ont été recueillis et analysés selon Carter et Gregorich (2007). Le coefficient d'uniformité moyen par bloc des pluies simulées était de 79 % et 78 % respectivement pour les deux intensités. De façon générale, le volume et le coefficient de ruissellement, ainsi que les exportations de sédiments présentent des accroissements significatifs respectivement de 47 %, 29 % et 48 % en climat futur. L'effet pluie s'avère être significatif sur ces variables pour les traitements Ch-SPLC et SD-PLC plus particulièrement. On a aussi noté une décroissance exponentielle et significative de l'érosion avec l'augmentation du couvert au sol. L'effet de la pluie a également été significatif pour les pertes de matière organique, d'azote total et de P-Mehlich3 (PM3), mais non significatif pour les autres formes de phosphore et l'ammonium. Les exportations de phosphore, de matière organique et d'azote total sont fortement corrélées ($r \geq 0,80$) avec les pertes de sol (t ha⁻¹). Une fraction du phosphore biodisponible (forme faiblement liée à la M.O.) et du PM3 exportés est également corrélée à la M.O. présente dans les érodats. Les ratios des pentes de régressions (phosphore, M.O., azote total dans les sédiments avec les pertes de sol) en conditions futures et conditions actuelles sont de 1,07 à 1,41 ce qui confirme les risques liés à l'intensification des pluies en climat futur.

Dans l'ensemble, la pluie future a provoqué plus d'exportations de sédiments et de nutriments que la pluie actuelle. Cependant, l'augmentation moyenne des exportations a été beaucoup plus significative sous l'effet du travail du sol (SD/Ch) que sous l'effet du couvert au sol (PLC/SPLC).

Références

Carter, M. R., Gregorich, E.G (eds). 2007. Soil Sampling and Methods of Analysis. Second Edition. CRC Press. Boca Raton, Florida. 1224 p.

Ouranos. 2015. Vers l'adaptation. Synthèse des connaissances sur les changements climatiques au Québec. Edition 2015. Montréal, Québec : Ouranos. 415 p.

Estimation des facteurs K et C sur un site agricole dans le Bassin Versant de la rivière Etchemin

ZITA ANTOINE ONDOA¹, CLAUDE BERNARD², ALAIN N. ROUSSEAU¹

¹ Centre Eau Terre Environnement (INRS ETE)

² Institut de recherche et de développement en agroenvironnement (IRDA)

Antoine_Zita.Ondoa@ete.inrs.ca

Mots clés : pluies simulées, systèmes culturaux, Indices K C

L'intensification de l'agriculture suite aux changements climatiques pressentis au Québec à l'horizon 2050 pourrait davantage fragiliser la structure des sols (Ketcheson, 1980), accroître son érodabilité (K), et augmenter l'indice cultural (C) des systèmes culturaux, conduisant à terme à l'augmentation de l'érosion et ce malgré l'adoption de pratiques de conservation (Bootsma et al., 2001). L'objectif de la présente étude était d'estimer les indices K et C de l'équation universelle des pertes de sol de 36 parcelles regroupées dans 6 blocs disposés sur un sol Le Bras à la ferme de l'IRDA à Saint-Lambert-de-Lauzon. La texture variait de Loam à Loam sableux, avec une pente moyenne de 4,5 %.

Six systèmes culturaux (6) ont été installés sur des parcelles de 6 m x 3 m : (i) sol nu, S-Nu; (ii) semis direct sans plante de couverture, SD-SPLC; (iii) semis direct avec plante de couverture, SD-PLC; (iv) labour au chisel sans plante de couverture, Ch-SPLC; (v) labour au chisel avec plante de couverture, Ch-PLC; et (vi) prairie, P. Deux pluies simulées de 30 minutes d'une récurrence de 1/20 ans de 61 mm h⁻¹ et 70 mm h⁻¹ ont été appliquées respectivement aux blocs du climat actuel et à ceux du climat futur. Les valeurs de K ont été estimées à partir du nomogramme de Wischmeier et al. (1971), et des pertes de sols mesurées. Ces dernières ont aussi servi à mesurer les valeurs C des traitements étudiés en utilisant l'érodabilité mesurée sur la parcelle de S-Nu comme référence.

L'érodabilité moyenne du site calculée avec le nomogramme est de 0,019 t.ha.h.(ha.MJ.mm)⁻¹. Les blocs recevant les pluies futures ont démontré une érodabilité 11% supérieure, bien que non statistiquement significative. L'érodabilité du site mesurée au simulateur de pluie variait de 0,0038 à 0,0085 t.ha.h.(ha.MJ.mm)⁻¹, soit 2 à 5 fois inférieure à la valeur obtenue avec le nomogramme. Salehi et al. (1993) ont rapporté de telles différences à Lennoxville.

Sous pluie actuelle, l'indice C a varié de 0,19 (P) à 1,00 (S-Nu). Sous pluie future, l'indice a varié de 0,14 (P) à 2,37 (Ch-SPLC). Ce résultat s'explique par une érosion plus forte sous Ch-SPLC que sur S-Nu. Sous pluie future, l'indice C a plus augmenté sous le travail au chisel que le semis direct. Ce dernier semble donc plus approprié pour réduire l'érosion sous pluie future que le travail primaire avec chisel.

Références

- Bootsma, A., Gameda, S. et al. 2001. Adaptation de la production agricole au changement climatique dans le Canada atlantique. Rapport final de l'équipe du projet A214 du Fonds d'action pour le changement climatique. Agriculture et Agroalimentaire Canada, Ottawa.
- Ketcheson, J.W. 1980. Long-range effects of intensive cultivation and monoculture on the quality of southern Ontario soils. *Canadian Journal of Soil Science*, **60**:403-410.
- Salehi, F., Pesant, A.R., Bérard, A and Lagacé, R. 1993. Preliminary estimates of the erodibility of ten Quebec Eastern Townships soil series. *Canadian Agricultural Engineering*, **35**:157-164.
- Wischmeier, W.H., Johnson, C.B. et al. 1971. A soil erodibility nomograph for farmland and construction sites. *Journal of Soil Water Conservation*, **26**:189-193.

ANALYSE ET DISCUSSION

L'utilisation des analyses de sols issues d'un laboratoire privé pour faire des inférences sur l'évolution de la santé des sols comporte ses avantages et ses inconvénients. Le peu d'information sur l'origine des sols (agricole, forestière, potager, etc.) soumis à l'analyse ainsi que leur géoréférence associée à un code postal engendre un certain niveau d'incertitude qui nécessite des méthodes de filtration efficaces. Le laboratoire privé qui vise à partager ou vendre ces analyses pourraient prévoir des formulaires de demande d'analyse pour améliorer la fiabilité de l'information. En revanche, même avec ces incertitudes, l'analyse de ces données de sols a permis de réaliser des constats sur la spatialisation des propriétés des sols à l'échelle de la zone agricole du Québec et de l'effet des changements climatiques, qui n'ont jamais été réalisés auparavant avec une telle résolution. Donc ces analyses de sols représentent une mine d'information sur les sols à valoriser dans le cadre d'une telle étude.

Les travaux réalisés en Ontario ont démontré que les informations contenues dans les études pédologiques peuvent aussi être exploitées et interpolées à des résolutions plus fines (grilles de 50 à 250 m) avec différentes techniques de désagrégation (rééchantillonnage par polygone pure ou surface, techniques d'IA) permettant de réduire les niveaux d'incertitudes. Ainsi, le grand groupe, la série de sol, ou la teneur en matière organique du sol (MOS) a pu être spatialisée à haute résolution à l'échelle de l'Ontario. Le même exercice a été réalisé à l'échelle du Québec avec les analyses de MOS cumulées à l'échelle des municipalités sous forme de quantiles et désagrégées sur des grilles de 250 m de résolution.

Le calcul et l'inclusion de plusieurs indicateurs agroclimatiques a fait ressortir les liens entre la teneur en MOS et certains indicateurs comme le *Climate decomposition index* (CDI), la température moyenne et le bilan hydrique, en plus d'examiner la relation entre la teneur en MOS et la définition de la longueur de saison de croissance définie par le seuil des degrés jours et la période de référence en années lors du calcul de la normale climatique du CDI. Cette relation était plus forte à un seuil de -5°C pour les degrés-jours et une normale climatique sur 30 ans. Ce genre d'analyse de la définition des indicateurs agroclimatiques et leur relation avec les indicateurs de sol (MOS) mériterait d'être approfondie, car toute inférence sur les changements climatiques en dépend.

La définition du nombre de zones agroclimatiques a été restreint à six, la méthode de classification ascendante hiérarchique utilisée n'en retenant que quatre. Il serait pertinent de comparer ce découpage à d'autres techniques pour augmenter le nombre de zones permettant de raffiner l'analyse sur des zones plus distinctes en termes de potentiel agricole et d'influence sur les sols.

L'application du modèle d'évolution de la MOS dans les conditions de l'Ontario a donné des résultats mitigés pour différentes raisons. Premièrement, certaines simulations ont été réalisées en dehors du domaine d'entraînement du modèle. Par exemple, les températures moyennes sont plus élevées en climat actuel dans les zones les plus chaudes en Ontario comme en climat futur pour les zones plus chaudes au Québec, par rapport au climat actuel qui a servi à la modélisation. Deuxièmement, le ratio de cultures pérennes très différent entre les deux provinces peut résulter soit d'une réalité agricole différente ou de la source de données différente utilisée pour calculer ces ratios. Cette deuxième raison mériterait d'être analysée, en même temps que l'inclusion d'autres ratios de cultures dans les modèles.

Une application a été développée et remise au laboratoire privé qui a contribué aux données pour les travaux de modélisation de l'évolution de la teneur en matière organique du sol (MOS) en climat futur. L'application permet de prédire la texture à partir des autres analyses courantes réalisées au laboratoire sur un échantillon de sol. Ces travaux ont également conduit à la production de niveaux probables de MOS par groupe textural de sol et par zone agroclimatique pour fin de référence dans des suivis de la MOS à la ferme. Les travaux de modélisation dans les conditions du Québec ont souligné l'importance de mettre en place des mesures de conservation des sols qui seront éventuellement convertis en cultures annuelles dans les régions plus froides. Selon les prévisions des modèles, les quantités perdues de MOS et de C émis dans l'atmosphère seront plus importantes dans les sols argileux des régions des Laurentides, du Témiscamingue et du pourtour du Lac-Saint-Jean et dans les sols sableux du nord du Lac-Saint-Jean et de l'Abitibi.

L'expérience conduite en parcelles expérimentales sous simulateur de pluie a démontré que sous le régime de précipitations en climat futur, le ruissellement et l'érosion des sols augmenteraient en moyenne de 48%, tandis que les pertes d'orthophosphates, de P biodisponible et de N total augmenteraient de 42%, 53% et 86% respectivement. Le sol travaillé au chisel aurait rendu le sol plus vulnérable à l'érosion que le sol laissé à nu sans travail de sol peu importe l'usage ou non d'une culture ou intercalaire de couverture, du moins en fin de saison dans les cultures à large interligne qui laisse peu de résidus au sol comme le soya.

CONCLUSION ET RECOMMANDATIONS

Différents modèles en apprentissage machine ou linéaires ont prédit sur un jeu indépendant de données, la répartition de la MOS à l'échelle du Québec avec des R^2 variant de 0,42 à 0,49. Le modèle Random Forest retenu pour faire ces prédictions inclut par ordre d'importance, la texture, le ratio de cultures pérennes, des indicateurs climatiques et finalement des variables liées au MNT. À l'orée de 2100, les sols connaîtront une diminution plus importante de leur teneur en MOS dans les régions plus froides du Québec, ainsi qu'une diminution des superficies en cultures pérennes. Mais les diminutions de la MOS dans ces régions seront davantage imputées aux réchauffements climatiques qu'aux changements de cultures. La texture des sols jouera également un rôle important. La teneur en MOS étant généralement plus faible dans les sols sableux que dans les sols argileux, la dynamique du climat futur les affectera différemment selon les régions. Les différents scénarios d'émissions de GES anticipés (RCP 4.5 et 8.5) auront moins d'effet et du même ordre de grandeur que le changement d'affectation des terres de cultures pérennes en annuelles. L'application du modèle climatique à l'échelle ontarienne n'a pas produit de résultats aussi crédibles en termes de teneurs initiales en MOS et d'évolution de MOS en raison des domaines climatiques mal représentés et des sources de données différentes pour l'évaluation du ratio de cultures pérennes.

Plusieurs travaux additionnels pourraient être réalisés pour améliorer la capacité prédictive des modèles dont l'inclusion d'autres ratios de cultures comme le soya reconnu pour laisser peu de résidus aux sols et qui a pris de l'expansion ces dernières années dans les régions plus froides à mesure que le climat se réchauffe. Des tentatives pour désagréger les analyses de sols relevées à l'échelle des municipalités (travaux de T.H. Easher de l'équipe d'Asim Biswas) ont été réalisées avec succès et les travaux pourraient être poursuivis soit pour associer les teneurs en MOS désagrégées à des grilles plus fines (250 m) aux unités cartographiques des cartes de sols (polygones), ou pour les associer à des covariables du MNT, ou du climat. Finalement, la calibration du modèle dans les conditions plus chaudes actuelles de l'Ontario permettrait de simuler des conditions plus chaudes rencontrées en climat futur au Québec avec plus d'assurance. Sur ce dernier point, l'incertitude des projections n'a pas été mesurée et rapportée. Seule la variabilité des 11 modèles de simulation du climat futur par RCP nous permet de mesurer l'incertitude générée par le climat futur. Le calcul d'indices comme le facteur d'érodabilité des sols K de l'USLE basés sur la MOS et d'autres paramètres comme la texture pourrait aussi conduire à anticiper comment les sols vont évoluer régionalement en termes de risque à l'érosion en climat futur. À ce sujet, les travaux en parcelles expérimentales sous simulateur de pluie ont toutefois démontré que le loam limoneux Le Bras à l'étude avait des valeurs de K nettement en dessous de ce que propose les nomogrammes de Wischmeier, mais plus proches que ce d'autres ont mesuré sous simulateurs de pluie.