



5^e Congrès du RQRAD

Développement d'un SAD pour déterminer le potentiel de MOS et les pratiques pour l'atteindre (appel 2)

Conférenciers :

Yacine Bouroubi, UdeS; Maxime Leduc, MSF; Rami Albasha, Ulaval & MSF;
Hamed Etezadi, McGill; Guy Armel Fotso Kamga, UdeS; Jérémie Durand, UdeS

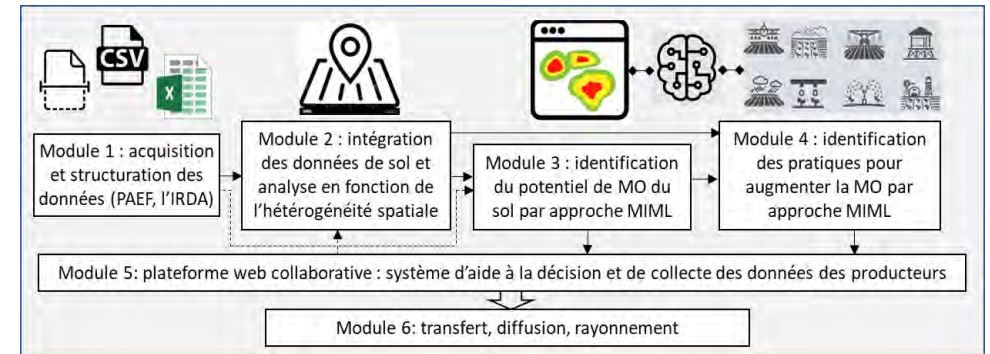
Autres co-auteurs:

Samuel Foucher, Mickaël Germain, UdeS; Marc-Olivier Gasser, IRDA;
Marie-Élise Samson Ulaval; Viacheslav Adamchuk, McGill



Contenu

- Équipe du projet
- Problématique et objectifs
- Méthodologie – Volets/Modules
- Réalisations (par Volet)
 - Volet 1 : Maxime Leduc
 - Volet 2 : Hamed Etezadi
 - Volet 3 : Rami Albasha
 - Volet 4 : Guy Armel Fotso Kamga
 - Volet 5 : Jérémie Durand
- Reste à faire - Activités de transfert
- Retombées attendues
- Questions



Équipe – organismes – expertise – rôle

Co-chercheurs

- Yacine Bouroubi - UdeS : OT, agriculture de précision
- Maxime Leduc - MSF : initiateur de l'idée du projet
- Samuel Foucher - UdeS : OT, IA
- Marc-Olivier Gasser - IRDA : science des sols
- Marie-Élise Samson - ULaval : science des sols
- Slava Adamchuk - McGill : agriculture de précision

PHQ

- Rami Albasha : professionnel de recherche, ULaval
- Hamed Etezadi : PhD, McGill
- Guy Armel Fotso Kamga : PhD & postdoc, UdeS
- Jérémie Durand : MSc, UdeS
- Autres

Collaborateurs

- Mickaël Germain - UdeS: infrastructure géospatiale
- Joann Whalen - McGill

Partenaires utilisateurs (collaborateurs)

- **Mon Système Fourrager MSF : Maxime Leduc**
- IRDA : Marc-Olivier et équipe
- LogiAg : Charles Nault et équipe

Transfert (collaborateurs)

- Louis Dionne - CRAAQ
- Marie-Claude Lapierre – Coord. services-conseils

Problématique et objectifs

- Maintenir et améliorer la MOS au Québec → priorité du PAD
- Connaître le potentiel de MOS de chaque parcelle → problème complexe (sol-climat-culture- ...)
- Connaître les meilleures pratiques pour atteindre ce potentiel pour chaque parcelle → autre problème complexe (pratiques de conservation des sols : travail réduit, cultures de couverture, rotations, chaulage, fertilisants organiques, etc.)
- Spatialiser les mesures de MOS : échantillons → cartes : variations intra-parcellaires



But : développer un SAD pour estimer le potentiel de MOS et les meilleures pratiques pour l'atteindre selon les conditions spécifiques à chaque parcelle en s'appuyant sur les connaissances scientifiques (modèles mécanistes), les données massives, les méthodes d'IA, et des technologies numériques de diffusion.

Objectifs spécifiques

- Volet 1 • Collecte, intégration et structuration des données multi-sources : données terrain (BDs Logiag, IRDA, ...).
- Volet 2 • Spatialisation → cartes de MOS produites par modélisation impliquant des données géospatiales.
- Volet 3 • Création d'un jeu de données massif à l'aide d'un simulateur (DNDC) → dynamique de la MOS en fonction des variables de sol (texture, physico-chimiques, ...), topographie, cultures, pratiques de gestion, climat, ...
- Volet 4 • Approche d'IA de type *Model-Informed Machine Learning* → entraînement sur le jeu de données massif simulé et *fine-tuning* sur les données réelles → recommandations (potentiel et pratiques) adaptées aux conditions des parcelles (choix des variables) et à la diversité du système agricole québécois.
- Volet 5 • Réalisation d'un SAD sous forme de plateforme cartographique : *back-end* (données géospatiales structurées) et *front-end* (interface usager).
- Volet 6 • Diffusion aux partenaires-utilisateurs (MSF, Logiag, IRDA) et au public (CRAAQ, Coord. services-conseils, presse).

Méthodologie : six Volets

- Six Modules / Volets

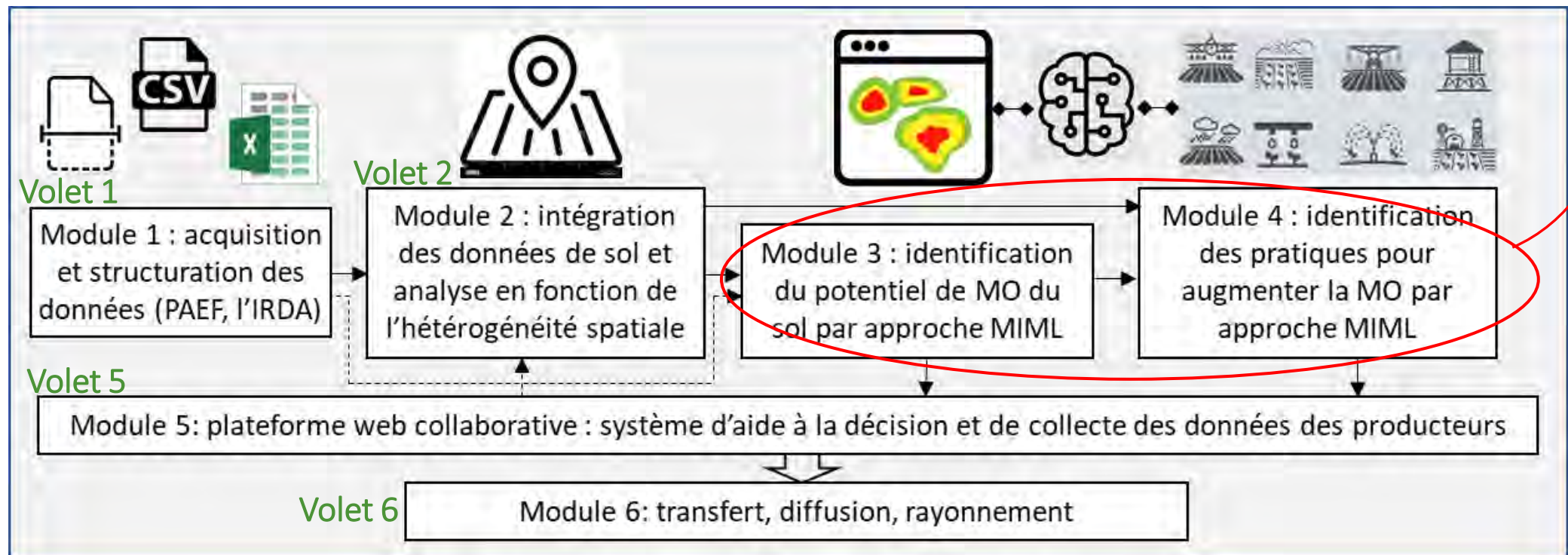
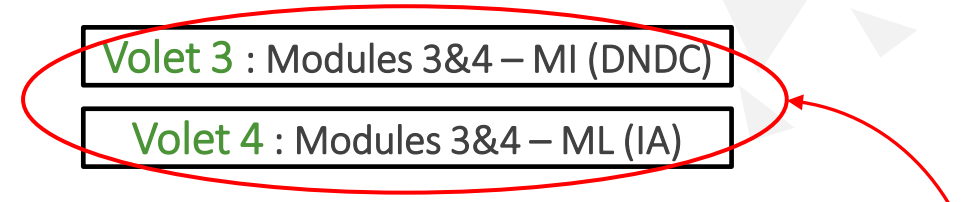
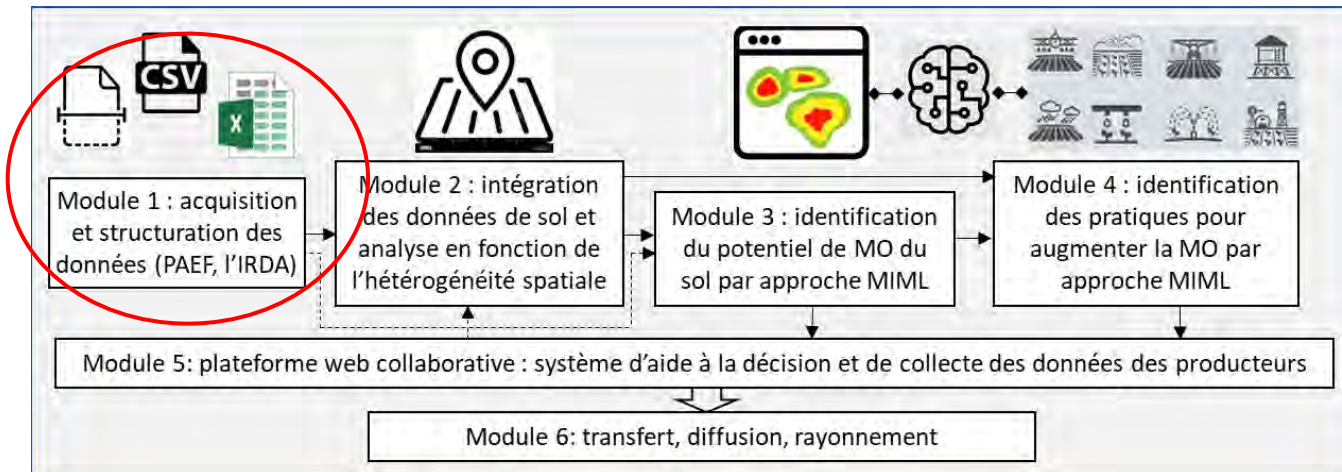


Schéma conceptuel de départ (proposition)

Réalisations – Volet 1

Collecte, intégration et structuration des données multi-sources



Dr Maxime Leduc, MSF, co-chercheur et partenaire

BDs : accès, prétraitements, structuration

- Données de producteurs
- LogiAg
- IRDA : données de recherche
- PAEF : FADQ

Réalisations – Volet 1

Collecte, intégration et structuration
des données multi-sources

Maxime



Réalisations – Volet 1 (Maxime)

Collecte, intégration et structuration des données multi-sources

Sources de données multi-fermes

Trois bases de données complémentaires à intégrer

Donnée de producteurs

- Analyses de sol géoréférencées
- Texture, MO, pH, CEC, éléments
- Coordonnées GPS des échantillons
- Couverture : Québec agricole

Format : fichiers structurés

IRDA

- Données expérimentales de sols
- Parcelles de recherche
- Historiques de cultures et pratiques
- Données physico-chimiques détaillées

Format : bases de données internes

PAEF (FADQ)

- Plans Agro-Environnementaux
- Analyses de sol par parcelle
- Applications fumier et engrais
- Cultures déclarées

Format : documents PDF

Objectif : Unifier ces 3 sources dans une base de données PostgreSQL structurée, nettoyée et géoréférencée

Réalisations – Volet 1 (Maxime)

Collecte, intégration et structuration des données multi-sources

Enjeux de la collecte multi-sources

Défis rencontrés lors de l'intégration des données

Formats hétérogènes

- CSV, Excel, bases propriétaires, PDF scannés et numériques
- Colonnes contenant plusieurs informations à splitter
- Structures de tables différentes entre Logiag et IRDA
- Encodage et séparateurs variables

Données manquantes ou incomplètes

- Cultures non renseignées dans certains PAEF
- Analyses de sol partielles (paramètres absents)
- Historiques de pratiques incomplets
- Coordonnées GPS absentes ou imprécises

Unités et nomenclatures incohérentes

- Unités de mesure variables entre sources (% , g/kg, ppm)
- Noms de cultures non standardisés
- Identifiants de parcelles incompatibles
- Systèmes de coordonnées différents

PAEF en format PDF

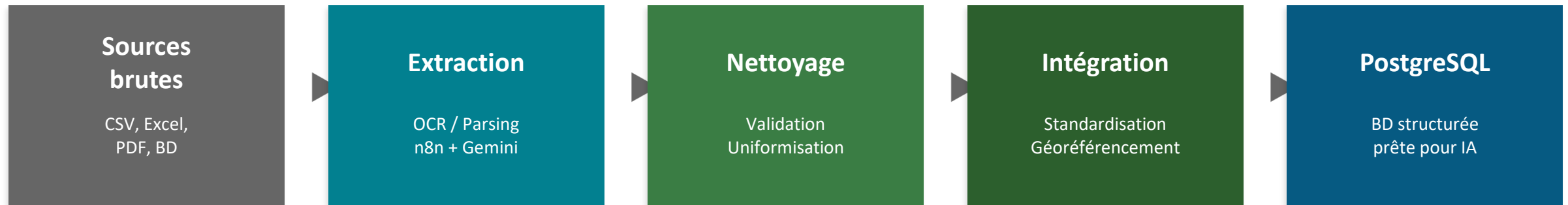
- Mix de PDFs numériques et scannés (images)
- Tableaux complexes avec fusions de cellules
- Mise en page variable selon les agronomes
- Extraction automatisée nécessaire (OCR + IA)

Réalisations – Volet 1 (Maxime)

Collecte, intégration et structuration des données multi-sources

Pipeline de traitement des données

Architecture générale du flux de données



Donnée de producteurs : données récupérées et prétraitées

IRDA : ententes finalisées, accès aux données

PAEF : pipeline n8n automatisé avec Vertex AI / Gemini

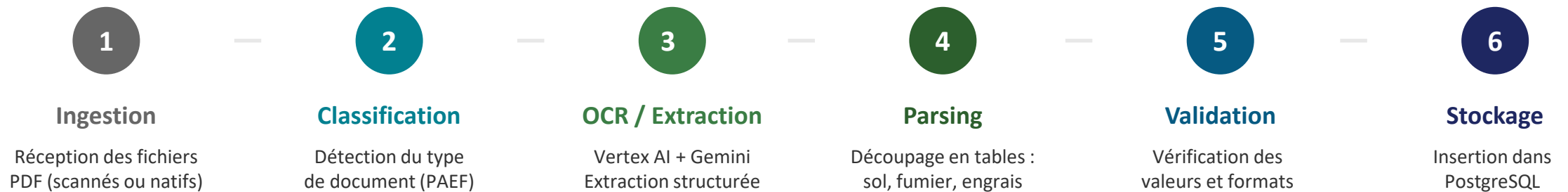
Réalisations – Volet 1 (Maxime)

Collecte, intégration et structuration des données multi-sources

Pipeline n8n pour les PAEF

Extraction automatisée des données depuis les documents PDF

Workflow n8n – Traitement automatisé des PAEF



Technologies utilisées

- n8n** : Orchestration du pipeline (open-source, auto-hébergeable)
- Vertex AI / Gemini** : OCR et extraction intelligente des tableaux PDF (scannés et natifs)
- PostgreSQL** : Stockage relationnel avec schéma standardisé
- Code disponible** : Node n8n reproductible et partageable pour recréer le pipeline

Réalisations – Volet 1 (Maxime-Jérémie)

Collecte, intégration et structuration des données multi-sources

Pour les données géospatial open source :

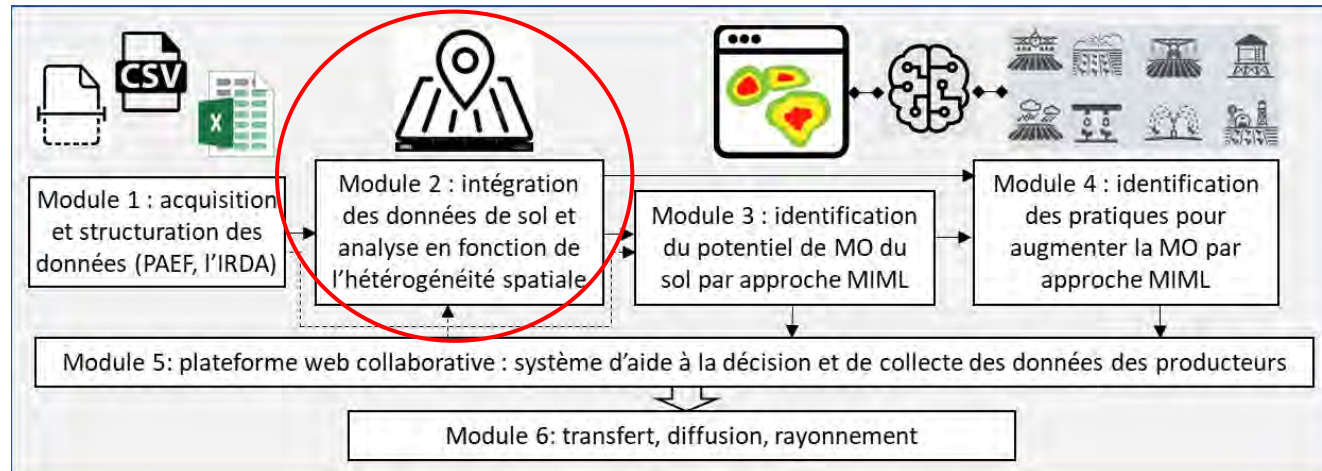
- Propriétés des sols, topographie, hydrologie, couverture du sol, imagerie satellite et données climatiques
- Collecte, prétraitement et harmonisation



Sera présenté par Jérémie Durand dans le volet 5 sur la plateforme cartographique

Réalisations – Volet 2

Spatialisation de la MOS



Hamed Etezadi, PhD supervisé par V. Adamchuk, M-O Gasser et Y. Bouroubi

- Article 1 : revue
- Article 2 : interpolation mathématique de la MO à partir de la texture
- Article 3 : ajout d'images satellites (soumis)
- Article 4 : aspects technologiques pour l'implémentation opérationnelle (en cours)

Réalisations – Volet 2

Spatialisation de la MOS

Hamed



American Society of Agricultural and Biological Engineers
2950 Niles Road | St. Joseph, MI 49085-9659 | USA
269.429.0300 | fax 269.429.3852 | hq@asabe.org | www.asabe.org

An ASABE Meeting Presentation
DOI: <https://doi.org/10.13031/aim.202500179>
Paper Number: 2500179

A Novel Method for Differentiating Soil Organic Matter in a field Using Soil Survey Maps

Hamed Etezadi¹, Viacheslav Adamchuk¹, Yacine Bouroubi², Maxime Leduc³, Marc-Olivier Gasser⁴, David Tittley-Peloquin¹



Quantifying intra-field soil variability using categorical data: A case study of predicting soil organic matter using soil survey maps

Hamed Etezadi^{*}, Viacheslav Adamchuk^{*}, Yacine Bouroubi^{*}, Maxime Leduc^{*}, Marc-Olivier Gasser^{*}, David Tittley-Peloquin^{*}

A Multi-Scale Framework for Predicting Continuous Soil Properties Using Ranked Sentinel-2 Images and Zone-Level Soil Data: From a Farm Case Study to a Regional Application.

Hamed Etezadi; Yacine Bouroubi; Viacheslav Adamchuk; Maxime Leduc; Marc-Olivier Gasser

Submitted to : Smart Agricultural Technology

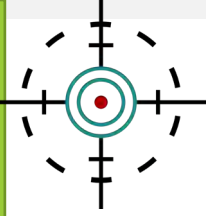
Development of a Multi-scale SOM Prediction Framework



Hamed Etezadi

Department of Bioresource Engineering
McGill University





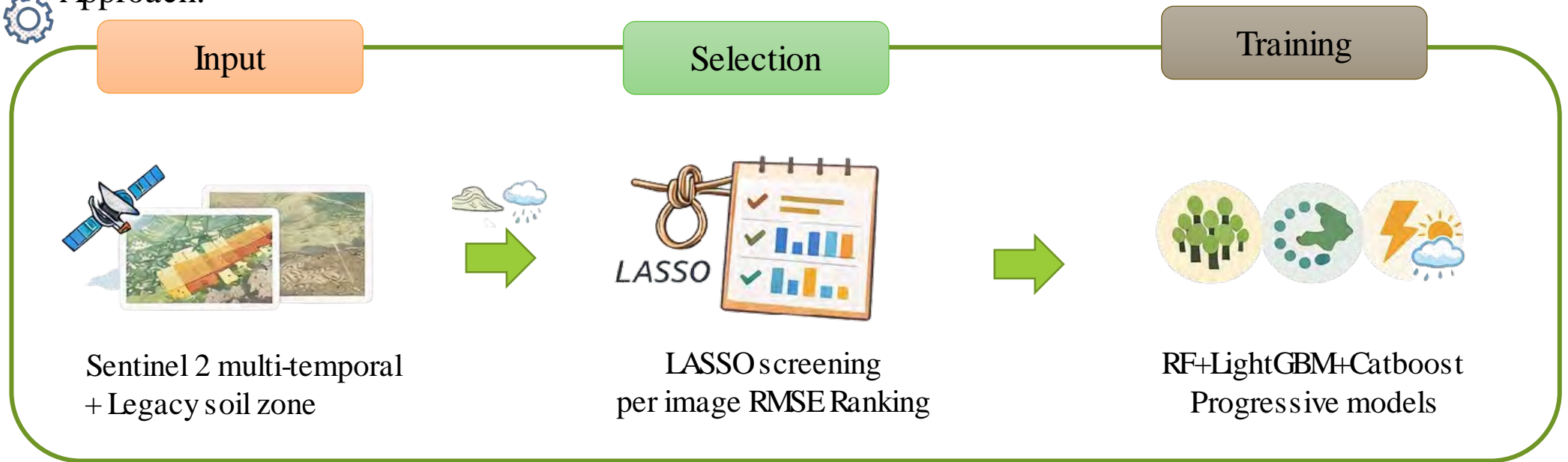
Multi-scale SOM prediction (Farm → Regional)



Rank images by predictive value, not by quantity



Approach:



Expected Outcome:

A scalable and adaptive SOM prediction model



Sentinel-2 preprocessing: snow-free & bare-soil pixels:

Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



2019 to 2023



Cloud <1%



Spring & Fall

Snow Masking (NDSI)



Threshold: > 0.42



Bare Soil Masking



$BSI_{pixel} > BSI_{mean_field}$



Clean Soil Signal

Votre logo ou nom ici

18



Feature Engineering:

Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



Soil signal
(Bare soil condition)



- Surface brightness & color
- Iron/redness response
- Organic matter sensitivity

BI . CI . RI . OMI . SI



Complementary
signals from
the same pixels

Vegetation context
(residual cover)

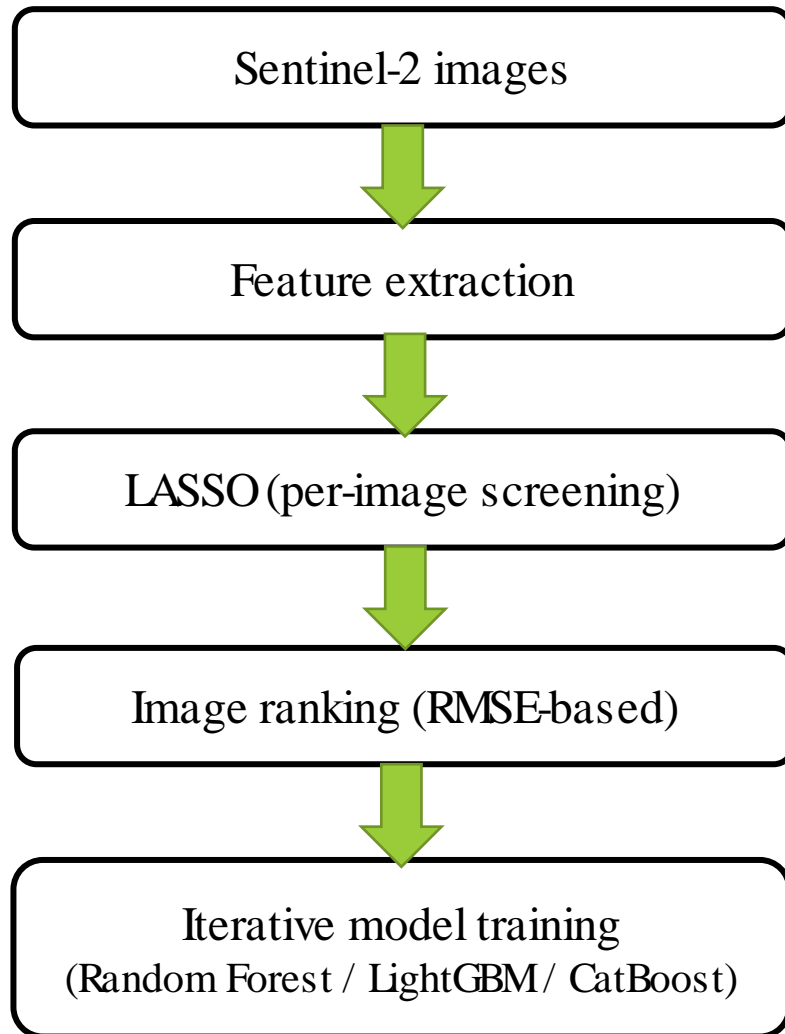


- Moisture condition
- Canopy influence
- Soil-vegetation contrast

NDVI . NDMI . EVI . SAVI



Modeling Framework



- Image-level informativeness (not pixel-level)
- Progressive inclusion of best images
- Model-agnostic ranking strategy
- Designed for scalability

Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation





Farm-scale



Regional / Province-scale



Two Evaluation Levels

Which images are informative?



Does it generalize to unseen fields?



S1 : RS+ Soil



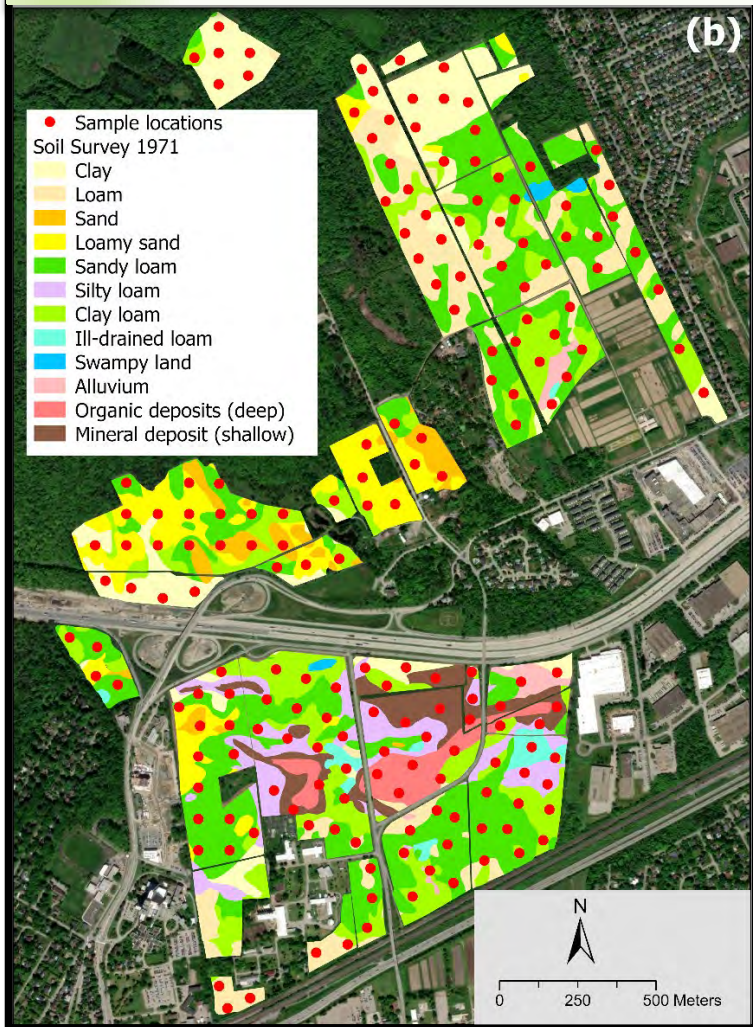
S2 : S1+ Topography



S3 : S2+ Climate

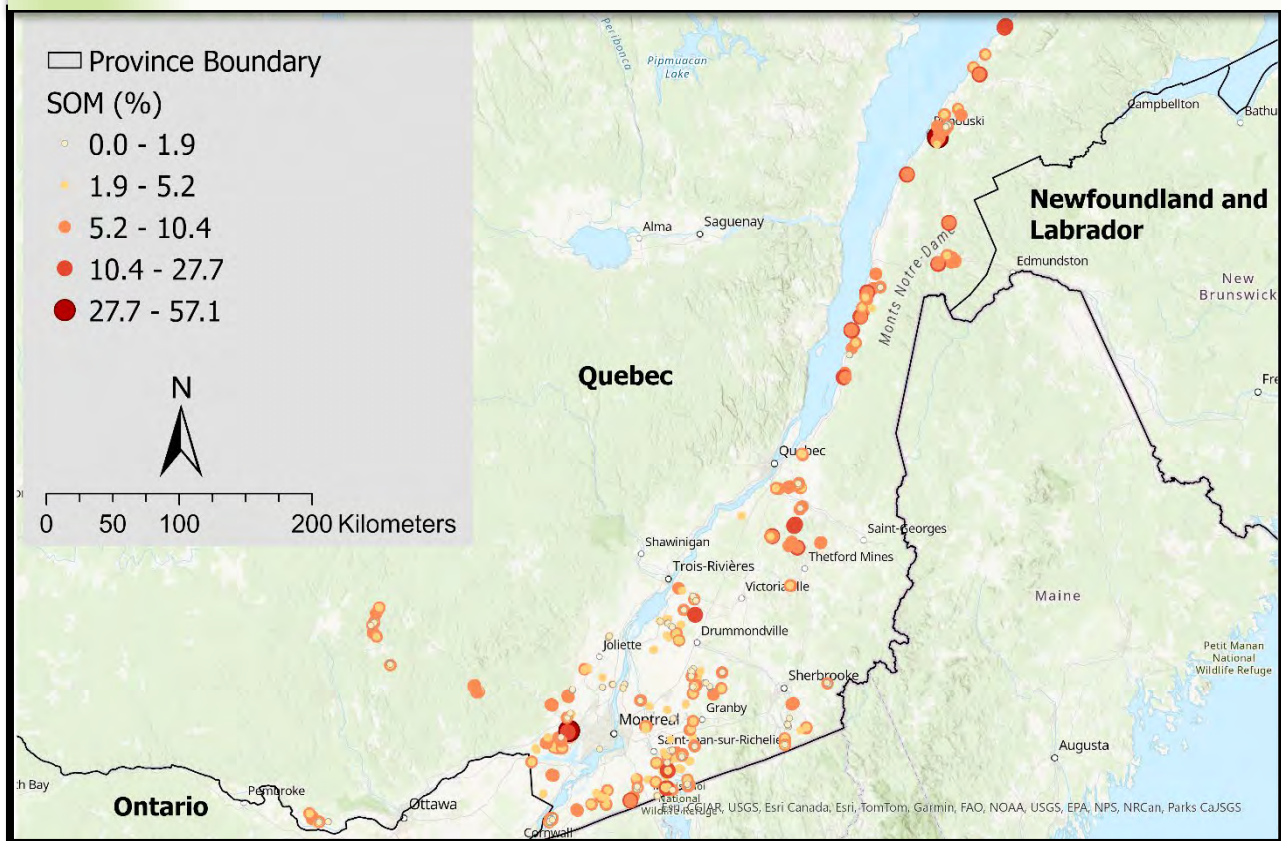


Field scale case study (Method validation)



Scaling up

Regional scale application



Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



Farm-scale SOM prediction (Case Study)

Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation

Baseline
(Field avg)



RMSE ≈ 0.14

$R^2 \approx 0.74$

RS only



RMSE $\approx 0.35-0.37$

$R^2 < 0.1$

Hybrid (RS + Soil)
(Zone level)



RMSE ≈ 0.16

$R^2 \approx 0.83$

Soil texture/zone information is essential; RS-only fails to capture subsurface SOM variability



Introduction

Preprocess

Feature Eng.

Methodology

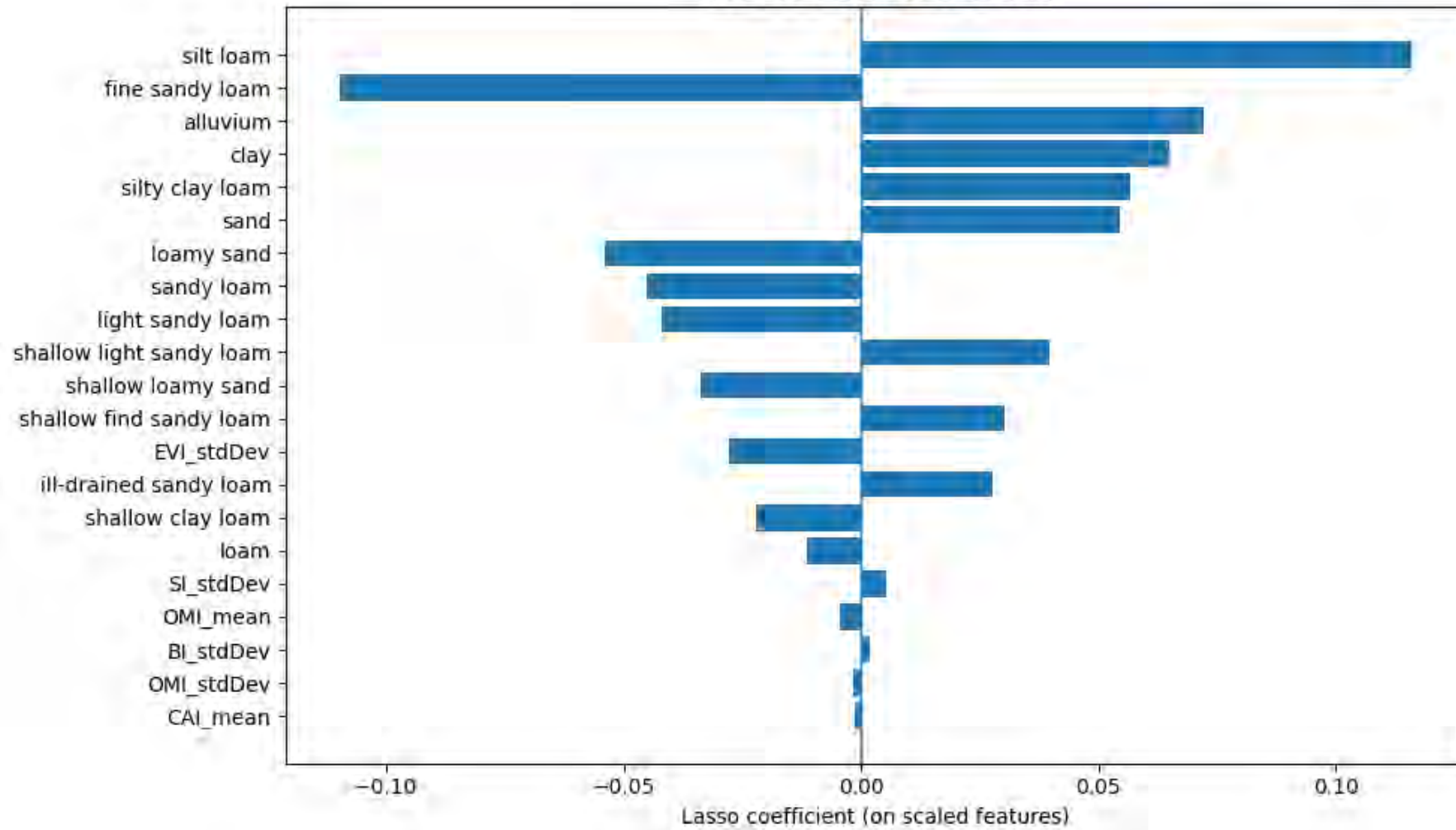
Farm-Scale

Québec-wide

Validation



Lasso Selected Features (All)



Votre logo ou nom ici



Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



RGB Sentinel-2 image (Best Image)



SI (Best LASSO Feature)



om ici



Introduction

Preprocess

Feature Eng.

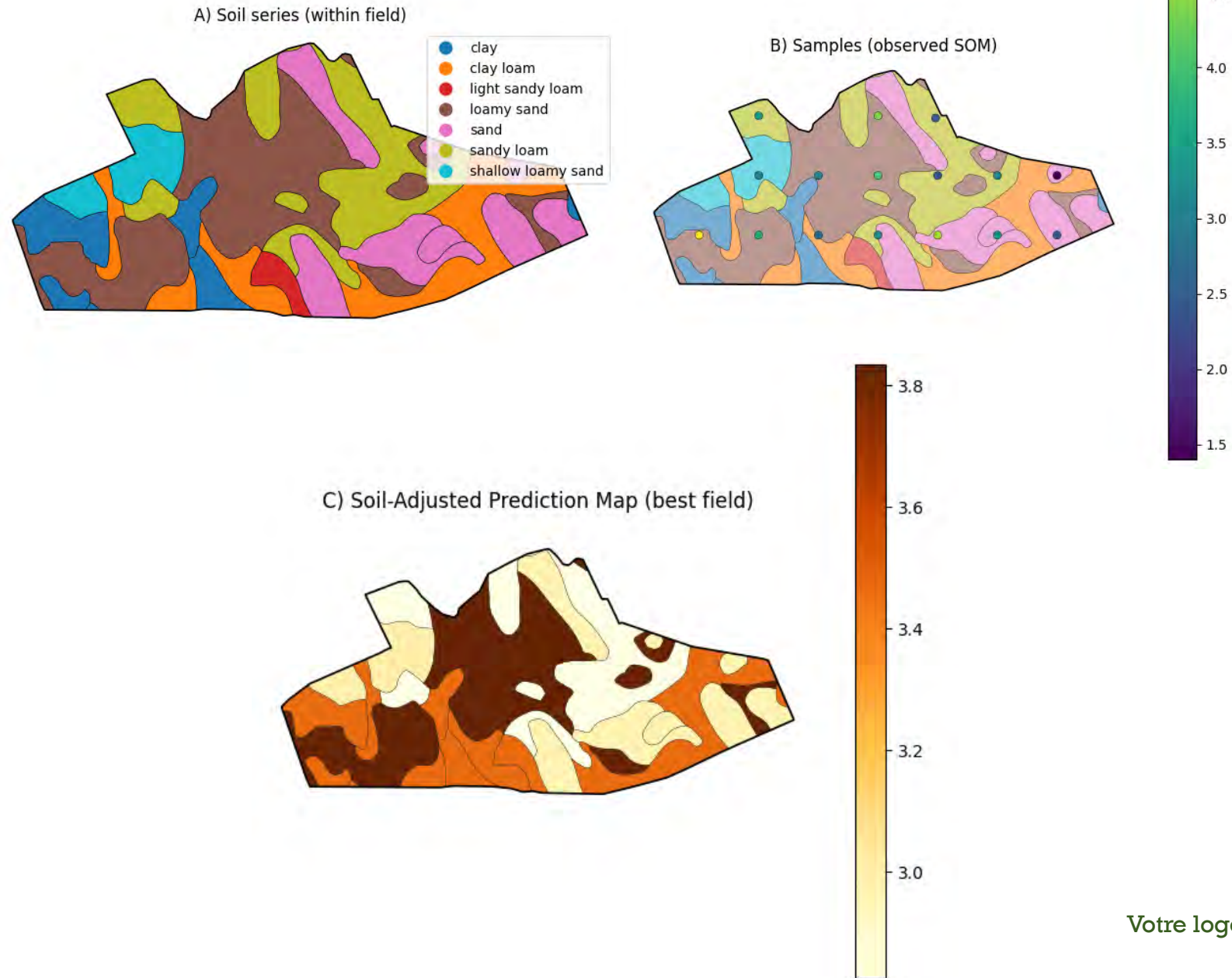
Methodology

Farm-Scale

Québec-wide

Validation

Soil Deviations from Field SOM Prediction



Spatial validation (Pattern preservation)

Introduction

Preprocess

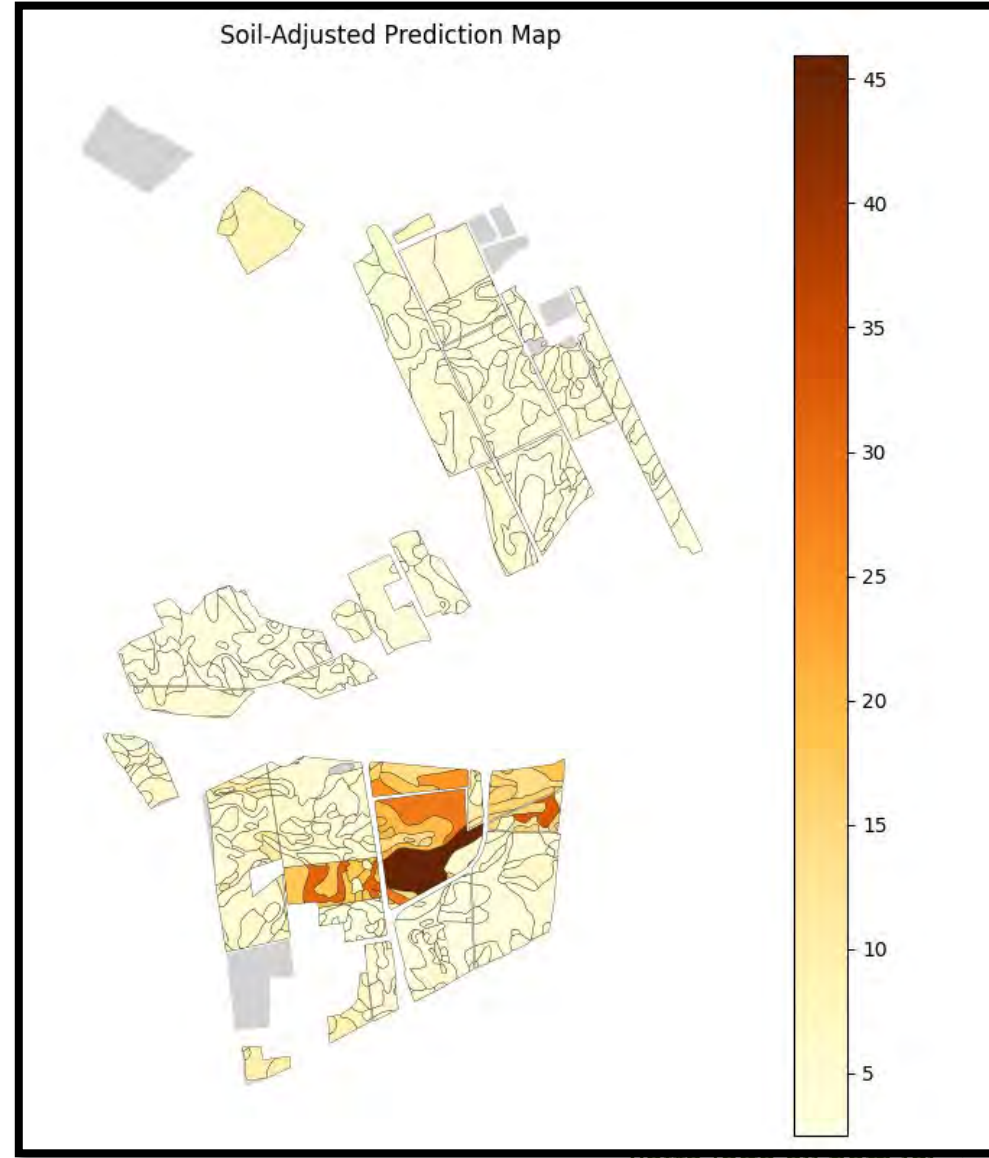
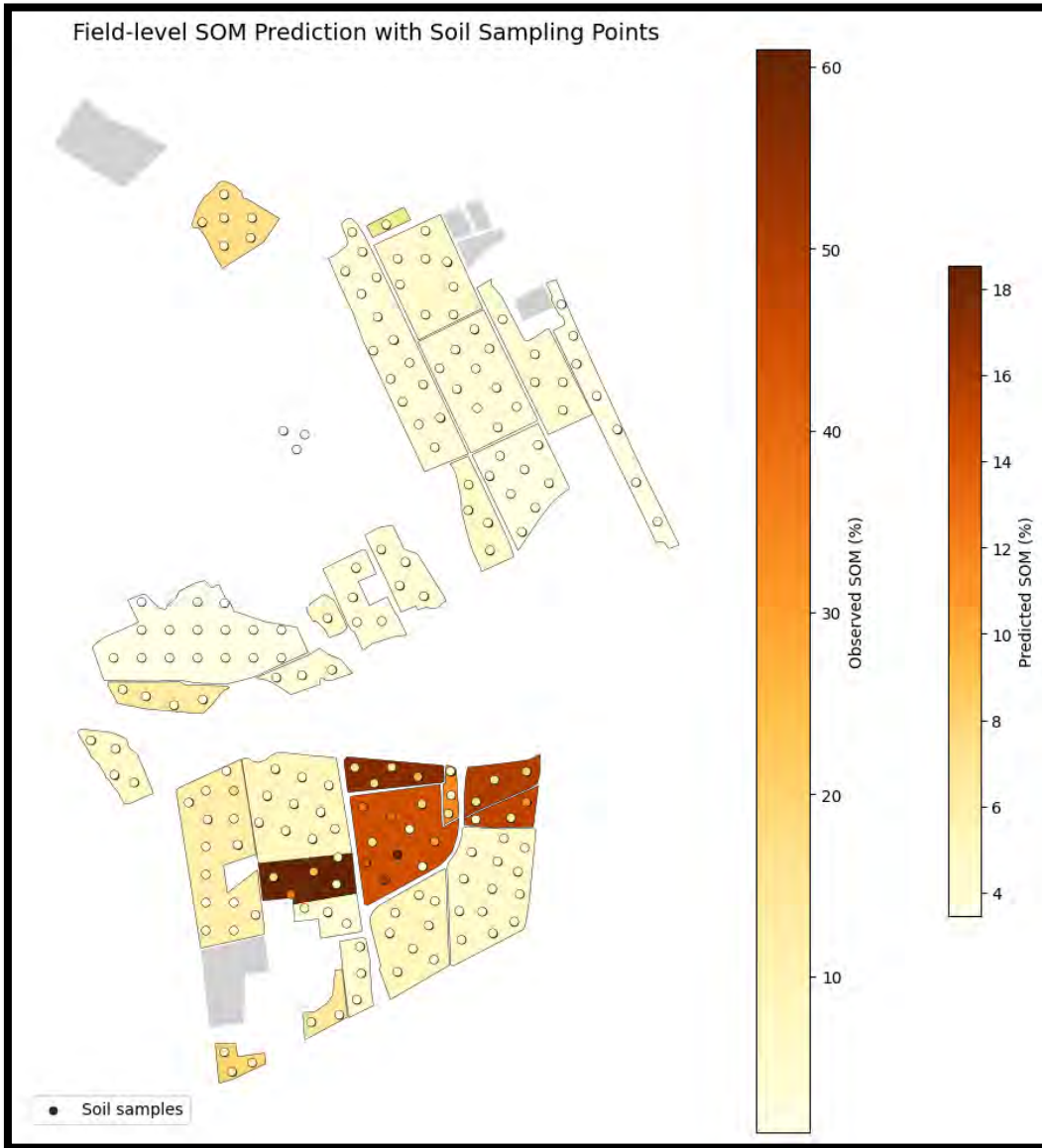
Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



voire logo ou nom ici

Spatial uncertainty (Model confidence)

Introduction

Preprocess

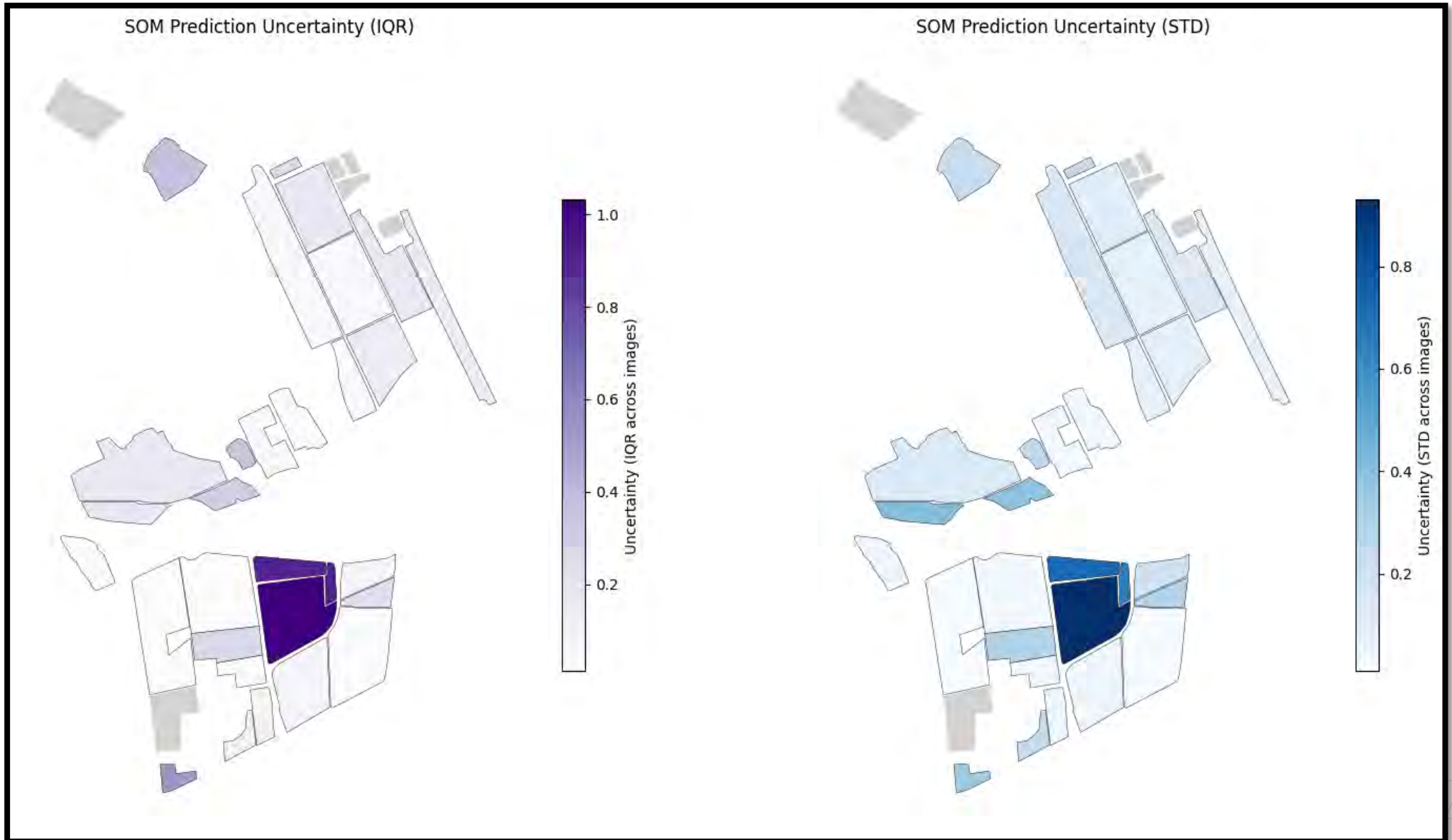
Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



Regional-scale result (Québec-wide)

A) Algorithm comparison (Image level vs Field-level)



Random Forest

Image:	RMSE(log) 0.113	R ² (log) 0.131
Field:	RMSE(log) 0.111	R ² (log) 0.218

LightGBM

Image:	RMSE(log) 0.118	R ² (log) 0.050
Field:	RMSE(log) 0.111	R ² (log) 0.217

Catboost

Image:	RMSE(log) 0.116	R ² (log) 0.083
Field:	RMSE(log) 0.114	R ² (log) 0.176

B) Scenario Progression (Random forest)

S1 : RS+Soil

Image:	RMSE(log) 0.113	R ² (log) 0.131
Field:	RMSE(log) 0.111	R ² (log) 0.218

S2: S1+Topography

Image:	RMSE(log) 0.113	R ² (log) 0.137
Field:	RMSE(log) 0.110	R ² (log) 0.239

S3: S2+Climate



Image:	RMSE(log) 0.102	R ² (log) 0.287
Field:	RMSE(log) 0.100	R ² (log) 0.364

Introduction

Preprocess

Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



Validation & Robustness (Regional scale)

Introduction

Preprocess

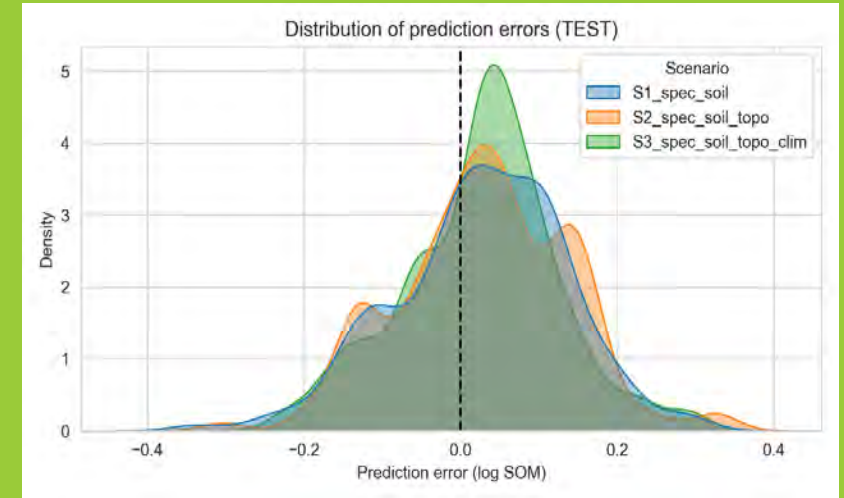
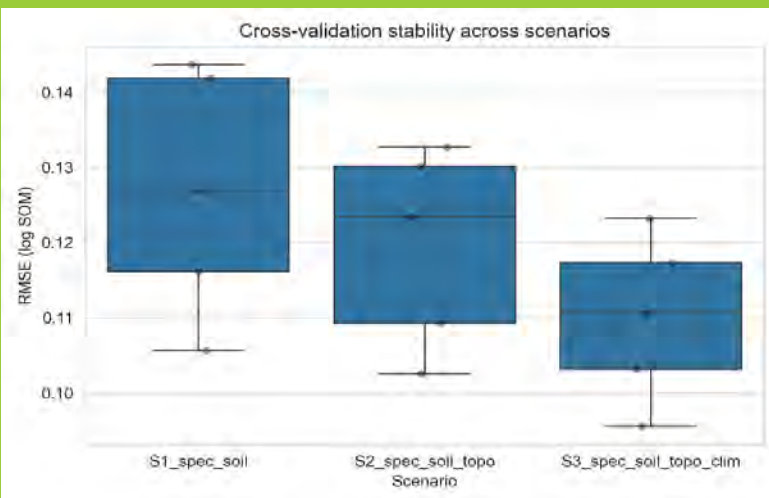
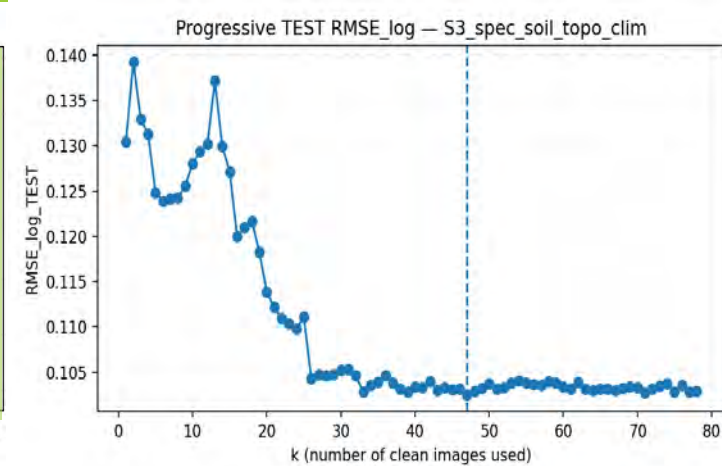
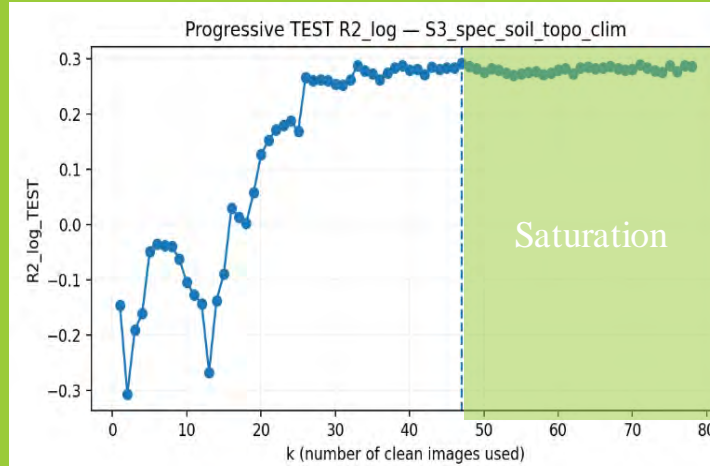
Feature Eng.

Methodology

Farm-Scale

Québec-wide

Validation



Thank you.



**Looking
For An
Answer?**

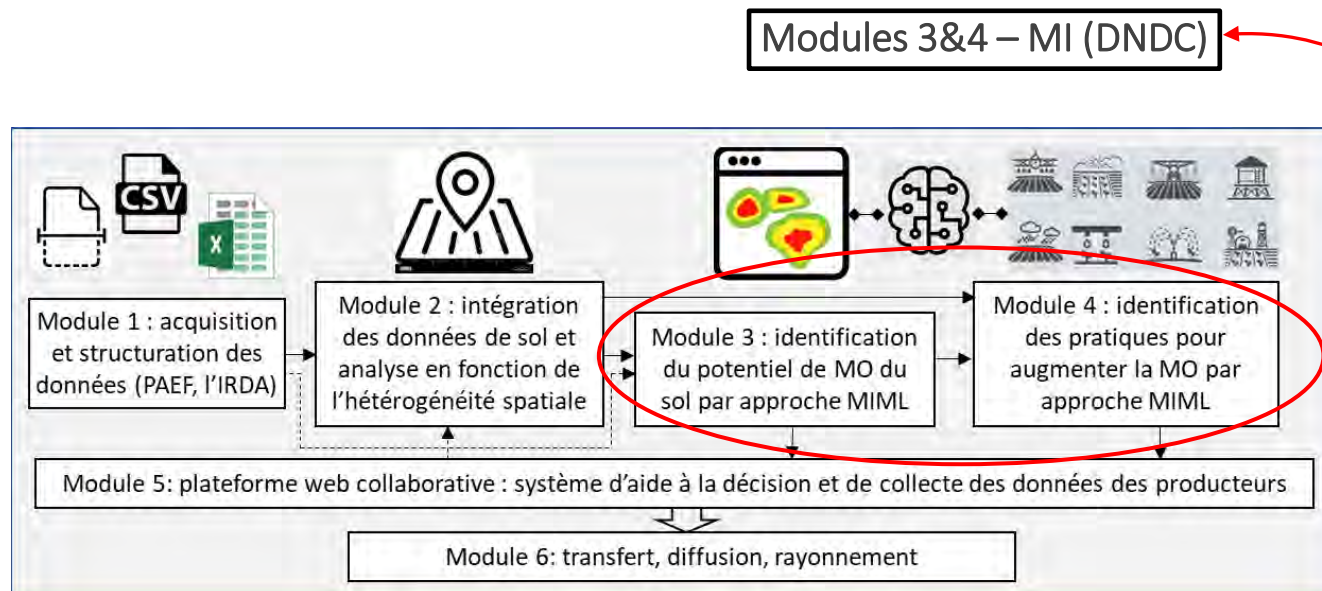


Réalisations – Volet 3 : DNDC

Création d'un jeu de données massif à l'aide d'un simulateur de MOS

Dr Rami Albasha, profess. de rech., supervisé par Pre M-É Samson; co-supervisé par M-O Gasser, M. Leduc et Y. Bouroubi

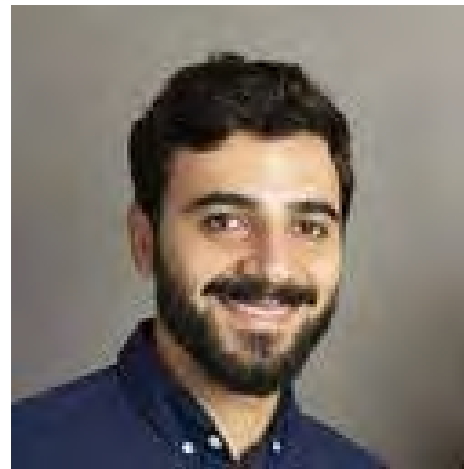
- Choix du modèle mécaniste DNDC pour produire des simulations de la dynamique de la MO
- Choix des variables d'input et d'output et des scénarii de combinaison



Réalisations – Volet 3 : DNDC

Création d'un jeu de données massif à l'aide d'un simulateur de MOS

Rami



Réalisations – Volet 3 : DNDC (Rami)

Méthodologie de la génération des données

Rappel du contexte

- Objectif :
 - Produire un jeu de données massif pour entraîner un réseau de neurones (Model-Informed Machine Learning, MIML).
- Etendue géographique :
 - Surfaces cultivées du Québec.

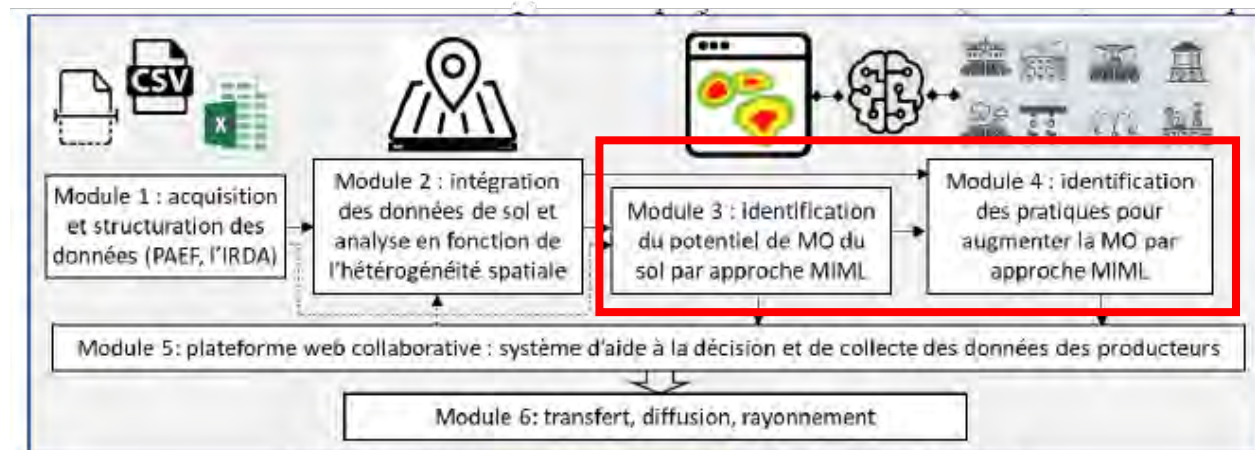


Figure 1. Diagramme illustrant la méthodologie du projet

Réalisations – Volet 3 : DNDC (Rami)

- Défis :
 - Scientifique : choisir le modèle le plus adéquat
 - Logique : créer des scénarios qui ont du sens
 - Technique : générer/gérer plusieurs (centaines de) milliers de simulations

Réalisations – Volet 3 : DNDC (Rami)

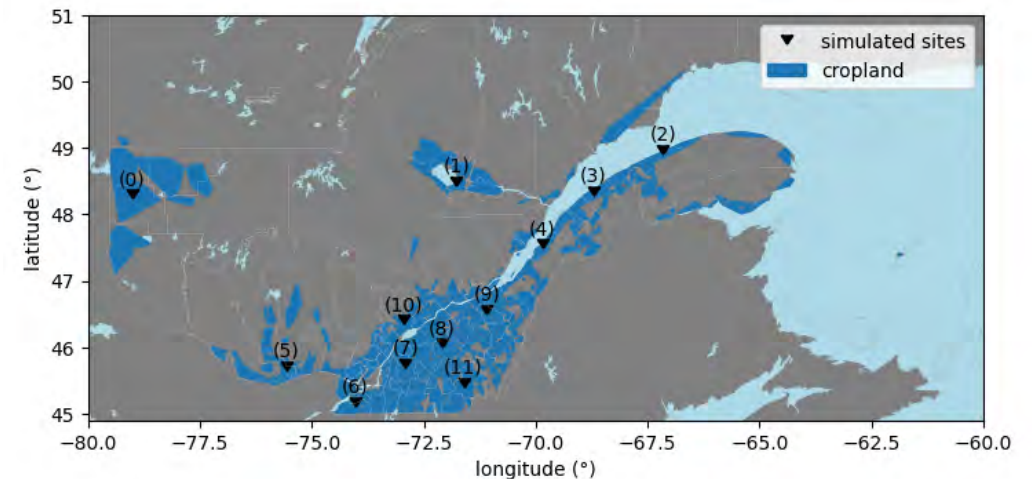
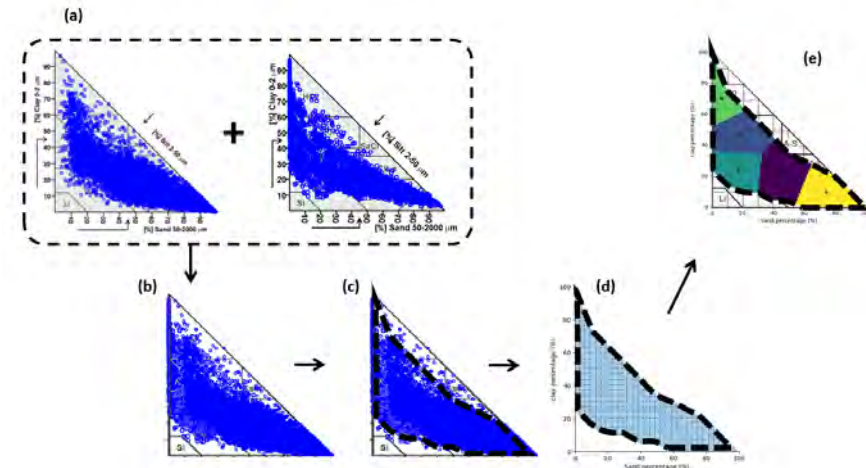
Modèle	Pas de temps	Spatialisation	Langage	GUI	Canada	Physiologie des plantes	Effet neige	Effet gel	Cycles sol	Nb de params	Notes
OGEMOS	• Jour • Année (outputs)	Oui	Python	Oui	Oui	Non	?	?	C	?	<ul style="list-style-type: none"> Utilise le modèle minimaliste CBM • Spatialisation du modèle CBM • Analyse de sensibilité: https://link.springer.com/article/10.1007/s10705-007-9156-9 • Analyse de sensibilité: https://doi.org/10.2319/isspacep07.2ed.c2 • Analyse de sensibilité: doi:10.1016/j.ecolmodel.2010.04.019 • Cultures & forêts • https://link.springer.com/article/10.1007/s10584-008-9493-y
CENTURY5 DayCent5	mois jour	CSMPI	Fortran, C++	?	not for croplands	Oui	Oui	Oui	C, N, P, K...	*150	<ul style="list-style-type: none"> • No simple pas d'érosion • analyse de sensibilité déjà fait (https://www.sciencedirect.com/science/article/pii/S136481523002189) • doc: https://www2.urel.colostate.edu/projects/nc/public/Documents/Software/Century5/Reference/ • bilé de printemps Canada Est (Dr. Bruno, St. Jean, and Ottawa) https://link.springer.com/article/10.1007/s10705-017-9880-8 • bilé de printemps dans l'Est (Dr. Bruno, St. Jean, and Ottawa) http://dx.doi.org/10.1016/j.fcr.2013.11.010 • une version alpha est des https://www.scribd.com/document/374014034/History • décrit en plusieurs versions https://fao-gsp.github.io/GSDCaq/loc-modelling.html#process-oriented-models • la version spatialisée sous R -> carte mondiale de la MOD (https://data.apps.fao.org/gloss/?lang=en) • https://www.scribd.com/document/374014034/History
RothC	mois	• Oui (en R) • Oui (Rothamsted Landscape Model)	Fortran77	Oui	Oui	Non	?	?	C, N, P, K...	*150	<ul style="list-style-type: none"> • la version canadienne https://bitbucket.org/ROTHC/CAN/ • comprehensive library of default settings for 62 crops and 12 soil types • très bien documenté dans la version DNDC v. CAN (https://github.com/brunobrunant/DNDCvCAN) • code bien maintenu • la physiologie des cultures est trop simplifiée: moitié du temps en stade de croissance végétative, l'autre moitié en stade reproductif • on n'a pas les composants can be entered, P and s are not actively simulated. • la lixiviation ne semble pas être simulée mécaniquement (déterminée par une fraction de N qui peut se lixivier, pas de mention de flux d'eau, de solubilité, etc.) • il ne semble avoir un bilan d'énergie pour le temps aère du sol • params for wheat, canola and pea in Saskatchewan https://www.sciencedirect.com/science/article/pii/S0959526210379268#appsec3 • bilé de printemps Canada Est (Dr. Bruno, St. Jean, and Ottawa) https://link.springer.com/article/10.1007/s10705-017-9880-8 • bilé de printemps dans l'Est (Dr. Bruno, St. Jean, and Ottawa) http://dx.doi.org/10.1016/j.fcr.2013.11.010 • origine de printemps dans plusieurs régions https://doi.org/10.1002/agr2.21482 • ans en compte de la préférence de drains https://www.sciencedirect.com/science/article/pii/S1364815219304438#tbl1 • érosion simulée
DNDC	jour	Oui (GIS/wcra4)		Oui	Oui, DNDC/CAN	Oui, simplifié	Oui	Oui	C, N, P	*150	<ul style="list-style-type: none"> • la version canadienne https://bitbucket.org/ROTHC/CAN/ • comprehensive library of default settings for 62 crops and 12 soil types • très bien documenté dans la version DNDC v. CAN (https://github.com/brunobrunant/DNDCvCAN) • code bien maintenu • la physiologie des cultures est trop simplifiée: moitié du temps en stade de croissance végétative, l'autre moitié en stade reproductif • on n'a pas les composants can be entered, P and s are not actively simulated. • la lixiviation ne semble pas être simulée mécaniquement (déterminée par une fraction de N qui peut se lixivier, pas de mention de flux d'eau, de solubilité, etc.) • il ne semble avoir un bilan d'énergie pour le temps aère du sol • params for wheat, canola and pea in Saskatchewan https://www.sciencedirect.com/science/article/pii/S0959526210379268#appsec3 • bilé de printemps Canada Est (Dr. Bruno, St. Jean, and Ottawa) https://link.springer.com/article/10.1007/s10705-017-9880-8 • bilé de printemps dans l'Est (Dr. Bruno, St. Jean, and Ottawa) http://dx.doi.org/10.1016/j.fcr.2013.11.010 • origine de printemps dans plusieurs régions https://doi.org/10.1002/agr2.21482 • ans en compte de la préférence de drains https://www.sciencedirect.com/science/article/pii/S1364815219304438#tbl1 • érosion simulée
CANDY											
DAISY	• Heure • Jour	?	• Fortran • C++	Non	?		Oui		C, N, P		<ul style="list-style-type: none"> • version non maintenue • ne semble plus en développement depuis 2015 ou 2016 • Biomasse non simulée -> la demande en azote est fonction, soit une courbe hypothétique (appelée courbe S) soit en fonction de l'évapotranspiration (spécifique aux prairies (grazlands)) • Code maintenu • semble se focaliser sur les problématiques agricoles - e.g. teneur en protéines des cultures céréalières/danaises, dégradation et lixiviation de pesticides, flux d'eau préférentiel sous l'effet de drains. • plusieurs couches de sol sont possibles
ECOSSE	• 30 mins • Heure • Jour • Semaine • Mois	?	Fortran			Non			C, N	*85	<ul style="list-style-type: none"> • basé sur les mêmes formules que RothC et SUNDIAL (deux modèles écosyst) mais se veut plus simple • a vocation d'être appliqué sur de grandes échelles
EPIC	jour		• Fortran • call Python	Oui (Windows)	Oui	Oui (80 crops)			Inspiré de CENTURY	>200	<ul style="list-style-type: none"> • V. Python (https://github.com/rivkashaj/EPIC) • échelle spatiale de la parcelle au bassin versant (pourvu que ce dernier représente une unité homogène). EPIC fonctionne par ailleurs à l'échelle de Hydrologic Landuse Unit (HLU)" • EPIC a son propre modèle de croissance des cultures qui est le même pour ~80 cultures différentes • Accrès mis sur l'érosion et son impact sur les rendements. • Testé au Canada pour la prédiction des rendements (Quart, Ontario, Québec) en 1998 • Testé à Chio lin des années 2000s • Mobile et code ne semblent plus maintenus • Call depuis Python https://groups.google.com/g/zall-rpm-monica/c/VARZCQaf2qI • Succède au modèle Hermes (cycle d'azote sol/plante) et l'ensemble des formalismes par ceux venant du modèle Daisy • Intégré au DSI LandCall • Validé en zones tempérées (Allemagne) et tropicales (Brésil) • Documentation à jour et détaillée • Code maintenu (https://github.com/zall-rpm/monica) • la sol est limité à 2 m de profondeur • 30 à 30 mins (processus courts, e.g. bilan d'énergie) et 1 (processus longs, e.g. LAI) • DROUIDE is the land surface component of the IPSU-CMS Earth System Model • Cutting-edge model • Used for IPCC simulations • voit pre-processing for point simulations (not global) (https://github.com/kashfahmad/IGRR_2021_SW_US_semiard_NEE_optimization/tree/main) • Requière l'utilisation des clusters de calcul de l'institut Pierre Simon Laplace • Trouvé sans hospitalisation agricole • Basé sur l'ICBM • Accrès sur l'estimation des GES, la séquestration de carbone et l'ACV (uel et cie) • la croissance de la culture n'est pas simulée, le rém est une entrée • ETR = le * ETP (FAO) bigleaf PML, ce pour le Canada pour 30 cultures -> pas d'effet de SOC sur la culture • Pour les zones forestières
MONICA	jour		• C++ • call Python		Non	Inspiré de Daisy			Inspiré de Daisy		<ul style="list-style-type: none"> • Call depuis Python https://groups.google.com/g/zall-rpm-monica/c/VARZCQaf2qI • Succède au modèle Hermes (cycle d'azote sol/plante) et l'ensemble des formalismes par ceux venant du modèle Daisy • Intégré au DSI LandCall • Validé en zones tempérées (Allemagne) et tropicales (Brésil) • Documentation à jour et détaillée • Code maintenu (https://github.com/zall-rpm/monica) • la sol est limité à 2 m de profondeur • 30 à 30 mins (processus courts, e.g. bilan d'énergie) et 1 (processus longs, e.g. LAI) • DROUIDE is the land surface component of the IPSU-CMS Earth System Model • Cutting-edge model • Used for IPCC simulations • voit pre-processing for point simulations (not global) (https://github.com/kashfahmad/IGRR_2021_SW_US_semiard_NEE_optimization/tree/main) • Requière l'utilisation des clusters de calcul de l'institut Pierre Simon Laplace • Trouvé sans hospitalisation agricole • Basé sur l'ICBM • Accrès sur l'estimation des GES, la séquestration de carbone et l'ACV (uel et cie) • la croissance de la culture n'est pas simulée, le rém est une entrée • ETR = le * ETP (FAO) bigleaf PML, ce pour le Canada pour 30 cultures -> pas d'effet de SOC sur la culture • Pour les zones forestières
ORCHIDEE	• 30 mins • Jour	Oui (nécessite cluster)	Fortran	Non	Non (pas spécifiquement)	Oui (VICI), BAB et Cie			Utilise CENTURY		<ul style="list-style-type: none"> • 30 à 30 mins (processus courts, e.g. bilan d'énergie) et 1 (processus longs, e.g. LAI) • DROUIDE is the land surface component of the IPSU-CMS Earth System Model • Cutting-edge model • Used for IPCC simulations • voit pre-processing for point simulations (not global) (https://github.com/kashfahmad/IGRR_2021_SW_US_semiard_NEE_optimization/tree/main) • Requière l'utilisation des clusters de calcul de l'institut Pierre Simon Laplace • Trouvé sans hospitalisation agricole • Basé sur l'ICBM • Accrès sur l'estimation des GES, la séquestration de carbone et l'ACV (uel et cie) • la croissance de la culture n'est pas simulée, le rém est une entrée • ETR = le * ETP (FAO) bigleaf PML, ce pour le Canada pour 30 cultures -> pas d'effet de SOC sur la culture • Pour les zones forestières
HOLOS	• Jour (bilan hydr.) • Mois & Année (flux C)			Oui	Oui	Non					<ul style="list-style-type: none"> • basé sur les mêmes formules que RothC et SUNDIAL (deux modèles écosyst) mais se veut plus simple • a vocation d'être appliqué sur de grandes échelles
GCBM DSSAT (CERES)	jour	Oui	Oui (CRAFT_pytha)		Oui	Oui	Oui (d'après Jégo et al., 2014 pour STICS)		N, P, K, SOM (CENTURY side)		<ul style="list-style-type: none"> • Utilise CENTURY SOM model (optionnel) • DSSAT: The CERES Regional Agricultural Forecasting Toolbox, ensemble modeling approach (ESAT, APIM, SAPIM, SAPIM-H) • MINK à l'échelle globale, par IFPRI https://eibrary.ifpri.org/collec/collec/15738c02/4213406/FileName/131617.pdf • pytha: utilise python pour diriger et paralléliser DSSAT, ne semble pas avoir été trop testé (https://github.com/DSSAT/pytha/tree/main) • DSSAT à l'échelle globale (code non accessible, car disponible sur demande https://models.org/receiv/1403171) • params for wheat, canola and pea in Saskatchewan (with Current) https://www.sciencedirect.com/science/article/pii/S0959526210379268#appsec3 • params pour canola à Ontario (West Niagara) https://www.scribd.com/document/374014034/History • params pour blé d'hiver et de printemps à Ottawa et Québec https://www.scribd.com/document/374014034/History • origine de printemps dans plusieurs régions https://doi.org/10.1002/agr2.21482 • params pour blé de printemps à Saskatchewan (Swift Current & Melton) https://www.scribd.com/document/374014034/History • les modules C & N sont propres à APSIM (pas pris de CENTURY ou autres modèles de ref) • le cycle d'azote est partiellement inspiré de CERES-Maize (DSSAT)
APSIM-X (APSIM obsolète)	jour	apsimRegions(pytho2.7)	• C# NET • VB.NET • R (version non officielle)		non	Oui	?	?	C, N, P...		<ul style="list-style-type: none"> • Utilise CENTURY SOM model (optionnel) • DSSAT: The CERES Regional Agricultural Forecasting Toolbox, ensemble modeling approach (ESAT, APIM, SAPIM, SAPIM-H) • MINK à l'échelle globale, par IFPRI https://eibrary.ifpri.org/collec/collec/15738c02/4213406/FileName/131617.pdf • pytha: utilise python pour diriger et paralléliser DSSAT, ne semble pas avoir été trop testé (https://github.com/DSSAT/pytha/tree/main) • DSSAT à l'échelle globale (code non accessible, car disponible sur demande https://models.org/receiv/1403171) • params for wheat, canola and pea in Saskatchewan (with Current) https://www.sciencedirect.com/science/article/pii/S0959526210379268#appsec3 • params pour canola à Ontario (West Niagara) https://www.scribd.com/document/374014034/History • params pour blé d'hiver et de printemps à Ottawa et Québec https://www.scribd.com/document/374014034/History • origine de printemps dans plusieurs régions https://doi.org/10.1002/agr2.21482 • params pour blé de printemps à Saskatchewan (Swift Current & Melton) https://www.scribd.com/document/374014034/History • les modules C & N sont propres à APSIM (pas pris de CENTURY ou autres modèles de ref) • le cycle d'azote est partiellement inspiré de CERES-Maize (DSSAT)
STICS	jour	Non (pas d'outil dédié)	• C (STICS)	non	Oui	Oui	Oui (Snow Maus)	Oui (Snow Maus)	C, N, P...		<ul style="list-style-type: none"> • bilé de printemps Canada Est (Dr. Bruno, St. Jean, and Ottawa) https://link.springer.com/article/10.1007/s10705-017-9880-8 • bilé de printemps dans l'Est (Dr. Bruno, St. Jean, and Ottawa) http://dx.doi.org/10.1016/j.fcr.2013.11.010 • soja et blé de printemps 10.1016/j.fcr.2010.03.008 (Jégo et al., 2010) • origine de printemps dans plusieurs régions https://doi.org/10.1002/agr2.21482 • STICS & Sentinel https://link.springer.com/article/10.1007/s11098-012-0097-w • STICS spatialisation in Canada https://www.sciencedirect.com/science/article/pii/S0303243414000926 • STICS spatialisation in Canada https://doi.org/10.1016/j.fcr.2013.11.010
SWAP SWAT(+) Bilan Humique AMG	SOC not simulated Annuel	Oui					Oui (sur l'évaporation du sol)		C, N, P		<ul style="list-style-type: none"> • Les principes de modélisation sont les mêmes que dans EPIC • pas d'effet du gel sur la mortalité des plantes • https://www.mspq.gov.ca/fr/productions/agroenvironnement/territoires/2pages/bilanhumique.aspx • https://www6.hautsdefrance.inrae.fr/agroimpact/Neos-dispositifs-outils/Modeles-et-outils-d-aide-a-la-decision/AMG-et-SMRCES-AMG/AMG-model-description

Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Texture du sol
- Densité du sol [1.2, 1.5]
 - Texture + densité => params hydrodynamiques (rosetta)
- Teneur initiale en MOS [0.01, 0.12]
- pH initial du sol [5, 7]
- Localisation
- Climat
(pour chaque site, 3 modèles RCP4.5 [Ouranos])



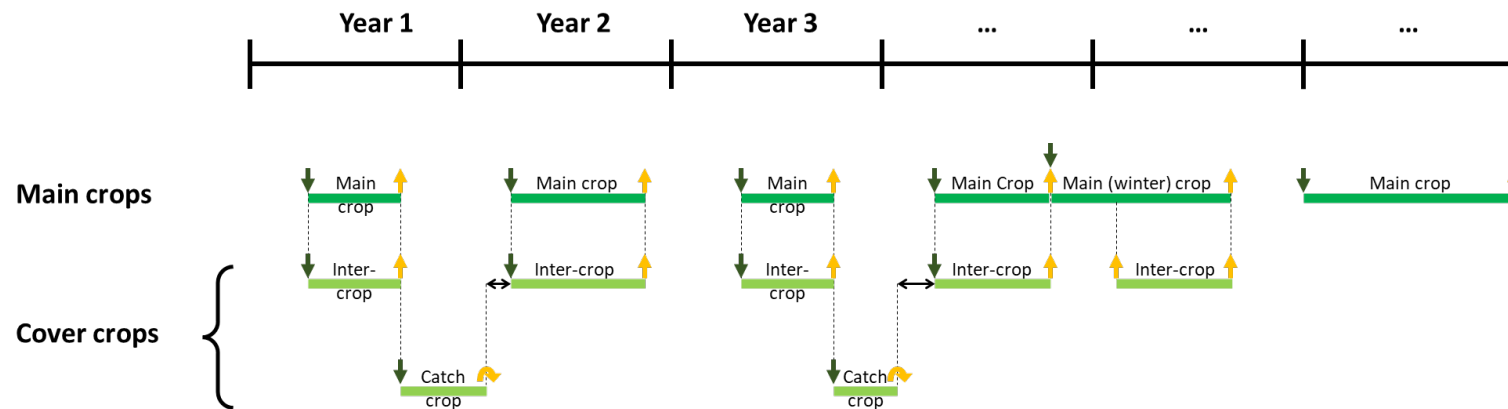
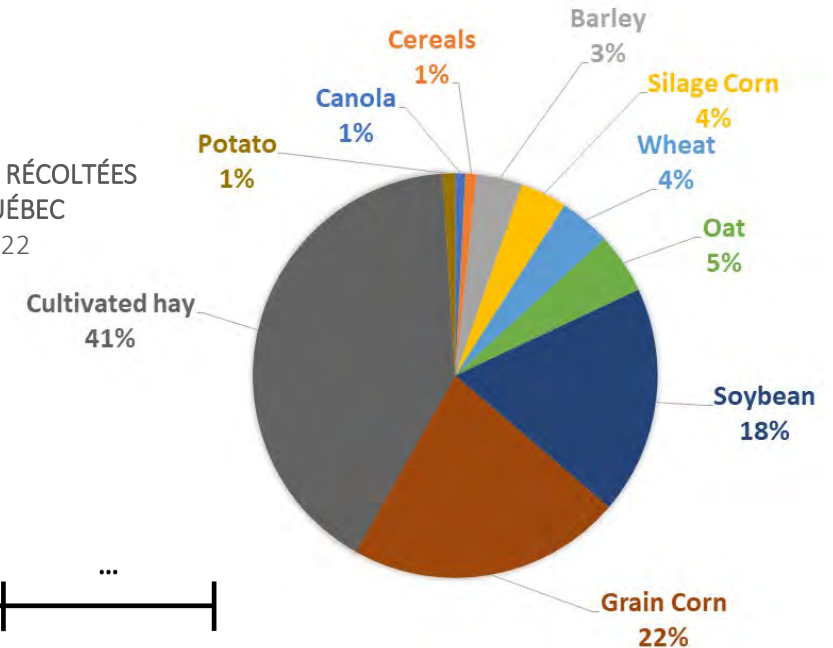
Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Rotation culturale
 - Culture principale
 - (optionnel) Culture secondaire en intercalaire ou en dérobé

RÉPARTITION DES SUPERFICIES RÉCOLTÉES EN GRANDES CULTURES AU QUÉBEC MOYENNES ENTRE 2007 ET 2022

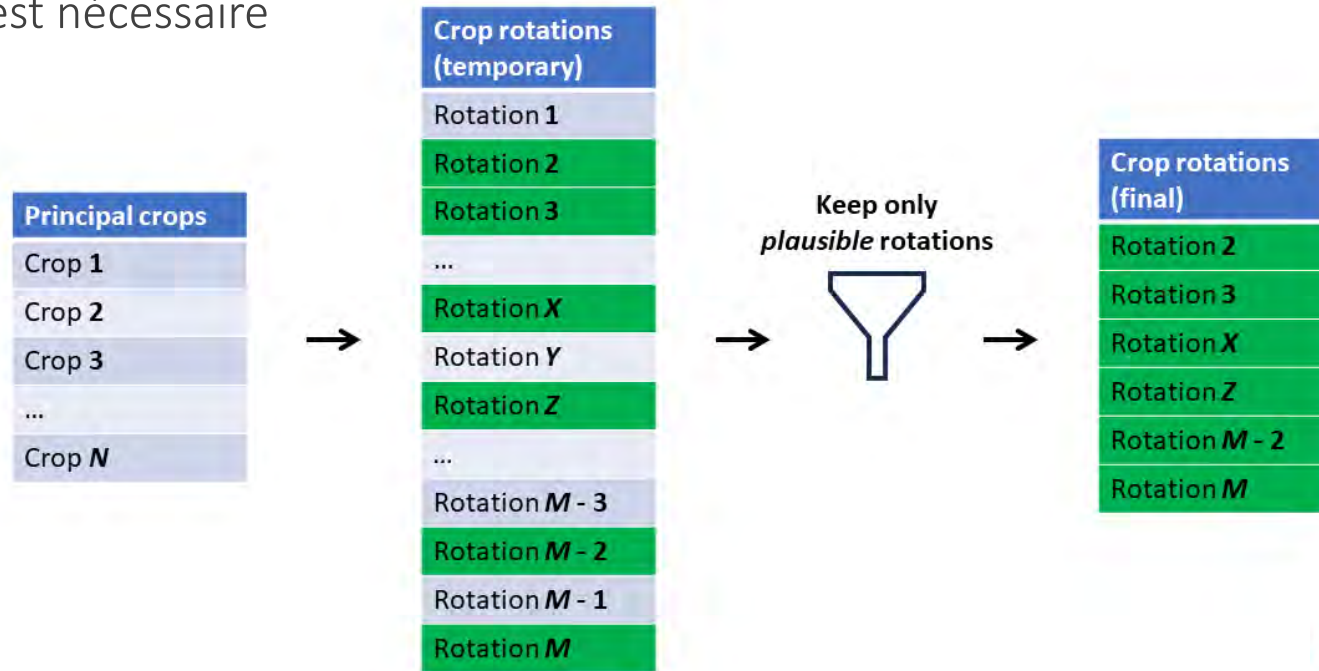


Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Rotation culturale
 - ... un peu de logique est nécessaire



Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Rotation culturale
 1. Une durée min & max fixée pour chaque culture
 2. ~~Blé de printemps suivi de blé d'hiver~~
 3. ~~Rotation qui commence par une culture d'hiver (début sim. en janvier)~~
 4. Associations plausibles (e.g. graminée & légumineuses)

Acceptable rotations



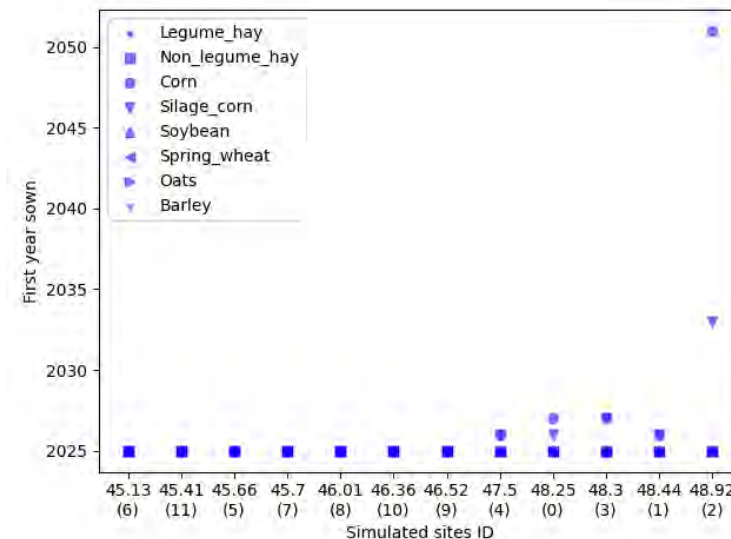
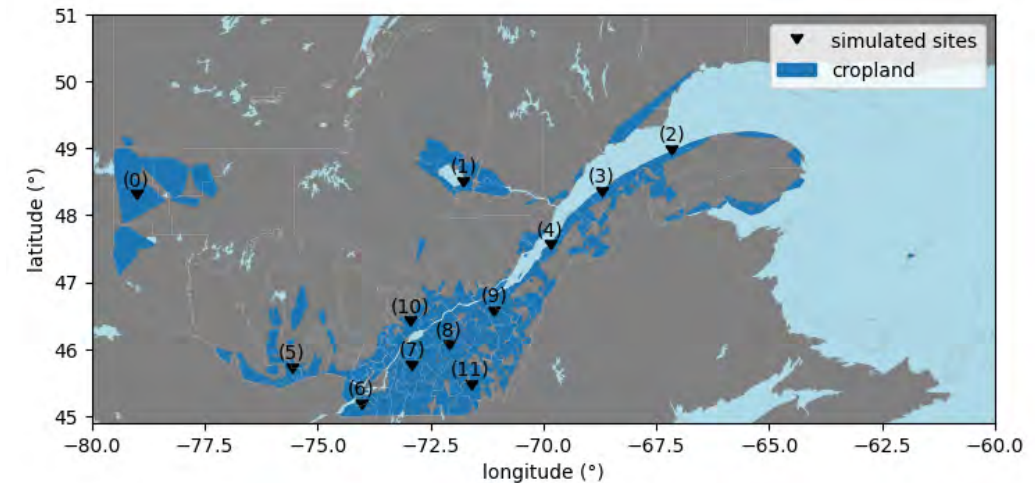
	Year 1	Year 2	Year 3	Year 4	Year 5
✓	Hay	Hay	Hay	Corn	Corn
✓	Hay	Hay	Corn	Corn	Corn
	Corn	Hay	Hay	Corn	Corn
	Corn	Hay	Hay	Corn	Hay
	Corn	Hay	Corn	Corn	Hay
✓	Hay	Corn	Corn	Corn	Hay

Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Rotation culturale
 - Des variétés qui changent avec la localisation (température) (biomasse max, WUE, T_{opt}, TDD, LAI max)
 - Des cultures non-praticables dans certains contextes (seuils TDD)

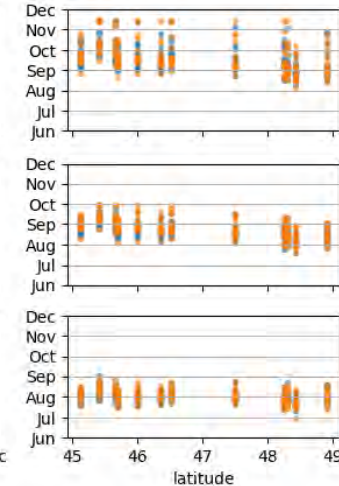
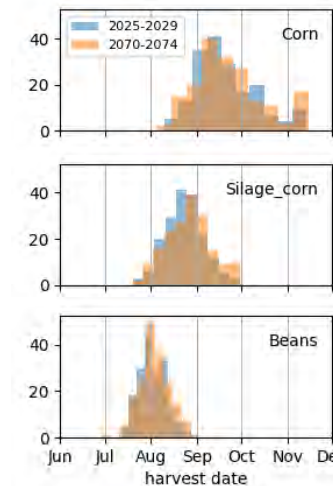
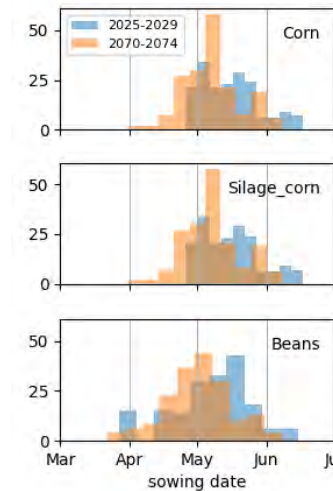


Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

- Rotation culturale
 - Dates de semis adaptées (premier jour où $T_{avg} > 12.8$ °C après dernier gel)
- Dates de récolte adaptées (grâce à la modif de TDD)

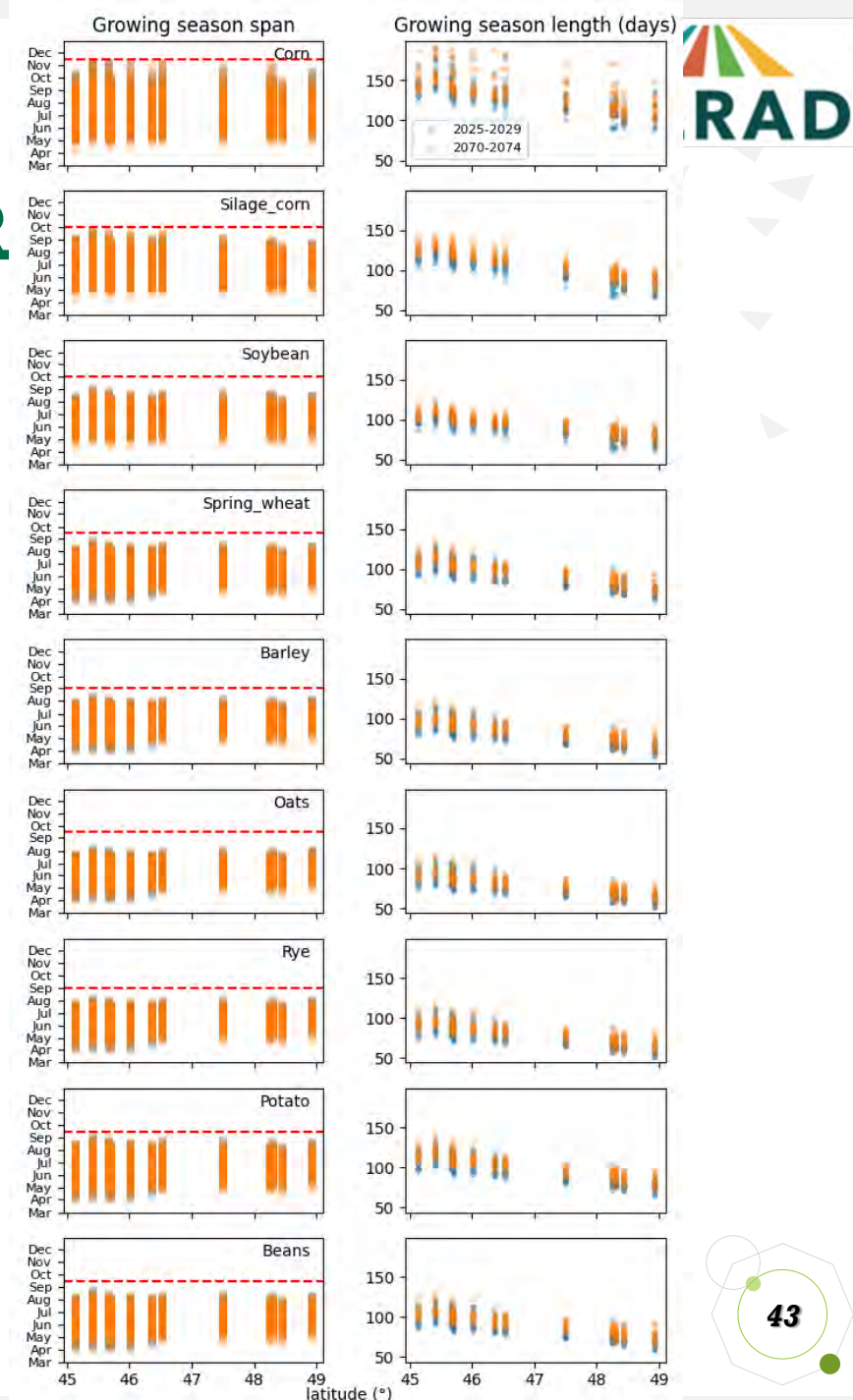


Réalisations – Volet 3 : DNDC (R)

Créer les scénarios

Paramètres:

- Rotation culturale
 - Dates de semis adaptées (premier jour où $T_{avg} > 12.8$ °C après dernier gel)
 - Dates de récolte adaptées (grâce à la modif de TDD)

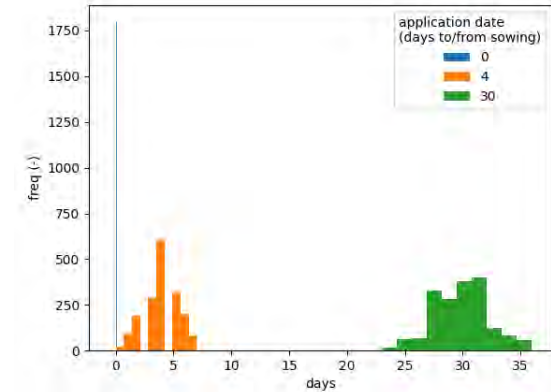
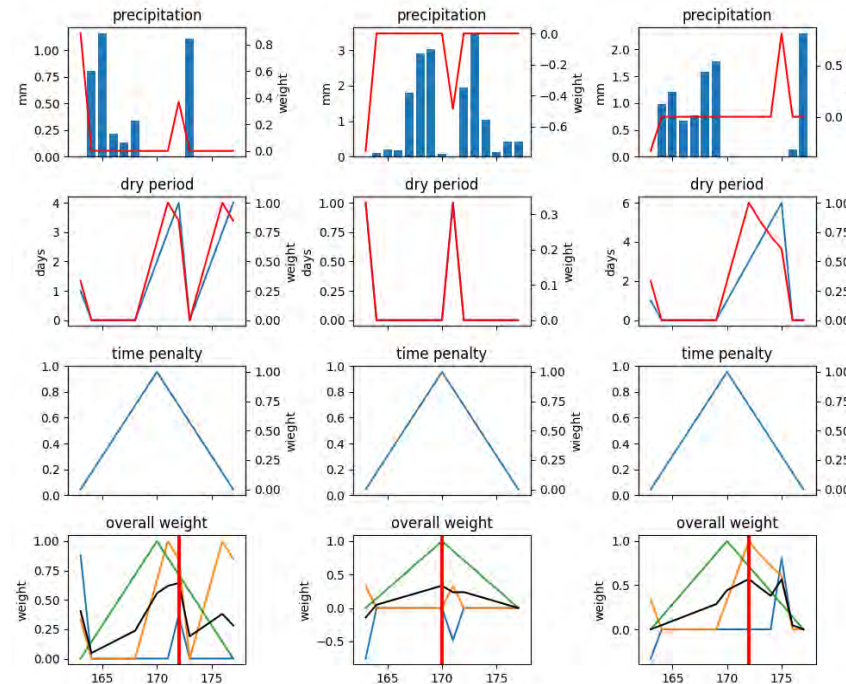
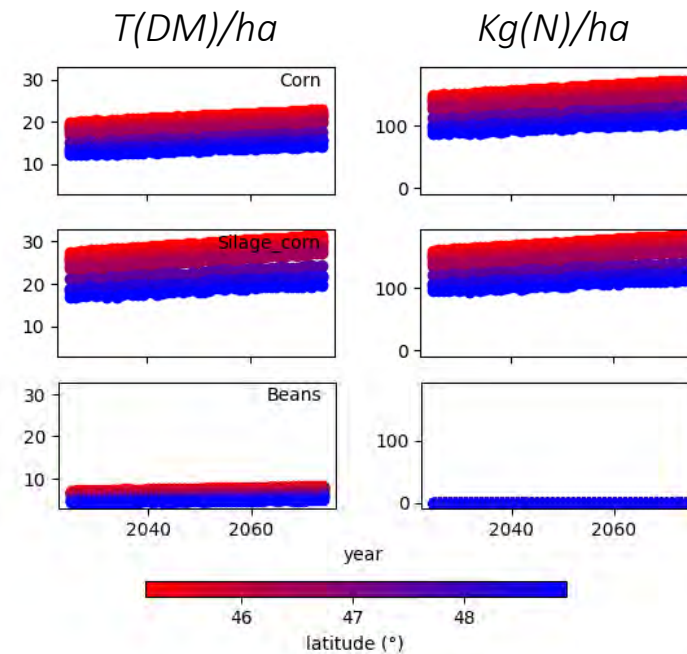


Réalisations – Volet 3 : DI

Créer les scénarios

Paramètres:

- Fertilisation azotée
 - Pour simplifier *Urée* ou *fumier*
 - Dose max = f(Biomasse max)
 - Doses variables
 - Dates variables et optimisées
 - En fonction de la météo
 - En fonction de la période de croissance



- weight precipitation
- weight dry period
- weight time penalty
- weight
- app doy

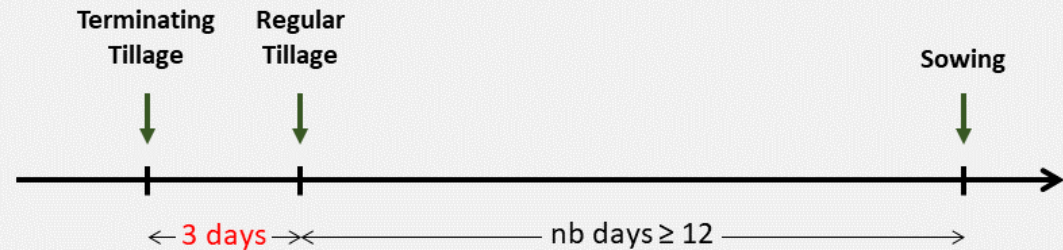
Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

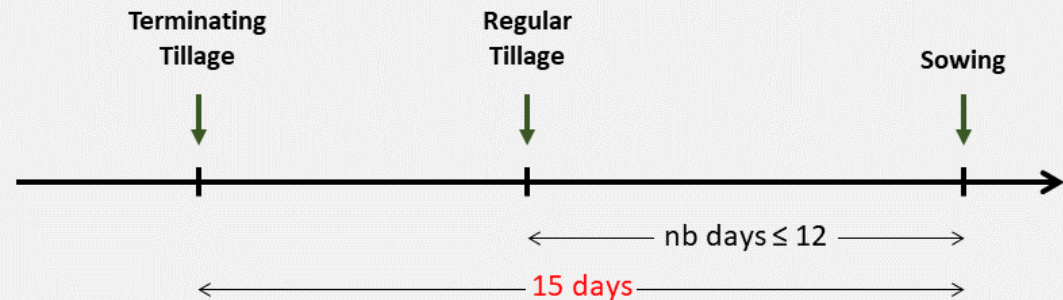
Paramètres:

- Travail du sol
 - Régulier : cultures annuelles
 - Profond (pour détruire la culture) : cultures pérennes

Case 1: a regular tillage event occurs at least 12 days before sowing



Case 2: a regular tillage event occurs less than 12 days before sowing

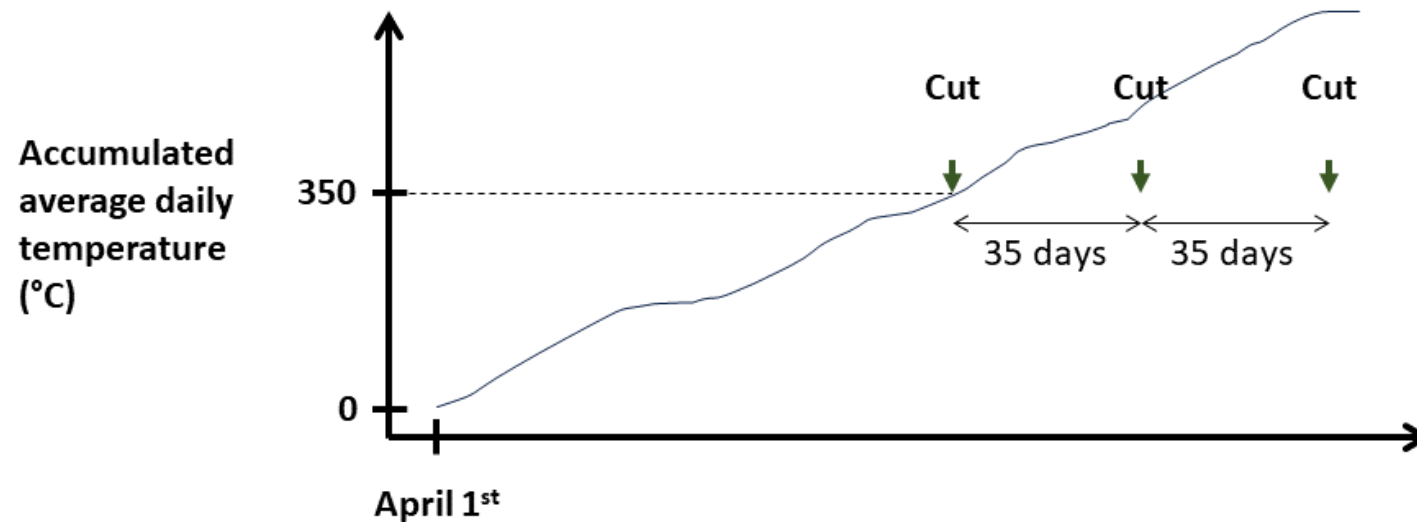


Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

Paramètres:

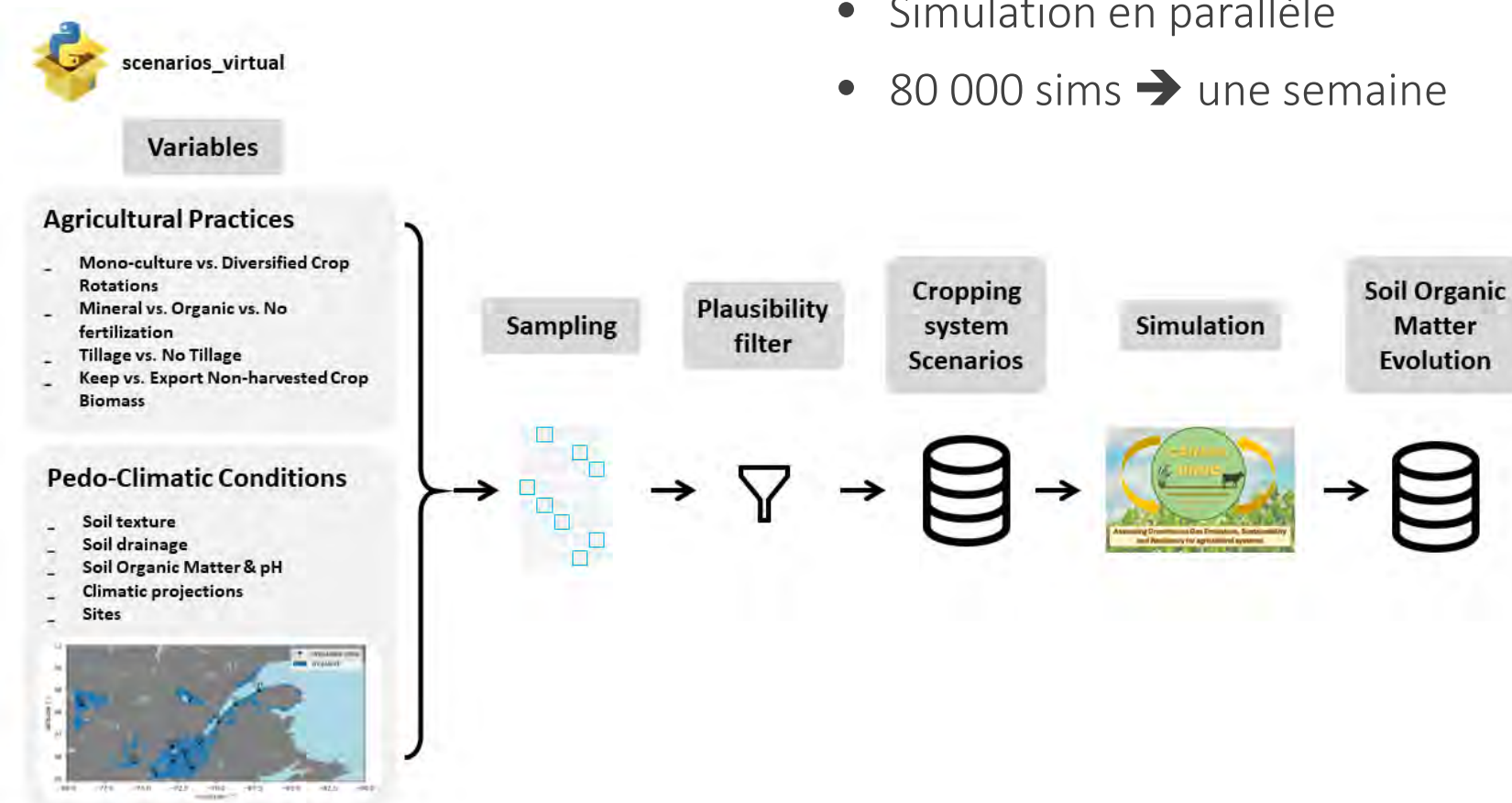
- Tonte



Réalisations – Volet 3 : DNDC (Rami)

Créer les scénarios

- Utilisation d'un serveur payant
- Simulation en parallèle
- 80 000 sims → une semaine



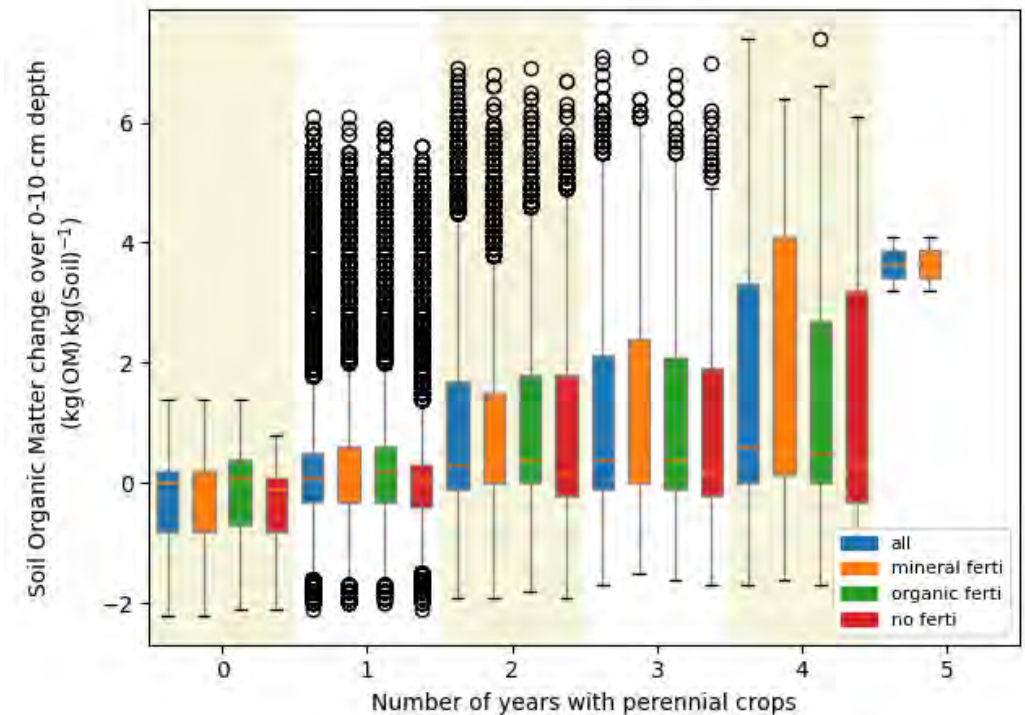
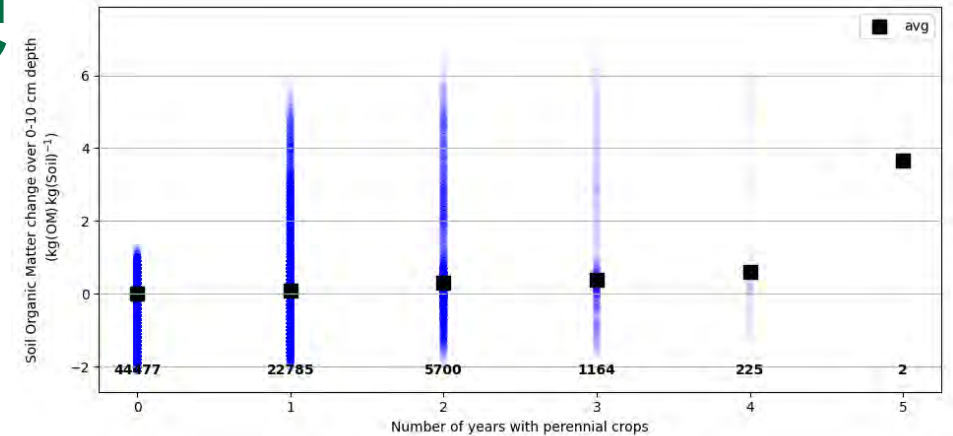
Réalisations – Volet 3 : DNDC

Résultats

(MOS à 10 cm - première simulation)

Effet de la présence de cultures pérennes dans la rotation

- ↗ Nb années avec cultures pérennes => ↗ MOS
- En présence des cultures pérennes, la fertilisation organique a un effet mitigé sur la MOS.



Réalisations – Volet 3 : DNDC (Rami)

Résultats

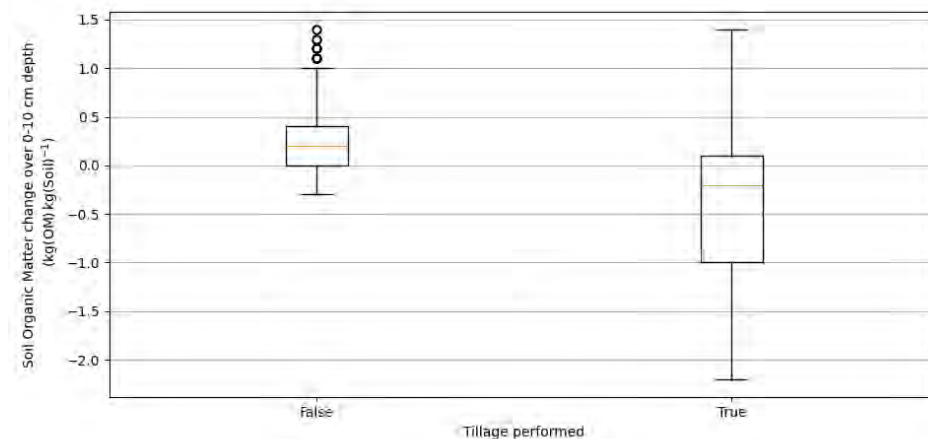
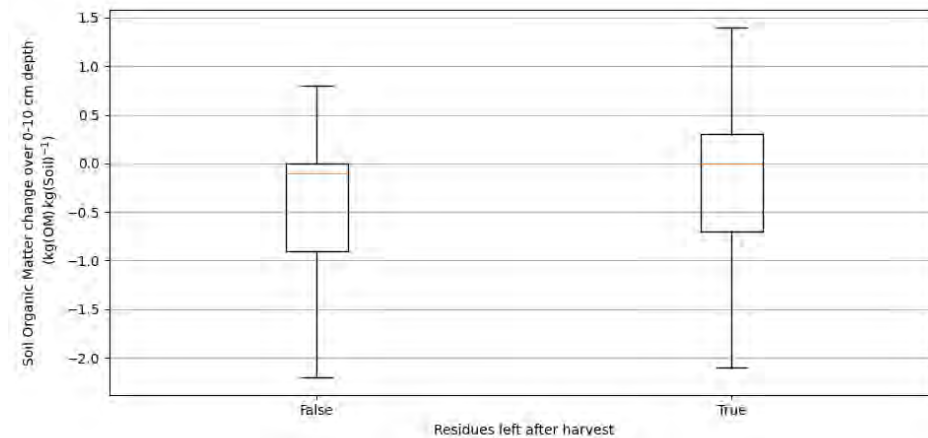
(MOS à 10 cm - première simulation)

Effet de l'exportation des résidus

- Exportation => baisse de la MOS

Effet du travail de sol

- Travail de sol => baisse de la MOS



Réalisations – Volet 3 : DNDC (Rami)

Résultats

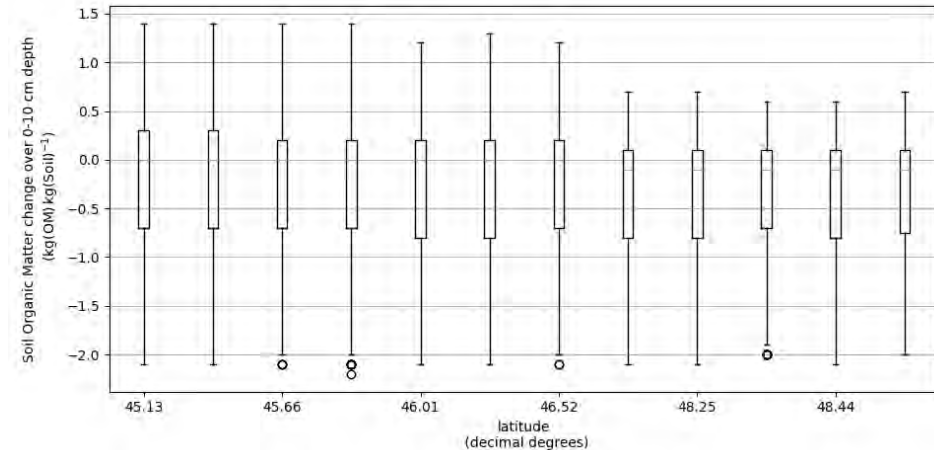
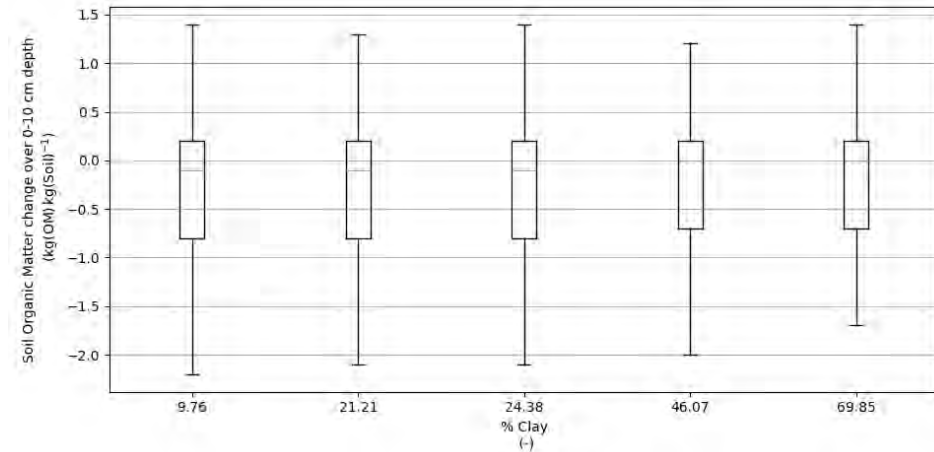
(MOS à 10 cm - première simulation)

Effet de la teneur en argile

- Baisse moins accrue avec les sols argileux

Effet de la latitude

- Baisse accrue de la MOS avec la latitude



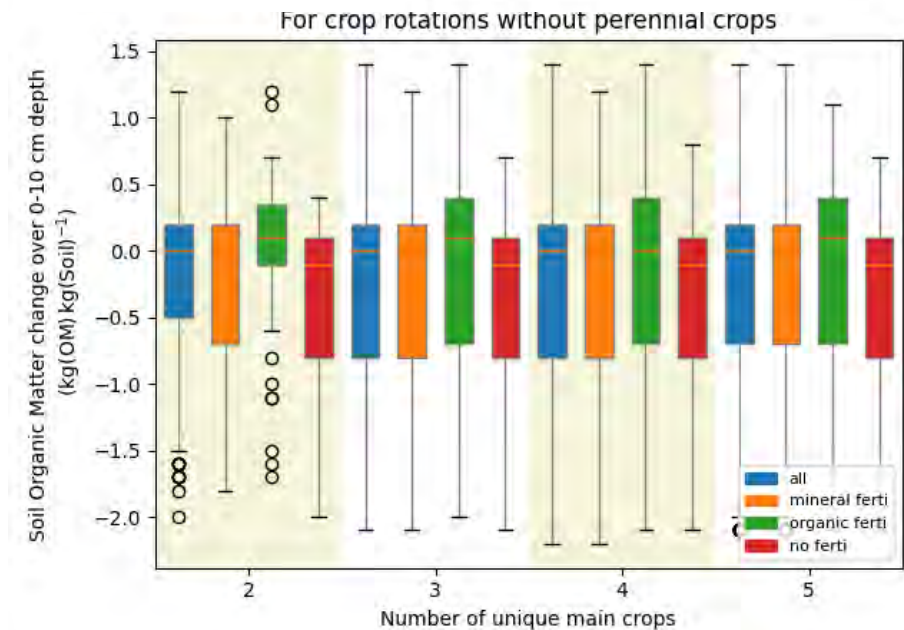
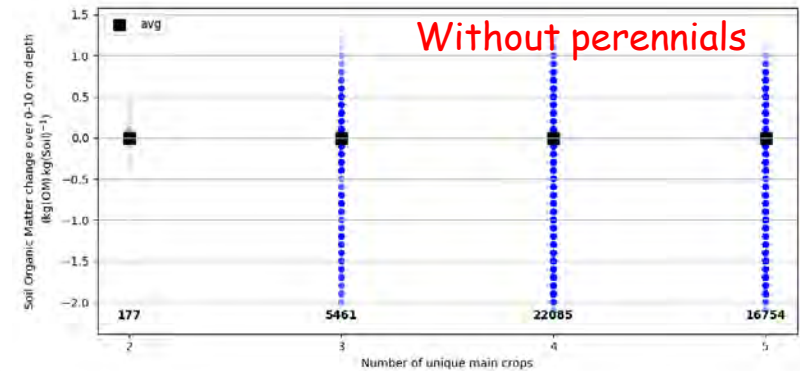
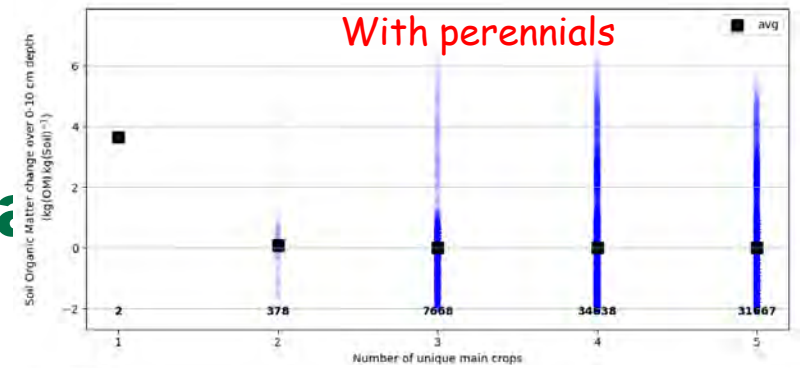
Réalisations – Volet 3 : DNDC (R)

Résultats

(MOS à 10 cm - première simulation)

Effet de la diversité des cultures

- ↗ diversité des cultures => ↔ MOS
- En l'absence de culture pérennes, la réduction de la MOS est plus faible avec la fertilisation organique comparée à la fertilisation minérale.



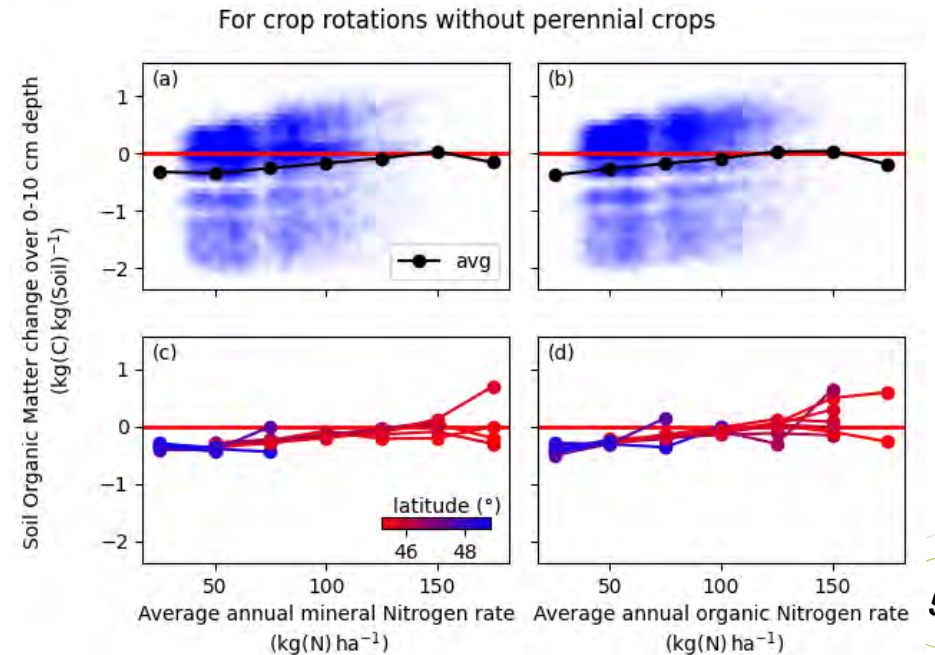
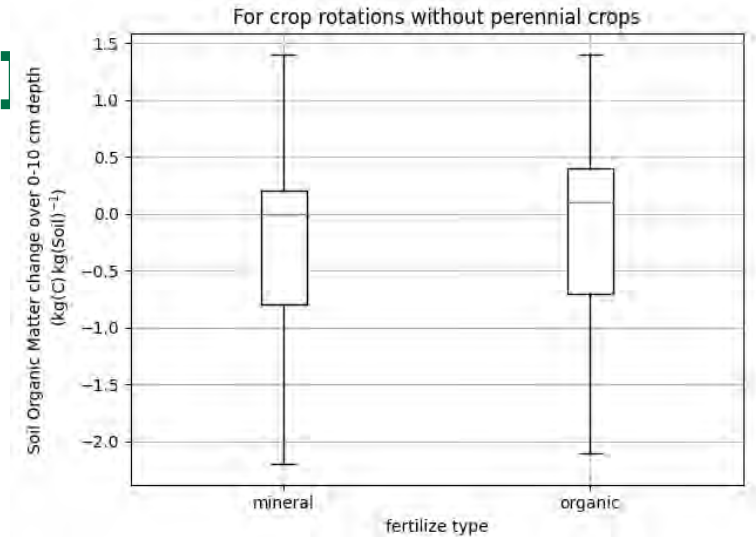
Réalisations – Volet 3 : DNDC

Résultats

(MOS à 10 cm - première simulation)

Effet de la fertilisation azotée

- Fertilisation organique => moindre baisse de la MOS
- La réduction de la perte de MOS commence à partir de :
 - ≈ 50 kgN/ha/an avec la fertilisation minérale
 - 0 kgN/ha/an avec la fertilisation organique
- La perte de la MOS est contrebalancée avec
 - ≈ 150 kgN/ha/an avec la fertilisation minérale
 - ≈ 120 kgN/ha/an avec la fertilisation organique
- La latitude ne semble pas perturber ces relations.

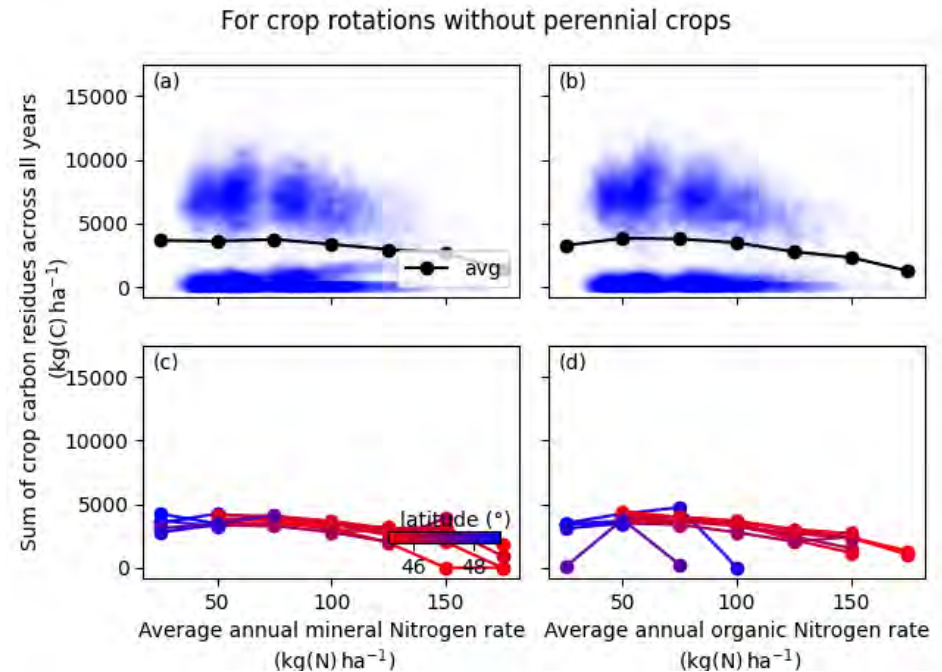
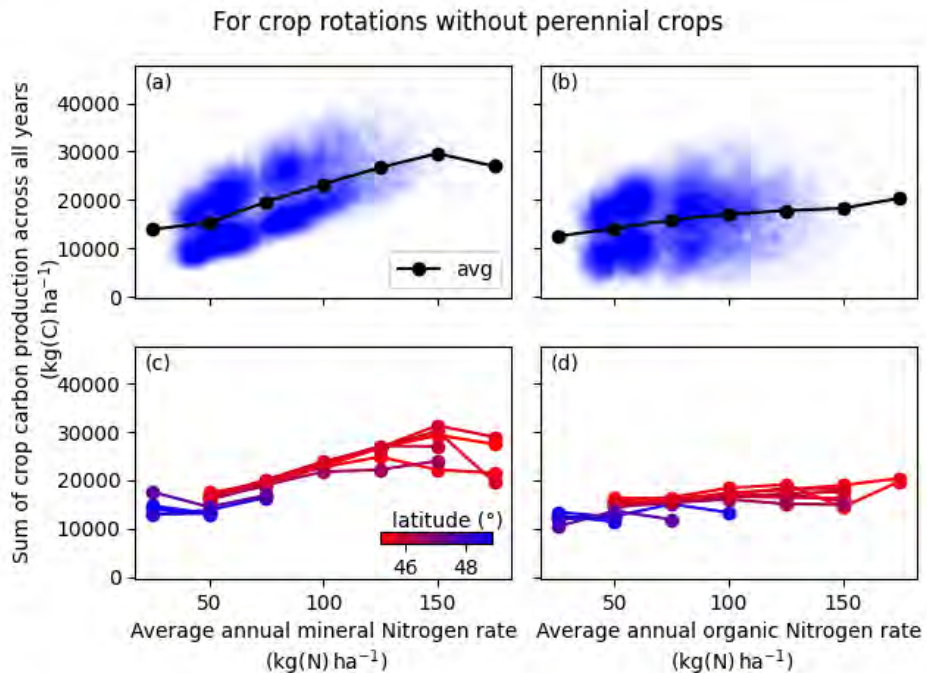


Réalisations – Volet 3 : DNDC (Rami)

Résultats

(MOS à 10 cm - première simulation)

Effet de la fertilisation azotée



Réalisations – Volet 3 : DNDC

Suite :

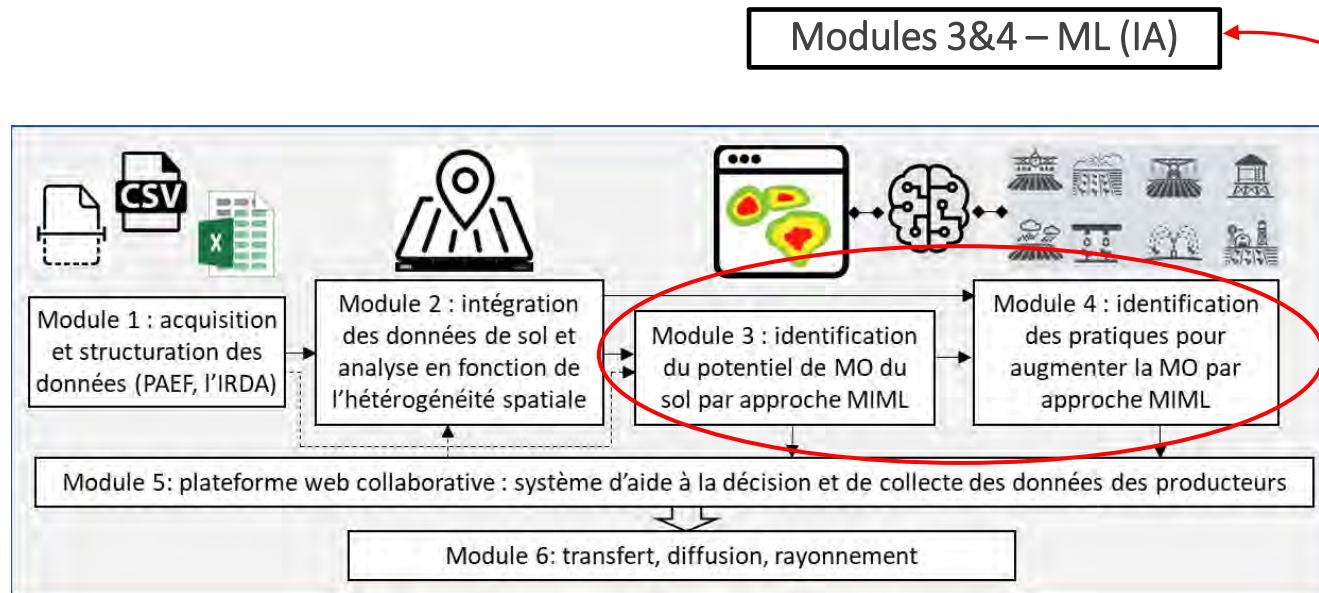
- Deuxième simulation suite aux tests de modélisation par IA et autres constatations :
 - Un plus grand nombre de simulations
 - Meilleure représentativité des variables du modèle d'IA
- Mise à jour de la documentation
- Publication

Réalisations – Volet 4 : IA

Modélisation de MOS par IA : potentiel et meilleures pratiques

Dr Guy Armel Fotso Kamga, PhD et postdoc, supervisé par Pr Y. Bouroubi

- Modélisation par IA de la dynamique de la MO
- Test de modèles de ML classiques et nouveaux



Réalisations – Volet 4 : IA

Modélisation de MOS par IA : potentiel et meilleures pratiques

Guy



Réalisations – Volet 4 : IA (Guy)

DNDC : simulation multi-annuelle intégrant sol - climat - pratiques de gestion



Données sur les propriétés physico-chimiques du sol

'sand_perc', 'clay_perc', 'bulk_density', 'ph', 'soc', 'drain_depth', 'drain_spacing'



Données climatiques (annuelles, années 1 à 5)

'precipitation_cum_annual_1', 'precipitation_cum_annual_2', 'precipitation_cum_annual_3',
'precipitation_cum_annual_4', 'precipitation_cum_annual_5',
'precipitation_shannon_diversity_index_annual_1', 'precipitation_shannon_diversity_index_annual_2',
'precipitation_shannon_diversity_index_annual_3', 'precipitation_shannon_diversity_index_annual_4' ...



Cultures et pratiques de gestion (années 1 à 5)

'catch_crop_1', 'catch_crop_2', 'catch_crop_3', 'catch_crop_4', 'catch_crop_5', 'inter_crop_1',
'inter_crop_2', 'inter_crop_3', 'inter_crop_4', 'inter_crop_5', 'main_crop_1', 'main_crop_2', 'main_crop_3',
'main_crop_4', 'main_crop_5'

Réalisations – Volet 4 : IA (Guy)

DNDC : simulation multi-annuelle intégrant sol - climat - pratiques de gestion

```
{'scenario_id': '3',
'latitude': '45.41',
'sand_perc': '19.19',
'clay_perc': '46.07',
'bulk_density': '1.35',
'ph': '7.0',
'soc': '0.0696',
'drain_depth': '1.0',
'drain_spacing': '10.0',
'precipitation_cum_annual_1': '121',
'precipitation_cum_annual_2': '102',
'precipitation_cum_annual_3': '127',
'precipitation_cum_annual_4': '122',
'precipitation_cum_annual_5': '115',
'precipitation_shannon_diversity_index_annual_1': '81',
'precipitation_shannon_diversity_index_annual_2': '81',
'precipitation_shannon_diversity_index_annual_3': '80',
'precipitation_shannon_diversity_index_annual_4': '82',
'precipitation_shannon_diversity_index_annual_5': '82',
'tdd_cum_annual_1': '2952',
'tdd_cum_annual_2': '3356',
'tdd_cum_annual_3': '3085',
'tdd_cum_annual_4': '3330',
'tdd_cum_annual_5': '3411',
'apps_mineral_1': '0.0',
'apps_mineral_2': '0.0',
'apps_mineral_3': '0.0',
'apps_mineral_4': '0.0',
'apps_mineral_5': '',
'apps_organic_1': '45.0',
'apps_organic_2': '69.0',
'apps_organic_3': '67.0',
'apps_organic_4': '49.0',
'apps_organic_5': '',
'catch_crop_1': 'Beans_green',
'catch_crop_2': 'Radish',
'catch_crop_3': '',
'catch_crop_4': 'Beans_green',
'catch_crop_5': '',
'inter_crop_1': 'Cover_crop',
'inter_crop_2': '',
'inter_crop_3': '',
'inter_crop_4': 'Cover_crop',
'inter_crop_5': '',
'is_residues_left': 'True',
'is_tillage': 'True',
'main_crop_1': 'Spring_wheat',
'main_crop_2': 'Oats',
'main_crop_3': 'Soybean',
'main_crop_4': 'Winter_wheat',
'main_crop_5': 'Legume_hay',
'dif_SOM0-30cm': '1.7'}
```



```
{'scenario_id': '3',
'latitude': '45.41',
'sand_perc': '19.19',
'clay_perc': '46.07',
'bulk_density': '1.35',
'ph': '7.0',
'soc': '0.0696',
'drain_depth': '1.0',
'drain_spacing': '10.0',
'precipitation_cum_annual_1': '121',
'precipitation_cum_annual_2': '102',
'precipitation_cum_annual_3': '127',
'precipitation_cum_annual_4': '122',
'precipitation_cum_annual_5': '115',
'precipitation_shannon_diversity_index_annual_1': '81',
'precipitation_shannon_diversity_index_annual_2': '81',
'precipitation_shannon_diversity_index_annual_3': '80',
'precipitation_shannon_diversity_index_annual_4': '82',
'precipitation_shannon_diversity_index_annual_5': '82',
'tdd_cum_annual_1': '2952',
'tdd_cum_annual_2': '3356',
'tdd_cum_annual_3': '3085',
'tdd_cum_annual_4': '3330',
'tdd_cum_annual_5': '3411',
'apps_mineral_1': '0.0',
'apps_mineral_2': '0.0',
'apps_mineral_3': '0.0',
'apps_mineral_4': '0.0',
'apps_mineral_5': '',
⋮
```

```
⋮
'apps_organic_1': '45.0',
'apps_organic_2': '69.0',
'apps_organic_3': '67.0',
'apps_organic_4': '49.0',
'apps_organic_5': '',
'catch_crop_1': 'Beans_green',
'catch_crop_2': 'Radish',
'catch_crop_3': '',
'catch_crop_4': 'Beans_green',
'catch_crop_5': '',
'inter_crop_1': 'Cover_crop',
'inter_crop_2': '',
'inter_crop_3': '',
'inter_crop_4': 'Cover_crop',
'inter_crop_5': '',
'is_residues_left': 'True',
'is_tillage': 'True',
'main_crop_1': 'Spring_wheat',
'main_crop_2': 'Oats',
'main_crop_3': 'Soybean',
'main_crop_4': 'Winter_wheat',
'main_crop_5': 'Legume_hay',
'dif_SOM0-30cm': '1.7'}
```

Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'intelligence artificielle



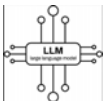
Approches d'apprentissage en ensemble (Ensemble Learning)

Random Forest, XgBoost, Extra Tree, Catboost, LightGBM



Approches basées sur les transformers pour données tabulaires

TabNet, TabTransformer



Approches basées sur les grands modèles de langage (LLMs)

Regress-llm, OmniPred

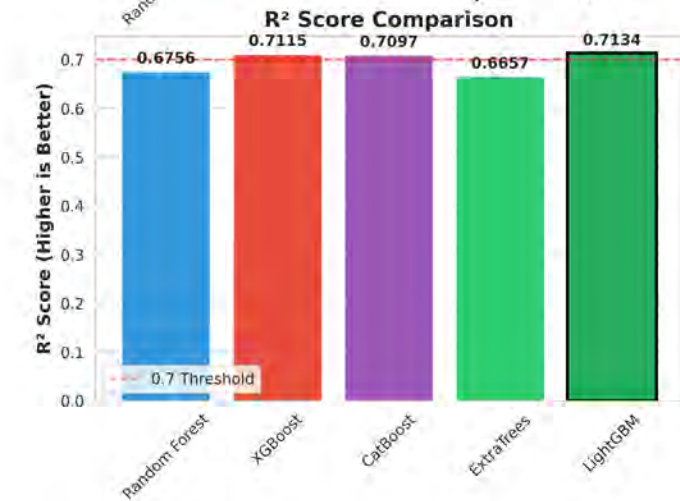
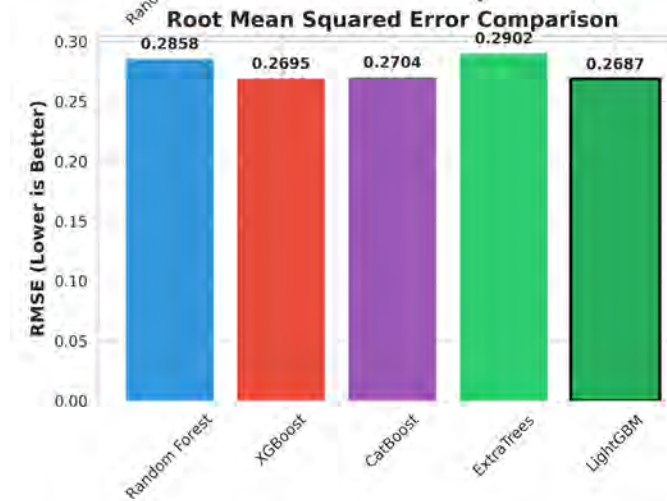
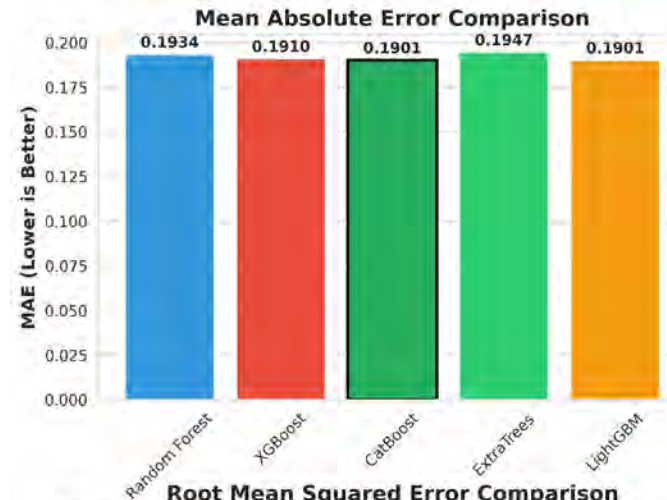
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

Model Performance Comparison - All Metrics



Approches d'apprentissage en ensemble (Ensemble Learning)



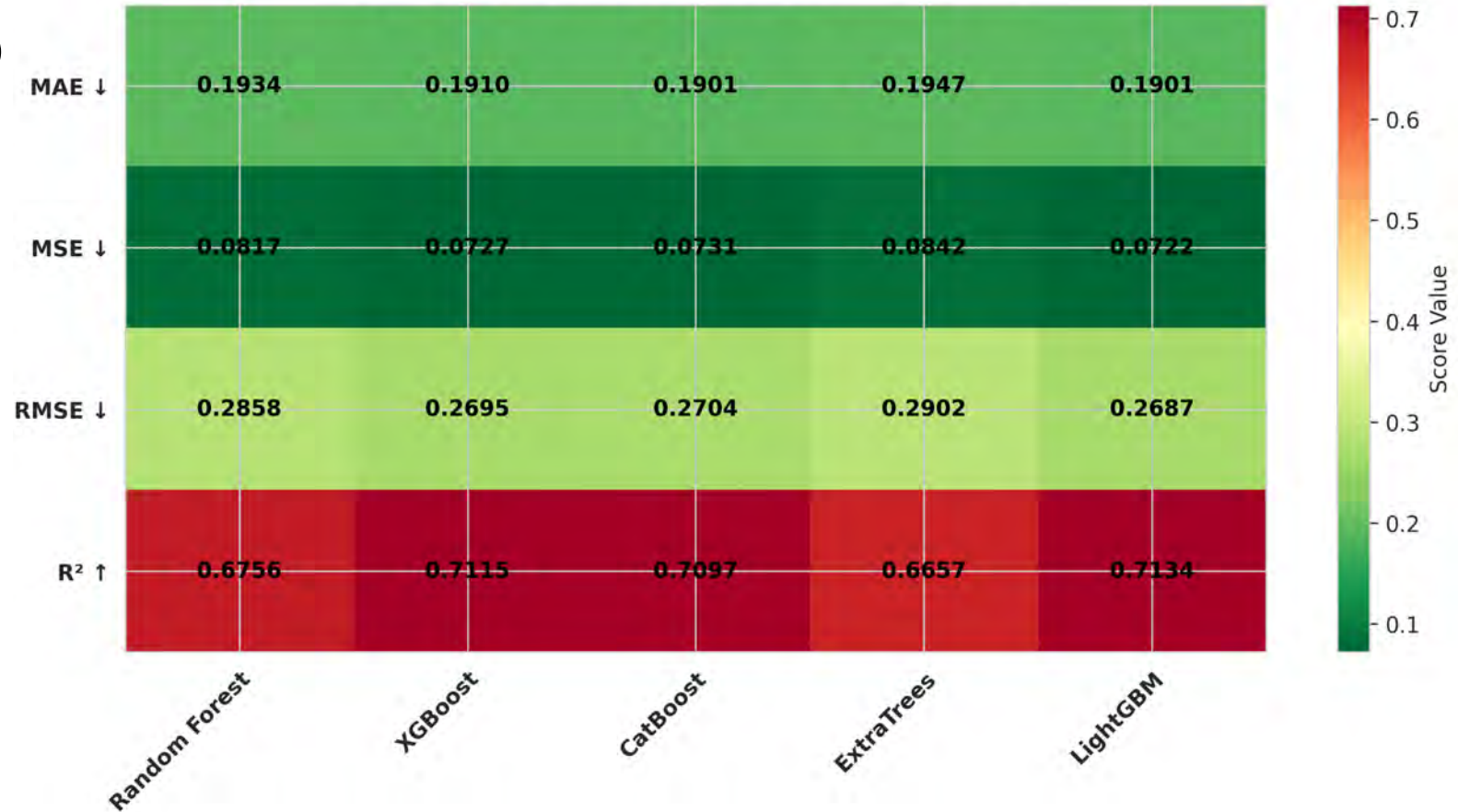
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches d'apprentissage en ensemble (Ensemble Learning)

Model Performance Heatmap
(↓ Lower is Better, ↑ Higher is Better)



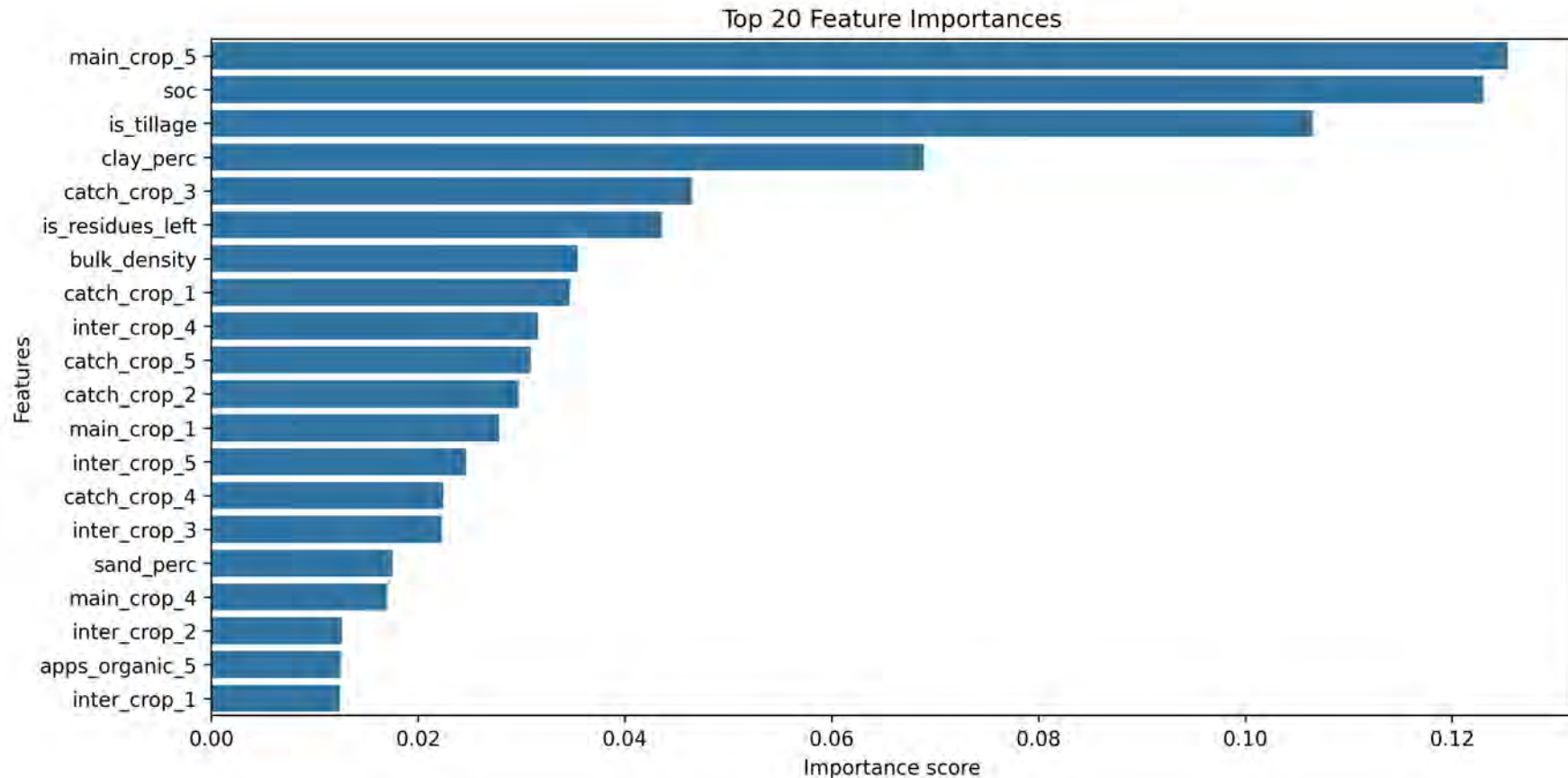
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

 Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Interprétabilité
du modèle



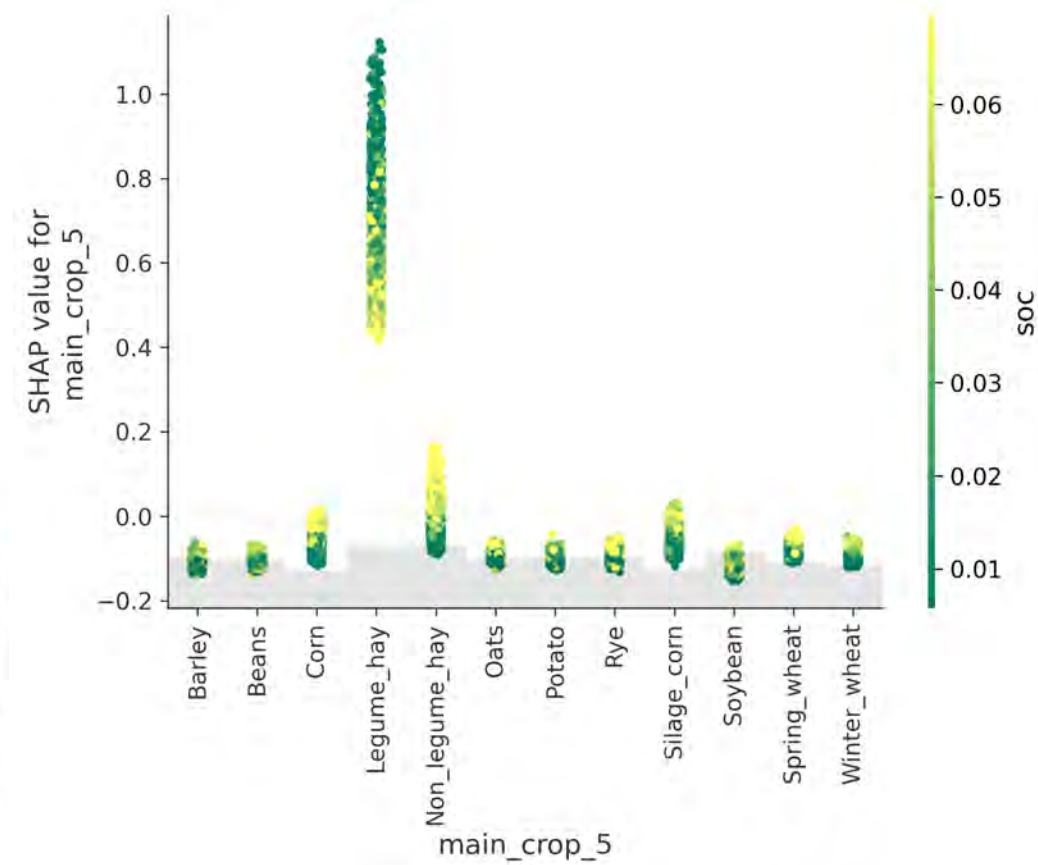
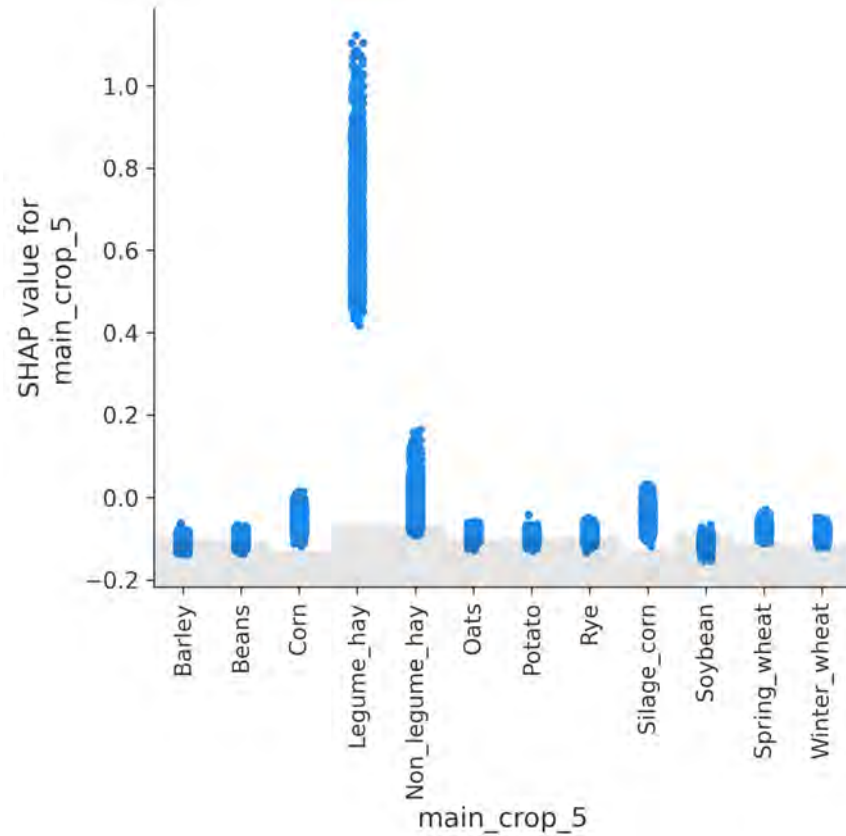
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

 Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Interprétabilité
du modèle



Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données physico-chimiques

'sand_perc', 'clay_perc', 'bulk_density', 'ph', 'soc', 'drain_depth', 'drain_spacing'

Évaluation du modèle

Métriques	R ²	RMSE	MAE
	0,2325	0,4397	0,3320

Réalisations – Volet 4 : IA (Guy)

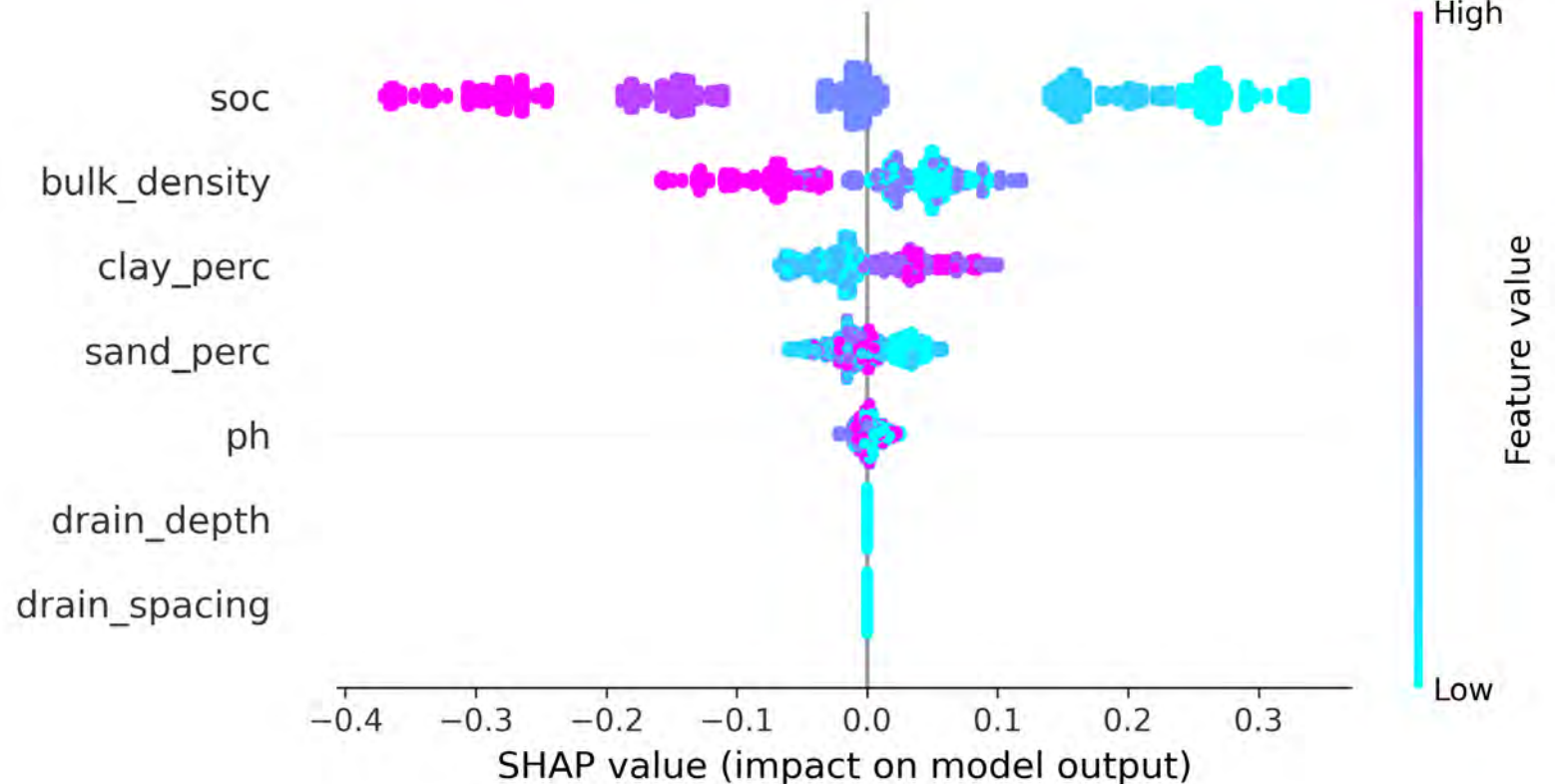
Modélisation par des algorithmes d'IA

Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données physico-chimiques

Interprétabilité du modèle



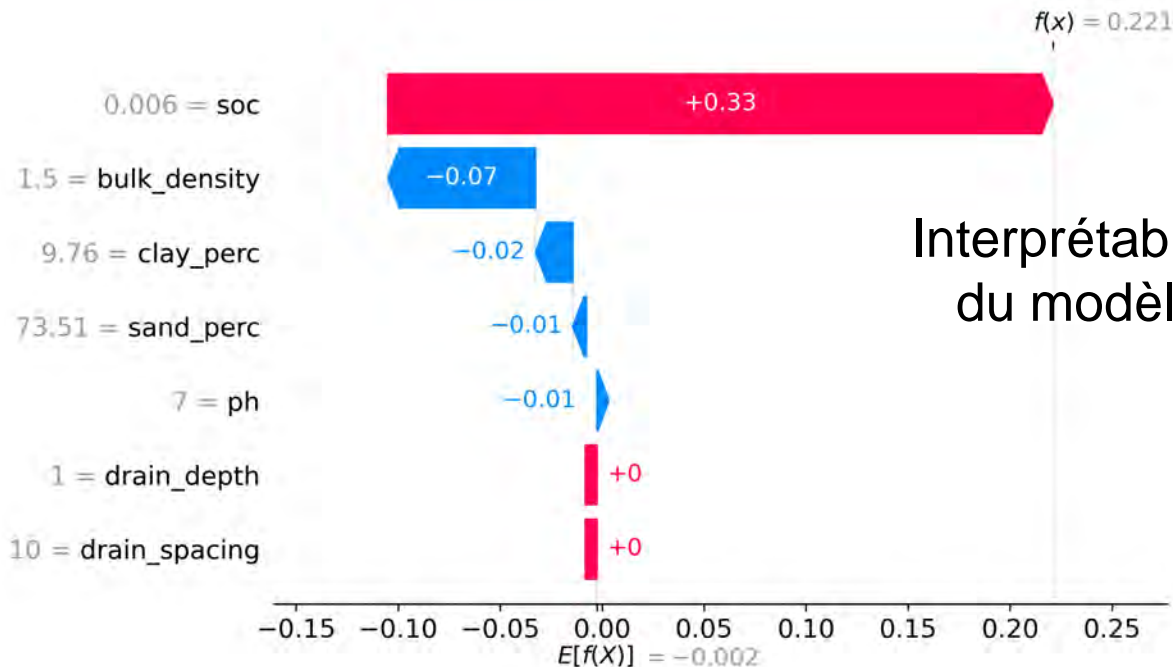
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

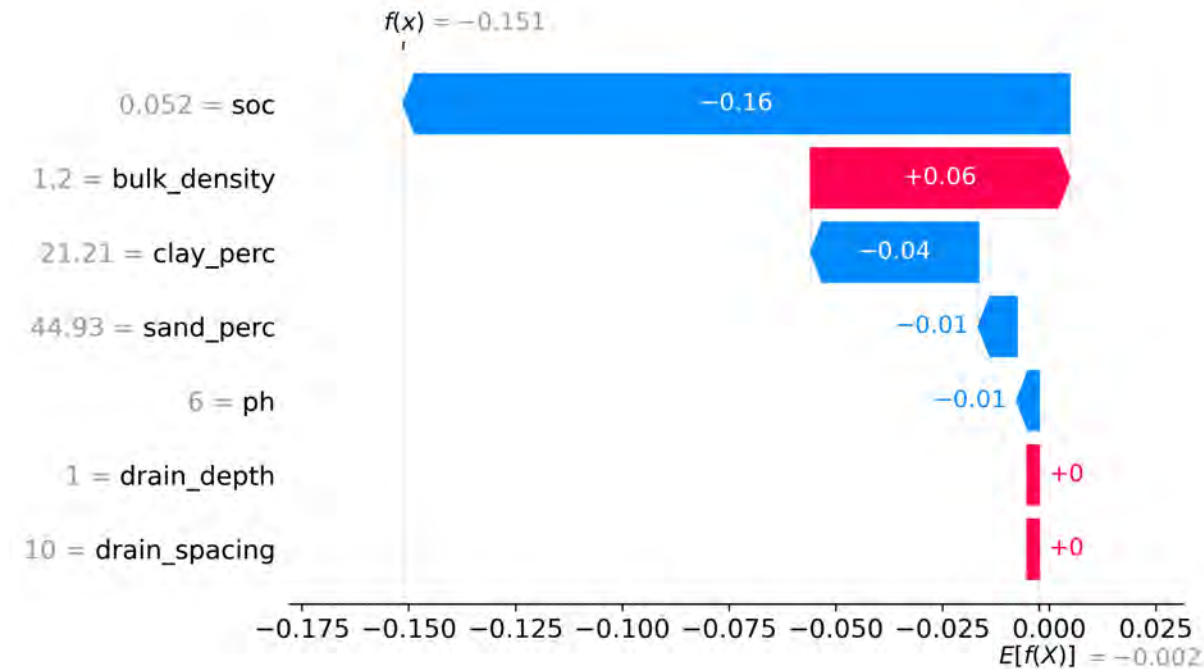
Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données physico-chimiques



Interprétabilité du modèle



Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données sur les pratiques agricoles

'catch_crop_1', 'catch_crop_2', 'catch_crop_3', 'catch_crop_4', 'catch_crop_5',
 'inter_crop_1', 'inter_crop_2', 'inter_crop_3', 'inter_crop_4', 'inter_crop_5',
 'main_crop_1', 'main_crop_2', 'main_crop_3', 'main_crop_4', 'main_crop_5'

Évaluation du modèle

Métriques	R ²	RMSE	MAE
	0,3886	0,3924	0,2835

Réalisations – Volet 4 : IA (Guy)

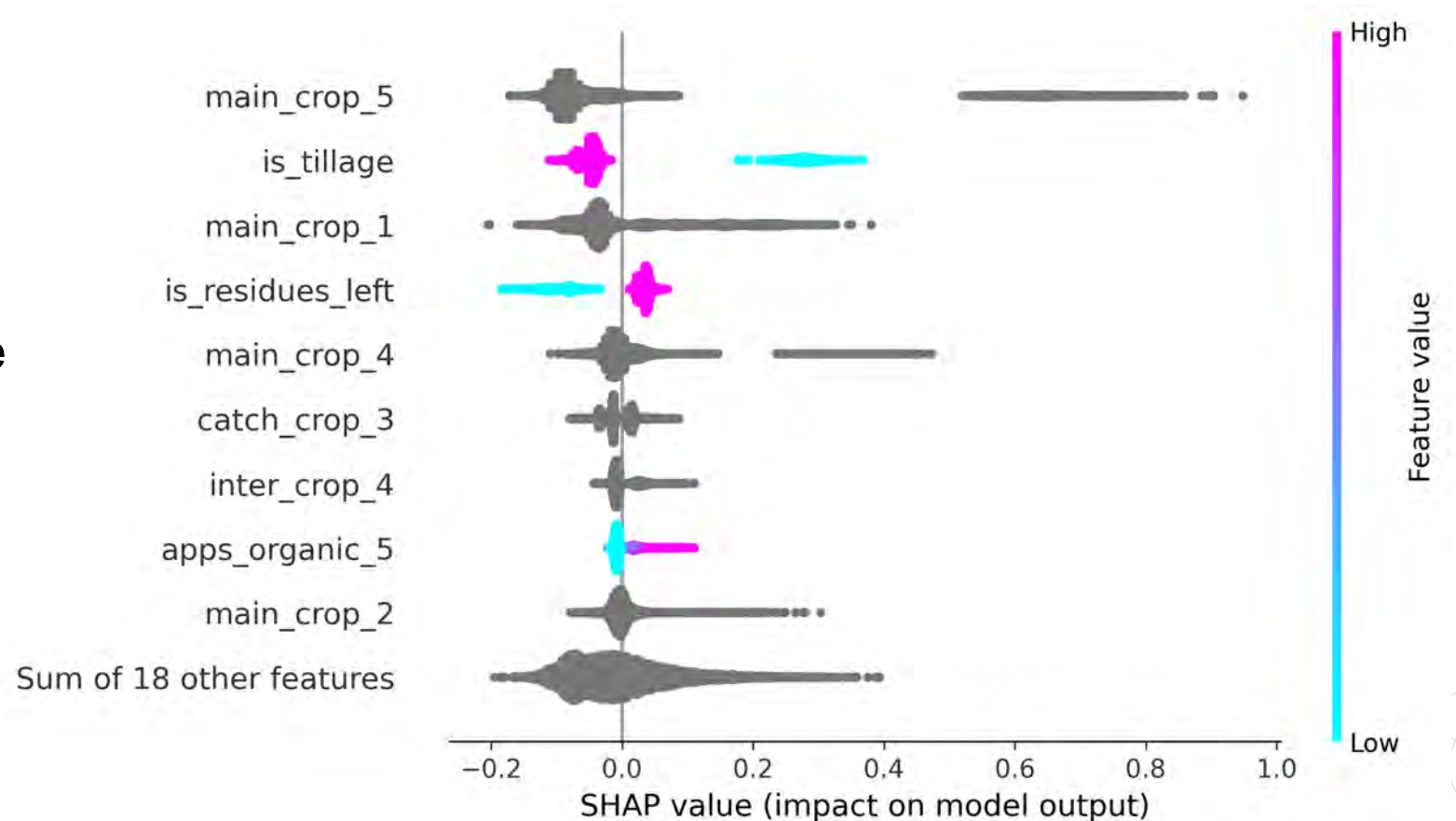
Modélisation par des algorithmes d'IA

 Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données sur les pratiques agricoles

Interprétabilité du modèle



Réalisations – Volet 4 : IA (Guy)

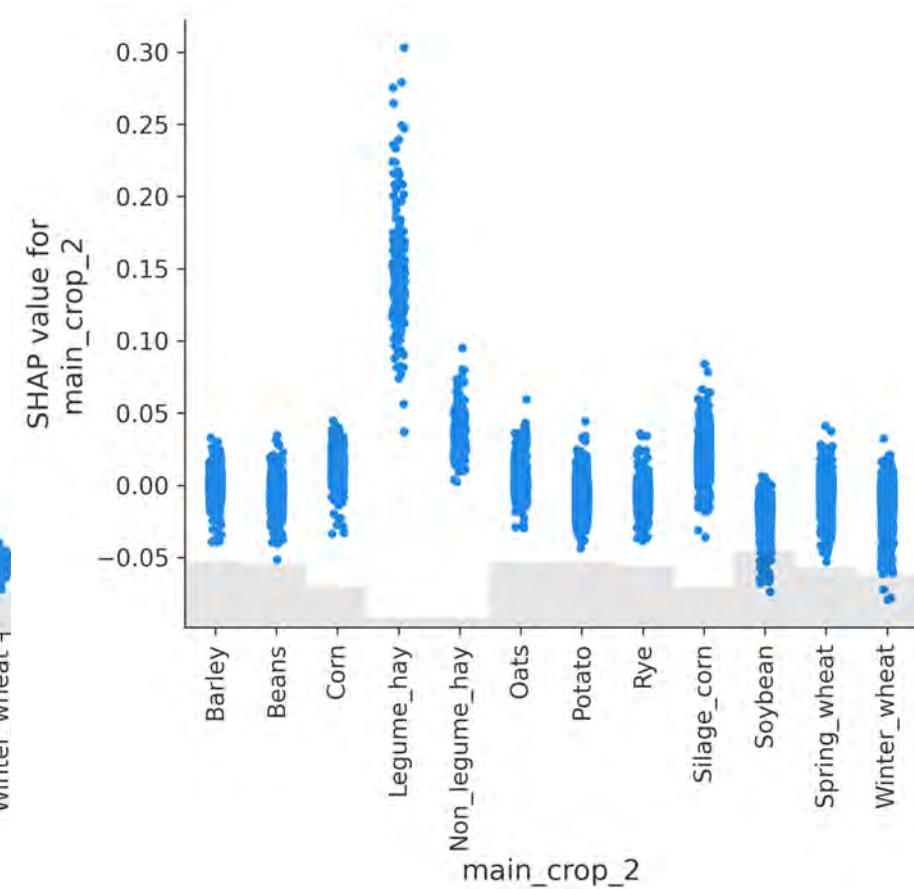
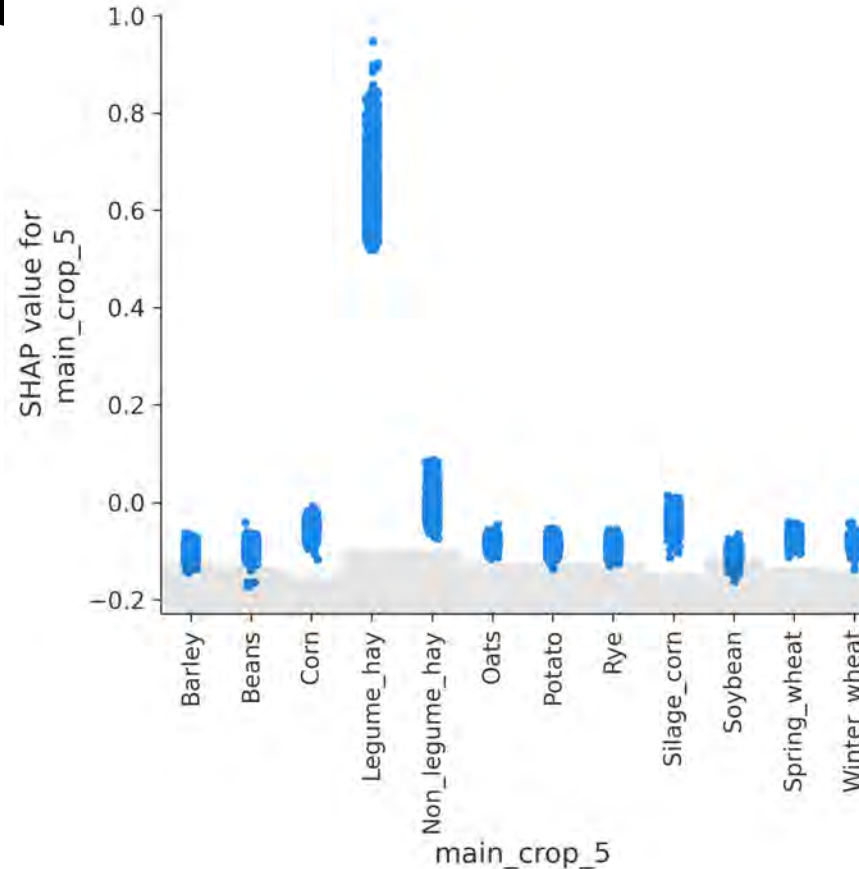
Modélisation par des algorithmes d'IA

 Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données sur les pratiques agricoles

Interprétabilité du modèle



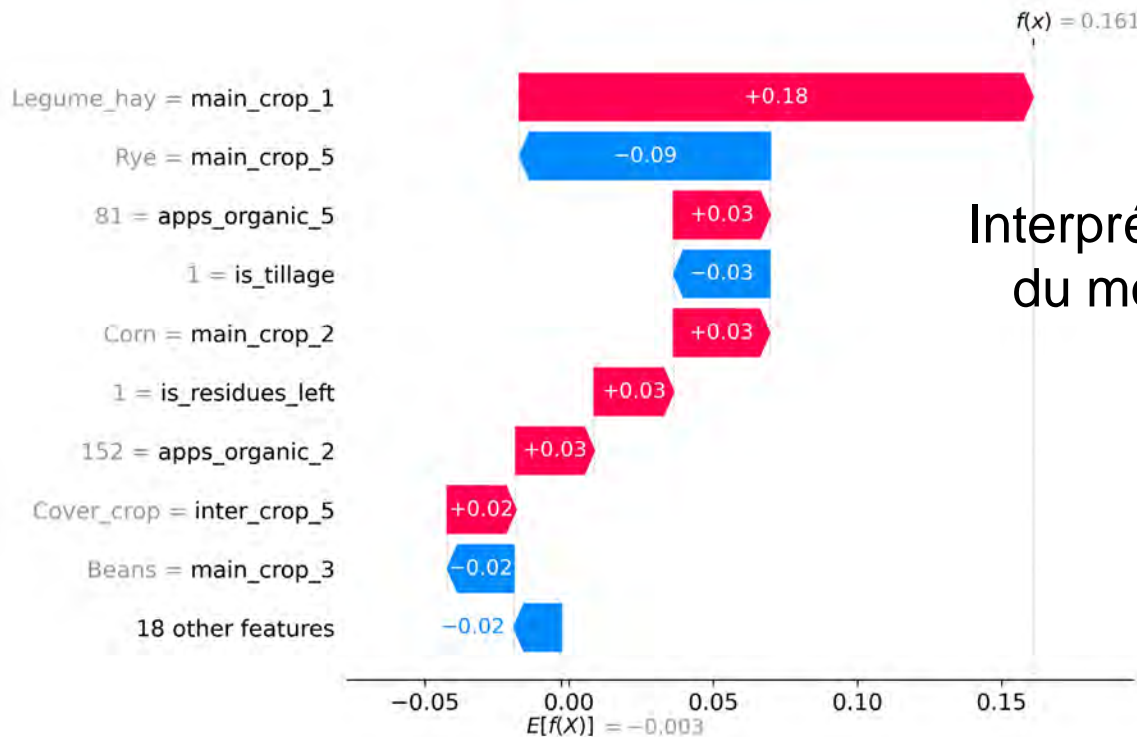
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

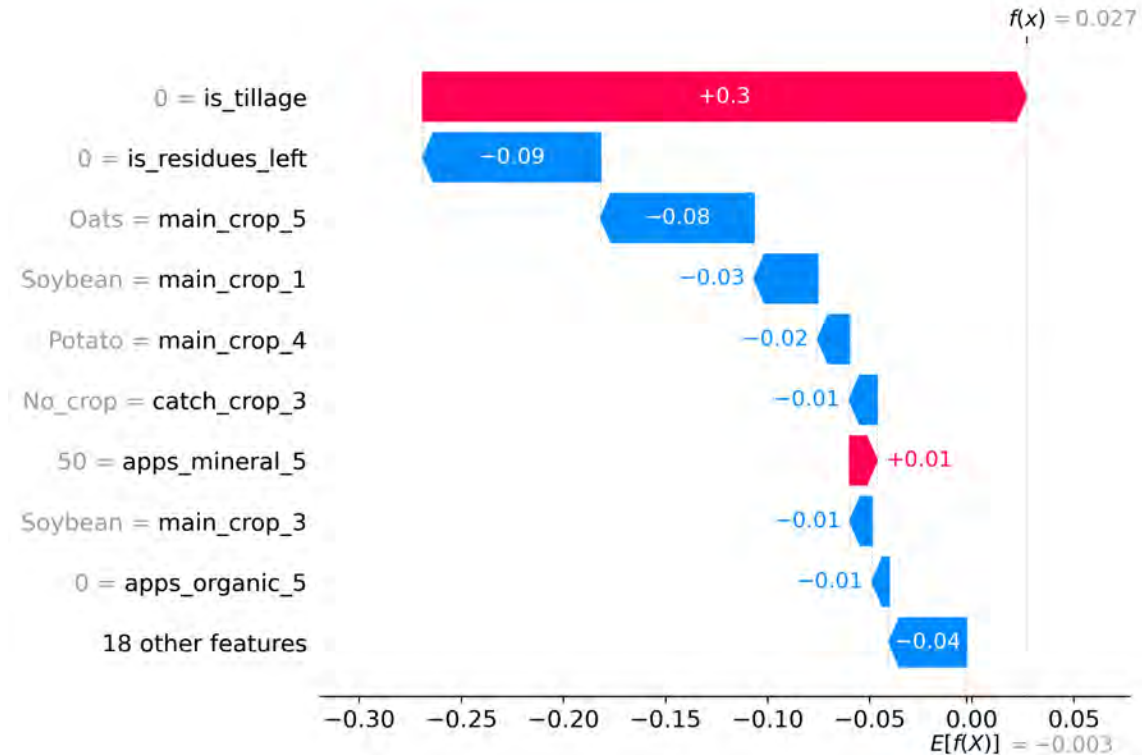
Approches d'apprentissage en ensemble (Ensemble Learning)

Algorithme XGBoost

Scénario de données sur les pratiques agricoles



Interprétabilité du modèle

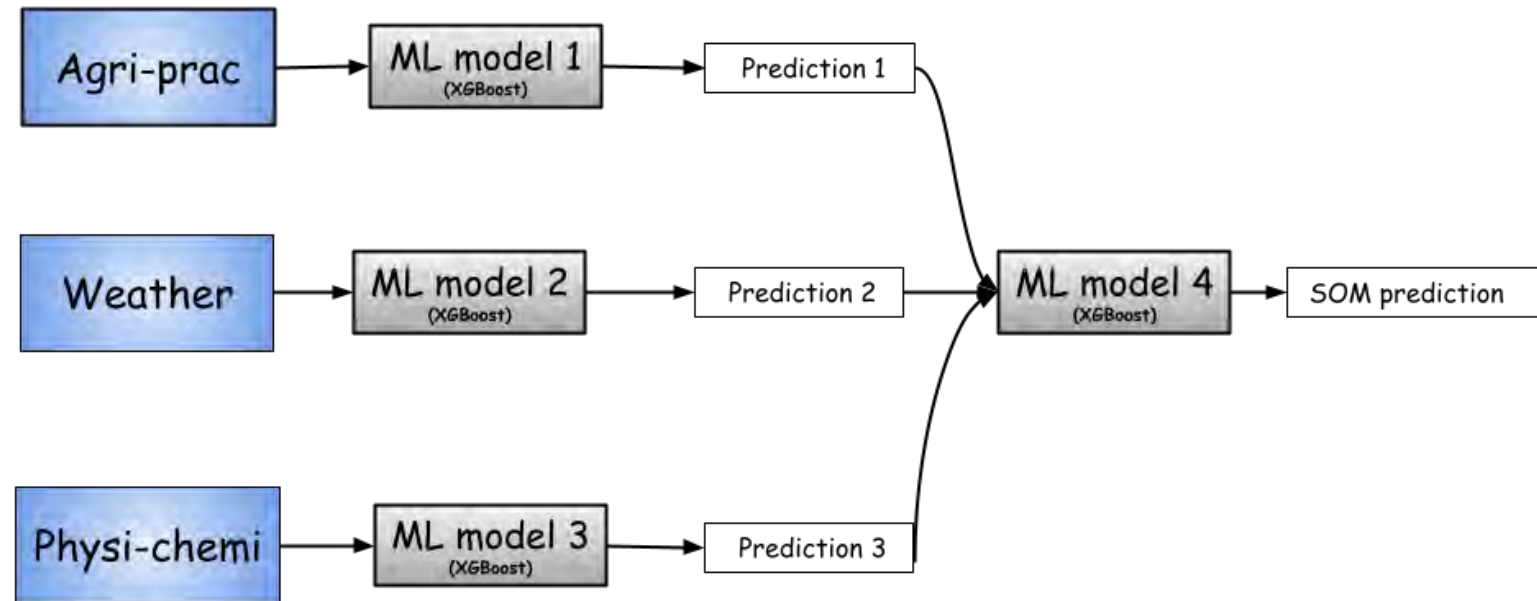


Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

 Approches d'apprentissage en ensemble (Ensemble Learning)

Architecture d'ensemble d'apprentissage multi-couches



Évaluation du modèle

Métriques	R ²	RMSE	MAE
	0,661	0,2918	0,2077

Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches basées sur les transformers pour données tabulaires

Architecture TabTransformer

The TabTransformer architecture comprises a column embedding layer, a stack of N Transformer layers, and a multi-layer perceptron. Each Transformer layer (Vaswani et al. 2017) consists of a multi-head self-attention layer followed by a position-wise feed-forward layer. The architecture of TabTransformer is shown below in Figure 1.

source : Huang, Xin, Ashish Khetan, Milan Cvitkovic, and Zohar Karnin. "Tabtransformer: Tabular data modeling using contextual embeddings." *arXiv preprint arXiv:2012.06678* (2020).

Évaluation du modèle

Métriques	R2	RMSE	MAE
	0,684	0,281	0,192

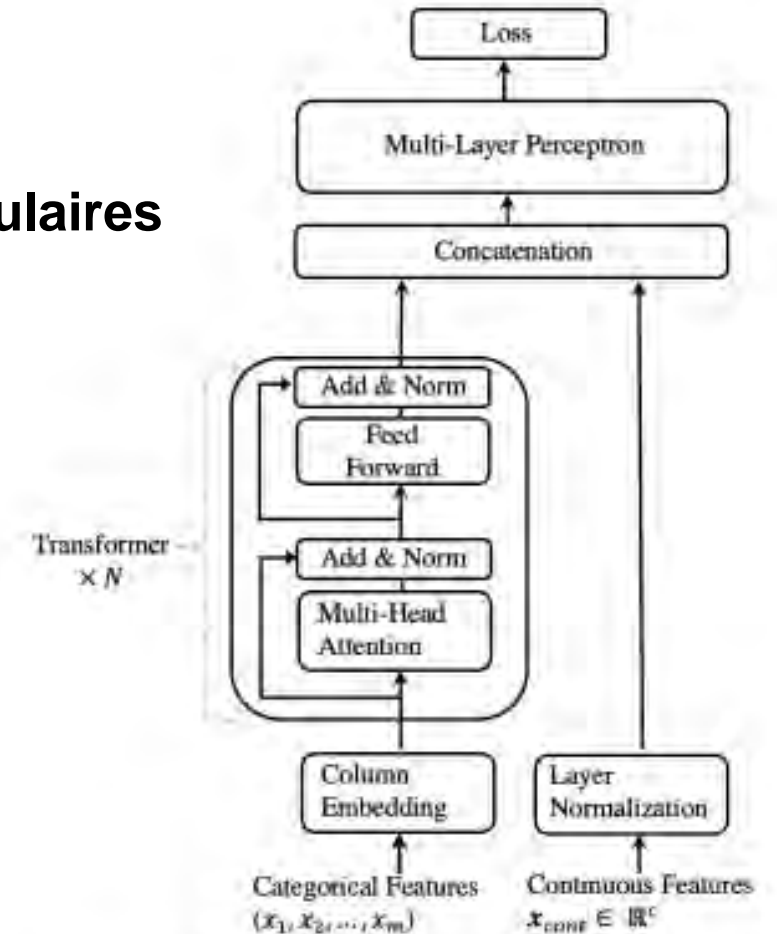


Figure 1: The architecture of TabTransformer.

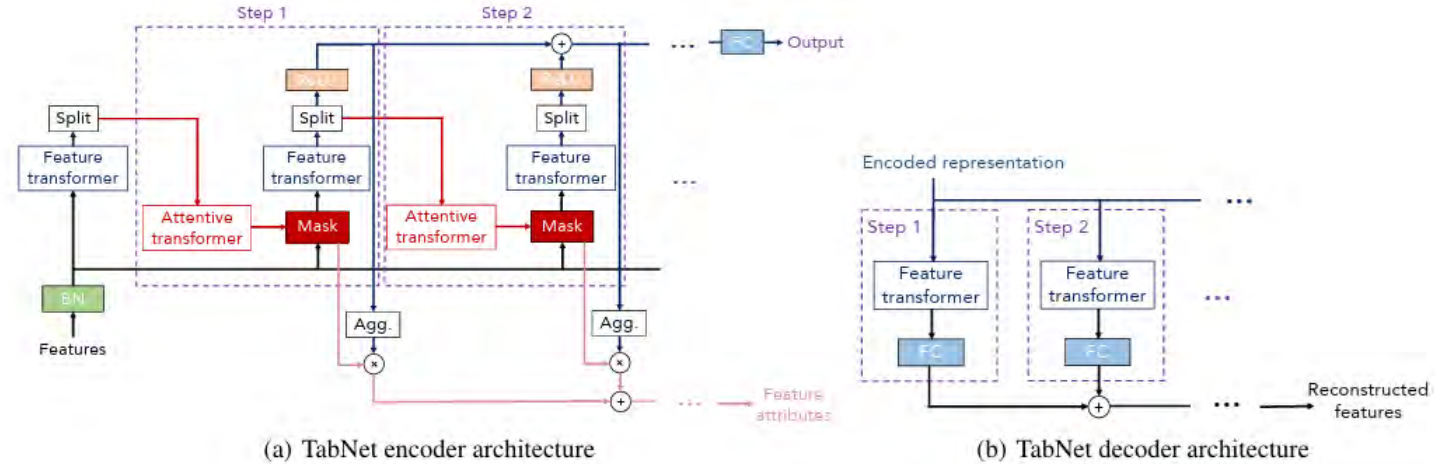
Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



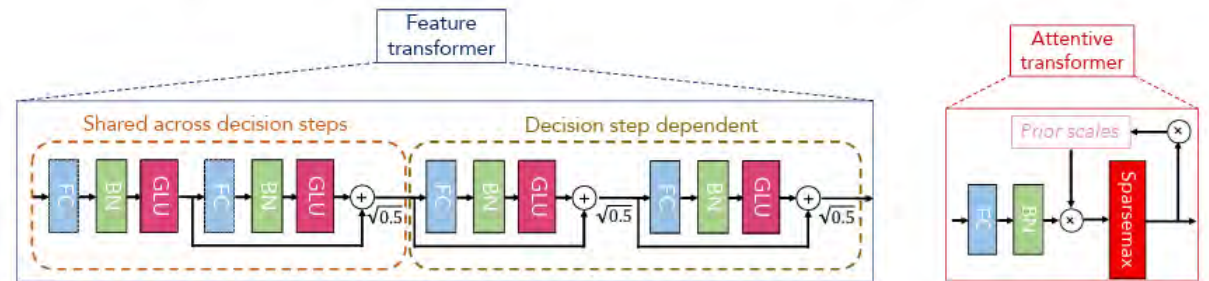
Approches basées sur les transformers pour données tabulaires

Architecture TabNet



Évaluation du modèle

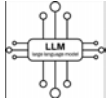
Métriques	R ²	RMSE	MAE
	0,712	0,273	0,197



source : Arik, Sercan Ö., and Tomas Pfister. "Tabnet: Attentive interpretable tabular learning." In Proceedings of the AAAI conference on AI, vol. 35, no. 8, pp. 6679-6687. 2021.

Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches basées sur les grands modèles de langage (LLMs)

RegressLM: Easy Text-to-Text Regression

```
x = '[SOIL] lat=45.66 sand=73.51 clay=9.76 bd=1.50 ph=7.0 soc=0.0696 [WEATHER] precip=[90,77,76,92,106] td
d=[3356,3391,3630,3426,3507] [ROTATION] y1=Corn|_|_ y2=Oats|_|_ y3=Rye|_|_ y4=Soybean|_|_ y5=Barley|_|_ [
MGMT] drain_d=1.0 drain_s=10.0 residues=T tillage=F mineral=[0,0,0,0,0] organic=[0,0,0,0,0]'
```

y =0.0

```
x = '[SOIL] lat=45.13 sand=16.27 clay=24.38 bd=1.50 ph=6.0 soc=0.0174 [WEATHER] precip=[93,74,71,91,90] td
d=[3791,3803,4048,3852,3959] [ROTATION] y1=Soybean|Rye|_|_ y2=Barley|Radish|_|_ y3=Corn|_|_ Annual_grass y4=Cor
n|_|_ Annual_grass y5=Soybean|Rye|_|_ [MGMT] drain_d=1.0 drain_s=10.0 residues=T tillage=T mineral=[0,0,0,0,0
] organic=[68,53,161,158,69]'
```

y =0.2

```
x = '[SOIL] lat=47.50 sand=44.93 clay=21.21 bd=1.50 ph=5.0 soc=0.0522 [WEATHER] precip=[109,91,102,103,88]
tdd=[2772,3042,2752,3063,3142] [ROTATION] y1=Non_legume_hay|_|_ y2=Rye|Beans_green|Cover_crop y3=Spring_w
heat|Beans_green|Cover_crop y4=Beans|Beans_green|_|_ y5=Legume_hay|_|_ [MGMT] drain_d=1.0 drain_s=10.0 resi
dues=T tillage=T mineral=[37,61,36,0,0] organic=[0,0,0,0,0]'
```

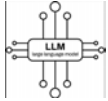
y =0.5

```
x = '[SOIL] lat=46.52 sand=44.93 clay=21.21 bd=1.35 ph=6.0 soc=0.0058 [WEATHER] precip=[123,93,105,121,108
] tdd=[3117,3194,3430,3330,3338] [ROTATION] y1=Spring_wheat|Beans_green|Cover_crop y2=Oats|Radish|_|_ y3=Be
ans|Beans_green|_|_ y4=Potato|Beans_green|_|_ y5=Non_legume_hay|_|_ [MGMT] drain_d=1.0 drain_s=10.0 residues=
T tillage=T mineral=[0,0,0,0,0] organic=[0,0,0,0,0]'
```

y =0.1

Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA



Approches basées sur les grands modèles de langage (LLMs)

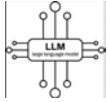
RegressLM: Easy Text-to-Text Regression

Modèle T5X de Google basé sur une architecture transformer encodeur- decodeur ([GitHub - google-research/t5x](https://github.com/google-research/t5x)) :

- Couches d'encodeur : 12, couches décodeur : 12 , têtes d'attention par couche : 12, dimension des têtes d'attention : 64, dimension de l'embeddings : 768, dimension du réseau MLP : 2048.
- Optimisation et entraînements : Adafactor avec taux d'apprentissage ($lr = 0,01$), avec scheduler linéaire; Taille du Batch size = 256
- Encodeur d'entrée : vocabulaire SentencePiece ([GitHub - google/sentencepiece: Unsupervised text tokenizer for Neural Network-based text generation.](https://github.com/google/sentencepiece))
- Décodeur de sortie (y-tokenization) : Tokenization de type P10Tokeniser, IEEEFloatTokeniser pour la représentation des sorties numériques

Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

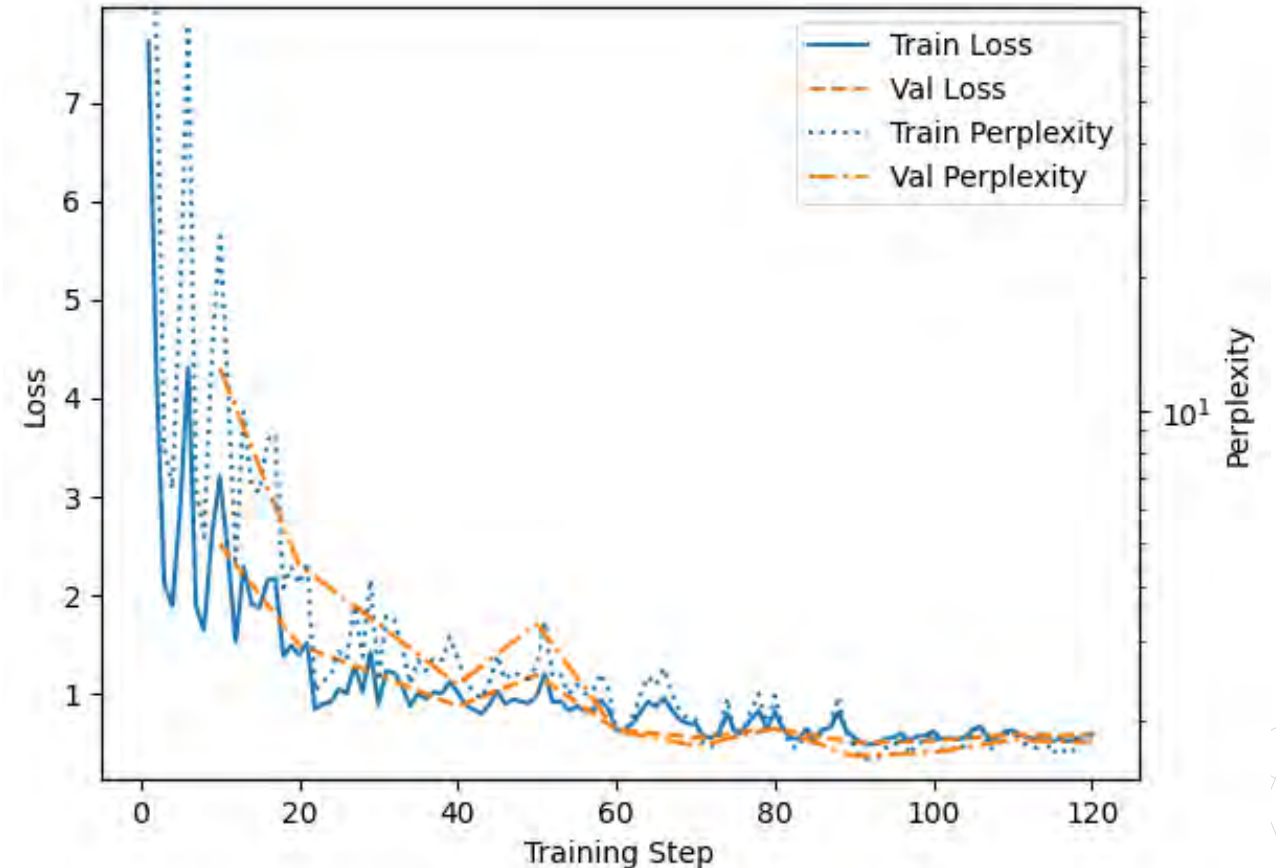


Approches basées sur les grands modèles de langage (LLMs)

RegressLM: Easy Text-to-Text Regression

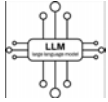
Étape 1 : pré-entraînement

Courbes d'apprentissage : perte d'entraînement vs perte de validation et perplexité avec **P10Tokeniser** comme décodeur se sortie



Réalisations – Volet 4 : IA (Guy)

Modélisation par des algorithmes d'IA

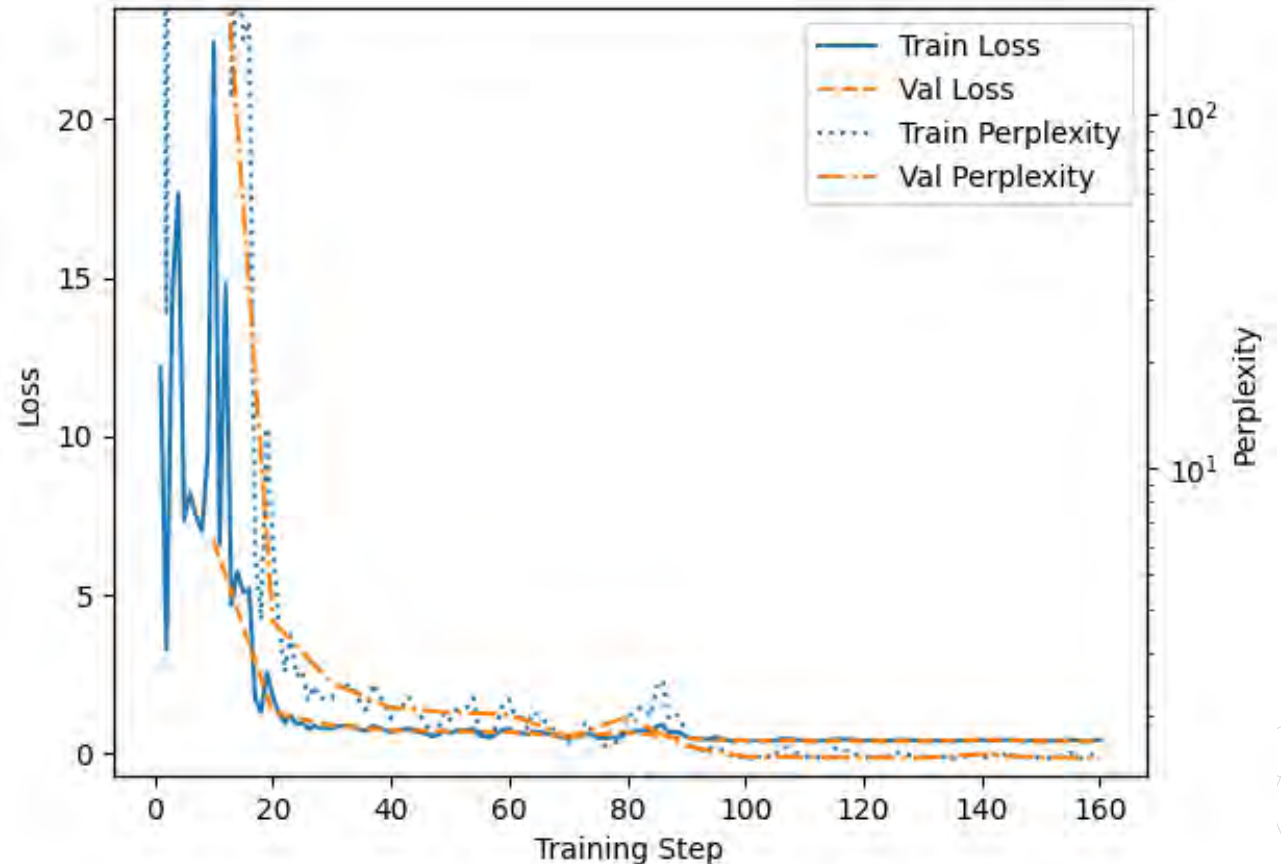


Approches basées sur les grands modèles de langage (LLMs)

RegressLM: Easy Text-to-Text Regression

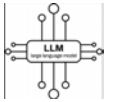
Étape 1 : pré-entraînement

Courbes d'apprentissage : perte d'entraînement vs perte de validation et perplexité avec **IEEFloatTokeniser** comme décodeur se sortie



Réalisations – Volet 4 : IA

Modélisation par des algorithmes d'IA



Approches basées sur les grands modèles de langage (LLMs)

Suite

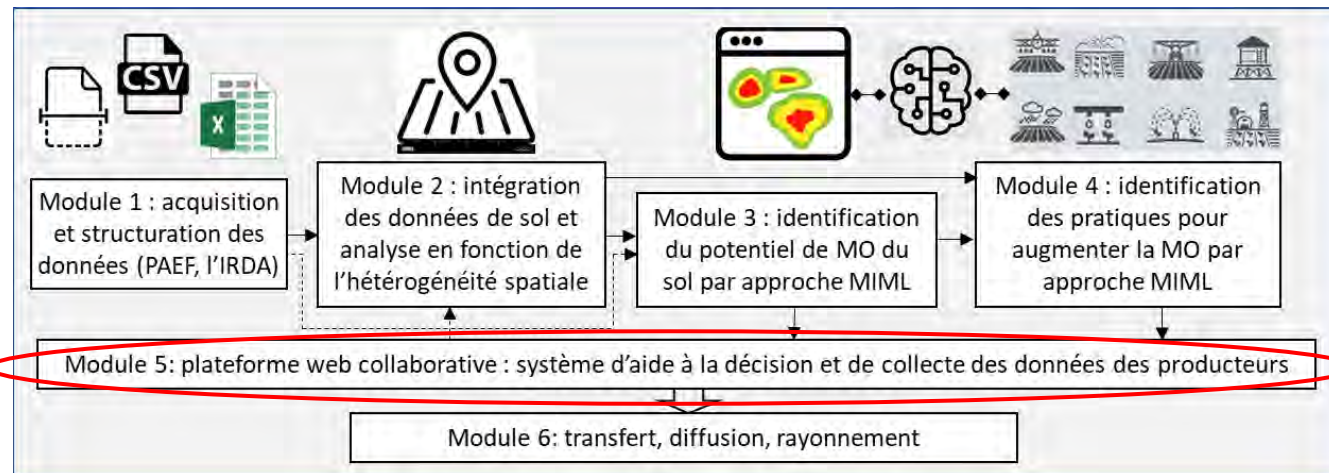
- Approfondir la « voie LLM »
- Utiliser les nouvelles simulations DNDC
- Fine-tuning avec les BDs réelles

Réalisations – Volet 5

SAD – plateforme cartographique

Jérémie Durand, MSc dirigé par Pr Mickaël Germain, Dr Maxime Leduc et Pr Yacine Bouroubi

- Réalisation des *back-end* et *front-end* du SAD
- Stagiaire chez MSF



Réalisations – Volet 5

SAD – plateforme cartographique

Jérémie

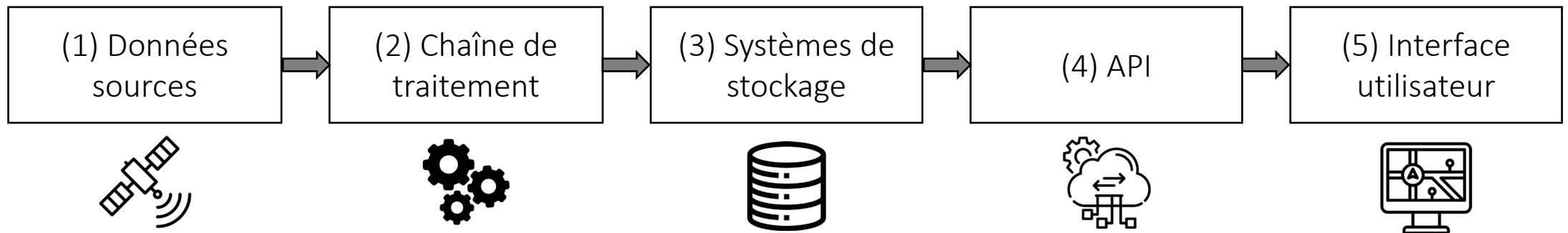


Réalisations – Volet 5 (Jéréemie)

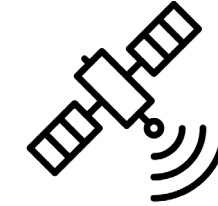
SAD – plateforme cartographique

Objectifs :

- SAD spatial : interopérable, modulaire et standardisé
- Contexte d'agriculture de précision
- Outils et données *open-source*
- Composantes haut-niveau :



Réalisations – Volet 5 (Jérémie)

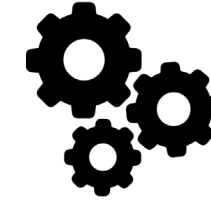


SAD – plateforme cartographique

Données : sources principales (*back-end*)

- Couverture pédologique 2022 et base de données sur les sols (IRDA)
- Base de données des parcelles et productions agricoles déclarées (FADQ)
- Géobase du réseau hydrographique et données topographiques (MRNF)
- Inventaire annuel des cultures, produits dérivés et indicateurs climatiques (AAC)
- Images Sentinel-2

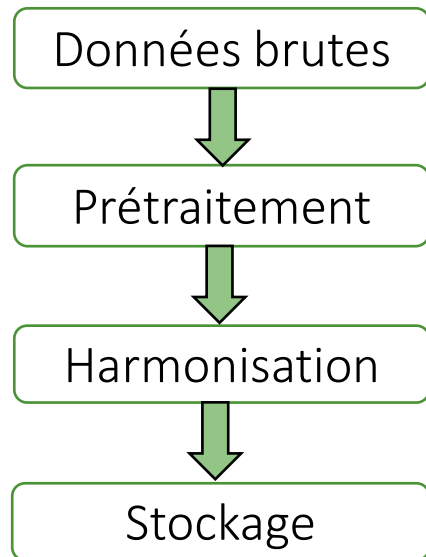




Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique

Chaîne de traitement (*back-end*)



Automatisation avec Python

(A) Vectorielles (points, lignes, polygones)

- Uniformation des attributs
- Reprojection des systèmes de coordonnées
- Exportation en tables spatiales **PostGIS** et fichiers **Parquet**

(B) Matricielles / Raster (GeoTIFF)

- Extraction des métadonnées
- Création de Cloud-Optimized GeoTIFF (**COG**)
- Standardisation des métadonnées en SpatioTemporal Asset Catalog (**STAC**)
- Exportation des métadonnées en table **PostGIS**

Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique



Systemes de stockage des données géospatiales (*back-end*)

- PostGIS (PostgreSQL)
 - Base de données relationnelle spatiale
 - Schémas multiples selon le type de données
- Stockage analytique orienté colonne (Parquet)
 - Données massives, compactes et haute performance
- Moteur de requêtes analytiques (DuckDB)
 - Exploration et agrégation spatio-temporelle
- Cloud-Optimized GeoTIFF (COG)
 - Visualisation et extraction dynamique des rasters
- SpatioTemporal Asset Catalog (STAC)
 - Catalogue de métadonnées (Items & Collections)
 - Découverte, indexation et interopérabilité

STAC – exemple visuel

Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique

Exemple STAC - Collections



The screenshot displays a STAC interface with a grid of data collections. At the top, there are controls for 'Catalogs' (441), 'Tiles', 'List', and sorting options (Ascending/Descending). A search bar is present with the text 'Filter catalogs by title, description or keywords'. Below the search bar, a grid of eight collection cards is shown, each with a thumbnail image and descriptive text:

- CopDEM COG (30 m)**: This Collection provides Copernicus DEM COG products, which represent the surface of the Earth including buildings, infrastructure and vegetation, in...
2010-12-12 00:00:00 UTC – 2015-01-16 00:00:00 UTC
- Sentinel-3 SLSTR Land Surface Temperature (NTC)**: This Collection provides Sentinel-3 SLSTR Level-2 Land Surface Temperature products containing data on land surface temperature measurements on a...
2016-04-19 1:35:17 UTC until present
- Sentinel-5P Level 1 Radiance Band 3 (NRTI)**: This Collection provides Sentinel-5P Level-1 RA BD3 products, which contains Earth radiance spectra for spectral band 3.
2018-04-30 0:19:50 UTC until present
- Sentinel-5P Level 2 Nitrogen Dioxide (NRTI)**: This Collection provides Sentinel-5P Level-2 NO2 products, which contains high-resolution imagery of nitrogen dioxide concentrations.
2018-04-30 0:18:51 UTC until present
- CopDEM COG (90 m)**: This Collection provides Copernicus DEM COG products, which represent the surface of the Earth including buildings, infrastructure and vegetation, in...
- Sentinel-3 SLSTR Radiance and Brightness Temperature (NRT)**: The Copernicus Sentinel-3 SLSTR RBT
- Sentinel-5P Level 1 Radiance Band 3 (OFFL)**: This Collection provides Sentinel-5P Level-1 RA BD3 products, which contains Earth radiance spectra for spectral band 3.
- Sentinel-5P Level 2 Nitrogen Dioxide (OFFL)**: This Collection provides Sentinel-5P Level-2 NO2 products, which contains high-resolution imagery of nitrogen dioxide concentrations.

Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique



Exemple STAC - Items

Description

This Collection provides Copernicus DEM COG products, which represent the surface of the Earth including buildings, infrastructure and vegetation, in resolution of 30 meters.

[Copernicus](#)
[ESA](#)
[Satellite](#)
[Global](#)
[DEM](#)
[EU](#)
[EC](#)

License Legal notice on the use of Copernicus DEM Data and Service Information
Temporal Extent 2010-12-12 00:00:00 UTC – 2015-01-16 00:00:00 UTC

+
Map

-
Map

Leaflet | © OpenStreetMap contributors.

Asset

> COPDEM COG Thumbnail THUMBNAIL JPEG

Providers

ESA PRODUCER

Items

< First
< Previous
Next >
Show Filters

<p>Copernicus_DSM_COG_10_N37_00_E051_00_DE M COG 2020-11-12 00:10:01 UTC – 00:11:01 UTC</p>	<p>Copernicus_DSM_COG_10_N39_00_E051_00_DE M COG 2020-11-12 00:10:07 UTC – 00:11:07 UTC</p>
<p>Copernicus_DSM_COG_10_N37_00_E052_00_DE M COG 2020-11-12 00:10:02 UTC – 00:11:02 UTC</p>	<p>Copernicus_DSM_COG_10_N40_00_E051_00_DE M COG 2020-11-12 00:10:08 UTC – 00:11:08 UTC</p>
<p>Copernicus_DSM_COG_10_N38_00_E050_00_DE M COG 2020-11-12 00:10:03 UTC – 00:11:03 UTC</p>	<p>Copernicus_DSM_COG_10_N41_00_E050_00_DE M COG 2020-11-12 00:10:09 UTC – 00:11:09 UTC</p>
<p>Copernicus_DSM_COG_10_N38_00_E051_00_DE M COG 2020-11-12 00:10:04 UTC – 00:11:04 UTC</p>	<p>Copernicus_DSM_COG_10_N41_00_E051_00_DE M COG 2020-11-12 00:10:10 UTC – 00:11:10 UTC</p>

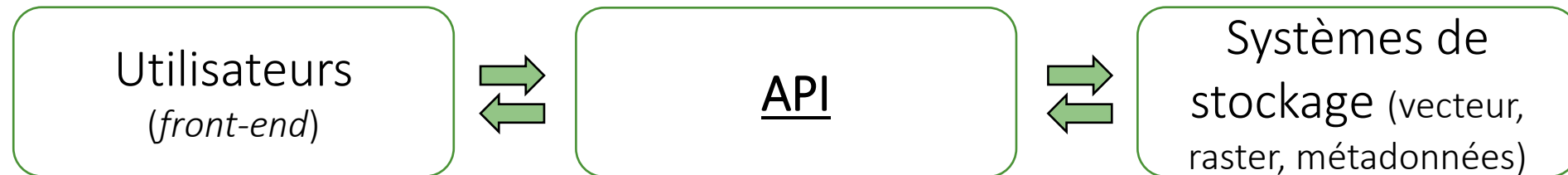
Réalisations – Volet 5 (Jérémie)



SAD – plateforme cartographique

API d'accès aux données géospatiales (*back-end*)

- Principe d'une API : liens entre les utilisateurs et les systèmes de stockage via un modèle requête-réponse
- *Open Geospatial Consortium* (OGC)
- Publication de services web spatiaux **interopérables** et **standardisés**



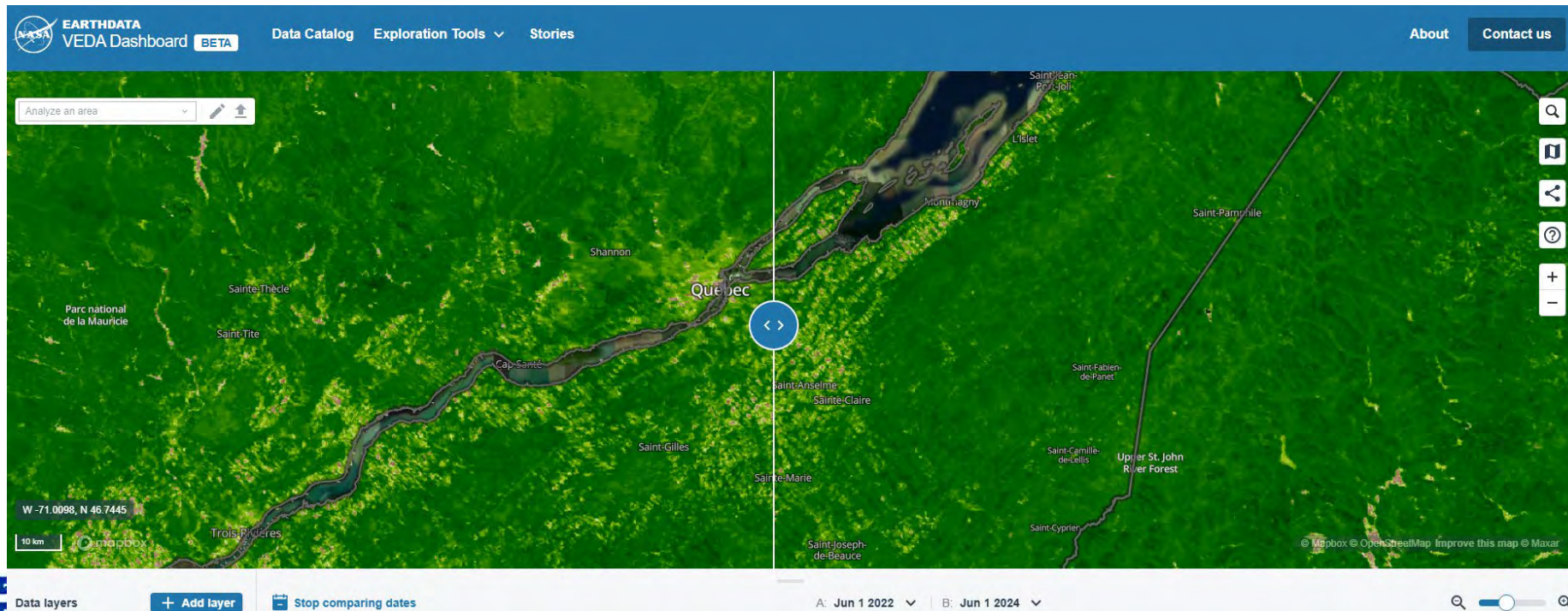
Réalisations – Volet 5 (Jérémie)



SAD – plateforme cartographique

Interface utilisateur (*front-end*)

- VEDA : Projet open-source de la NASA pour plateforme web infonuagique
- Intégration des données et APIs



Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique

Intégration modèle MOS

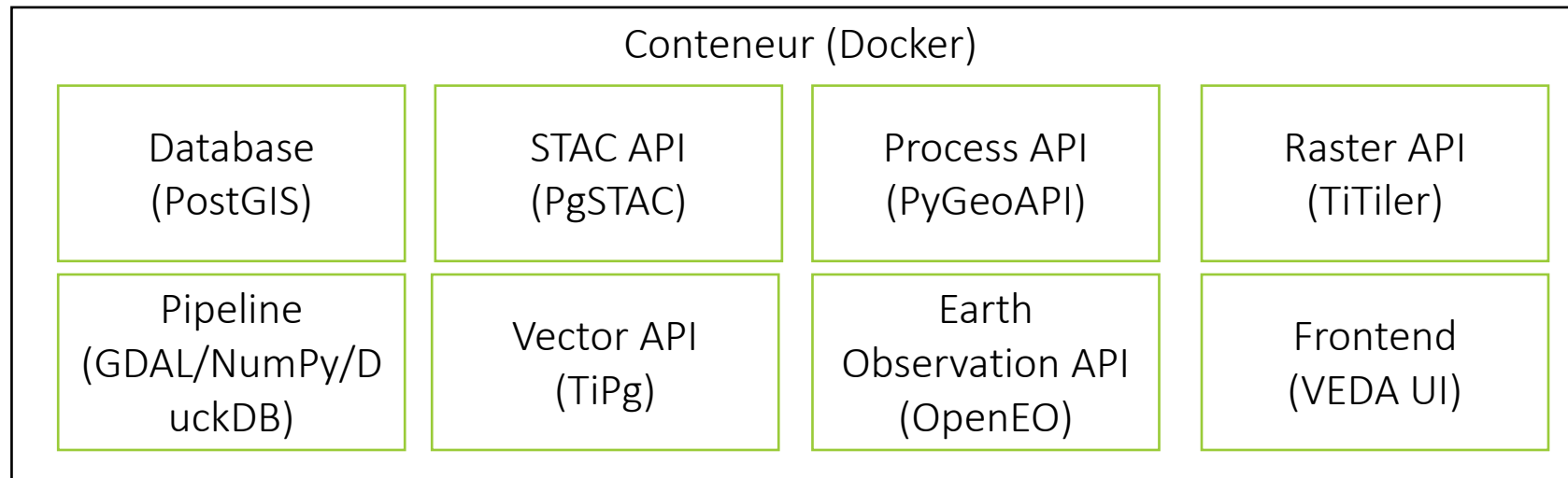
- Modèle exposé sous un service de traitement géospatiaux OGC API – Processes
- Utilisation de l'API :
 1. Requête d'exécution du modèle avec valeurs d'entrées spécifiées par l'utilisateur
 2. Récupération des données via APIs
 3. Calcul et stockage (GeoJSON) à la volée
 4. Résultat attendu : MOS estimée en fonction des valeurs d'entrées
- Solution interopérable, réutilisable et reproductible

Réalisations – Volet 5 (Jérémie)

SAD – plateforme cartographique

Architecture modulaire

- Microservices : déploiement et maintenance indépendante
- Application conteneurisée : communications entre services
- Basé sur eoAPI



Réalisations – Volet 5 (Jérémie)


SAD – plateforme cartographique

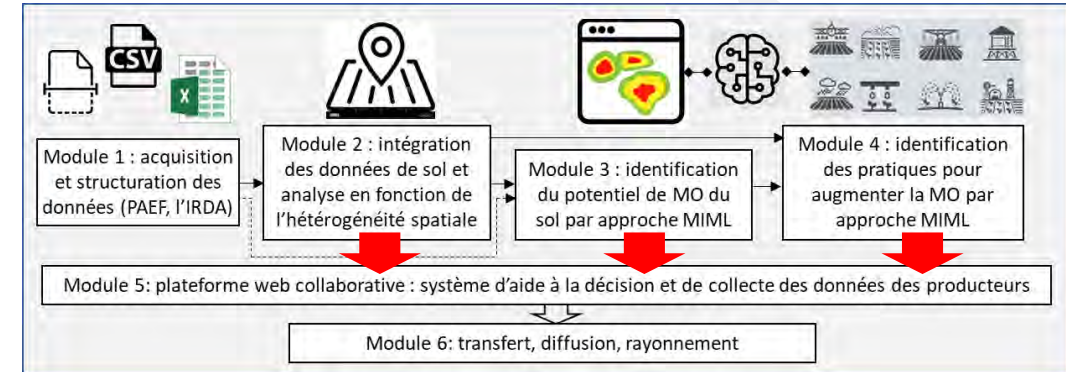
Suite

- Finalisation du *front-end*
- Intégration *back-end*– *front-end*
- Tests et diffusion

Fin du projet : mars 2027 au lieu de mars 2026

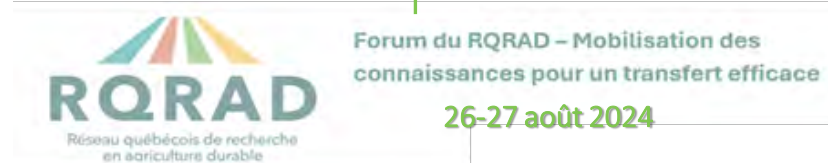
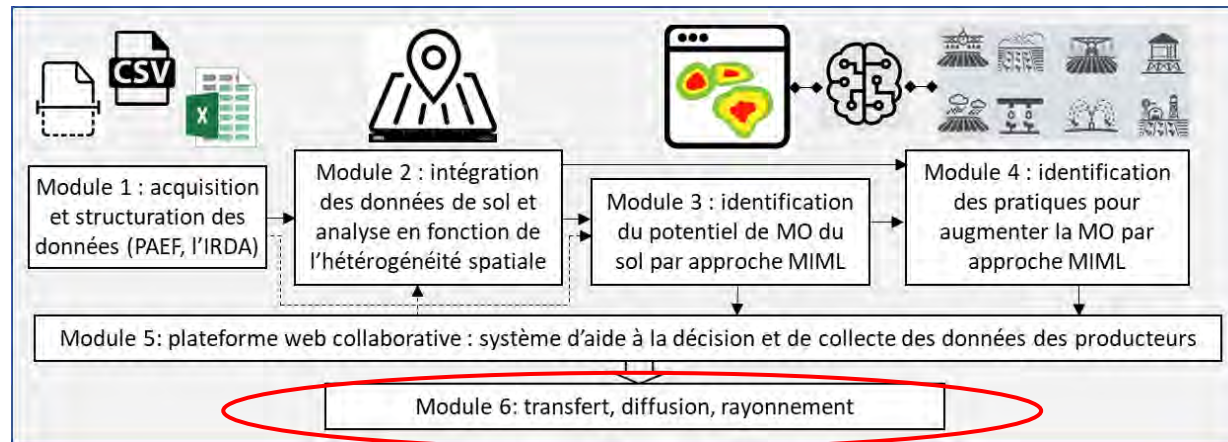
Reste à faire

- Plateforme : connecter les modules (en cours)
 - Spatialisation de la MOS à l'échelle des parcelles : modèle de Hamed avec les données de Jérémie
 - Potentiel de MOS et meilleures pratiques : modèle de Guy avec les données de Jérémie
- Adaptation des modules aux besoins des partenaires
 - MSF : déjà avancé
 - IRDA et LogiAg
- Aspect économique
 - Rendement généré par DNDC
 - Autre information économique : CRAAQ
- Transfert 



Reste à faire

Activités de transfert – Volet 6



- Plateforme numérique
 - En étroite collaboration avec MSF
- Articles scientifiques
 - 3 publiés, 2 en cours, autres prévus
- Articles de vulgarisation
 - Plusieurs prévus (ex.: Forum d'août 2024)
- Conférences
 - 6 pour le moment (RQRAD, St-Rémi, CRAAQ)
- Formations
 - 1 pour le moment (RQRAD, atelier modélisation)
- Partenaires
 - MSF: très avancé
 - IRDA et Logiag : contacts établis
 - CRAAQ, Coord. services-conseils : bientôt

Retombées attendues (selon les objectifs du PAD)

Retombées technologiques, environnementales, économiques et sociales :

- Création de nouvelles connaissances intégrées dans une plateforme numérique pour fournir des recommandations de pratiques d'amélioration de la MOS.
- Les partenaires du projet bénéficieront des outils générés pour enrichir leurs portfolios de services.
- Réduire incertitude sur les pratiques de conservation ainsi que les freins à l'adoption par les producteurs de bonnes pratiques grâce à des solutions adaptées aux parcelles.
- La carte de potentiel d'augmentation de MOS utile au gouvernement du Québec afin d'implanter de nouvelles politiques d'utilisation des sols agricoles comme puits de carbone.
- Les PHQ formés contribueront à renforcer le marché du travail de l'agriculture durable au Québec.

Merci



**Yacine Bouroubi; Maxime Leduc;
Rami Albasha; Hamed Etezadi; Guy
Armel Fotso Kamga; Jérémie Durand**

