

Agence spatiale canadienne

1. FICHE DOCUMENTAIRE DE RAPPORT

ASC

Date du rapport : 28 octobre 2016

Titre : Rapport sommaire

Développement d'une méthodologie opérationnelle de suivi des rendements à l'aide de l'imagerie RADARSAT-2 dans deux cultures stratégiques pour le Canada : maïs et soja

Auteurs : Karem Chokmani, Simon Perreault, Yves Gauthier, Jimmy Poulin, Monique Bernier, Lucie Vallée, Dominic Thériault, Alain Brassard et Hicham Bencharki

Nom et adresse de l'organisme réalisant les travaux :

Centre Eau Terre Environnement, Institut national de la recherche scientifique
490, rue de la Couronne, Québec (Québec) Canada G1K 9A9

N° et titre du contrat : PDAOT no. 9F043-130644/012/MTB

Agence spatiale canadienne

6767, route de l'Aéroport
Saint-Hubert (Québec) Canada J3Y 8Y9

Autorité technique : Anna Pacheco

Autorité du projet : Pierre Nadeau

Résumé : Au Canada comme au Québec, les cultures du maïs et du soja occupent des places importantes dans notre économie agricole. Ce projet a pour but de développer une méthode qui permettra aux agriculteurs canadiens de maïs et de soja, d'obtenir des prévisions du rendement potentiel de ces cultures à des moments clés durant la saison de croissance à partir de l'imagerie RADARSAT-2. La zone d'étude est le sud du Québec (Montérégie, Centre du Québec, Lanaudière et Mauricie). L'originalité de la recherche proposée ici découle de l'utilisation potentielle d'un très grand nombre de données (images RS-2 et rendements agricoles), à l'échelle fine de la parcelle, sur un grand territoire et sur plusieurs années. Les données de rendements moyens à la parcelle acquises sur toute la zone d'étude pendant les saisons de croissance végétatives de 2008-2013 ont été fournies par la Coop Fédérée. Ceux-ci ont également récolté des données terrain sur une quarantaine de parcelles en 2014 et 2015. Près de 400 images RADARSAT-2 d'archives ont été fournies par l'Agence spatiale. Celles-ci ont été intercalibrées afin de réduire l'effet de l'angle d'incidence. Toutes ces données ont été structurées dans une base données à référence spatiale (PostgreSQL) « open source ». Le modèle STICS a été calibré et utilisé pour estimer le stade phénologique de chaque parcelle à chacune des dates d'acquisition des images d'archives RADARSAT-2. Après analyse des paramètres d'acquisition (orbite, polarisation, résolution), des cultures (maïs et soja), du stade phénologique, des conditions au sol (pluie, rosée) et de la taille des parcelles, des modèles de prévision du rendement sont proposés. Les meilleures corrélations entre le coefficient de rétrodiffusion et le rendement final surviennent pour le soja, durant la fenêtre de stades phénologiques 55-71, correspondant au maximum de l'indice de surface foliaire. Une procédure de validation croisée a été menée sur les modèles. La qualité d'ajustement et la qualité d'estimation des modèles proposés pour le soja vont de acceptable à souhaitable. Les résultats pour le maïs sont pour l'instant non concluants mais un travail d'harmonisation des données de rendement doit être effectué. Les routines pour appliquer le modèle ont été automatisées.

Mots clés : Rendement, Estimation, Maïs, Soja, RADARSAT-2, STICS, Phénologie, Intercalibration

Notes supplémentaires :

Distribution/Disponibilité :

2. Introduction

Les deux plus grandes productions végétales québécoises, en termes de recettes monétaires, sont le maïs et le soya. À elles-seules, elles ont représenté 63 % des recettes des productions végétales en 2012 (MAPAQ *et al.*, 2013). Ceci témoigne de l'importance de ces cultures pour la production agricole du Québec. En 2012, les agriculteurs québécois ont dépensé 321,4 millions de dollars en achat d'engrais et de chaux afin de maintenir les sols fertiles. Ils ont également investis 116,4 millions de dollars dans l'achat de pesticides. La facture totale de ces dépenses s'avère couteuse pour l'ensemble des agriculteurs désirant maximiser le rendement de leurs productions. En plus de ces coûts économiques, l'application de ces produits a notamment des conséquences écologiques. Il s'avère alors essentiel que des outils d'aide à la décision soient développés et utilisables pour chaque agriculteur afin d'optimiser le rendement de sa production.

Depuis plusieurs années l'agriculture de précision s'est donc développée de plus en plus dans le milieu agricole du Québec. Contrairement aux pratiques traditionnelles de l'agriculture où les opérations au champ s'effectuaient de manière uniforme ou encore selon la meilleure des connaissances de l'agriculteur, l'agriculture de précision vise à connaître la variabilité au champ et à limiter les interventions de l'agriculteur aux endroits jugées nécessaires. L'intérêt porté par les agriculteurs envers ces produits concerne deux aspects essentiels : la gestion de la variabilité intra-parcellaire à des moments clés de la saison pour l'optimisation du rendement ainsi que le point de vue administratif de la performance de l'entreprise (prévision des ventes, stockage de la récolte, transformation du produit, etc.).

L'imagerie satellitaire multi-spectrale est déjà utilisée pour produire des cartes de rendement. Les canaux du proche-infrarouge et de l'infrarouge sont sensibles à la concentration de chlorophylle présente dans les plantes. Les images permettent de quantifier avec un certain succès, la biomasse (LAI) en fonction du stade phénologique de la plante. La COOP Fédérée offre déjà certains produits à ses membres. Afin de compléter les produits optiques qui sont toujours soumis aux aléas de la couverture nuageuse, les données radar représentent une avenue potentielle qui intéresse fortement ce partenaire. En effet, il est connu que la rétrodiffusion du signal radar répond en fonction de la structure de la plante, ses tiges et ses feuilles, particulièrement avec les polarisations VV et croisées (HV ou VH). Des études ont aussi prouvé qu'il existe une relation entre le LAI (Leaf Area Index), et par conséquent le rendement d'une culture et le coefficient de rétrodiffusion du signal radar en bande C. Par contre, ces travaux ont généralement été faits en milieu contrôlé et à petite échelle.

3. Objectifs techniques

L'objectif premier du projet est de tirer profit de ces connaissances pour développer une méthode permettant aux producteurs canadiens d'obtenir des prévisions du rendement potentiel à des moments clés de la saison et l'estimation du rendement à la récolte pour les cultures du maïs et du soya à partir d'images RADARSAT-2. Les objectifs spécifiques sont :

- Colliger un jeu de données suffisamment grand pour pouvoir dégager des relations significatives entre la rétrodiffusion radar et le rendement moyen à la parcelle.
- Déterminer quelles configurations du capteur permettent la meilleure relation entre le rendement et le signal radar.
- Déterminer à quel(s) stade(s) phénologique(s) de croissance on peut avoir la meilleure estimation du rendement potentiel.

La zone d'étude inclue la Montérégie, le Centre du Québec, Lanaudière et la Mauricie (basses-terres du St-Laurent).

4. Approches et tâches du projet

Le projet contient deux volets principaux. Le premier concerne le développement et la calibration du modèle liant éventuellement le signal radar au rendement des cultures de maïs et de soya. Ce travail est effectué à partir d'images RADARSAT-2 d'archive. Le second volet concerne la validation du modèle développé. Ce travail s'effectue à partir de nouvelles acquisitions d'images RADARSAT-2 coordonnées avec des visites sur le terrain.

L'approche proposée se base sur un jeu de données inédit. En effet, la plupart des travaux de recherches sont effectués à des échelles allant de locale à régionale (principalement par comtés aux États-Unis). Lorsque les études sont faites au niveau du champ, les rendements sont mesurés dans un contexte restreint de quelques parcelles expérimentales contrôlées. Le jeu de données de rendements de ce projet, fourni par la COOP Fédérée, a des caractéristiques uniques :

- Échelle parcellaire;
- Couverture de l'ensemble du sud du Québec;
- Milieu non-contrôlé qui représente la réalité des rendements des cultures du maïs et du soya du sud du Québec;
- Pluriannuelle (2008-2013).

Pour la même période, le sud du Québec est couvert par des centaines d'images RADARSAT-2, en archives et qui sont rendues disponibles par l'Agence spatiale canadienne. Ainsi, ce projet vise à tirer profit du croisement de ces deux sources de données de masse afin d'établir des liens statistiques

entre la rétrodiffusion radar et le rendement agricole à la parcelle. Les approches statistiques employant des régressions reposant sur des estimations de la moyenne sont sensible aux extrêmes et par le fait même au nombre de données.

Sept tâches principales ont été nécessaires afin de mener le projet à terme :

1) Mise sur pieds d'une base de données

Le traitement et la gestion de cette grande quantité de données a nécessité un travail colossal en amont du développement des modèles. La première tâche du projet était donc de structurer cette information dans une base de données à référence spatiale (PostgreSQL) « open source ». Ceci permettant ensuite de pouvoir ultérieurement faire des requêtes rapides qui permettent de visualiser, de comprendre et d'analyser les données et leurs interrelations. L'unité de base est donc la parcelle. On y adjoint d'abord les informations provenant du producteur (culture, variété, date de semis, rendement moyen à la fin de l'année) lorsqu'elles sont disponibles. On ajoute ensuite pour chaque parcelle, les données radar (date d'acquisition, heure, azimuth, angle d'incidence, polarisation, résolution, rétrodiffusion moyenne). On complète enfin avec les données environnementales (type de sol dominant, température et précipitation journalières). Il n'y manque alors qu'une information qui est essentielle pour le développement des modèles : le stade phénologique de la culture sur la parcelle, lors de chaque acquisition d'image.

2) Estimation du stade phénologique

Évidemment, il n'existe aucune information sur l'état végétatif des plantes sur chaque parcelle, au moment des acquisitions d'images de 2008 à 2013. Toutefois, dans l'optique d'établir des corrélations statistiques, on se doit de mettre en lien des données comparables, c'est-à-dire dans le cas qui nous occupe, des parcelles qui sont au même stade phénologique. Ici, cette variable a été estimée à partir du modèle français STICS (Simulateur multIdisciplinaire pour les Cultures Standard), un modèle de culture dynamique, générique et robuste, reconnu internationalement et permettant de simuler le système sol-atmosphère-culture. Le modèle a dû être adapté pour la zone d'étude, paramétré avec les données disponibles, appliqué pour chaque année et pour chaque parcelle de maïs et de soya et validé.

3) Travaux de terrain (2014-2015)

En parallèle au développement des modèles, il est aussi essentiel de préparer un jeu de données de validation. Ainsi, pour les saisons de croissance 2014 et 2015, des campagnes de terrain exhaustives ont été menées par la COOP Fédérée chez quatre producteurs de la Montérégie. Les mesures récoltées incluaient le pourcentage de recouvrement de l'inter-rang par les plantes, la hauteur de la plante, l'humidité du sol (qualitatif), la quantité de mauvaises herbes (qualitatif) et la caractérisation du

stade de développement phénologique. Ces visites étaient menées selon le calendrier de nouvelles acquisitions RADARSAT-2, qui prévoyait une vingtaine d'images réparties entre juillet et novembre à chaque année. Pour ces parcelles, le rendement en fin de saison a été récupéré auprès du producteur, lorsque possible.

4) Traitement des images radar

Les images RADARSAT-2 récupérées en archive et certaines des nouvelles acquisitions devaient d'abord être corrigées géométriquement et radiométriquement. Le premier traitement est une orthorectification basée sur les paramètres d'orbite et le modèle numérique de terrain. Le second traitement vise une fois de plus, à rendre les données comparables entre elles. En effet, les images d'archive peuvent avoir été acquises selon une variété de paramètres différents (orbite, faisceau, polarisation, résolution). Une intercalibration a ainsi été effectuée afin de réduire la dépendance du signal à l'angle d'incidence entre les images.

5) Développement des modèles

Tout le travail de développement statistique des modèles consiste à tester et sélectionner les combinaisons idéales pour obtenir la meilleure estimation du rendement à venir. La première étape vise à regrouper les couples selon la culture, l'intervalle de stades phénologiques estimé par STICS d'après le jour julien de l'image, par groupe de polarisation (HV-VH et VV) et par orbite (ascendante et descendante) et par résolution spatiale si nécessaire. Pour les couples d'une même parcelle/année/culture/phénologie ayant une image avec les polarisations VV et VH, le rapport de polarisation VV/VH est également généré en plus des polarisations disponibles. La régression linéaire « Stepwise » est ensuite appliquée sur chacun des regroupements de manière à ne sélectionner que les variables ayant une contribution significative à l'estimation du rendement (seuil de signification fixé à 5%). Le développement des modèles s'est fait de façon itérative, en retenant ou en éliminant des couples, jusqu'à l'obtention des meilleures corrélations. Entre chaque itération, les étapes précédentes étaient revues afin d'améliorer la qualité des données.

6) Validation des modèles

Pour obtenir un véritable portrait de la qualité d'estimation des modèles retenus, il faut utiliser à une méthode de validation dont les couples utilisés n'ont pas servis à la calibration. Les couples des données 2014-2015 ont été jugés plus utiles à la calibration des modèles qu'à leur validation car au final, leur nombre n'était pas suffisant pour en tirer des conclusions. Une procédure de validation croisée a donc été sélectionnée puisqu'elle permet d'exploiter toutes les données dans la calibration sans en sacrifier et ce, tout en n'étant pas biaisée dans la validation. La stratégie consiste à retirer aléatoirement des couples, puis de calculer un modèle avec les couples restants et ensuite d'estimer

les rendements avec ce modèle pour les couples qui ont été retirés. Dans le cas présent, 10% des couples sont retirés et servent à la validation alors que 90% servent à l'entraînement. L'exercice a été répété 10 000 fois avec des tirs aléatoires avec remise. À chaque fois les rendements mesurés et estimé sont enregistrés. À la toute fin des itérations, les R^2 , RMSE et biais sont calculés.

7) Automatisation des procédures

Comme il s'agit d'un projet de recherche et de développement d'une application en observation de la terre, il est important que les modèles développés se concrétisent sous une forme facilitant le transfert technologique et l'application du modèle. Dans le cas présent, la solution concrète envisagée correspond à une interface graphique pour l'utilisateur. Deux interfaces ont été développées pour ce projet : soit une version s'intégrant dans une boîte à outils dans ArcGIS et une autre version s'exécutant à l'extérieur d'ArcGIS (Standalone) mais qui fait quand même appel aux bibliothèques de ce SIG. Ces interfaces effectuent l'ensemble des prétraitements et traitements requis pour aboutir à l'obtention d'une valeur de rendement estimée en appliquant un modèle de rendement.

5. Défis et réalisations

1) Quantité, diversité et qualité des données

La gestion d'une grande quantité de données est toujours un défi. Le premier accomplissement du projet réside donc dans la structuration de cette information, dans une base de données à référence spatiale (PostgreSQL) « open source ». Mais cet avantage du projet quant à la quantité de données disponibles s'est aussi transformé en un autre défi : une très grande diversité en termes de caractéristiques et en termes de qualité. Ainsi, il nous est apparu au fil des traitements et des analyses, que les données de rendement ne pouvaient toutes être utilisées de façon optimale. Deux types de données de rendement ont été utilisés soit des rendements déjà moyennés à la parcelle (2011 à 2013) et des données de capteurs de rendement (2008 à 2010 et 2014-2015). Les données de capteurs de rendement de 2008 à 2010 ce sont avérées inutilisables dû à d'importantes discontinuités spatiales alors que celles de 2014-2015 ont pu être converties en rendements moyens. Ces dernières ont été intégrées aux jeux de données de calibration déjà composé des rendements moyens de 2011 à 2013. Une stratégie visant à éliminer les valeurs extrêmes a aussi été utilisée afin de nettoyer les rendements moyens (2011 à 2013). Toutefois, même après des travaux de mise en forme et de nettoyage, nous avons dû laisser de côté une partie significative des données, ce qui a réduit le nombre de couples disponibles pour le développement des modèles. Puisqu'une telle base de données historique de rendements est inédite, il serait dommage de ne pas l'exploiter de façon optimale. Il est donc suggéré de poursuivre le travail avec la COOP Fédérée afin de réduire l'incertitude sur les données de

rendement en archive et de mettre en place des moyens de contrôler cette incertitude sur les données de rendement des années à venir.

2) L'estimation du stade phénologique

La nécessité d'estimer le stade phénologique pour toutes les images d'archive a été expliquée précédemment. Cela représentait un défi énorme, qui a été relevé par l'adaptation et l'utilisation du modèle STICS. Entre autres, nous avons mis beaucoup d'attention à optimiser le paramètre d'entrée « date de semis » puisque le modèle y est très sensible. Dans l'ensemble, la prédiction de STICS reflète la croissance des cultures sur le terrain, avec toutefois un niveau d'incertitude plus élevé pour le soya. En général, le modèle retarde de quelques jours sur la réalité. Mais cela nous assure que sur les images retenues, les cultures ont bel et bien atteint le stade phénologique recherché.

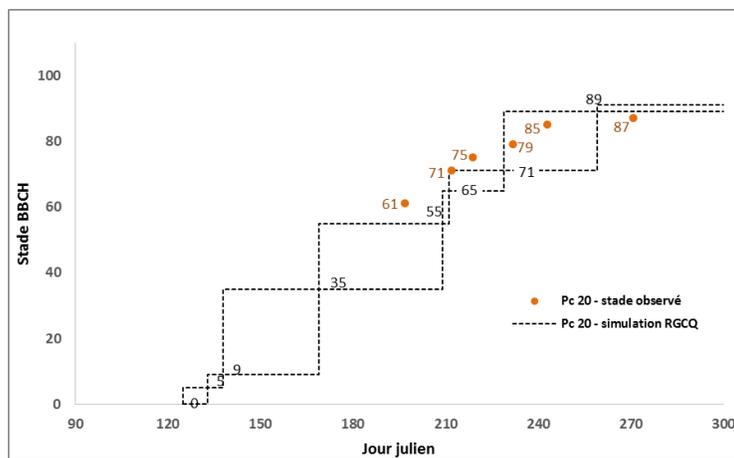


Figure 1: Comparaison des fenêtres de stades phénologiques estimés par la simulation STICS au cours de la saison de végétation 2015 et des stades observés lors des visites sur le terrain pour la parcelle maïs-grain # 20.

3) L'intercalibration des images RADARSAT-2

L'utilisation d'images radar d'archives (paramètres d'acquisition variés) a aussi posé un défi afin de pouvoir comparer entre eux, des coefficients de rétrodiffusion extraits d'images acquises selon des orientations et des angles d'incidence différents. Dans un premier temps, l'analyse de nos données ne montrait aucun biais du signal en fonction de l'orientation (orbite ascendante ou descendante), pour des conditions au sol similaires. Par contre, ces mêmes analyses montraient une dépendance à l'angle d'incidence. Le troisième accomplissement du projet a donc été l'application d'une approche d'intercalibration, afin de réduire la dépendance du signal à l'angle d'incidence. Celle-ci consiste en premier lieu, à extraire pour chacune des images, les signatures radar de l'eau dite « pure » et « calme ». À chaque polygone, la moyenne de rétrodiffusion et le coefficient de variation sont calculés. Un seuil maximal de coefficient de variation de la rétrodiffusion est fixé. La procédure d'intercalibration vise à ramener les histogrammes de la signature de l'eau extraite, à une valeur de référence. Il s'agit

d'une translation de l'histogramme de l'eau dans l'espace dB. Les résultats sont très bons avec les données de polarisation croisée mais pour la polarisation VV, il reste encore de l'incertitude après l'application de l'intercalibration. Nous avons donc conservé un levier (coefficient de variation) nous permettant de retirer une image de l'analyse, si la correction de l'angle d'incidence s'avérait inefficace. Notons aussi que nous avons observé une plus grande sensibilité des images du matin (orbite descendante) à la présence de rosée. Nous avons donc priorisé les images de fin de journée (orbite ascendante) pour le développement des modèles.

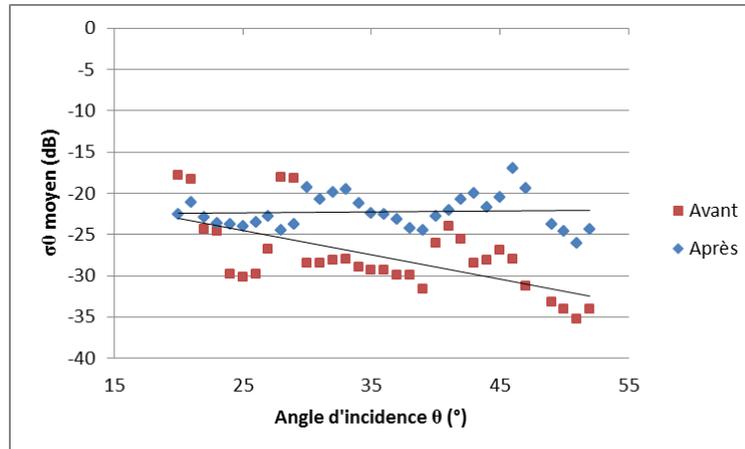


Figure 2: Illustration du résultat de l'intercalibration des images (polarisations HV, VH) sur la dépendance du coefficient de rétrodiffusion radar à l'angle d'incidence.

4) Développement des modèles

Pour le développement des modèles, le plus grand défi a été de déterminer des relations significatives, tout en conservant un nombre de couples « parcelle/rétrodiffusion/rendement » statistiquement représentatif. L'analyse a toutefois révélé en fonction des données utilisables, que seule la fenêtre de stade phénologique 55-65 pour le maïs et 55-71 pour le soya (épiaison et floraison) permettait d'établir une corrélation entre le signal radar et le rendement final de la parcelle. Les résultats pour les modèles d'estimation du **soya** se sont avérés concluants, avec une qualité d'ajustement du modèle allant de acceptable (indice Nash-Sutcliffe > 0.4) à souhaitable (indice Nash-Sutcliffe > 0.6), et une qualité d'estimation allant aussi de acceptable (RMSEr inférieur à 20%) à souhaitable (RMSEr inférieur à 10% (environ 300 kg-ha-1)). De très bons résultats sont obtenus avec la polarisation VV ou avec le rapport de polarisation VV/VH. Quant aux modèles développés pour la culture du **maïs**, il en est autrement. Dans ce cas, la qualité d'ajustement du modèle est faible et la qualité d'estimation est passable. Nous avons toutefois décidé de conserver le modèle puisqu'à la calibration, il montrait une corrélation significative. Il est toutefois très clair qu'il ne peut être utilisé que comme base à de futurs travaux et devrait fortement bénéficier d'un travail de fond sur la qualité des données historiques de rendement pour le maïs. Il n'a pas été possible de différencier les différents cultivars. Les données d'orbite

ascendante (soir) donnent les meilleurs résultats. Les modèles intègrent les résolutions de 30m et de 10m. Il n'a pas été possible de développer des modèles à l'échelle intra parcellaire.

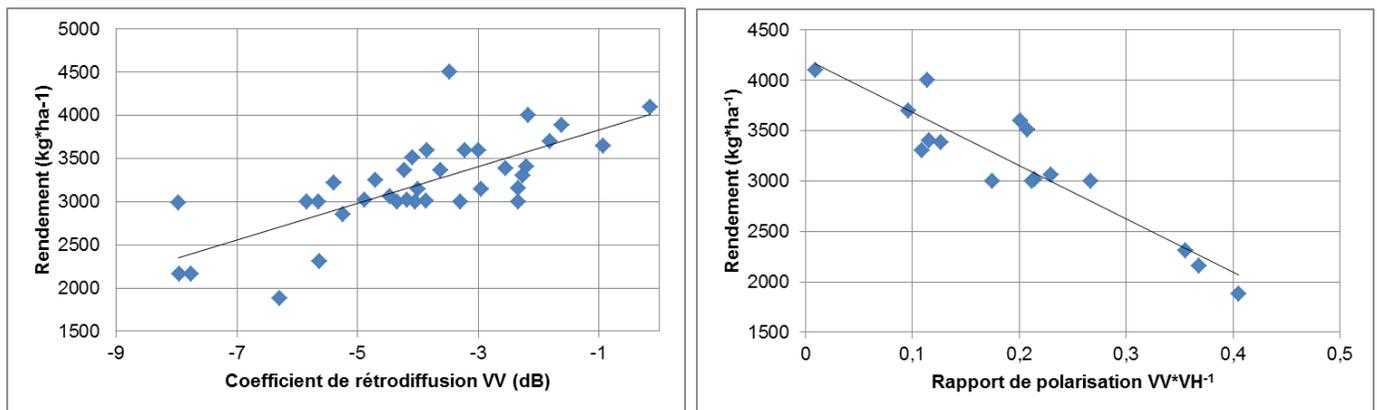


Figure 3: Calibration des modèles d'estimation du rendement Soya-2-VV-n+ et Soya-1-VVH-n+.

6. Technologies

Description/état de la technologie

Les modèles proposés pour le soya sont prêts à être testés en mode opérationnel. Les procédures de traitement sont automatisées mais la séquence opérationnelle reste à définir. Les modèles pour le maïs doivent être recalibrés avec de nouvelles données. Des travaux supplémentaires sont aussi nécessaires pour élargir les fenêtres d'application.

Aspects innovateurs

Une approche de développement de modèles statistiques basée sur un jeu de données inédit, à l'échelle de la parcelle et pour tout le sud du Québec. Une structuration des données permettant une analyse pointue des meilleures combinaisons entre caractéristiques radar, culture et conditions au champ.

Champ(s) d'application

Prévision du rendement agricole permettant d'envisager des mesures de mitigation ou de prévoir la gestion des récoltes.

Répercussion sur l'entreprise ou l'organisme utilisateur

Valeur ajoutée à l'offre de services de la COOP Fédérée.

7. Échantillons de produits illustrant le projet (c'est-à-dire des images contextuelles ou des produits d'OT)

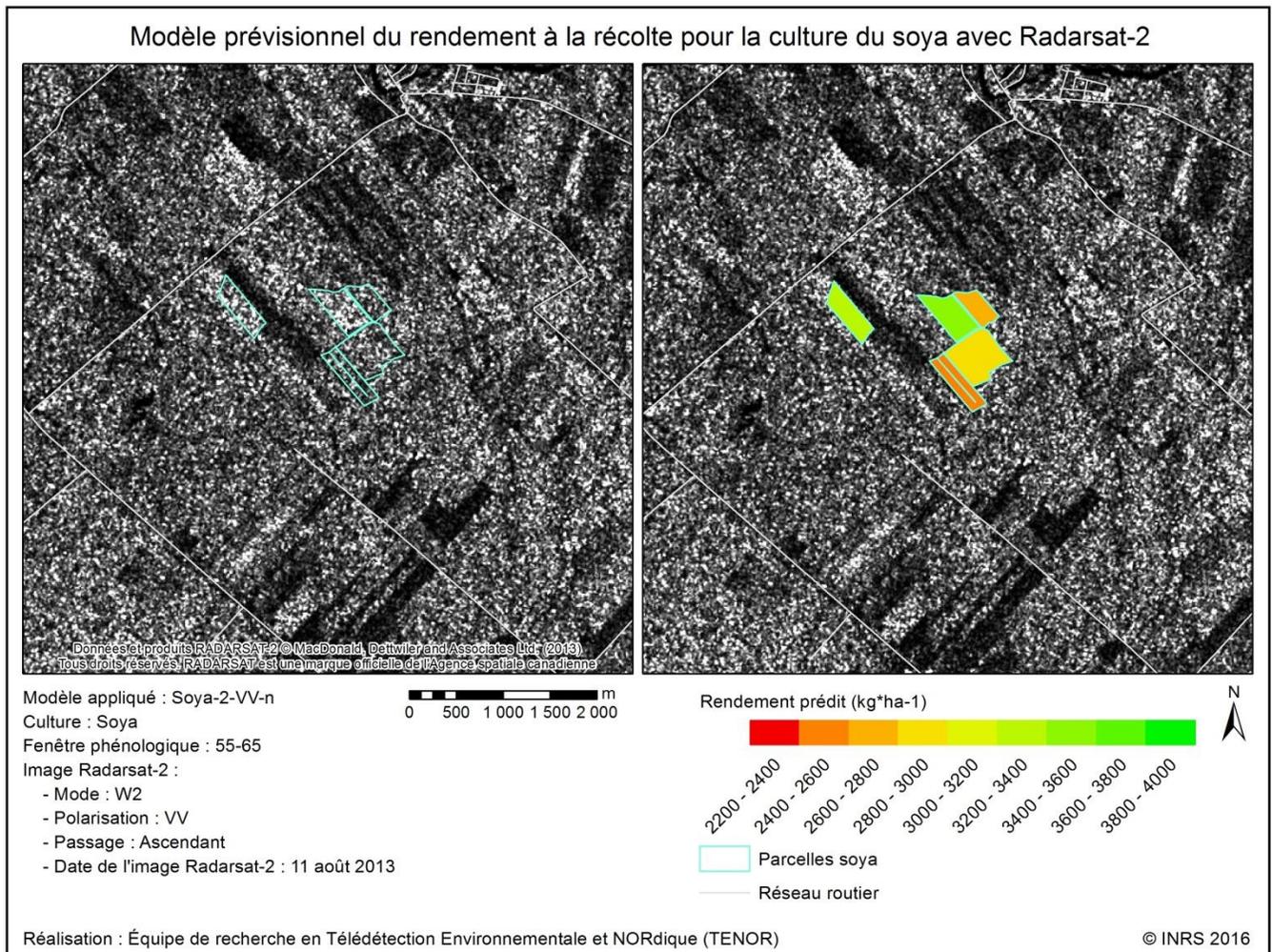


Figure 4: Exemple de résultats du modèle de prévision de rendement pour le soya (Soya-2-VV-n).

8. Titre de propriété intellectuelle

Les algorithmes et interfaces issus de ce projet (FIP) sont regroupés sous l'appellation « Télérndement » et la propriété intellectuelle appartient à l'INRS.

9. Publications et références

Aucune publication scientifique n'a été produite à l'intérieur de la période du contrat.