

Utilisation possible de la télédétection pour produire des statistiques sur les grandes cultures à Statistique Canada

James Brisbane et Chris Mohl¹

Résumé

Dans l'optique de réduire le fardeau de réponse des exploitants agricoles, Statistique Canada étudie d'autres approches que les enquêtes par téléphone pour produire des estimations des grandes cultures. Une option consiste à publier des estimations de la superficie récoltée et du rendement en septembre, comme cela se fait actuellement, mais de les calculer au moyen de modèles fondés sur des données par satellite et des données météorologiques, ainsi que les données de l'enquête téléphonique de juillet. Toutefois, avant d'adopter une telle approche, on doit trouver une méthode pour produire des estimations comportant un niveau d'exactitude suffisant. Des recherches sont en cours pour examiner différentes possibilités. Les résultats de la recherche initiale et les enjeux à prendre en compte sont abordés dans ce document.

Mots clés : Télédétection; agriculture; contrôle des cultures; statistiques sur les cultures.

1. Introduction

Chaque année, Statistique Canada mène six enquêtes sur les grandes cultures, en vue d'estimer les superficies et les rendements. Chaque enquête est menée au moyen d'interviews téléphoniques auprès des exploitants agricoles. Les estimations sont utilisées par le secteur des cultures et par les décideurs gouvernementaux. Parmi les statistiques importantes pour le secteur des cultures figure la prédiction pendant la saison du niveau de production des cultures individuelles. Ces statistiques sont nécessaires pour faire des prédictions justes des prix, lesquelles servent à leur tour à prendre des décisions commerciales.

Comme de nombreux autres instituts nationaux de statistique, Statistique Canada fait l'objet de pressions croissantes en vue de trouver des approches de rechange aux enquêtes traditionnelles pour la production de statistiques, afin de réduire le fardeau de réponse et le coût des enquêtes. On examine l'utilisation de la télédétection et des données administratives comme option de rechange pour remplacer un cycle des enquêtes sur les grandes cultures. Dans le contexte du présent document, la télédétection désigne l'utilisation de données et d'images par satellite comme données d'entrée dans des modèles qui servent à estimer les paramètres en question. Statistique Canada a mené quelques essais à petite échelle sur des cultures individuelles au moyen de cette technologie au cours des 20 dernières années. Toutefois, cela n'a pas été utilisé en production, ni pour couvrir une vaste gamme de cultures en même temps.

Nous abordons certains des objectifs de la recherche sur l'utilisation de la télédétection et les défis qu'elle comporte. Dans la section 2, nous présentons la stratégie actuelle de collecte de données sur les grandes cultures. La section 3 aborde certaines des sources de données potentielles qui peuvent servir à produire des statistiques sur les grandes cultures. Dans les sections 4 et 5, nous examinons l'utilisation de la télédétection et des données administratives pour la modélisation d'estimations du rendement et des superficies des cultures, respectivement. Certains résultats d'un modèle de rendement initial sont présentés. La section 6 décrit certains des travaux à venir qui sont prévus et elle est suivie par des conclusions.

¹Statistique Canada, Division des méthodes d'enquêtes auprès des entreprises, 150, promenade Tunney's Pasture, Ottawa (Ontario) Canada K1A 0T6.

2. Méthodes actuelles de production de statistiques sur les grandes cultures à Statistique Canada

Le secteur des cultures profite des meilleures estimations possible de la production de cultures. Ces estimations jouent un rôle important pour prédire l'offre de cultures à partir de la récolte à venir, et contribuent à fixer le prix que les exploitants agricoles peuvent s'attendre à recevoir pour leurs cultures. Pour ces estimations, on a besoin de deux composantes : la superficie récoltée et le rendement. Au cours d'une saison, au fur et à mesure de la croissance des cultures, de nouveaux événements météorologiques se produisent et de nouvelles données deviennent disponibles, ce qui fait qu'il est important de mettre à jour les estimations au fil du temps. Par conséquent, Statistique Canada mène son enquête téléphonique sur les grandes cultures six fois par année, et pose des questions aux exploitants agricoles concernant environ 52 cultures différentes. Le tableau 2-1 montre le calendrier des six cycles de l'enquête, la taille approximative de l'échantillon et les données recueillies dans chaque cycle.

Tableau 2-1
Enquêtes téléphoniques de Statistique Canada pour la production de statistiques sur les grandes cultures

Mois de la collecte des données	Taille approximative de l'échantillon (en 2014)	Quantités estimées
Mars	11 500	<ul style="list-style-type: none">• Superficie prévue d'ensemencement pour chaque culture• Quantité de chaque culture en entreposage
Juin	24 500	<ul style="list-style-type: none">• Superficie réelle ensemencée pour chaque culture
Juillet	13 100	<ul style="list-style-type: none">• Superficie réelle ensemencée pour chaque culture• Superficie récoltée prévue pour chaque culture• Rendement prévu pour chaque culture• Quantité de chaque culture en entreposage
Septembre	9 300	Mêmes quantités qu'en juillet, sauf pour les cultures en entreposage
Novembre	26 800	Mêmes quantités qu'en septembre
Décembre	8 700	Uniquement la quantité de chaque culture en entreposage.
Total	93 900	

Chaque enquête recueille des données de façon indépendante par rapport aux autres cycles. Les exploitations agricoles font l'objet d'enquêtes au plus deux fois pendant l'année. Les estimations les plus précises de la superficie récoltée et du rendement sont produites durant l'enquête de novembre, une fois les récoltes terminées, mais ces estimations ne sont pas disponibles avant le début de décembre, et il est important de produire des estimations précises avant ce moment là. On les appelle les estimations en saison. Pendant l'été, Statistique Canada pourrait utiliser d'autres sources de données pour estimer les superficies récoltées et les rendements prévus. Parmi les mesures envisagées figure l'annulation de l'enquête téléphonique de septembre et la publication d'estimations modélisées en septembre, sur la base de données de télédétection, de l'enquête téléphonique de juillet et d'autres sources disponibles, comme les stations météorologiques.

3. Sources de données possibles pour la production d'estimations des cultures

Statistique Canada évalue comment différentes sources de données peuvent être utilisées pour produire des statistiques sur les grandes cultures. Parmi les sources possibles figurent les suivantes :

1. Enquête téléphonique conventionnelle auprès des exploitants agricoles
2. Observation directe des champs (par des employés en voiture ou en avion)
3. Données de l'assurance-récolte
4. Données par satellite
5. Données de stations météorologiques.

Chacune de ces cinq sources de données pose des défis particuliers.

Enquête téléphonique traditionnelle auprès des exploitants agricoles. Le programme des enquêtes agricoles de Statistique Canada impose un fardeau important aux exploitants agricoles. La gestion du fardeau de réponse fait partie des priorités pour Statistique Canada.

Observation directe des champs. Elle est coûteuse en raison des coûts de déplacement et du besoin de personnel qualifié pour la mener à bien. Il faut aussi une base aréolaire de champs dont Statistique Canada ne dispose plus.

Données de l'assurance-récolte. Elles montrent uniquement combien de cultures ont été assurées. Des estimations des superficies non assurées devront être obtenues à partir d'autres sources.

Données par satellite. Lorsque l'on utilise des images par satellite pour estimer les superficies cultivées, chaque image nécessite l'élaboration d'un algorithme de classification qui lui est propre. L'exactitude de cette classification doit être vérifiée, grâce à une comparaison avec des petits échantillons de terres, c'est-à-dire la « réalité de terrain », selon le principe que la nature de la culture est connue sans erreur.

Données de stations météorologiques. Ces données peuvent être utilisées de concert avec d'autres sources pour estimer le rendement. Leur utilité dépend du niveau d'exactitude avec lequel elles représentent les conditions atmosphériques dans les champs où poussent les cultures.

Statistique Canada a l'obligation de produire des statistiques fiables, d'être rentable dans tous ses projets et de réduire le fardeau de réponse. La présente recherche vise à concevoir un programme comportant un fardeau de réponse et des coûts plus faibles, grâce à l'augmentation de l'utilisation des données de l'assurance-récolte, des données recueillies par satellite et des données des stations météorologiques, de concert avec les enquêtes traditionnelles, tout en maintenant un niveau acceptable de qualité des données. L'annulation de l'enquête de septembre et la tenue de cinq enquêtes téléphoniques seulement par année représenteraient une façon de réduire le fardeau. Selon un tel plan, Statistique Canada continuerait de publier des estimations du rendement en septembre, mais celles-ci seraient calculées à partir de données par satellite et de données météorologiques, ainsi que des estimations de l'enquête téléphonique de juillet. À court terme, les estimations de la superficie seraient reportées de l'enquête de juillet. À plus long terme, on pourrait adopter des approches de télédétection pour les estimations de la superficie.

4. Estimation du rendement des cultures à partir de données par satellite, de données météorologiques et de données d'enquête téléphonique

Statistique Canada envisage l'utilisation de modèles de régression pour déterminer à l'avance l'estimation du rendement de l'enquête de novembre. Des modèles et des estimations sont élaborés séparément et de façon indépendante pour chaque région agricole de recensement (RAR) au pays, et les résultats sont agrégés au niveau provincial. Une RAR est une zone infraprovinciale de terres composée (habituellement) de divisions de recensement comportant les mêmes propriétés agricoles, c'est-à-dire le type de climat et de sol.

Une des variables explicatives du modèle est l'estimation du rendement à partir de l'enquête téléphonique de juillet. Les autres prédicteurs proviennent de divers types de mesures par satellite et de stations météorologiques. Les mesures par satellite comprennent un indice de végétation par différence normalisée (IVDN) mesuré à différents moments pendant la saison de croissance et réparti en moyenne sur la superficie cultivée de la RAR. L'IVDN est une mesure par satellite courante de la superficie de végétation, établie au moyen d'un capteur de luminosité. Les autres variables sont des mesures recueillies par les stations météorologiques, y compris les précipitations totales, les degrés-jour de croissance, l'indice de stress hydrique et l'indice d'hydratation du sol. Les deux premières mesures sont cumulées sur différentes périodes pendant la saison de croissance, et les deux autres font l'objet d'une moyenne sur différentes périodes. Puis, dans les quatre cas, les chiffres font l'objet d'une moyenne pour toutes les stations de la RAR, afin de produire la variable explicative. Il est important de tenir compte de différentes périodes, la sensibilité du rendement à la météo variant au fil du temps.

Dans les résultats présentés ici, environ 60 variables explicatives ont été produites à partir de données du début de la saison de croissance au début de septembre, et les cinq meilleures ont été sélectionnées au moyen de la méthode de sélection pas à pas pour la régression linéaire multiple. Au total, 27 années de données (1987 à 2013) sont disponibles pour produire les modèles. Le modèle de régression a par la suite servi à estimer le rendement de l'enquête de novembre. Les graphiques présentés ci-après comparent cette estimation avec l'estimation de rendement de l'enquête de novembre à partir des enquêtes téléphoniques de juillet et de septembre.

La figure 4-1 montre le rendement du blé de printemps en Saskatchewan au cours des 27 dernières années, selon l'estimation de l'enquête de novembre. La Saskatchewan est la province la plus importante du point de vue de la production des grandes cultures, et le blé de printemps est l'une des principales cultures. Le rendement moyen par acre est d'environ 30 boisseaux. Même si la tendance générale est à la hausse, les rendements peuvent varier significativement d'une année à l'autre. Par exemple, le rendement était d'environ 15 boisseaux inférieurs à la moyenne en 1988 et d'environ 18 boisseaux supérieurs à la moyenne en 2013. Ces points de données atypiques rendent plus complexe la modélisation du rendement.

Figure 4-1
Rendement estimé au moyen de l'enquête téléphonique de novembre pour 27 années de blé de printemps en Saskatchewan

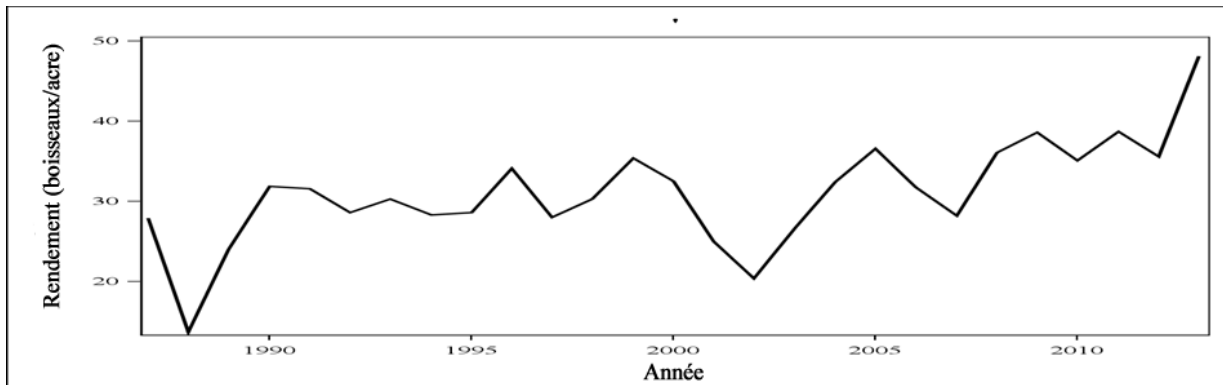
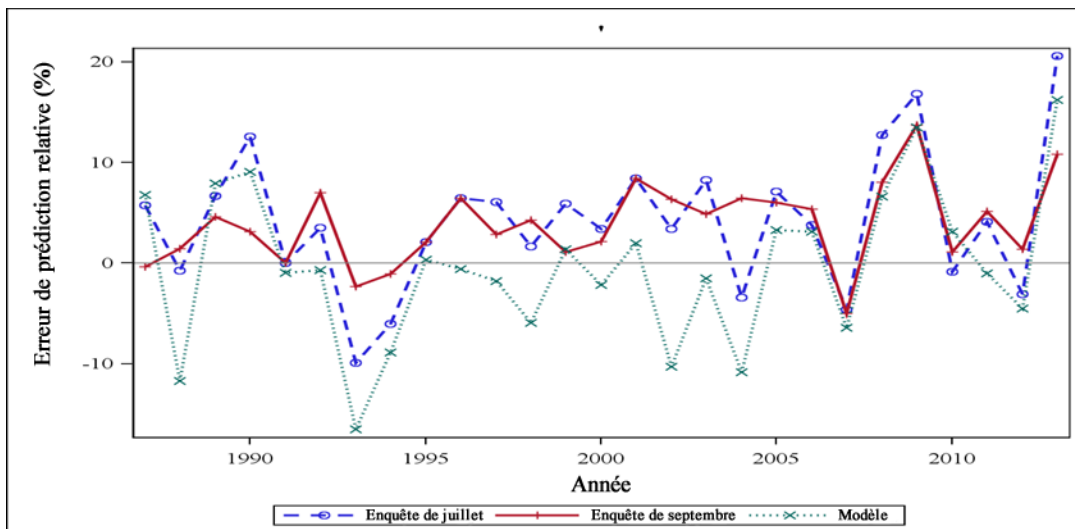


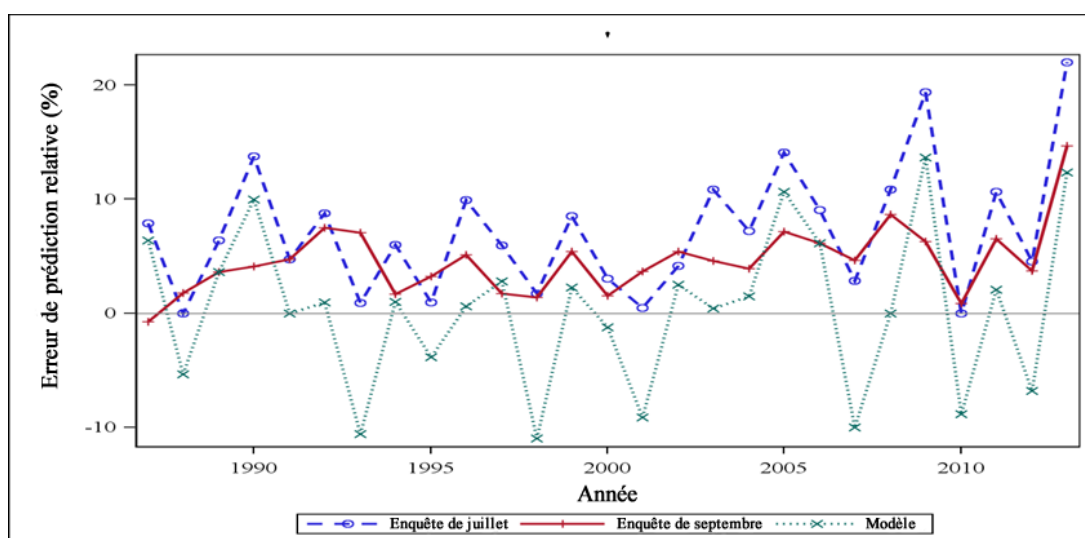
Figure 4-2
Erreurs de prédiction relatives aux estimations de l'enquête de novembre pour 27 années de blé de printemps en Saskatchewan



La figure 4-2 montre les différences relatives entre l'estimation de l'enquête de novembre du rendement du blé de printemps et les trois prédictions au cours des 27 dernières années. Une valeur positive signifie que l'estimation d'enquête est supérieure à la prédiction, et une valeur négative signifie qu'elle est inférieure à la prédiction.

Le modèle de régression linéaire multiple initial a produit des estimations qui étaient de toute évidence différentes de celles des enquêtes de juillet ou de septembre. Alors que les données d'enquête avaient tendance à sous-estimer le rendement par rapport aux estimations de novembre, le modèle l'a surestimé, un peu plus de la moitié du temps. L'ampleur des différences est aussi digne de mention. D'un point de vue absolu, les chiffres modélisés s'écartaient des rendements estimés de novembre davantage que les chiffres d'enquête. Sans surprise, on a eu de la difficulté à prédire le rendement des années atypiques à partir du modèle, lorsque le rendement était beaucoup plus faible (comme en 1988 ou en 2002) ou plus élevé (comme en 2013) que d'habitude.

Figure 4-3
Erreurs de prédiction relatives aux estimations de l'enquête de novembre pour 27 années de blé dur en Saskatchewan



La figure 4-3 montre un graphique similaire pour le blé dur en Saskatchewan. Même s'il s'agit aussi d'une culture importante en Saskatchewan, elle représente généralement moins de la moitié de la superficie du blé de printemps. Le graphique qui en résulte montre des tendances similaires à celles du blé de printemps. Les enquêtes de juillet et septembre ont eu tendance à sous-estimer la production, tandis que le modèle a affiché une répartition plus égale de la surestimation et de la sous-estimation. Toutefois, les pics indiquent que le modèle comportait beaucoup plus d'instabilité d'une année à l'autre et avait tendance à moins bien prédire les chiffres de novembre que l'enquête de septembre. Il s'agit de la tendance typique qui a été observée dans les cinq provinces et pour sept cultures importantes qui ont été examinées dans le modèle initial. Cela montre que certaines recherches concernant l'amélioration de la modélisation sont nécessaires pour résoudre certaines des faiblesses du modèle actuel. Un modèle de régression linéaire multiple standard pourrait être insuffisant pour résoudre certaines des valeurs aberrantes des ensembles de données historiques.

5. Estimation de la superficie cultivée à partir des données d'images par satellite

Cela exige la saisie d'une image du terrain par un satellite. Le terrain est échantillonné au moyen d'une base aréolaire, et les cultures poussant dans les zones échantillonnées sont consignées. Il s'agit de la réalité de terrain parce que l'on présume que les données sont exemptes d'erreurs, de la même façon qu'on présume que les répondants fournissent des réponses correctes dans une enquête téléphonique. Deux échantillons de terrain sont sélectionnés, et toutes les cultures dont les superficies doivent être estimées doivent être présentes dans les deux.

Avec le premier échantillon, les données sur le terrain sont comparées aux données de l'image par satellite, et la comparaison sert à élaborer un algorithme de classification appelé simplement « analyseur d'images ».

L'analyseur d'images représente un ensemble de règles de décision mettant en correspondance les valeurs de réflexion de la lumière à différentes longueurs d'onde et une culture. Il existe différents algorithmes pour élaborer l'analyseur d'images. Statistique Canada utilise l'algorithme See5 fourni par le logiciel RuleQuest (www.rulequest.com) et les images gratuites provenant du satellite Landsat 8 de la US Geological Survey. Landsat 8 fournit des mesures de la réflexion sur six longueurs d'onde, ce qui améliore l'exactitude de l'analyseur d'images par rapport à la mesure d'un moins grand nombre de longueurs d'onde. L'utilisation de plusieurs images prises différents jours peut aussi améliorer l'exactitude. Les données de production comprenant la nature prévue de la culture dans chaque zone de terres sont appelées « données de classification ». La force de la méthode vient de ce que l'analyseur d'images produit des données de classification pour l'ensemble de la superficie couverte par l'image par satellite, même si la réalité de terrain couvre une superficie relativement petite.

Si la méthode est bien mise en œuvre, la classification peut être assez exacte. Toutefois, la superficie totale classifiée pour une culture donnée représente toujours une estimation biaisée, dans une certaine mesure, en raison des erreurs appelées erreurs d'omission et de commission (voir Congalton et Green, 1999). Ce biais est supprimé en utilisant le deuxième échantillon de terrain comme échantillon d'enquête traditionnel, et les données de classification comme données auxiliaires pour l'estimation. L'étalonnage ou l'estimation fondée sur un modèle peut servir à supprimer le biais et à produire des estimations comportant une variance d'échantillonnage connue. Pour obtenir des renseignements à ce sujet ou d'autres renseignements, voir Commission européenne, Joint Research Centre, MARS (1999).

Statistique Canada a déterminé deux défis majeurs liés à l'utilisation des données d'images par satellite. Tout d'abord, l'analyseur d'images a de la difficulté à faire des distinctions entre les cultures qui ont une apparence similaire de loin. Par exemple, il est impossible de distinguer le blé et l'orge et, dans certains cas, il est difficile de distinguer le soya et les pommes de terre. Étant donné que les superficies selon le type de culture sont nécessaires, il faudra résoudre ce problème. Une solution simple consisterait à répartir la superficie estimée des cultures impossibles à distinguer en cultures individuelles, à partir des proportions observées dans l'enquête de juillet. En deuxième lieu, il est difficile d'estimer les superficies des cultures rares avec un bon degré d'exactitude relative sans un échantillon important sur le terrain. Si la culture rare ne fait pas partie de l'échantillon sur le terrain, l'estimation de la superficie est égale à zéro. Une estimation de zéro comporte une faible erreur absolue, mais une erreur relative importante parce que la superficie réelle est faible et non nulle. Une solution pourrait être d'obtenir des données sur les superficies de cultures rares auprès de l'assurance-récolte ou dans le cadre d'une enquête sur le terrain recueillant des données cartographiées par GPS auprès d'exploitants connus de cultures rares. Ces données échantillonnées de façon non aléatoire pourraient alors être utilisées pour élaborer un analyseur d'images pour les données d'images par satellite.

L'assurance-récolte peut jouer un rôle dans l'estimation des superficies des cultures, mais son utilisation comporte des limites. Comme il est mentionné précédemment, les données de l'assurance-récolte fournissent des renseignements uniquement sur la superficie des cultures assurées. Elles ne comprennent pas de renseignements concernant les superficies non assurées. Elles peuvent servir comme premier échantillon sur le terrain pour élaborer l'analyseur d'images, mais elles ne peuvent pas servir comme principal échantillon d'enquête pour estimer les superficies non assurées. Toutefois, elles peuvent être utiles pour l'estimation en tant que données auxiliaires. Elles peuvent servir à corriger les erreurs dans les données de classification par satellite, avant que celles-ci soient utilisées dans l'estimation.

6. Travaux à venir à Statistique Canada

Vers la fin de 2014, on prendra une décision quant à l'annulation de l'enquête de septembre 2015 et son remplacement par des estimations modélisées. Le cas échéant, les estimations du rendement de l'enquête seront remplacées par des estimations du rendement d'un modèle de régression. Dans l'intervalle, des recherches additionnelles seront effectuées pour améliorer l'exactitude des modèles initiaux et corriger certaines des faiblesses observées. Jusqu'à maintenant, la recherche a été axée uniquement sur sept grandes cultures. Le modèle devra être aussi suffisamment robuste pour produire des estimations suffisantes pour certaines cultures plus rares. Agriculture

et Agroalimentaire Canada (AAC) a aussi examiné la meilleure façon d'utiliser les données par satellite pour la prédiction des récoltes (Newlands et coll., 2014). Statistique Canada continuera de collaborer avec AAC avec comme objectif commun de pouvoir calculer des estimations exactes et opportunes au moyen de la technologie de télédétection.

La deuxième phase du projet consiste à examiner de façon plus approfondie l'utilisation possible de la télédétection pour prédire la superficie des cultures. Les travaux effectués jusqu'à maintenant indiquent qu'il sera difficile de distinguer les différentes grandes cultures. Ce problème ne s'est pas posé dans certaines des petites recherches spécialisées qui ont été menées à Statistique Canada par le passé. Dans le cadre de cette étude, à l'été 2014, Statistique Canada a recueilli un échantillon de réalité de terrain dans trois zones infraprovinciales au Canada. Ces données ont été recueillies par du personnel qualifié de Statistique Canada qui a visité les champs individuels et a pris note des cultures s'y trouvant. Ces données peuvent être utilisées pour mieux comprendre les limites possibles de la télédétection, ainsi que la mesure dans laquelle les données de l'assurance-récolte peuvent être intégrées dans le processus d'estimation de la superficie.

À plus long terme, Statistique Canada devra être vigilant pour examiner d'autres options de rechange pour la production de ses statistiques sur les grandes cultures. Les exploitations agricoles adoptent continuellement des technologies plus sophistiquées pour leurs activités au quotidien. Parmi les sources d'information pouvant être exploitées figurent les données produites maintenant par certains équipements agricoles et qui permettent de totaliser la superficie des terres en cultures et, par la suite, la production de ces terres.

7. Conclusion

L'utilisation de la télédétection et de la technologie par satellite peut jouer un rôle dans la production de statistiques sur les grandes cultures à Statistique Canada. Le fait qu'elles peuvent être utilisées sans imposer un fardeau trop lourd aux exploitants agricoles les rend attrayantes, Statistique Canada étant à la recherche de façons de produire des statistiques au moyen d'autres méthodes que les enquêtes traditionnelles. Par ailleurs, un examen approfondi des avantages et des inconvénients de ces approches doit être effectué avant qu'on les utilise en remplacement des données de l'enquête. La qualité des estimations découlant de ces approches doit continuer d'être suffisante pour répondre aux besoins des utilisateurs des données. Les données de l'assurance-récolte représentent une autre source de données qui, à elle seule, est insuffisante pour produire des statistiques complètes sur les cultures, mais qui peut jouer un rôle complémentaire important dans le processus.

8. Remerciements

Les prédictions du rendement des cultures décrites dans la section 4 du présent document ont été produites par Frédéric Bédard, de la Section de la télédétection et de l'analyse géospatiale de la Division de l'agriculture de Statistique Canada. Les auteurs souhaitent remercier Frédéric Bédard, Gordon Reichert et Martin Renaud pour leurs commentaires utiles concernant ce document.

Bibliographie

Congalton, R.G. et K. Green (1999). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Boca Raton, Floride : Lewis.

Commission européenne, Joint Research Centre, MARS (1999). Best Practices for Crop Area Estimation with Remote Sensing. URL: <http://mars.jrc.ec.europa.eu/mars/Bulletins-Publications> (consulté en décembre 2014).

Newlands, N.K., D. Zamar, L.A. Kouadio, Y. Zhang, A. Chipanshi, A. Potgeiter, S. Toure et H.S.J. Hill (2014), « An Integrated Probabilistic Model for Improved Seasonal Forecasting of Agricultural Crop Yield Under Environmental Uncertainty », *Frontiers in Environmental Science*, 2(17).