

**UNE ESTIMATION PRÉCISE ET RAPIDE DE LA SUPERFICIE
PLANTÉE EN POMMES DE TERRE GRÂCE À LANDSAT:
RÉSULTATS D'UNE DÉMONSTRATION¹**

R.A. Ryerson, J.-L. Tambay, R.J. Brown

et

L.A. Murphy, B. McLaughlin²

Cette communication décrit les procédures et les résultats d'un projet conjoint de Statistique Canada et du Centre canadien de télédétection (CCT) visant à fournir une estimation rapide de la superficie plantée en pommes de terre au Nouveau-Brunswick, province qui est un important producteur de pommes de terre au Canada. Le projet a démontré que l'imagerie obtenue par satellite et combinée aux méthodes plus usuelles d'estimation des superficies plantées en pommes de terre peut alléger le fardeau des répondants, fournir rapidement des cartes de la distribution des cultures et permettre l'établissement d'estimations fiables pour les sous-régions.

1. INTRODUCTION

Les premières applications de la télédétection (par satellite) dans la vallée de la rivière Saint-Jean (Nouveau-Brunswick) par le Centre canadien de télédétection et le ministère de l'Agriculture du Nouveau-Brunswick ont démontré que les données obtenues par satellite pouvaient fournir des estimations précises et peu coûteuses de la superficie plantée en pommes de

¹ Communication initialement présentée au Quinzième symposium international sur la télédétection de l'environnement, Ann Arbor (Michigan), mai 1981.

² R.A. Ryerson et R.J. Brown, Centre canadien de télédétection (CCT), EMR; J.-L. Tambay, Division des méthodes d'enquête-entreprises (ce travail a été fait lorsque l'auteur était à la Division des méthodes d'enquête - institutions et agriculture), Statistique Canada; L.A. Murphy, Division de la statistique agricole, Statistique Canada; et B. McLaughlin, Division de la statistique agricole, Bureau régional de Truro (Nouvelle-Écosse), Statistique Canada.

terre (Mosher et al, 1978; Ryerson et al, 1979; Ryerson et al, 1980). L'intérêt manifesté dans cette étude et dans d'autres travaux du CCT sur le colza (Brown et al, 1980) a amené Statistique Canada à mettre sur pied un projet de démonstration en temps réel utilisant les données recueillies par le satellite Landsat au cours de l'année-récolte 1980. Statistique Canada, qui est l'organisme fédéral chargé d'obtenir des données sur les récoltes, désirait comparer les estimations des superficies plantées obtenues par des méthodes traditionnelles avec les estimations obtenues par satellite pour une même région. On a retenu la pomme de terre comme récolte-cible, et la vallée de la rivière Saint-Jean comme région d'étude.

La télédétection par satellite présente plusieurs attraits pour Statistique Canada: précision accrue des estimations établies à partir d'enquêtes régulières (probablement à des niveaux plus locaux), allègement du fardeau imposé aux répondants (ce qui pourrait se traduire par une réduction du nombre de questionnaires et des questions) et possibilité de dresser des cartes pour de petites régions où poussent des récoltes spéciales, ceci afin de mieux planifier les méthodes d'échantillonnage.

La section suivante résume les principaux résultats de cette démonstration, et le reste de la communication indique en gros les méthodes de télédétection qui ont été utilisées dans ce projet. On décrit aussi le système actuel de saisie des données de Statistique Canada, la région visée par l'étude, l'échantillon au sol et la saisie des données sur le terrain, l'analyse des données obtenues par télédétection et, en dernier lieu, la vérification et l'analyse des résultats.

2. PRINCIPAUX RÉSULTATS

Les données recueillies par satellite ont servi à estimer la superficie des cultures de pommes de terre dans la vallée de la rivière Saint-Jean, au Nouveau-Brunswick. Ces estimations, extrapolées au niveau provincial, correspondaient à 2 % près aux estimations de 52,000 acres publiées par Statistique Canada. Ces estimations étaient basées sur les résultats de trois enquêtes indépendantes menées dans la province.

L'analyse des données obtenues par satellite a été faite en temps réel (c'est-à-dire presque instantanément) au CCT, car la plupart des travaux de préparation avaient été effectués avant l'acquisition des données. La zone d'échantillonnage retenue pour l'Enquête descriptive sur l'agriculture (EDA) a fourni les données au sol nécessaires pour l'étalonnage du système; ces données ont permis d'établir des estimations par le quotient et par régression pour la correction des erreurs systématiques dans la classification par télédétection des champs de pommes de terre. Bien que cette démonstration n'ait pas eu lieu dans un cadre opérationnel, on aurait pu produire les estimations définitives moins de deux semaines après le passage du satellite au-dessus de la région d'étude.

La classification par satellite a été entravée par la présence de nuages (non-réponse au niveau du satellite) et par la similitude entre la signature spectrale des pommes de terre et celle d'autres objets présents dans la région étudiée. Le premier de ces problèmes s'est traduit par une perte de données, et sa correction a nécessité une certaine dose d'imputation. Quant au second problème, on l'a résolu en partie en corrigeant la classification et en utilisant des estimateurs par le quotient et par régression.

La comparaison des données recueillies au sol (par les enquêteurs), à partir des photographies aériennes des champs échantillonnés, a révélé que certains champs ont été omis par les enquêteurs de l'EDA, car les données sur ces champs n'étaient pas requises pour les besoins de l'EDA. Aussi a-t-on dû se servir des données tirées de photographies aériennes au lieu des données de l'EDA pour les estimations par satellite en 1980. En 1981, les procédures d'énumération de l'EDA ont été modifiées pour satisfaire tant aux exigences de l'EDA qu'à celles de la télédétection.

À la suite du succès rencontré par cette démonstration, l'expérience a été reprise en 1981. De plus, une expérience similaire a aussi été entreprise en 1981 afin d'estimer la superficie des récoltes de colza dans le district de Peace River, à la frontière de l'Alberta et de la Colombie-Britannique.

3. TÉLÉDÉTECTION À L'AIDE DES SATELLITES LANDSAT

La télédétection consiste à mesurer les caractéristiques d'un objet à distance, habituellement à partir d'un aéronef ou d'un satellite. Avec les données captées par satellite, on peut obtenir rapidement la couverture complète de grandes superficies, et ce, à un coût relativement faible. Parmi les différents domaines d'application de ces données, mentionnons l'agriculture, l'exploitation forestière, l'utilisation des terres, la formation des glaces et la cartographie en général.

Les satellites américains de la série Landsat, lancés par la NASA, ont fourni les données pour cette expérience, et celles utilisées auparavant pour une expérience précédente au Nouveau-Brunswick. Chaque satellite Landsat fait le tour de la terre 14 fois par jour, en orbite héliosynchrone (ce qui permet ainsi de couvrir les différents points de la terre à la même heure solaire locale). La lumière réfléchiée par le sol est enregistrée dans quatre bandes spectrales étroites, à l'aide du balayeur multispectral (MSS). Les données transmises au Canada sont captées à l'une des deux stations de réception situées, respectivement, à Prince Albert (Saskatchewan) et à Shoe Cove (Terre-Neuve). Chaque point de la terre est survolé tous les 18 jours par un satellite Landsat (si deux satellites sont utilisés, cette fréquence de couverture est de neuf jours).

Les données sont analysées par le Système d'analyse d'images du CCT (CIAS); ces données étant codées sur des produits standard comme les bandes d'ordinateur, chaque scène imagée couvrant une superficie de 25,600 kilomètres carrés. La plus petite unité de surface imagée est appelée élément d'image, ou pixel (de l'anglais "picture element"). Chaque pixel est identifié par une signature spectrale qui lui est propre, cette signature étant une mesure de la réflectance de la zone imagée dans les quatre bandes spectrales. La signature spectrale dépend des objets présents dans le pixel (routes, cultures, etc.), chaque objet ayant une signature particulière. La signature des terres cultivées dépend de la structure des plantes, du type de sol apparent, de la maturité des récoltes, de la hauteur des plantes, de la densité des feuilles et de nombreux autres facteurs.

Pour estimer la superficie des récoltes, on doit identifier chaque pixel qui appartient à un type de récolte donné. Des grands champs, que l'on sait appartenir au type de récolte qui nous intéresse sont localisés pour "l'apprentissage" du système. Les données de ces champs permettent d'identifier la signature particulière à cette récolte. Tous les pixels sont ensuite classés comme appartenant ou n'appartenant pas à cette récolte, selon leurs signatures spectrales.

Pour déterminer la superficie d'une récolte dans une région donnée, on compte le nombre de pixels se trouvant à l'intérieur de cette région et ayant été identifiés comme appartenant à la récolte. Un "apprentissage" additionnel peut être exécuté afin d'identifier les pixels "manqués" dans la classification initiale, ou encore pour réduire la confusion entre les récoltes, c'est-à-dire pour mieux séparer les récoltes dont la signature spectrale ressemble fortement à celle de la récolte d'intérêt.

Des données au sol précises sont requises pour deux raisons: en premier lieu, pour localiser les champs d'apprentissage étendus et, en second lieu, pour corriger toute erreur systématique dans la classification des données obtenues par satellite. Ces données au sol peuvent être obtenues par des recenseurs spécialisés travaillant sur le terrain, ou encore en utilisant une imagerie aérienne qui est interprétée par des analystes d'image.

4. SAISIE DES DONNÉES STATISTIQUES

Au fil des ans, Statistique Canada a utilisé les données obtenues par les enquêtes postales annuelles comme intrants principaux pour son système d'estimation des récoltes. Bien que ces enquêtes soient relativement peu coûteuses et qu'elles peuvent être réalisées rapidement, elles sont limitées par les taux de réponses variables et par la non-représentativité toujours possible des répondants. Les enquêtes descriptives probabilistes ont été introduites au cours des années 1970, afin de pallier à certains de ces problèmes. Ces procédures font appel à des enquêteurs sur le terrain pour le dénombrement d'un échantillon aléatoire d'agriculteurs. En 1980, Statistique Canada a fondé ses estimations des superficies plantées en pommes de terre au Nouveau-Brunswick sur les résultats de trois enquêtes:

l'enquête postale, l'enquête objective sur le rendement des pommes de terre (EORPT) et l'enquête descriptive sur l'agriculture (EDA).

Les questionnaires de l'enquête postale sont expédiés au début de juin à tous les agriculteurs répertoriés dans le registre des fermes dressé par la Division de la statistique agricole. Les réponses sont compilées par comté, et l'on obtient des estimations par comté en reliant la variation annuelle des superficies indiquées à la superficie totale de pommes de terre déterminée d'après le recensement pour chaque comté. Les estimations par comté sont ensuite additionnées afin de donner les estimations provinciales pour la fin juin.

L'enquête objective sur le rendement des pommes de terre est une enquête spécialisée, à la fois postale et descriptive, conçue pour estimer la superficie des récoltes de pommes de terre, ainsi que le rendement de la pomme de terre dans son habitat du Nouveau-Brunswick. L'enquête est réalisée à la mi-juillet sur un échantillon aléatoire de producteurs de pommes de terre obtenu du registre des fermes; les estimations des superficies de pommes de terre sont produites vers la mi-août.

L'enquête descriptive sur l'agriculture (EDA) est, comme son nom l'indique, une enquête descriptive et polyvalente conçue pour estimer les récoltes, le bétail et les dépenses agricoles au niveau provincial. L'EDA est une enquête à bases multiples réalisée à partir d'un échantillonnage aléatoire d'agriculteurs choisis dans le registre des fermes et à partir d'un échantillon aréolaire et aléatoire de segments (parcelles de terre). Les agents recenseurs visitent les fermes échantillonnées à la fin juin et au début de juillet. Les estimations des superficies sont disponibles au début d'août. Chaque année, environ 20 % des segments sont renouvelés.

Au cours de la saison de croissance, on publie deux estimations de la superficie plantée en pommes de terre. La première, publiée au début de juin, est basée sur les résultats de l'enquête postale. La seconde estimation, publiée au début de septembre, est basée sur l'examen des estimations provenant des trois enquêtes et de discussions avec les autorités provinciales. Comme date cible pour la publication des estimations obtenues par

satellite, on a retenu la date de publication de la seconde estimation.

5. RÉGION D'ÉTUDE

La région visée par l'étude est située dans la vallée supérieure de la rivière Saint-Jean, au Nouveau-Brunswick. Cette région prend naissance au sud de Woodstock, dans le comté Carleton, et suit la rivière Saint-Jean en direction nord-ouest sur environ 200 kilomètres, à travers le comté Victoria, jusqu'à Claire, dans le comté de Madawaska.

La région est densément boisée, de topographie variée et accidentée. Le drainage et la nature pierreuse du sol présentent certains problèmes dans cette région. On y compte quelque 70,000 hectares de terres agricoles améliorées, dont environ 20,000 servent habituellement à la culture de la pomme de terre. Les autres cultures importantes de ces régions sont les céréales, le foin et les légumes de traitement comme les pois, le brocoli et le chou de Bruxelles. La taille des terrains varie des parcelles de semences de 0.1 hectare aux champs de 40 hectares.

6. SAISIE DES DONNÉES ET ÉCHANTILLONNAGE AU SOL EN 1980

Dans le cadre de l'EDA au Nouveau-Brunswick, l'échantillon aréolaire a été jugé approprié pour obtenir les données au sol requises pour l'interprétation des données de télédétection. Cet échantillon a été composé en deux phases. Dans la première phase, les secteurs de dénombrement (SD) du recensement (lesquels comprenaient les sièges des fermes du recensement de 1976 - appelés SD agricoles du recensement) ont été stratifiés selon la superficie des champs de pommes de terre, l'importance du bétail et le nombre de porcs (données du recensement de 1976). À l'intérieur de chaque strate, on a choisi deux échantillons aléatoires simples par réplicat de secteurs de dénombrement. Chaque secteur échantillonné a été ensuite segmenté à l'aide de cartes, en superficie identifiable d'environ 5 à 8 kilomètres carrés, et on a choisi, comme échantillon aléatoire simple, un segment sur 10 par secteur de dénombrement. On a fourni aux agents recenseurs de l'EDA travaillant dans la région d'étude des agrandissements d'anciennes photo-

graphies aériennes (échelle de 1" pour 832') de chaque segment échantillonné. Les photographies provenaient de sources provinciales. La plupart d'entre elles avaient été prises en 1976. Ces agents recenseurs devaient visiter les agriculteurs exploitant des terres à l'intérieur de chaque segment afin de leur présenter les photographies, pour qu'ils puissent identifier tous leurs champs de pommes de terre et de maïs³; les agents recenseurs devaient noter les superficies indiquées par les agriculteurs. Le manuel des enquêteurs contenait une section sur les procédures à suivre, les procédures faisant partie de leur entraînement pour l'EDA.

7. ANALYSE DES DONNÉES OBTENUES PAR TÉLÉDÉTECTION

7.1 TRAVAUX PRÉCÉDENTS

Les travaux réalisés dans cette même région à l'aide de données de 1975 ont été rapportés ailleurs (Mosher et al) et la description détaillée de la méthodologie utilisée a été publiée (Ryerson et al, 1980). Dans les travaux de 1975, on a retenu une zone d'étude contenant environ 20 % de toutes les récoltes de pommes de terre de la province. À l'appui de cette étude, on a obtenu des données de terrain pour toute la région d'étude.

La région d'étude de 125 kilomètres carrés et deux sous-régions ont été localisées sur l'écran couleurs du Système d'analyse d'images du CCT (CIAS - Goodenough, 1979). On a eu recours à une méthode très simple d'apprentissage dirigé pour dresser les données des pixels dans trois champs de pommes de terre sous forme de quatre histogrammes unidimensionnels. Un parallépipède quadridimensionnel a été défini par les limites de chaque histogramme et a servi à borner les zones de décisions. Tous les points qui se trouvent à l'intérieur du parallépipède ont été classés comme des pommes de terre, et ceux qui se trouvent à l'extérieur de cette région ont été classés sous la rubrique "autres".

Un des principaux problèmes rencontrés dans l'analyse des données était la classification adéquate des pixels situés sur la bordure des champs. Ces

³ On devait aussi identifier les champs de maïs, car les travaux antérieurs en télédétection ont indiqué que le maïs pouvait être confondu avec les pommes de terre (Ryerson et al, 1980).

pixels présentent un problème spécial, car, comme nous venons de le mentionner, ils chevauchent deux champs différents. La réflectance de ces pixels dépend de la proportion de la superficie du pixel couverte par chaque champ et de la réflectance du matériau superficiel dans chaque champ. Il est normalement très difficile de calculer le pourcentage de chaque type de matériau présent dans ces pixels. Toutefois, on a pu obtenir des estimations acceptables des superficies en modifiant la limite originale des zones de décision: on y a joint un second parallélogramme formé par apprentissage sur un certain nombre de ces pixels de chevauchement qui semblaient appartenir aux champs de pommes de terre. Pour choisir les pixels de chevauchement appropriés et les inclure dans la classe des pommes de terre, on s'est fondé sur une interprétation visuelle et subjective de la scène affichée à l'écran (les données de trois des quatre bandes spectrales étaient fusionnées pour former une image dont les couleurs ressemblaient à celle d'une émulsion de couleurs pour l'infrarouge).

L'estimation de la superficie pour toute la ceinture de la pomme de terre a nécessité moins de quatre heures d'utilisation du système CIAS. La localisation, l'affichage et l'analyse du secteur primaire et des sous-secteurs ont pris un peu plus d'une heure. La sélection, l'affichage et l'analyse des cinq sous-scènes choisies ultérieurement ont pris deux heures et demi, et une autre heure a été nécessaire pour localiser la frontière du Nouveau-Brunswick et éliminer les données appartenant à l'extérieur de la province.

Par rapport à la superficie totale plantée de pommes de terre qui avait été déterminée et mesurée à partir des photographies aériennes à basse altitude prises en même temps, les estimations obtenues par satellite en 1975 étaient précises à 95 % (c'est-à-dire qu'elles correspondaient à 95 % de la valeur véritable estimée) dans la sous-région contenant le site d'apprentissage, à 80 % dans la deuxième sous-région et à 88 % pour l'ensemble de la région d'étude. Lors d'essais répétés utilisant des champs d'apprentissage différents, la précision est passée de 85 à 97 % pour la région d'étude primaire, alors que la précision pour l'ensemble de la province s'établissait à 84.5 %. Une partie de la marge d'erreur, dans l'estimation provinciale, était due au fait qu'un certain pourcentage des champs de pommes de terre se trouvaient à l'extérieur de la ceinture dite de la pomme de

terre. D'autres facteurs ayant contribué à cette marge d'erreur sont discutés ci-dessous.

7.2 AMÉLIORATION DES ESTIMATIONS

Bien que les travaux précédents aient connu un certain succès, on a identifié les sources potentielles d'erreur pour des applications nécessitant une précision supérieure à, disons, 85 %. Les principaux problèmes sont associés au choix subjectif des pixels de chevauchement des champs de pommes de terre, à la manipulation des champs de petite taille et à l'élimination des récoltes présentant des caractéristiques spectrales similaires à celles des pommes de terre.

En ce qui concerne ce dernier problème, l'idéal serait de connaître précisément la réflectance spectrale des pommes de terre et celle des autres récoltes tout au cours de la saison de croissance. À l'aide de ces renseignements, il serait alors possible d'identifier la fenêtre phénologique pendant laquelle les pommes de terre peuvent être distinguées avec certitude des autres récoltes. Malheureusement, un tel ensemble de données n'existe pas, bien que les connaissances actuelles sur les récoltes et les pratiques agricoles de la région fournissent des indications générales. Dans ce cas particulier, l'expérience des auteurs sur le terrain leur a permis de supposer que la date optimale pour distinguer les pommes de terre des autres récoltes dans cette région serait entre la mi-juillet et la mi-août. Pour vérifier cette hypothèse et fournir une indication quant au degré de séparabilité des pommes de terre et des autres récoltes, on a analysé des données MSS (Landsat) sur bande d'ordinateur obtenues au dessus de la vallée de la rivière Saint-Jean le 8 août 1975. La Figure 1 indique les valeurs relatives de radiance dans chaque bande Landsat pour les pommes de terre, le maïs, les pois, le foin, le brocoli, le pâturage, le sarrasin, le sol nu et les céréales. On voit sur cette figure que les pommes de terre sont aisément séparables des autres récoltes, sauf les pois qui sont habituellement récoltés entre la mi-août et la fin août. Il semble donc que l'analyse des données recueillies vers la fin de la saison de croissance permettent la séparation et l'identification des pommes de terre.

Quant aux problèmes des champs de petite taille et des pixels de chevauchement, on pourrait les résoudre en utilisant davantage de données au sol dans quelques petites surfaces, afin d'obtenir des estimations plus précises sur la superficie des récoltes dans une région. Vu l'étendue de la région où la pomme de terre est cultivée, on doit connaître la superficie des champs de pommes de terre à l'intérieur de dix à quinze segments, chaque segment ayant une superficie qui varie de cinq à huit kilomètres carrés. Toute la région serait alors classifiée le mieux possible, mais sans se servir de la classe subjective des pixels de chevauchement. La classification de chaque segment serait ensuite utilisée, avec les données au sol disponibles (les photographies aériennes prises en 1980, les données de l'EDA pour les années à venir), pour obtenir une relation de régression appliquée par la suite à l'estimation de toute la région afin de produire une estimation révisée (Hanuschak et al, 1979). On établirait aussi une estimation par le quotient, basée sur les estimations de toute la région obtenues par photographies aériennes et par données de satellite.

7.3 PRODUCTION DES ESTIMATIONS DE 1980 À PARTIR DES DONNÉES LANDSAT

On peut décrire l'estimation de la superficie plantée en pommes de terre, à partir des données satellite, comme un processus en trois étapes: l'obtention des données anciennes et nouvelles, la prélocalisation des frontières des segments de l'EDA, et, en dernier lieu, l'analyse des images.

Les données Landsat utilisées pour l'estimation étaient codées sur des bandes pour ordinateur et traitées sur le système de correction des images numériques par l'interpolation ($\sin x/x$) pour la correction géométrique (Guertin et al, 1979) et par la correction radiométrique CAL 3 (Ahern et Murphy, 1978). Chaque bande pour ordinateur couvre quatre cartes du Système topographique national au 1:50,000^e et comporte des pixels carrés de 50 m. Quatre bandes pour ordinateur sont nécessaires pour couvrir la région. Des données existantes avaient été commandées et elles devaient être livrées en mai 1980, alors que de nouvelles données avaient été commandées pour les passages du satellite au-dessus de la région entre la mi-juillet et la mi-août. La commande n'a posé aucun problème pour les données existantes, mais elle s'est avérée plus compliquée pour les données

en temps réel en raison de la panne de Landsat-III. Un passage de Landsat-II le 17 août a été utilisé pour produire une bande d'ordinateur traitée sur le système DICS et livrée au Centre d'analyse le 22 août, bien avant la date prévue.

La prélocalisation des segments de l'EDA a été réalisée au printemps de 1980 à l'aide de la fonction "curseur de polygonation" sur le Système d'analyse des images du CCT (Goodenough, 1977). Les frontières des segments de l'EDA ont été fournies par Statistique Canada sur des cartes au 1:50,000^e et sur des photocopies d'agrandissement de photographies aériennes, remises aux agents recenseurs de l'EDA pour leur énumération. Bien que les limites de certains segments furent faciles à localiser (ruisseaux, lisières de boisés, lacs, etc.), d'autres frontières géographiques ou de recensement, étaient beaucoup plus complexes. Pour les segments dont les limites consistaient en une combinaison de routes principales et (ou) des rivières importantes, on pouvait les localiser, les délimiter et les mettre en mémoire en moins de cinq minutes de travail sur l'image couleurs de 128×128 pixels de 50m agrandie sur un écran de 512×512 lignes. Les segments les plus complexes ont nécessité jusqu'à une heure de travail - la moyenne s'établissant entre 20 et 25 minutes. Une fois localisé, la frontière du segment était mise en mémoire selon des coordonnées de pixel propres au système DICS, de sorte qu'elle pouvait être superposée sur de nouvelles données lorsque venait le temps de localiser les données d'apprentissage et les intrants des estimateurs. Lors de la phase préparatoire, on n'a pu localiser qu'un certain nombre de frontières de segment, car les autres segments chevauchaient deux bandes d'ordinateur DICS ou présentaient d'autres problèmes similaires. On travaille actuellement sur un logiciel qui permettra de réduire le temps requis pour tout le projet, particulièrement pour la phase de prélocalisation. L'utilisation des photographies originales au lieu des photocopies, prévue pour le volet 1981 du projet (avec 15 nouveaux segments en 1981), devrait aussi permettre de réduire le temps requis.

Une fois les données au sol de 1980 acheminées au centre d'analyse, on procéda au choix des sites potentiels d'apprentissage (selon la taille des champs). On a retenu plusieurs champs dans un segment situé au nord de la

région (près de Saint-André), ainsi que plusieurs autres champs dans un segment situé au sud de la région (près de Hartland).

Après la réception des données satellite, il fut relativement aisé de rappeler les frontières de segment, de les superposer, de localiser les champs d'apprentissage et de commencer la classification selon les méthodes décrites à la section 7.1. En plus du choix des régions d'apprentissage, on a retenu un autre groupe de champs de grande superficie, ainsi que des secteurs de champs connus de pommes de terre qui apparaissaient plus rouges sur l'écran que les champs faisant partie de l'ensemble d'apprentissage. À mesure que les résultats de la classification devenaient disponibles, les superficies cultivées étaient enregistrées pour chaque sous-scène de 512 x 512 pixels et pour 17 segments de l'EDA.

Les analystes ont rencontré quatre problèmes au cours de la classification; un de ces problèmes avait trait à l'interpolation des récoltes se trouvant sous des nuages épars et sous l'ombre des nuages, alors que les trois autres problèmes avaient trait au brouillage par d'autres récoltes. La méthode d'interpolation pour les champs de pommes de terre se trouvant sous les nuages était relativement simple. On a supposé que la proportion de récoltes sous les nuages était similaire à celle de la région adjacente qui semblait contenir des pommes de terre. Pour déterminer la superficie des champs de pommes de terre sous les nuages (la valeur PC), on a utilisé la formule suivante:

$$PC = \frac{P_M}{T_A - C_A} \times C_A$$

où: P_M = superficie des champs de pommes de terre mesurée dans la région sans nuage; C_A = superficie recouverte par les nuages et l'ombre des nuages dans la région et T_A = superficie totale de la région. Les estimations totales par satellite ont incorporé ces régions, alors qu'on n'a inclus aucun segment EDA sous les nuages dans les analyses pour les

estimations par le quotient et par régression. Quant aux problèmes du brouillage par d'autres récoltes, on les a, du moins pour la plupart, résolus par une légère modification des paramètres de classification. Dans un cas, le brouillage était dû à une forme non-identifiée de repousse naturelle fortement diffusée dans des régions de coupes forestières. Dans d'autres cas, des champs de foin dans lesquels se trouvaient des amoncellements de pierres espacés régulièrement ont présenté une signature spectrale similaire à celle des pommes de terre. Après modification de la classification, les seuls éléments de brouillage qui demeuraient étaient dus à quelques parcelles de trèfles dans un segment et à certaines allées d'un terrain de golf.

Le calcul des superficies est présenté plus en détail dans les paragraphes suivants.

8. RÉSULTATS ET ANALYSE DU PROJET D'ÉVALUATION DE LA SUPERFICIE PLANTÉE EN POMMES DE TERRE AU NOUVEAU-BRUNSWICK (1980)

8.1 INTRODUCTION

Dans le cadre du projet d'évaluation de la superficie plantée en pommes de terre au Nouveau-Brunswick, réalisé en 1980, on a analysé les données au niveau de la région, des segments et des champs. On a obtenu deux types d'estimation pour la superficie totale des récoltes de pommes de terre dans la région cible de la vallée de la rivière Saint-Jean. Les estimations proviennent des données recueillies par satellite et de données obtenues par photographies aériennes en haute altitude, ces dernières ayant fait l'objet de vérification au sol (les photographies aériennes ont été obtenues et analysées par le CCT). Les estimations et leurs variances ont ensuite été comparées aux autres estimations obtenues par Statistique Canada lors d'enquêtes réalisées au Nouveau-Brunswick. On a ensuite comparé les superficies plantées en pommes de terre calculées par l'EDA et par satellite à celles qui furent calculées à partir des photographies aériennes (et qui représentaient, dans ce cas-ci, la meilleure approximation des valeurs réelles) afin de déterminer le degré de concordance de ces deux es-

timations au niveau du segment. Comme complément d'analyse, on a examiné la variation des superficies indiquées lors de l'EDA, pour chaque secteur assigné à un enquêteur.

8.2 ESTIMATION DE LA SUPERFICIE PLANTÉE EN POMMES DE TERRE À L'AIDE DES DONNÉES OBTENUES PAR SATELLITE.

Pour les estimations par le quotient et par régression, on a estimé la superficie totale des champs de pommes de terre dans la région d'étude du Nouveau-Brunswick, recourant pour ce faire à des données obtenues par satellite et à des données sur les segments obtenues par photographie aérienne. Les estimations, ainsi que leurs variances, ont été calculées selon le plan d'échantillonnage de l'EDA. Pour le Nouveau-Brunswick, l'EDA est une enquête à base multiple utilisant un échantillon stratifié à deux degrés par réplicat de segments, cet échantillon étant conçu pour donner des estimations précises sur divers caractères au niveau provincial (voir les sections 4 et 6). Comme les strates de l'EDA ne coïncidaient pas avec les limites de la région d'étude, on a eu recours à la technique de la post-stratification afin de calculer les estimations, traitant les secteurs de dénombrement comme des échantillons avec remplacement. On n'a pas incorporé dans l'échantillon les segments pour lesquels les données par satellite étaient manquantes, ni un segment représentant une valeur aberrante (voir la Figure 3). Ces estimations ont été basées sur 40 segments. En dernier lieu, comme les agents recenseurs de l'EDA ne recueillaient pas toujours les données sur toutes les fermes se trouvant à l'intérieur du segment (voir la section 8.3), on a utilisé la superficie du segment calculée d'après les photographies aériennes.

Soit "x" la donnée sur les pommes de terre obtenue par satellite et "y" une donnée obtenue par photographie aérienne. Les estimations par le quotient et par régression de Y, qui représente la superficie totale des champs de pommes de terre dans la région d'essai, ont été calculées à l'aide des formules suivantes:

$$\hat{Y}_{\text{quot.}} = \hat{R} X = \frac{\hat{Y}}{\hat{X}} X, \text{ et}$$

$$\hat{Y}_{\text{rég.}} = \hat{Y} + \hat{B} (X - \hat{X}),$$

où
$$\hat{Y} = \sum_{h=1}^L \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij}$$
 est l'estimation théorique de Y;

\hat{X} est l'estimation théorique de X, qui représente la superficie totale des champs de pommes de terre obtenue par satellite et non corrigée pour la région d'essai. \hat{X} est obtenu en substituant x_{hij} pour y_{hij} dans la formule pour \hat{Y} ;

y_{hij} et x_{hij} sont les valeurs observées pour le j-ème segment choisis dans la i-ème unité (SD) de premier degré choisie pour la strate h;

N_h et n_h sont respectivement le nombre total et le nombre échantillonné de secteur de dénombrement dans la strate h, h prenant les valeurs de 1 à L;

M_{hi} et m_{hi} sont, respectivement, le nombre total et le nombre échantillonné d'unités (segments) de second degré dans le i-ème SD choisi dans la strate h;

\hat{R} est une estimation de Y/X ; et, finalement,

$\hat{B} = \text{cov}(\hat{Y}, \hat{X}) / \text{var}(\hat{X})$ est le coefficient de régression linéaire.

Les estimations de variance de $\hat{Y}_{\text{quot.}}$ et $\hat{Y}_{\text{rég.}}$ sont donnés par les relations suivantes:

$$v(\hat{Y}_{\text{quot.}}) = v(\hat{Y}) - 2\hat{R} \text{cov}(\hat{Y}, \hat{X}) + \hat{R}^2 v(\hat{X}), \text{ et}$$

$$v(\hat{Y}_{\text{Rég.}}) = v(\hat{Y}) - 2\hat{B} \text{cov}(\hat{Y}, \hat{X}) + \hat{B}^2 v(\hat{X}),$$

$$\text{où } v(\hat{Y}) = \sum_{h=1}^L \frac{N_h^2}{n_h(n_h-1)} \sum_{i=1}^{n_h} \left[\frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{M_{hi}}{m_{hi}} \sum_{j=1}^{m_{hi}} y_{hij} \right]^2$$

et $v(\hat{X})$ est calculé en substituant x_{hij} pour y_{hij} dans l'expression pour $v(\hat{Y})$.

On peut voir que $\hat{B} = \text{cov}(\hat{Y}, \hat{X})/v(\hat{X})$ est la valeur de B qui minimise $v(\hat{Y}_{\text{Rég.}})$.

Le Tableau I illustre les estimations par le quotient et par régression, ainsi que les estimations de Statistique Canada pour la superficie plantée en pommes de terre au Nouveau-Brunswick. Les estimations par le quotient et par régression, calculées au prorata du niveau provincial, sont très près de la superficie calculée par Statistique Canada, laquelle se chiffre à 52,000 acres. En fait, il y a très peu de différence entre les deux estimations, et elles possèdent toutes deux le même coefficient de variation. Afin de donner une idée du meilleur rendement obtenu grâce à l'estimation par le quotient, la variance de cette estimation ne représentait que le 1/5 de celle de \hat{Y} , qui est l'estimation théorique de Y basé sur le plan d'échantillon avec post-stratification (échantillon aréolaire seulement). On peut noter que le coefficient de variation des estimations par le quotient et par régression sont du même ordre que celui de l'estimation à bases multiples obtenu par l'EDA, bien que cette dernière soit basée sur un échantillon plus important.

8.3 COMPARAISON DES DONNÉES AU NIVEAU DES SEGMENTS ET DES CHAMPS

Les Figures 2 et 3 contiennent les graphes des superficies plantées en pom-

mes de terre par segment et calculés selon les données recueillies par les agents recenseurs de l'EDA et par satellite, respectivement en fonction des superficies obtenues d'après les photographies aériennes. On n'a pas utilisé tous les segments d'échantillons dans l'analyse. Il n'y avait pas de données par satellite pour 16 segments, en raison de la couverture nuageuse et de la localisation des images. De plus, on n'a pas utilisé 8 segments de l'EDA, car ces segments contenaient des non-répondants à l'enquête et des fermes d'une certaine importance pour lesquelles les agents recenseurs n'avaient pas à recueillir de données sur les champs. Pour les calculs de cette section et ceux de la section 8.2, nous n'avons pas utilisé les deux valeurs aberrantes de ces graphes.

Ces graphes indiquent une très forte relation linéaire entre les données de l'EDA et celles obtenues par les photographies aériennes, mais il en est de même pour les données par satellite comparées aux données obtenues par photographie aérienne au niveau de segment: les coefficients de corrélation⁴ sont respectivement⁵ de 0.991 et de 0.968. On note une tendance à la sous-estimation des superficies dans le cas des agents recenseurs de l'EDA et dans celui des données obtenues par satellite. Cette tendance est cependant moins marquée pour l'EDA. On a déjà expliqué à la section 7 les causes des écarts dans le cas des superficies calculées par les données satellite. La superficie de certains segments contenant peu ou pas de pommes de terre a été surestimée, à cause du brouillage par d'autres récoltes (certaines données satellite ont été rejetées, car elles confondaient un grand champ de foin parsemé de pierre avec un champ de pommes de terre - on aurait pu éliminer cette erreur en modifiant la classification). Quant à la sous-estimation dans le cas de l'EDA, une des principales causes était que les enquêteurs n'avaient pas énumérés certaines fermes se trouvant dans les segments à cause des procédures à bases multiples (parmi les fermes

⁴ Les coefficients de corrélation ont été calculés en utilisant la pondération du plan d'échantillonnage.

⁵ La représentation graphique des données obtenues par satellite par rapport aux données de l'EDA a été très similaire à celle de la Figure 3, le coefficient de corrélation s'établissant dans ce cas à 0.957.

omises, on comptait les fermes qui étaient exclues de la base aréolaire, ainsi que des fermes dont les terrains couvraient plus d'un segment d'échantillon). Pour corriger cette situation, le guide des procédures sur les terrains pour l'EDA de 1981 contenait des instructions spécifiques pour s'assurer que toutes les fermes soient répertoriées, en vue des estimations de superficie qui seront réalisées l'an prochain, lors de la reprise du projet de démonstration. On croit que ces modifications permettront à l'EDA de fournir des superficies correspondant davantage aux superficies estimées d'après les photographies aériennes. La forte relation entre les données de l'EDA et celle des photographies aériennes, pour les segments échantillonnés, est encourageante et justifie la correction des estimations par satellite grâce aux données recueillies par les agents recenseurs de l'EDA.

Une description erronée des limites et de la superficie des champs fut une autre cause de l'écart entre les données EDA et celles des photographies aériennes. Les données de l'EDA sur la superficie des champs ont été obtenues en interrogeant les exploitants agricoles, et ainsi, elles ont été souvent indiquées en multiples de 5 acres. Les graphes des données EDA en fonction des superficies estimées d'après les photographies aériennes, par zone assignée aux enquêteurs, indiquent une différence possible dans les déclarations des répondants, selon le secteur d'affectation. Ces différences pourraient être attribuables aux enquêteurs, mais aussi à d'autres facteurs comme l'emplacement géographique, la structure des champs à l'intérieur des segments d'échantillon, ou bien à la variabilité des erreurs dans les déclarations. On a observé davantage de variations dans la région sise à l'est de la rivière Saint-Jean, où la superficie moyenne des champs était généralement plus grande. La relation entre les données EDA et les données recueillies par photographie aérienne était plus forte au niveau des segments qu'au niveau des champs.

9. RÉSUMÉ ET CONCLUSIONS

En 1980, on a utilisé des données recueillies par satellite pour estimer

assez précisément la superficie plantée en pommes de terre dans trois importants comtés producteurs de cette culture, au Nouveau-Brunswick. Tout au long du projet décrit dans cette communication, les auteurs ont raffiné les méthodes d'analyse et les procédures sur le terrain, afin d'accroître la précision des estimations, de réduire le fardeau imposé aux répondants et de fournir des données spatiales détaillées qui n'étaient disponibles auparavant que pour les années de recensement.

10. REMERCIEMENTS

Les auteurs désirent remercier pour leurs conseils techniques et leurs commentaires rédactionnels Mme N. Chinnappa, de Statistique Canada, et J. Cihlar et F. Ahern du Centre canadien de télédétection.

11. BIBLIOGRAPHIE

- [1] Ahern, F.J. et J. Murphy, 1978, "Radiometric Calibration and Correction of Landsat 1, 2, 3, MSS Data", CCRS Research Report 78-4, Énergie, Mines et Ressources, Ottawa, Canada.
- [2] Brown, R.J., Ahern, F.J., Ryerson, R.A., Thomson, K.P.B., Goodenough, D.G., McCormick, J.A. et Teillet, P.M., 1980, "Rapeseed: Operational Monitoring", 6ième Symposium canadien sur la télédétection, Halifax, Nouvelle-Écosse, Canada pp. 321-330.
- [3] Cochran, W.G., 1977, Sampling Techniques (3rd Edition), John Wiley and Sons Inc., New York.
- [4] Goodenough, D.G., 1979, "The Image Analysis System (CIAS) at the Canada Centre for Remote Sensing". Canadian Journal of Remote Sensing 5(1), mai 1979, 3-17.
- [5] Guertin, F.G., 1979, Butlin, T.J. and Jones, R.G., 1979, "Correction géométrique des images Landsat au Centre canadien de télédétection", Canadian Journal of Remote Sensing, 5(2), décembre 1979.
- [6] Hanuschak, G., R. Sigman, M. Craig, M. Ozga, R. Luebbe, P. Cook, D. Kleweno, and C. Miller, 1979, "Crop - Area Estimates with Landsat Transition from Research and Development to Timely Results", Proceeding of the Fifth Machine Processing of Remote Sensed Data Symposium, Lafayette, Indiana.

- [7] Mosher, P.N., R.A. Ryerson et W.M. Strome, 1978, "New Brunswick Potato Area Estimation from Landsat", 12th International Symposium on Remote Sensing of Environment, Manille, Philippines avril 1978, pp. 1415-1419.

- [8] Ryerson, R.A., P. Mosher, V. Wallen et N. Stewart, "Three Tests of Agricultural Remote Sensing in Eastern Canada: Results, Problems and Prospects", Canadian Journal of Remote Sensing, 5(1) 1979, pp. 53-66.

- [9] Ryerson, R.A., P.N. Mosher et Harvie J., 1980, "Potato Area Estimation Using Remote Sensing Methods", CCRS Users Manual 80-2, Énergie, Mines et Ressources, Ottawa, février 1980.

TABLEAU I

ESTIMATIONS DE LA SUPERFICIE TOTALE PLANTÉE EN POMMES DE TERRE
DANS LA RÉGION D'ÉTUDE ET POUR LA PROVINCE DU NOUVEAU-BRUNSWICK¹

Enquête et/ou moyen d'estimation	Estimation pour la région d'étude (en acres)	Estimation pour le Nouveau-Brunswick (en acres)	Coefficient de variation (%)
Satellite			
Données non corrigées	42,354	n.d.	n.d.
Estimation par le quotient	49,504	51,524	5.5
Estimation par régression	49,115	51,119	5.5
Enquête postale	n.d.	50,800	n.d.
EDA à bases multiples	n.d.	53,854	5.2
EORPT	47,203	49,129	4.59
Estimation de Statistique Canada publiée le 5 septembre	n.d.	52,000	n.d.

¹ La région d'essai, regroupant les comtés de Carleton, de Madawaska et le comté Victoria, représente environ 96.08 % des terrains producteurs de pommes de terre au Nouveau-Brunswick.

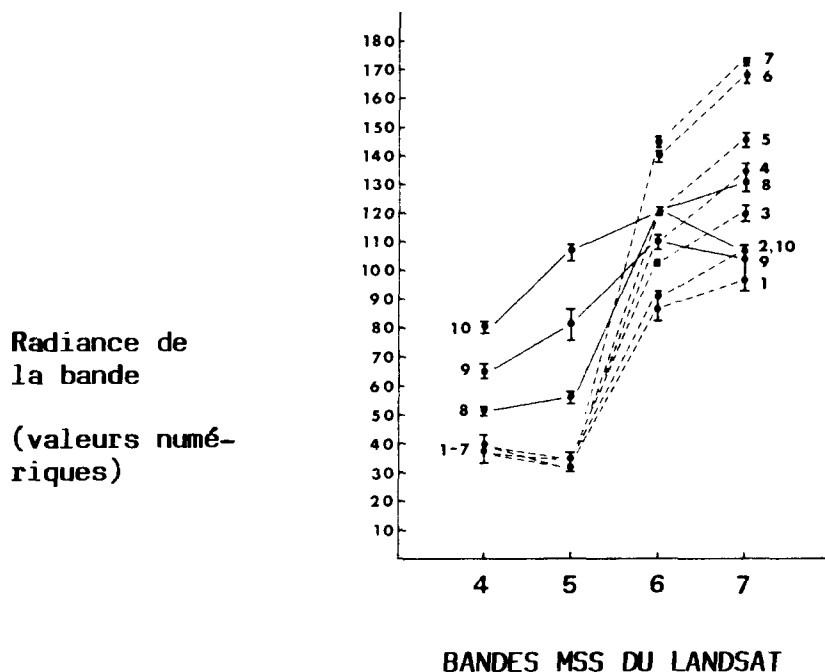
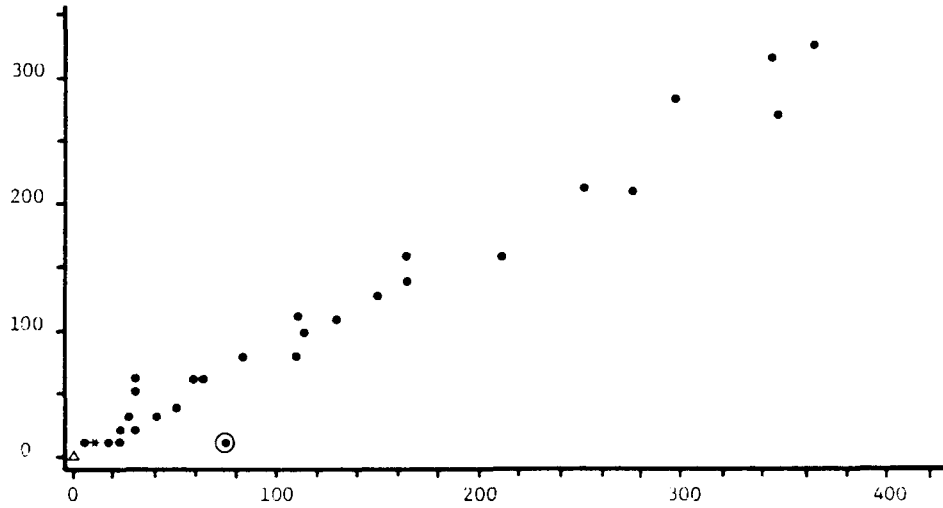


FIGURE I COMPARAISON DES DIVERSES BANDES LANDSAT POUR DIFFÉRENTES RÉCOLTES

Radiance des bandes MSS de Landsat (8 août 1975) pour les céréales (1), le sarrasin (2), les pâturages (3), le foin (4), le maïs (5), les pois (6), les pommes de terre (7), le brocoli (8), le sol nu (9), les herbes (10). Les bandes sont numérotées de 4 à 7. L'espacement relatif entre les valeurs numériques dans chaque colonne indique le degré de séparabilité des récoltes pour une bande donnée.

SUPERFICIE D'APRÈS
L'ENQUÊTE EDA

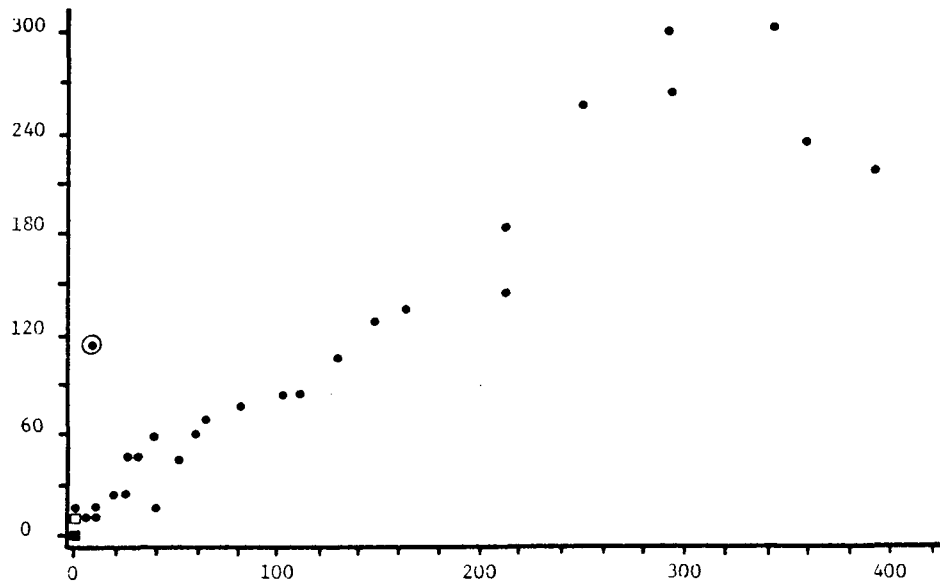
FIGURE 2



SUPERFICIE ESTIMÉE D'APRÈS LES PHOTOGRAPHIES AÉRIENNES

SUPERFICIE ESTIMÉE
PAR SATELLITE

FIGURE 3



SUPERFICIE ESTIMÉE D'APRÈS LES PHOTOGRAPHIES AÉRIENNES

FIGURES 2 et 3 - Graphes des superficies estimées, (en acres), respectivement, par l'EDA et par satellite en fonction des superficies estimées d'après les photographies aériennes à basse altitude, pour des segments échantillonnés du Nouveau-Brunswick. Les points encerclés représentent les valeurs aberrantes.

Légende: • = 1 observation, □ = 2 observations,
* = 3 observations, ■ = 11 observations et Δ =
15 observations