Département de géomatique appliquée

Faculté des lettres et sciences humaines

Cartographie extensive d'attributs forestiers en milieu boréal à partir d'images satellitaires et de photographies aériennes

Antoine Leboeuf
03 322 461

Thèse présentée pour l'obtention du grade de Philosophiae Doctor (Ph.D.) en télédétection, cheminement physique de la télédétection

Avril 2013

UNIVERSITÉ DE SHERBROOKE

Composition du jury

Cartographie extensive d'attributs forestiers en milieu boréal à partir d'images satellitaires et de photographies aériennes

Antoine Leboeuf

Cette thèse a été évaluée par un jury composé des personnes suivantes :

Richard Fournier, Directeur de recherche
(Département de géomatique appliquée, Faculté des lettres et sciences humaines)

Kalifa Goïta, Membre du jury interne
(Département de géomatique appliquée, Faculté des lettres et sciences humaines)

Jérôme Théau, Membre du jury interne
(Département de géomatique appliquée, Faculté des lettres et sciences humaines)

François Gougeon, Membre du jury externe
(Centre de foresterie du Pacifique, Ressources naturelles Canada)

Résumé

Les forêts boréales nordiques sont peu habitées et le réseau de transport dans ces régions est peu étendu, ce qui rend difficile l'accès en forêt pour sonder l'état de la ressource forestière. Ces milieux nordiques représentent néanmoins un très grand potentiel au niveau des ressources et de leur biodiversité. Leur homogénéité en terme de composition d'essences d'arbres favorisent l'utilisation des outils de télédétection pour leur cartographie ou monitoring sur de grands territoires. Le présent projet de recherche vise d'abord à développer une méthode de cartographie des attributs forestiers à échelle locale à partir d'images à très haute résolution spatiale (QuickBird ou WorldView) et de photographies aériennes acquises à une résolution spatiale de 30 cm. Une carte matricielle d'attributs forestiers (hauteur de peuplement, densité, surface terrière et volume marchand) à échelle locale est calculée à partir d'images à très haute résolution spatiale et basée sur une méthode de fraction d'ombre. Ensuite, la méthode de fraction d'ombre est étendue pour s'appliquer à des photographies aériennes dans le but de calculer des cartes d'attributs forestiers. Ces cartes locales d'attributs forestiers servent à enrichir la base de données des placettes d'inventaire. Enfin ces données locales sont spatialement généralisées à l'aide des méthodes de kNN (k Nearest Neighbor) et KMSN (k Most Similar Neighbors) pour cartographier les attributs forestiers à une échelle régionale. Les résultats obtenus de la cartographie locale et régionale démontrent le fort potentiel de l'approche développée avec des coefficients de régressions de 36 à 92% pour le modèle (local) de fraction d'ombre ainsi que des erreurs quadratiques moyennes de 18 à 58% pour la carte régionale d'attributs forestiers. Le couplage de ces méthodes permet donc de cartographier de grandes superficies du paysage forestier québécois avec très peu de ressources financière, humaine et technique.

Mots-clés : cartographie, télédétection, photographies aériennes, forêt boréale, fraction ombre .

Tables des matières

Liste des figures

Article 1.

Article 2.

Liste des tableaux

Article 1.

Article 2.

Article 3.

Avant-propos

Cette thèse par articles est composée de trois articles scientifiques rédigés en anglais. Chacun des articles représente un chapitre du document menant aux conclusions finales. Pour chacun de ces articles, ma contribution scientifique est celle d'un premier auteur. J'ai donc été le contributeur principal tant au niveau de l'élaboration des projets, le développement des méthodes et la rédaction scientifique.
Leur référence exacte est :

(1) Leboeuf, A., Fournier, R.A., Luther, J.E., Beaudoin, A. et Guindon, L. 2012. Forest attribute estimation of northeastern Canadian forests using Quickbird imagery and a shadow fraction method. *Forest Ecology and Management.* 266: 66-74. Soumis en juin 2011. Accepté en novembre 2011 Publié en février 2012.

(2) Leboeuf, A. et Fournier, R.A. 2012. Estimating stand attributes of boreal forest using digital aerial photography and a shadow fraction method. Soumis au journal: *Canadian Journal of Remote Sensing*. Soumis en décembre 2012. Soumis à nouveau en avril 2013.

(3) Leboeuf, A. et Fournier, R.A. 2012. Forest attribute estimation over large areas of northern boreal forests using KNN and shadow fraction method. À soumettre au journal: *Environmental Monitoring and Assessment.*

Remerciements

J'aimerais d'abord remercier mon directeur de recherche Richard Fournier pour le support financier, méthodologique et de rédaction tout au long du travail. Ensuite, je remercie André Robitaille et Robert Jobidon de la Direction de la Recherche Forestière du Ministère des Ressources naturelles et de la Faune du Québec (MRNF) pour avoir cru à l'apport scientifique de ces travaux et pour l'appui en matériel et ressources humaines. Cet apport a considérablement facilité l'aboutissement de ce projet. Je remercie le Service Canadien des Forêts (Luc Guindon et André Beaudoin) pour leur apport scientifique et pour la base de données en appui du premier article. Je souligne également l'appui des mes gestionnaires de la Direction des Inventaires Forestiers, principalement Élisabeth Bossert, Serge Gagné, Luc Tellier et Frédéric Dufour pour leur soutien et encouragements à terminer ce mandat. Enfin, je remercie Louise Bourque, Hakim Ouzenou, Jean-Gabriel Élie et Bastien Ferland-Raymond pour leurs précieux appuis techniques et statistiques.

1. Introduction

La forêt boréale du Québec est assez bien connue et inventoriée dans sa portion aménagée située approximativement au sud du 51^{ième} parallèle. Une cartographie écoforestière à l'échelle 1 : 20 000 (Robert et Robitaille, 2009a) est produite par le Ministère des Ressources naturelles du Québec (MRNQ) et couvre l'ensemble de cette portion du territoire. Cette carte est réinterprétée à un intervalle d'environ 10 à 15 ans, ce qui maintient l'information forestière relativement à jour. La forêt boréale localisée au nord du 51^{ième} parallèle est, quant à elle, très peu documentée. Pour palier ce manque de connaissance, le MRNQ a effectué, entre 2005 et 2009, une cartographie écoforestière nordique à l'échelle 1 : 100 000 qui couvre la limite nordique de l'actuel inventaire forestier (51^{ième} parallèle en moyenne) jusqu'au 53^{ième} parallèle (MRNF, 2009a). Ce projet visait à cartographier deux milieux : les milieux physique et forestier. La cartographie du milieu physique s'est faite à partir d'une interprétation de photographies aériennes 3D en archive. La cartographie forestière s'est effectuée, quant à elle, à partir d'images satellitaires à moyenne résolution (30 m) du capteur Landsat-TM. Ce type d'image satellitaire a été choisi à cause de son faible coût d'acquisition et le haut niveau d'information disponible. Contrairement à la cartographie écoforestière à l'échelle 1 : 20 000 de la zone aménagée, l'approche employée pour effectuer la cartographie à l'échelle 1 : 100 000 ne permettait pas d'interpréter les hauteurs de peuplement. Or, cet attribut forestier était nécessaire au mandat d'analyse du comité scientifique de la limite nordique des forêts attribuables du Québec pour un aménagement forestier durable (MRN, 2012). Ce comité était chargé de revoir la limite nordique des forêts attribuables sur la base de critères biophysiques de l'aménagement forestier durable soient : le milieu physique, la productivité forestière, les perturbations naturelles et la biodiversité du milieu.

Aussi, le faible taux d'échantillonnage, par placette d'échantillonnage terrain (PET), rendait difficile les calculs de volume marchands par tarifs de cubage. Ces deux attributs forestiers (hauteur et volume marchand) sont néanmoins essentiels à une meilleure connaissance de ces milieux forestiers.

Différentes approches peuvent donc être utilisées pour cartographier la hauteur et le volume marchand des peuplements forestiers sur un grand territoire. La photographie aérienne s'avère un

outil opérationnel incontournable. Ces photographies, acquises avec un équivalent de 21 ou 30 cm de résolution spatiale, servent généralement à effectuer de la photo-interprétation forestière (Robert et Robitaille, 2009b). Cette approche présente l'avantage que la hauteur du peuplement peut être interprétée avec une précision de l'ordre du mètre, par le biais de système de visionnement 3D (par exemple avec l'usage du logiciel Summit Evolution ©). Aussi, le volume marchand peut être calculé par des tarifs de cubage déjà bien documentés tel que celui de Perron (1986). Néanmoins, ces approches sont critiquées à cause de leur subjectivité liée à l'interprétation, ce qui peut affecter l'homogénéité des résultats dans l'espace et dans le temps. De plus, ces approches reliées aux photographies aériennes sont dépendantes des coûts élevés de prise de vue, d'interprétation de photographies aériennes et de réalisation des PET principalement en zones éloignées des zones habitées.

Dans les régions éloignées, les images satellitaires à très haute résolution spatiale (moins de 1 m/pixel) telles que celles provenant de QuickBird, GeoEye, Ikonos peuvent devenir une alternative aux photographies aériennes. Ces images offrent en effet, des résolutions spatiales s'approchant de celles disponibles par photographie aérienne. Elles procurent aussi quatre bandes spectrales et une géométrie de prise de mesure favorables à l'analyse avec un coût d'acquisition relativement compétitif. Plusieurs projets ont d'ailleurs démontré le fort potentiel de ces images dans une optique de mesure ou de cartographie d'attributs forestiers (Gougeon et Leckie, 2003; Mora *et al.*, 2010; Wulder *et al.*, 2012a). Par exemple, de nouvelles techniques de classification ont été développées et sont basées sur les composantes d'ombre de l'image. La fraction d'ombre est l'aire couverte par l'ombre des arbres sur une surface donnée divisée par l'aire de cette surface. Plusieurs études ont lié la fraction d'ombre, par régression statistique, à plusieurs attributs forestiers. C'est le cas de Greenberg *et al.* (2005) qui ont trouvé une forte relation entre l'aire de l'ombre des arbres mesurée à partir d'images Ikonos et le diamètre à hauteur de poitrine (DHP) des arbres, la densité des couronnes et la biomasse forestière. D'autres études ont aussi établi des régressions entre la fraction d'ombre et certains attributs biophysiques en forêt boréale. Peddle et Johnson (2000) affirment que la fraction d'ombre permet de prédire les attributs forestiers pour différentes variétés de densité de couvert et pour une variété d'espèces couvrant différentes régions forestières boréales. Ces auteurs ont utilisé la fraction d'ombre extraite à partir d'une méthode d'analyse de démixage spectral pour estimer la surface foliaire (*Leaf Area Index* -

LAI). Leur méthode donne un coefficient de régression (R^2) de l'ordre de 20% de plus que celui obtenu d'une méthode basée sur un indice de végétation (NDVI). En utilisant cette même approche, Peddle *et al.* (2001) ont réussi à relier la fraction d'ombre à plusieurs attributs forestiers différents tels que la biomasse forestière, la productivité forestière, le LAI, le DHP, la densité de cime par unité de surface ainsi que la surface terrière. Toutefois, en complément, Seed et King (2003) n'ont trouvé qu'une faible relation entre la fraction d'ombre et le LAI d'un site en forêt boréale mixte. L'étude révèle néanmoins une relation plus forte pour les sites dont la densité est plus faible de 70%. Leboeuf *et al.* (2007) et Ozdemir (2008) ont respectivement trouvé une forte relation entre la fraction d'ombre calculée à partir d'images satellitaires à très haute résolution spatiale (QuickBird) et ce pour la biomasse et le volume marchand. Les résultats de ces études suggèrent donc que la fraction d'ombre des images satellitaires à très haute résolution spatiale peut être utilisée pour le traitement d'image afin de trouver les valeurs d'attributs forestiers tels que la hauteur du peuplement et le volume marchand pour fin d'inventaire forestier. Bien que le fort potentiel des approches basées sur les photographies aériennes et les images satellitaires à très haute résolution spatiale soit démontré, l'applicabilité de ces produits à un grand territoire éloigné reste non réaliste à cause des coûts élevés d'acquisition de ces images.

Les images satellitaires à moyenne résolution, telles que les images Landsat-TM ou Landsat-ETM, deviennent donc intéressantes pour cartographier les attributs forestiers sur un grand territoire puisqu'elles sont facilement disponibles. De plus, ces images possèdent une résolution spatiale à 30 m, plusieurs bandes spectrales, et couvrent une zone importante (180 km × 180 km) à un coût d'acquisition extrêmement faible, voire nul pour les images en archives. Ces images ont d'ailleurs démontré un fort potentiel pour la cartographie d'attributs forestiers et ce, à l'aide de méthodes KNN (K Nearest Neighbor). Ces méthodes ont été, au départ, proposées par Hart (1967) pour ensuite être utilisées à plusieurs applications dont la classification d'images satellitaires (McRoberts *et al.* 2002 ; McRoberts and Tomppo, 2007 ; Tomppo *et al.*, 2008). Ces méthodes intuitives et versatiles servent à estimer des variables qui sont basées sur la similarité des covariables entre des attributs à prédire et des échantillons observés.

La méthode de KMSN (k Most Similar Neighbor) développée par Moeur et Stage (1995) est appropriée lorsque plusieurs variables doivent être estimées simultanément et de façon cohérente

comme dans le cas des inventaires forestiers où différents attributs sont calculés (volume par essence, volume total, surface terrière, etc.). La méthode est non paramétrique et ne repose donc sur aucun modèle de probabilité pour la réalisation des prédictions. La valeur prédite est donc directement mesurée d'une autre observation ou d'une moyenne de différentes observations mais assume que la variabilité de la population est bien représentée au sein de l'échantillon.

Plusieurs programmes d'inventaire forestiers nationaux sont basés sur les approches KNN et KMSN permettant une mise à jour régulière et à moindre coût de leur inventaire forestier (Norvège : Gjertsen, *et al.*, 1999; États-Unis : McRoberts, et Tomppo *et al.*, 2007). La précision des résultats obtenus à partir de ces méthodes dépend néanmoins de la quantité et de la qualité des données d'entraînement (Franklin et Wulder, 2002). Les régions nordiques du Québec sont couvertes par peu de PET et celles existantes sont localisées dans des zones forestières assez denses (25% et plus). De plus, la disponibilité de PET pour les peuplements ayant une hauteur de 7 m ou moins est donc quasi inexistante. Or, la mise en place d'une méthode complète de cartographie de la forêt boréale au Québec nécessite la disponibilité de l'ensemble de la variabilité d'attributs forestiers : par exemple des mesures dans des peuplements de 0 à 25 m de hauteur. De plus, les PET disponibles dans le nord du Québec ont été mesurées sur une période de plus de dix ans, ce qui rend le lien avec une image satellitaire plus difficile. Ces deux problématiques sont donc des obstacles à la cartographie des deux attributs forestiers (hauteur de peuplement et volume marchand) sur un grand territoire nordique du Québec en utilisant uniquement le jeu de données actuel de PET. C'est donc pour pallier à ces défis techniques et permettre d'effectuer la cartographie de ces attributs forestiers sur le Québec nordique que ce projet de recherche a été développé.

2. Objectifs de travail

Le présent travail de recherche vise donc comme objectif principal le développement et l'adaptation de nouvelles méthodes de cartographie d'attributs forestiers (hauteur du peuplement et volume marchand) sur de grands territoires au Nord du Québec à l'aide d'images satellitaires et de photographies aériennes. Plusieurs contraintes sont intrinsèques aux milieux nordiques du Québec et au jeu de données disponibles. Avec son apport aux inventaires actuels, l'objectif principal du projet sera atteint par le biais de trois objectifs spécifiques :

(1) Une méthode de cartographie existe (Leboeuf *et al.* 2007) pour la cartographie de la biomasse des peuplements forestiers à l'aide d'images satellitaires à très haute résolution spatiale (ISTHRS). Dans le cadre de ce projet, nous visons d'étendre la méthode de fraction d'ombre au calcul de quatre attributs forestiers du peuplement : hauteur de peuplement, densité, surface terrière et volume marchand. Ces relations seront validées et pourront ensuite être appliquées aux ISTHRS disponibles pour produire plusieurs cartes locales d'attributs forestiers (L'article 1 soit Leboeuf *et al.* (2012) traite de cet objectif) ;

(2) Nous visons aussi d'étendre la méthode de fraction d'ombre aux photographies aériennes pour mesurer les attributs forestiers de la hauteur et du volume marchand des peuplements. Vue la grande hétérogénéité inter- et intra-photographie par rapport aux ISTHRS, les défis méthodologiques sont très importants pour l'application de la méthode de fraction d'ombre (L'article 2 soit Leboeuf et Fournier (2012) traite de cet objectif);

(3) Nous proposons enfin de développer une méthode pour produire une carte régionale des attributs forestiers à l'aide des cartes locales (produites lors des deux objectifs précédents). Les résultats cartographiques locaux viennent donc enrichir la base de données d'entraînement de la cartographie régionale pour permettre de spatialiser sur de grands territoire. Or, la problématique de délai d'acquisition entre les mesures terrains (PET) et les images satellites à l'échelle régionale persiste. Nous visons donc à porter une attention spéciale pour réduire les impacts négatifs entre les temps d'acquisition des bases de données. Enfin, il faudra tester les méthodes de KNN et de KMSN pour établir laquelle est la mieux adaptée pour la cartographie régionale (L'article 3 soit Leboeuf et Fournier (2012) traite de cet objectif).

3. Données

3.1. Description du territoire

Le territoire de cartographie couvre 470 000 km² et comprend la totalité du domaine bioclimatique de la pessière à mousse et la portion de la pessière à lichen au sud du 53^{ième} parallèle. Ces deux domaines bioclimatiques sont définis par MRN (1998). Le territoire est couvert par une cartographie écoforestière nordique à l'échelle 1 : 100 000 sur le 240 000 km² localisé dans la portion nord et une cartographie écoforestière à l'échelle 1 : 20 000 dans la portion sud. Le contour de ce territoire apparaît en rouge sur la figure 1. L'altitude y varie entre

1 135 m et le niveau de la mer et les massifs d'altitude sont localisés au centre du territoire (Monts Otish, Monts Groulx, etc.). Les milieux forestiers diffèrent grandement suivant les contextes climatiques et géographiques. Les espèces forestières les plus communes sur le territoire sont l'épinette noire (*Picea mariana* Miller), le sapin baumier (*Abies balsamea* Linné), le pin gris (*Pinus banksiana* Lamb.), bouleau à papier (*Betula papyrifera* Marshall.), le mélèze laricin (*Larix laricina* Du Roi), le peuplier (*Populus* spp.) et l'épinette blanche (*Picea glauca* Moench) couvrant respectivement 71%, 13%, 12%, 3%, 1%, 1% et 0,4% des espèces mesurées à partir des PET. Quelques autres espèces non commerciales comme le bouleau nain (*Betula pumila* L.) et glanduleux (*Betula glandulosa* Michx.), le saule (*Salix* spp.) et aulnes (*Alnus* spp.) sont aussi communs.

Un total de 22 sites d'études ont été utilisés pour cartographier localement les attributs forestiers. Leur étendue spatiale correspond à l'étendue des ISTHRS (QuickBird ou WorldView) pour 12 sites d'étude ou de photographies aériennes pour 10 sites d'études (figure 1).

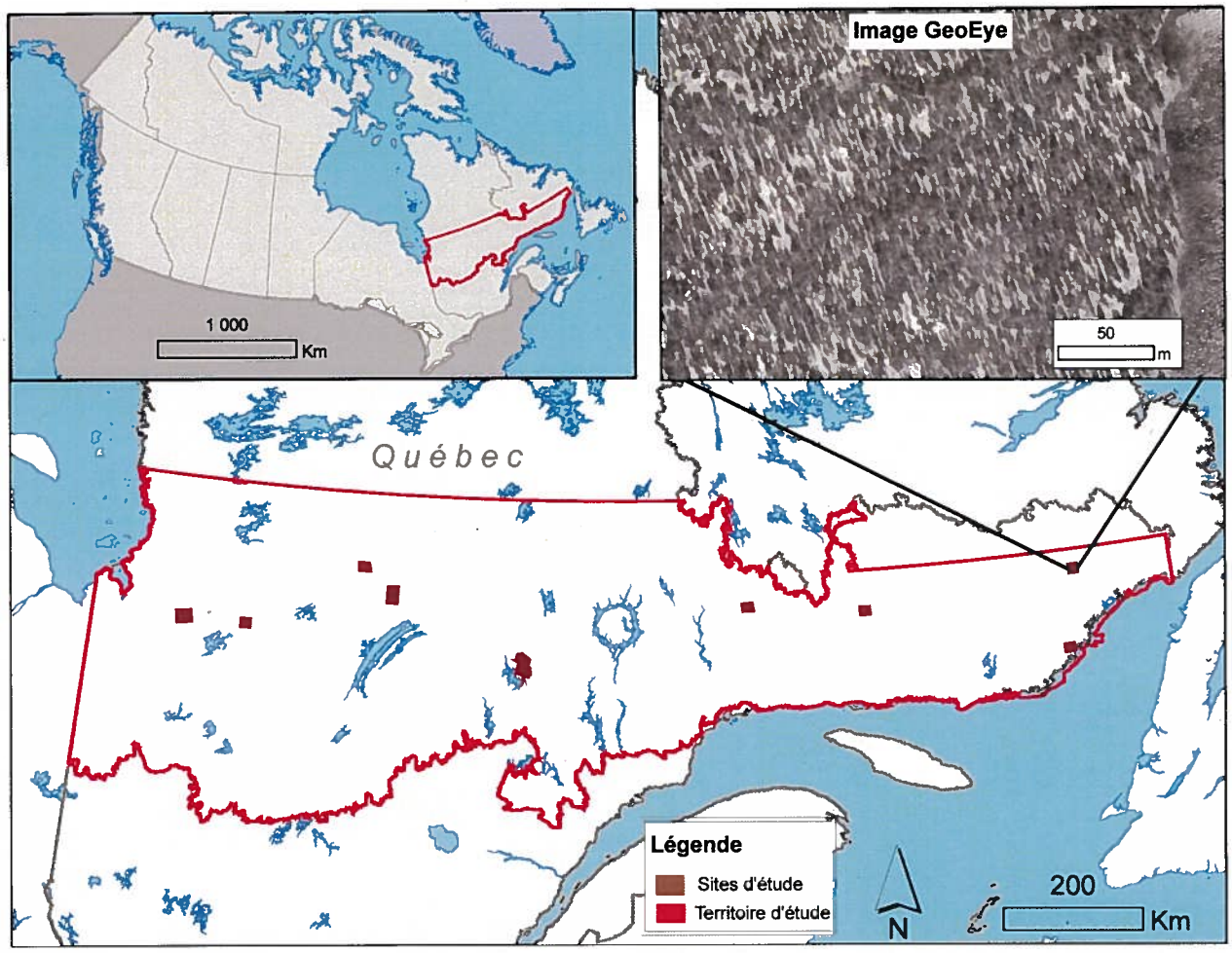


Figure 1 - Le territoire d'étude (contour rouge) ainsi que les 22 sites d'étude (contours bruns) réfèrent aux images satellitaires à très haute résolution spatiale et aux photographies aériennes.

3.2. PET et données auxiliaires

Deux ensembles distincts de PET ont été utilisés pour calibrer et valider les modèles : un premier ensemble de 108 PET mesuré entre 2002 et 2004 par le Service Canadien des Forêts (SCF) est lié au calcul et à la validation des modèles de fraction d'ombre à partir d'images QuickBird (premier sous-objectif). Un second ensemble de 7 579 PET mesuré entre 1996 et 2009 par le MRNQ est lié au calcul et à la validation des modèles KNN à partir de photographies aériennes et d'images Landsat-TM (2^{e} et 3^{e} sous-objectifs). Ces deux ensembles de PET ont été mesurées en suivant les

normes écodendrométriques du Nord pour les PET acquises dans le cadre de la cartographie du Nord et la norme écodendrométrique du 3ième programme d'inventaire pour le reste des PET (MRNF, 2009b). Ces normes écodendrométriques étaient similaires en ce qui a trait aux mesures de bases nécessaires à ce présent projet. Ainsi, tous le DHP et hauteur d'arbres ont été mesurées selon les même spécifications. Comme le jeu de données précédent, deux types d'attributs ont été mesurés pour chaque PET : (i) des attributs au niveau de l'arbre et (ii) des attributs au niveau de la PET (calculés à partir des attributs de l'arbre). Au niveau de l'arbre, le DHP et l'essence de chaque tige ont été notés pour tous les arbres de la PET ayant un DHP > 9 cm. Au niveau de la PET, trois attributs ont été calculés : la hauteur, la surface terrière (m^2/ha) et le volume marchand (m^3/ha) (MRNF, 2009b). Dans cet ensemble de données, la densité du couvert n'a pas été calculée puisqu'elle est déjà disponible à l'échelle régionale avec les cartes produites par photo-interprétation. Premièrement, la hauteur d'arbre individuelle est calculée à partir de la méthodologie proposée par Fortin *et al.* (2009) qui a été adaptée à notre territoire d'étude soit le domaine bioclimatique de la pessière à mousse. Le jeu de données visant à générer le modèle était constitué de 37 258 mesures de DHP et de hauteurs d'arbres. La hauteur de la PET est donc une moyenne de hauteur des 4 arbres ayant les DHP les plus élevés. Deuxièmement, la surface terrière de la PET a été obtenue par la somme des surfaces de tronc d'arbre à hauteur de 1,3 m pour les arbres dont le DHP > 9 cm. Les valeurs de hauteur d'arbre individuel utilisées pour ce modèle sont celles calculées précédemment. Troisièmement, le volume marchand (m^3/ha) au niveau de la PET a été obtenu en sommant le volume marchand de chaque tige d'arbre ayant un DHP > 9 cm. Ces volumes d'arbres individuels ont été calculés en utilisant la méthodologie de tarifs de cubage développée par Fortin *et al.* (2007) mais aussi adaptée spécifiquement à notre territoire d'étude. Un modèle a été calculé pour chaque essence forestière présente en utilisant 3 912 arbres échantillons répartis à travers 1 026 PET.

3.3. Images satellitaires à très haute résolution spatiale (ISTHRS)

Chacun des 12 sites d'étude comporte une image de la bande panchromatique des capteurs QuickBird ou World-View (450-900 nm), de 0,6 m de résolution spatiale, couvrant des superficies de 93 à 147 km^2. Ces images ont été acquises entre 2002 et 2009 durant la saison de croissance. Toutes les images satellitaires ont une résolution radiométrique de 16 bits. Les conditions d'acquisition impliquaient : (i) un ciel clair (moins de 10 % de nuage) ; (ii) un angle

de visée similaire d'une image à l'autre et près du nadir ; et (iii) des angles zénithaux solaires similaires (31° à 41°). Les images ont été corrigées géométriquement à l'aide de 12 points de contrôle de précision au sol de précision de 5 m ou moins. Un rééchantillonnage par la méthode du plus proche voisin et une correction polynomiale de premier ordre ont permis d'obtenir une précision de moins de 15 m. Le tableau 1 résume les caractéristiques et conditions d'acquisition des images satellitaires à très haute résolution spatiale.

Tableau 1 - Caractéristiques et conditions d'acquisition des images satellitaires à très haute résolution spatiale

	Date, UTC (jj/mm/aaaa, hh:mm)	**Soleil zénithal / azimuthal** (°)	**Capteur zénithal / azimuthal** (°)	**Aire** (km²)
QuickBird	10/07/2003, 15:33	31,4/143,2	11,5/96,4	147
QuickBird	12/08/2003, 16:22	41/152,9	13,3/172,3	93
QuickBird	10/08/2002, 16:18	38,9/157,7	9,6/225,6	107
QuickBird	10/06/2006, 16:15	30,6/159,8	10,3/145	114
QuickBird	20/08/2006, 18:59	39,8/167,4	19,2/339,9	89
QuickBird	12/07/2008, 18:23	29,3/163,9	22,9/322,1	34
WorldView	11/07/2008, 16:10	31,6/150,4	14,5/132,5	151
WorldView	16/06/2009, 15:12	28,6/157,8	3,3/100,8	26
WorldView	17/09/2009, 15:28	50,0/166,9	18,6/49,2	25
WorldView	13/08/2009, 15:49	37,8/164,4	17,2/324,6	27
WorldView	28/06/2008, 16:12	30,0/158,2	19,4/292,9	36
WorldView	06/09/2009, 16:24	45,5/166,2	4,3/176,6	30

3.4. Photographies aériennes

Chacune des 10 parcelles de photographies aériennes du territoire était constituée de 4 orthophotographies individuelles acquises en 2006 et dont les paramètres d'acquisition sont donnés au tableau 2. Chacune de ces photographies a été corrigée géométriquement à l'aide d'au moins 10 points de contrôle acquis à partir des couches de l'hydrographie à l'échelle 1 : 100 000. Un rééchantillonnage par la méthode du plus proche voisin et une correction polynomiale de premier ordre ont permis d'obtenir une précision au sol de moins de 10 m. Enfin, ces tuiles de photographies aériennes ont été assemblées en une mosaïque.

Tableau 2 - Caractéristiques de prise de vue des photographies aériennes utilisées

Numéro de rouleau	Numéro de parcelle de photographies	Date d'acquisition	Type de caméra	Distance focale nominale
Q06175	1	2006-09-02	Wild RC20	153 mm
Q06164	2, 3, 4	2006-09-01 et 2006-09-02	Wild RC10	153 mm
Q06161	5, 6, 7, 8, 9, 10	2006-09-01	RMK TOP 15	153 mm

3.5. Images satellitaires Landsat-TM

Une mosaïque composée de 68 images Landsat-TM a été utilisée pour couvrir le territoire à cartographier de 470 000 km². Ces images ont été acquises entre le 3 juillet 2005 et le 18 octobre 2009 bien que la grande majorité des images ont été acquises durant la saison de croissance de 2009. La mosaïque a été assemblée de façon à ce qu'elle ne présente que très peu de zones non utilisables à cause de couverture nuageuse ou voile atmosphérique. Trois bandes spectrales sont disponibles : la bande rouge (0,63 à 0,69 µm) qui est sensible à la présence ou à l'absence de chlorophylle, la bande proche infrarouge (0,76 à 0,90 µm) qui est sensible à la structure des couverts végétaux et la bande moyen infrarouge (1,55 à 1,75 µm) qui est sensible à la présence de d'eau dans les plantes (Jensen, 2005). Ces trois bandes ont été choisi parmi les 8 bandes disponibles de Landsat pour leur potentiel de différenciation des classes forestières du territoire et pour leur plus grande facilité d'assemblage des tuiles entre elles (Robitaille *et al.*, 2009). La résolution spatiale de ces images est de 30 m au sol et leur résolution radiométrique est de 8 bits. Les images ont été corrigées géométriquement à l'aide de la carte hydrographique qui est à l'échelle 1 : 100 000. Un rééchantillonnage par la méthode du plus proche voisin et une correction polynomiale de deuxième ordre ont permis d'obtenir une précision de moins de 30 m. Des étalonnages entre les images ont été nécessaires et ont été réalisés à l'aide des algorithmes de mariage d'histogramme « histogram matching » du logiciel PCI Geomatica (PCI Geomatics Enterprises Inc., 2003). Cet algorithme permet à l'utilisateur de spécifier des zones communes entre les différentes tuiles afin d'arrimer les teintes de ces tuiles.

4. Description sommaire des méthodes des articles 1, 2 et 3.

Pour répondre aux besoins d'inventaire forestier extensif dans les régions nordiques du Québec, une méthode de cartographie basée sur des images satellitaires a été développée en trois étapes, répondant à chacun des trois objectifs spécifiques du projet. La première étape, visait à établir une relation statistique entre la fraction d'ombre calculée à partir d'ISTHRS (par exemple, QuickBird) et les attributs forestiers : hauteur du peuplement, densité, surface terrière et volume marchand. La deuxième étape visait aussi à établir de nouvelles relations statistiques entre la fraction d'ombre et deux attributs forestiers (hauteur de peuplement et volume marchand) mais calculées à partir de photographies aériennes. Enfin, la troisième étape visait à établir une méthode de généralisation spatiale pour la cartographie régionale de la hauteur et du volume des peuplements de la région boréale. Pour pallier aux manquements de la base de données disponible, il a fallu développer une procédure adaptée pour la sélection des PET d'entraînement et à utiliser les PET et les placette satellitaire (PET_s) disponibles comme ensemble d'entraînement. Ces données, dorénavant complètes, ont permis l'application des méthodes de KNN et KMSN résultant à une carte des attributs forestiers pour l'ensemble du territoire d'étude.

4.1. Cartographie par fraction d'ombre avec les ISTHRS (Article 1 : Leboeuf et al. 2012)

Plusieurs méthodes existent pour la cartographie des attributs de peuplements forestiers à l'aide des ISTHRS. Les travaux de Peddle *et al.* (2001), Seed et King (2003) et Leboeuf *et al.* (2007) ont montré qu'il était possible de calculer certains attributs forestiers tels que la biomasse et la surface foliaire. La méthode proposée ici innove avec le calcul d'attributs pertinents à la foresterie, notamment la hauteur du peuplement, la densité, la surface terrière et le volume marchand. Trois étapes sont nécessaires pour le développement de la méthode par fraction d'ombre pour ces nouveaux attributs :

(i) Calculer la fraction d'ombre pour chaque PET à partir d'une approche basée sur le pixel et normaliser les valeurs résultantes ;

(ii) Calculer et valider, pour chaque site d'étude, une régression linéaire mettant en relation les valeurs de fraction d'ombre et les valeurs d'attributs forestiers (hauteur du peuplement, la densité, la surface terrière et le volume marchand). Ensuite, dans un souci de généralisation,

il a été important de déterminer si une régression globale ne pouvait pas être calculée pour son application sur l'ensemble des sites d'étude ;

(iii) Produire une carte d'attributs forestiers (matricielle et polygonale) à l'échelle du site d'étude pour la zone couverte par l'ISTHRS et la valider à partir des PET non utilisées pour le calcul des régressions.

Les travaux effectués par Leboeuf *et al.* (2007) procurent quatre directives méthodologiques spécifiques à suivre. Premièrement, la bande panchromatique seule semble suffisante pour établir des relations au même niveau de précision qu'avec l'utilisation des quatre bandes multi-spectrales ce qui simplifie le processus : réduction des coûts d'acquisition et du temps de traitement. Deuxièmement, le seuil spectral pour distinguer les pixels ombragés et non-ombragés peut être établi visuellement car il procure des relations similaires et cela même si ce seuil diffère quelque peu d'un interprète à l'autre. Troisièmement, la régression linéaire a été favorisée lorsque comparée à d'autres structures de régressions pour sa simplicité tout en procurant des résultats similaires. Finalement, une dimension de 30 m pour les cellules de la carte est judicieuse car elle procure des résultats similaires à des tailles plus réduites (10 et 20 m) et plus étendues (60 et 90 m). Ces conclusions ont été tirées des travaux de Leboeuf et al., 2007. Avec ces bases de travail, nous pouvons proposer une démarche pour établir les relations de fraction d'ombre pour les ISTHRS.

La première étape consiste à établir les zones d'ombre pour le calcul de la fraction d'ombre. Une couche binaire, ayant des valeurs soit de 0 (non-ombre) ou de 1 (ombre), est donc créée pour chaque image à l'aide d'un seuil spectral appliqué aux valeurs de niveau de gris des pixels de l'image. Le seuil spectral est établi visuellement par un interprète pour chaque site d'étude selon les deux considérations suivantes : (i) éviter une surestimation des zones d'ombres et (ii) assurer une proportion réaliste des zones d'ombres sur les multiples types de végétation de sous-couvert (lichen, mousse ou arbuste). Un carré de référence de 30 m a été centré au dessus de chaque PET. La fraction d'ombre a été déterminée pour chacun de ces carrés de référence, à l'aide de la couche binaire (ombre/non-ombre) en calculant la proportion de surface ombragée par rapport à la surface totale :

$$FO = A_{OA} / A_{CR} \quad (1)$$

où FO est la fraction d'ombre, A_{OA} est l'aire couverte par l'ombre des arbres et A_{CR} est l'aire totale du carré de référence. Les valeurs de fraction d'ombre sont normalisées sur une géométrie

commune afin de tenir compte des différences d'angles du satellite et du soleil selon les sites d'étude. Le processus de normalisation est basé sur une fonction trigonométrique qui modifie l'aire de l'ombre projetée au sol ou sur une cime vue par le satellite. Chaque PET a donc été associée à une valeur de fraction d'ombre compatible les unes aux autres.

Comme deuxième étape, une régression linéaire locale, c'est-à-dire par site d'étude, a été calculée entre la fraction d'ombre et chacun des attributs forestiers sélectionnés. Pour cela, 70% des PET (PET_{CAL}) ont été sélectionnées aléatoirement. Il en résulte un modèle de calcul des attributs selon :

$$AF = a + b*FO + \varepsilon \qquad (2)$$

où AF est l'attribut forestier mesuré sur la PET_{CAL}, b est la pente de la régression, a est l'ordonnée à l'origine et ε la valeur résiduelle. Une régression a ensuite été calculée à l'aide des PET_{CAL} de tous les sites d'étude pour produire des relations globales entre la fraction d'ombre et chaque attribut forestier. Le coefficient de régression ajusté (R^2_{adj}), l'erreur quadratique moyenne (RMSE) et le biais ont été calculés à l'aide de 30% des PET (PET_{VAL}) pour établir la précision de cette régression. Enfin, des tests statistiques ont été faits afin d'évaluer s'il y avait des différences significatives entre les trois régressions locales. Le test de Fisher permet d'évaluer les hypothèses de convergence (appliqué à l'ordonnée à l'origine) et de parallélisme (appliqué à la pente) des jeux de données d'une PET_{CAL} à l'autre en utilisant un seuil de $\forall$=0,05 (Milton et Arnold, 2003).

Les cartes d'attributs forestiers ont été produites à l'aide des régressions établies pour l'étendue des sites d'étude. Une grille ayant des carrés de 30 m a été appliquée à l'ISTHRS afin de créer des cellules compatibles à la résolution spatiale des images Landsat-TM. Un seuil de différenciation entre les pixels ombre et non-ombre est appliqué à l'ISTHRS. Il a donc été possible de calculer la valeur de fraction d'ombre pour chaque cellule de 30 m de la grille. Comme les relations statistiques par régression sont connues, la valeur de chaque attribut forestier est finalement calculée pour chaque cellule en appliquant la valeur de fraction d'ombre à la régression globale. Une opération de nettoyage est effectuée pour éliminer les cellules qui ne doivent pas servir dans les calculs, telles que celles dominées par des essences feuillues, celles affectées par l'ombre des nuages, ou les cellules couvrant des lacs ou des tourbières. Un calcul de RMSE et de biais a enfin été effectué en comparant les valeurs des PET_{VAL} avec celles des

cellules de la grille. Comme validation supplémentaire de ces cartes, nous avons comparé la somme des valeurs du volume marchand de toutes les cellules de la grille avec le volume marchand obtenu par la carte écoforestière, et ce pour toute la surface commune à ces deux cartes. Chaque cellule de 30 m de ces cartes locales constitue donc une placette satellitaire (PET_s). Les cartes locales produites deviennent donc disponibles pour permettre la spatialisation à l'échelle régionale en utilisant les PET_s. Les résultats statistiques obtenus pour les modèles et pour les cartes produites permettent de conclure que ces résultats peuvent être utilisés comme entraînement de la cartographie régionale de la section 7.

4.2. Cartographie locale d'attributs forestiers par fraction d'ombre et photographies aériennes (Article 2 : Leboeuf et Fournier, 2012)

La méthode développée et testée à l'étape précédente a permis de cartographier les attributs forestiers à partir d'ISTHRS. Les résolutions spatiales et spectrales similaires entre ces images et les photographies aériennes nous ont amené à évaluer leur potentiel pour le calcul d'attributs forestiers à l'échelle locale. La méthode proposée innove en ajoutant l'utilisation des photographies aériennes à la méthode de fraction d'ombre pour le calcul d'attributs forestiers alors qu'aucune autre étude n'a encore démontré cette possibilité de façon opérationnelle.

Le calcul d'attributs forestiers par fraction d'ombre nécessite trois étapes méthodologiques similaires à celles de la section 5 qui consistent à: (i) calculer la fraction d'ombre pour chaque PET à partir d'une approche basée sur le pixel; (ii) générer et valider une régression linéaire reliant les valeurs de fraction d'ombre et les valeurs d'attributs forestiers (hauteur de peuplement et volume marchand) ; (iii) analyser les approches permettant de réduire les biais induits par la bordure de la photographie comme les modèles de normalisation, des distances et angles optimaux. Les 144 PET d'entraînement et de validation utilisés pour le développement de cette méthode ont été tirés du second ensemble de 7 579 PET mesurées par le MRNQ.

Les conclusions tirées de cette étude permettent d'abord d'affirmer que les méthodes de fraction d'ombre peuvent s'appliquer aux photographies aériennes bien que certaines contraintes méthodologiques doivent être prises en compte. En effet, le fait que la photographie aérienne soit acquise à basse altitude déforme la surface de l'ombre observée. Ainsi, l'ombre est surestimée

dans la direction du soleil et sous-estimée à l'opposé. Pour stabiliser les valeurs de fraction d'ombre sur l'ensemble de la photographie aérienne, différents aspects ont été explorés : (i) la normalisation de l'image comme celle proposée par Leboeuf *et al.* (2007), (ii) la distance entre les PET et le centre de la photographie aérienne et (iii) l'angle entre la PET et le centre de la photographie aérienne. Les résultats obtenus indiquent que l'approche de normalisation n'améliore pas les résultats statistiques. Ceci est probablement causé par le fait que cette approche de normalisation est basée sur une base trigonométrique et est sans doute mal adaptée à une structure de peuplement forestier. Cependant, les résultats indiquent clairement l'impact de la distance de la PET au centre de la photographie aérienne. Une distance maximale de 1 000 m du centre de la photographie aérienne a ainsi pu être établie. L'angle entre la PET par rapport au centre de la photographie aérienne affecte aussi considérablement les résultats. Les quartiers est et ouest de la photographie aérienne sont plus favorables. Ces paramètres spécifient donc les zones de la photographie aérienne où les résultats obtenus par le modèle de fraction d'ombre sont les plus élevés. Ainsi, à l'intérieur de ces limites méthodologiques, il est possible de cartographier les deux attributs forestiers à l'aide de photographies aériennes panchromatiques. Ceci ouvre donc la voie à une utilisation étendue de cette méthode puisque les photographies aériennes sont très répandues sur le territoire à travers différents projets de cartographie.

Bien que les résultats obtenus à travers cette section soient encourageants, les sites disponibles pour cette étude étaient trop éloignés de notre territoire d'application pour pouvoir les utiliser comme PET_s dans la cartographie régionale de la section 7. Les résultats obtenus ouvrent néanmoins la voie à une application de cette méthode à de grands territoires dans la mesure où les contraintes d'utilisation soulevées dans cette étude sont respectées.

4.3. Sélection des PET d'entraînement et cartographie régionale d'attributs forestiers par méthode KNN et mosaïque d'images Landsat-TM (Article 3 : Leboeuf et Fournier, 2012)

L'objectif principal du projet de recherche visait le développement de nouvelles méthodes de cartographie des attributs forestiers sur de grands territoires au nord du Québec. Vu que plusieurs contraintes sont intrinsèques à ces milieux, différentes approches ont dû être développées afin de les surmonter. Une première contrainte reliée à la faible représentativité de la donnée d'entraînement (PET mesurées fournissent des hauteurs de peuplement de 7 m et plus), est

surmontée avec l'apport de la cartographie locale utilisant des images satellitaires à très haute résolution pour générer de nouvelles PET_s pour les peuplements ayant 7 m ou moins de hauteur. Néanmoins, une autre contrainte majeure liée à l'écart entre la mesure des PET et l'acquisition d'images demeurait importante. Dans certains cas, l'écart était de 13 ans. Cette faible représentativité entre l'image et la mesure terrain peut affecter considérablement le résultat de cartographie KNN ou KMSN (Franklin et Wulder, 2002). En conséquence, une méthode de sélection de ces 7 579 PET a donc été développée afin de retirer les PET dont la valeur d'attribut forestier n'est pas représentative de la situation de l'image. Trois facteurs principaux peuvent expliquer le manque de représentativité de certaines PET : (1) le changement de statut entre le temps de mesure de la PET et de l'acquisition de l'image causé par des travaux d'aménagement forestier, des feux, des épidémies, des chablis ; (2) des erreurs de positionnement des PET ou de l'image ; (3) une altération du milieu causé par différents éléments qui modifient la valeur spectrale de l'image comme une inondation. Ces facteurs peuvent amener un biais important aux modèles utilisés puisque ces PET « bruitées » réduisent la tendance entre la variable d'image et l'attribut forestier mesuré. Deux méthodes de sélection des PET les plus représentatives ont donc été développées et testées dans le cadre de ce travail pour l'ensemble du jeu de données. La première méthode visait à éliminer les PET non représentatives de la classe d'attribut forestier. La valeur spectrale de chaque bande de la mosaïque d'images Landsat (composée de 3 bandes : moyen infrarouge, proche infrarouge et rouge) a été associée à chaque PET. Ensuite, les valeurs de volume des PET ont été classées par dizaines (0-5 m^3/ha; 5-15 m^3/ha ; etc.). La moyenne des valeurs spectrales a été calculée pour chaque classe et pour chaque bande. Enfin, la distance euclidienne (distance entre la valeur en volume de la PET et la moyenne de sa classe) a permis d'identifier les PET qui sortent du lot en terme de valeurs spectrales. Cette discrimination s'est réalisée par itération de différents seuils (0,5, 1, 1,5 et 2 fois l'écart type) permettant de déterminer quelles PET devaient être utilisées dans l'ensemble d'entrainement dans la méthode de cartographie par KNN.

Une seconde méthode de sélection visait à éliminer les PET perturbées et non représentatives. Sur la base des valeurs spectrales associées aux PET, une classification non supervisée ISODATA a été effectuée pour déterminer 40 classes spectrales uniformes. Cette classification a permis de cibler quelles classes correspondent à des sites perturbés pour éliminer ces PET de la cartographie KNN. La détermination du statut des classes a été réalisée avec l'appui de données

terrain comme des photographies au sol ou aériennes. Ensuite, la valeur moyenne et l'écart type de chacune des classes restantes (non perturbées) a été calculé pour le volume. Les PET, dont la valeur de volume se trouve au-delà du seuil optimal déterminé (0,5, 1, 1,5 et 2 fois l'écart type), ont été éliminées du calcul de KNN. Les résultats obtenus nous ont permis de déterminer deux modes opérationnels pour la sélection des PET pertinentes pour l'entraînement de notre méthode de cartographie à l'échelle régionale.

La faible densité de données d'entraînement qui couvre le vaste territoire d'étude n'a pas permis d'utiliser les tarifs de cubage traditionnels pour calculer les volumes marchands des polygones écoforestiers (Robert et Robitaille, 2009a). Aussi, les approches de cartographie développées pour le Nord du Québec ne permettaient pas de calculer la hauteur de peuplements puisque les images utilisées (Landsat-TM) ne permettaient pas d'effectuer l'interprétation en 3D. Ces limites du processus de cartographie du territoire d'étude forcent le développement de méthodes adaptées basées sur des données d'entraînement à différentes échelles. La cartographie régionale s'est donc appuyée sur les méthodes KNN et KMSN qui ont largement démontré leur efficacité. Nos développements ont établi quelles méthodes de calcul, KNN (équation 3) ou KMSN (équation 4) et quelles couches d'informations (bandes spectrales, données du milieu physique, données climatologiques) permettent de cartographier à une échelle régionale les attributs forestiers le plus précisément possible.

Le modèle KNN a utilisé l'ensemble final des PET d'entraînement déterminés lors de l'étape méthodologique précédente (méthode de sélection des PET). Les couches d'informations pertinentes ont été déterminées par un test de corrélation de Pearson. Ainsi, plusieurs éléments cartographiés pour l'ensemble du territoire ont ainsi été évalués comme l'altitude, la pente, le drainage, les degrés-jours, etc. Sur la base de l'ensemble final des couches pertinentes, la valeur d'attributs forestiers ($AFknn_i$) a été calculée pour chaque pixel de la mosaïque d'image en suivant cette équation :

$$AFknn_i = \frac{\sum_{k=1}^{k} W_k AF_{PET\_k}}{\sum_{k=1}^{k} W_k} \; pour \; k \neq 1 \;\; où \;\; W_k = \frac{1}{d_k^j} \qquad (3)$$

où $AFknn_i$ est la valeur d'attributs forestiers calculé par le modèle KNN pour la cellule i ; k est le $k^{ième}$ plus proche voisin spectral ; AF_{PET_k} est la valeur d'attributs forestiers provenant des cartes générées par fraction d'ombre et des ISTHRS pour cette section-ci; W_k est le coefficient de poids ; d_k^j est la distance euclidienne spectrale sur laquelle est appliquée la puissance j. La valeur $AFknn_i$ résultante du modèle KNN est donc une valeur moyenne des AF_{PET} qui sont au niveau spectral le plus près des valeurs d'une cellule donnée. Le paramètre j = 0 génère une valeur moyenne simple, le paramètre j = 1 génère une valeur moyenne pondérée à la puissance 1 et le paramètre j = 2 génère une valeur moyenne pondérée à la puissance 2.

La cartographie d'attributs par KMSN se calibre également à partir de l'ensemble des PET sélectionnées. L'approche diffère toutefois de celle utilisée par le KNN puisqu'elle est basée sur les attributs des polygones écoforestiers et non uniquement sur les valeurs de pixel (Crookson *et al.*, 2002). L'approche a nécessitée trois étapes méthodologiques qui consistent à : (i) étalonner les données numériques de la carte écoforestière (ex. densité du couvert); (ii) calculer la moyenne de valeurs spectrales, altitudes, données de climat provenant de différentes cartes numériques (image satellitaire, données climatologiques, altitude, etc.) et ce, pour chacun des polygones écoforestiers ; (iii) faire des essais KMSN pour déterminer quels sont les intrants optimaux. La valeur d'attributs forestiers ($AFkmsn_i$) a donc été calculée pour chacun des polygones écoforestiers selon cette équation :

$$AFkmsn_i = observation\ de\ référence\ j\ avec\ d_{ij}^2 \quad minimum \qquad (4)$$

$$d_{ij}^2 = (X_i - X_j)W(X_i - X_j)\ pour\ tous\ j = 1,...,n$$

où $AFknn_i$ est la valeur d'attributs forestiers du voisin le plus similaire de la $i^{ième}$ observation cible; i est la valeur cible; j est la valeur de référence (parmi toutes les valeurs de référence) qui minimise la distance euclidienne pondérée du jeu de variable (Moeur et Stage, 1995); d^2_{ij} est la distance au carré entre la $i^{ième}$ observation cible et la $j^{ième}$ observation parmi les n observations de référence ; X_i est le vecteur des X variables normalisée pour la $i^{ième}$ observation cible; X_j est le vecteur des X variables normalisé pour la $j^{ème}$ observation de référence ; W est la matrice de poids. Les résultats obtenus de ces méthodes de calculs d'attributs forestiers permettront ainsi de

déterminer quelles sont les méthodes et les couches de bases pertinentes pour cartographier les attributs forestiers à une échelle régionale.

5. Article 1

FOREST ATTRIBUTE ESTIMATION OF NORTHEASTERN CANADIAN FORESTS USING QUICKBIRD IMAGERY AND A SHADOW FRACTION METHOD

A. Leboeuf [ab], R.A. Fournier [a], J.E. Luther [c], A. Beaudoin [d], and L. Guindon [d]

[a] Centre d'Applications et de Recherches en Télédétection (CARTEL), Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Quebec, J1K 2R1, Canada

[b] Ministère des Ressources Naturelles et de la Faune, Direction des Inventaires Forestiers, 880 Ch. Ste. Foy, Quebec City, Quebec, Canada G1S 4X4

[c] Natural Resources Canada, Canadian Forest Service - Atlantic Forestry Centre, P.O. Box 960, Corner Brook, Newfoundland and Labrador, A2H 6J3, Canada

[d] Natural Resources Canada, Canadian Forest Service - Laurentian Forestry Centre, P.O. Box 3800, Ste-Foy, Quebec, G1V 4C7, Canada

E-mail of corresponding author: antoine.leboeuf@mrnf.gouv.qc.ca

Telephone: 418-627-8669 ext. 4306 Fax: 819-821-7944

Manuscript accepted by Forest Ecology and Management

November 2011

Leboeuf, A., Fournier, R.A., Luther, J.E., Beaudoin, A., and Guindon, L. (2012) Forest attribute estimation of northern Canadian forests using QuickBird imagery and a shadow fraction method. Forest Ecology and Management, Vol. 266, pp. 66-74.

ABSTRACT

This study demonstrates a method to map forest stand polygons based on four forest attributes (volume, basal area, height, and crown closure) using shadow fraction values estimated from high

spatial resolution QuickBird panchromatic images. The method was tested over three test sites in northeastern Canada that were largely dominated by black spruce (*Picea mariana* (Mill.) B.S.P.). The method involved four sets of procedures: (i) estimating shadow fraction from the panchromatic band of QuickBird images, (ii) generating site-specific and global regression models linking shadow fraction with each of the four forest attributes, (iii) mapping the forest attributes as a grid layer (30 m × 30 m) for each test site using the global regression models, and (iv) generating stand polygons from the raster layers. Each test site was covered by a QuickBird satellite image with 0.6 m x 0.6 m of spatial resolution. Between 2002 and 2004, 108 ground sample plots were acquired to develop local regression models. Goodness of fit (R^2) of the linear regression models between shadow fraction and stand attributes ranged from 0.55 to 0.79. Relative root mean square error and bias estimates were, respectively, 0.32 and 0.15 for volume; 0.29 and 0.14 for basal area; 0.32 and -0.04 for crown closure; and 0.17 and 0.05 for height. Stand volume maps were produced from both the shadow fraction method and conventional forest stand maps (derived from aerial photo-interpretation) for a test site. Volume patterns were similar, and total volume for the test site differed by only 5.6% between the two maps. Lastly, the raster images derived from the shadow fraction method were used to produce a stand map following guidelines similar to those used by provincial inventory. In all cases, our results suggest that the shadow fraction method is a reliable and convenient way to map forest stand polygons and related attributes of black spruce stands of northeastern forests of Canada.

Keywords: boreal forest, forest inventory, mapping, remote sensing, stand attributes, volume.

1. Introduction

Forest inventory programs are central to forest resource management and monitoring. In Canada, typical forest inventories include: (i) stand maps in polygon format made from visual interpretation of aerial photographs at scales generally of 1:15 000 and (ii) ground sample plots (GSP). In most Canadian provinces, photo-interpreters are trained to identify stand polygons with similar attributes (species classes, crown closure, mean height, and age) with a minimum mapping unit of around 2 ha (Avery and Burkhart, 2001; Létourneau *et al.*, 2003). Tree-level measurements such as species, diameter at breast height (DBH) measured at 1.3 m, height, crown closure, and density are obtained for GSP. These measurements provide more detailed and more

accurate information than stand maps but are costly as they require substantial time and investment. Information from both the stand maps and GSP enable the estimation of other forest attributes. For instance, stand volume tables or models are often derived from the GSP and subsequently applied to attributes contained in the stand maps to produce stand volume maps (Avery and Burkhart, 2001; Perron, 1986; Ker, 1974).

Conventional forest inventory programs based on aerial photo-interpretation and GSP, such as Quebec's, are expensive to implement, especially in remote areas where infrastructures are not well developed. For that reason, forests north of the 52nd parallel in eastern Canada and north of the 58th parallel in western Canada currently have limited inventories compared with southern forests, even though there is an increased need for reliable information on forest status. Northern Canadian forests are largely located within the boreal ecoregion, which represents 67% of Canada's total forest land (Lowe *et al.*, 1994). Most of the northeastern Canadian forests are homogeneous in term of species and structure and are dominated by black spruce (*Picea mariana* (Mill.) B.S.P.) stands. These conditions dictate a need for new forest inventory methods to deal with, in particular, the inaccessibility and high cost of inventories in these regions.

The recent availability of high spatial resolution satellite imagery (HSRSI) for northern regions offers an interesting alternative to aerial photography for forest inventory. Aerial photography and HSRSI have similar spatial resolutions, but HSRSI offers distinct advantages for data processing, particularly for acquisition geometry. Moreover, progress has been made in recent years using HSRSI to assess forest stand attributes (Hyyppä *et al.*, 2000; Lévesque and King, 2003). Strong potential of such images was shown using segment-based approaches to overcome spatial heterogeneity (Pekkarinen, 2002). Also, image processing methods based on the analysis of individual tree crowns were recently developed especially for high crown closure stands (Gougeon and Leckie, 2003; Wulder *et al.*, 2000).

Generally, HSRSI acquired over boreal regions displays a visual gradient of intensities where light-colored areas correspond to open stands and dark areas correspond to stands with high crown density. This gradient of intensities is largely due to tree shadow. Tree shadow is composed of a shadowed portion within the crown and the shadow cast on the ground by crowns

(Li and Strahler, 1985). The two shadow types are not always distinct on the images or photographs. Tree shadow fraction (SF) is the sum of all tree shadow areas divided by the total ground area (Seed and King, 2003). The relationships between SF and forest attributes (FA) have been confirmed by several studies, which led to the development of regression models for predicting forest attributes from SF. For example, at the tree level, Greenberg *et al.* (2005) developed regression models between tree shadow and several tree variables: DBH, crown area, and biomass. Other regression models were established between SF and stand-level attributes such as leaf area index (LAI) (Peddle and Johnson, 2000; Seed and King, 2003). Leboeuf *et al.* (2007) also developed regression models between SF and stand biomass. The results of these and other studies suggest that SF from HSRSI may be useful for estimating and mapping a suite of FA in northern Canadian forests.

Consequently, the objectives of this study were to expand and evaluate the utility of the SF method of Leboeuf *et al.* (2007): (i) to map four additional FA of black spruce stands in Canada, i.e., merchantable volume (m^3/ha) (hereafter referred to as volume), basal area (m^2/ha), crown closure (%), and height (m), and (ii) to use the SF method to generate forest polygon maps representing homogeneous areas with respect to the suite of FA values. The expansion of the SF method was made possible by: (i) testing regression model relationships between SF and FA over the three test sites, (ii) developing global regression models for the three test sites, (iii) producing and validating FA raster maps for the three test sites using the global regression models, and (iv) producing FA polygonal maps representing areas with homogeneous FA values to assess the relative impact of spatial generalization.

2. Test sites and data

2.1. Test site description

Three test sites were selected within the eastern taiga shield and boreal shield ecozones (Fig. 1) of northeastern Canada. The sites were near the towns of Chibougamau (CH) and Radisson (RA) in Quebec, and Wabush (WA) in Labrador. The spatial extent of each site corresponded to the coverage of a single QuickBird (QB) image. Total overall coverage of the three sites was 347 km^2. The locations and extents of the test sites were selected to provide variability of FA range and stand conditions that are typical of northeastern Canadian forests. Forest stands in the

test sites were largely composed of black spruce with the presence of a few other species such as jack pine (*Pinus banksiana* Lamb.) and balsam fir (*Abies balsamea* (L.) Mill.), which is typical for the region (Rowe, 1972). The understory of black spruce stands is generally covered by lichen, moss, and shrubs in various proportions. Mean stand volume and annual growth were 97 m^3/ha and 1.2 m^3/ha/yr, respectively, for the CH test site and 42 m^3/ha and 0.5 m^3/ha/yr for both RA and WA (Lowe *et al.*, 1994).

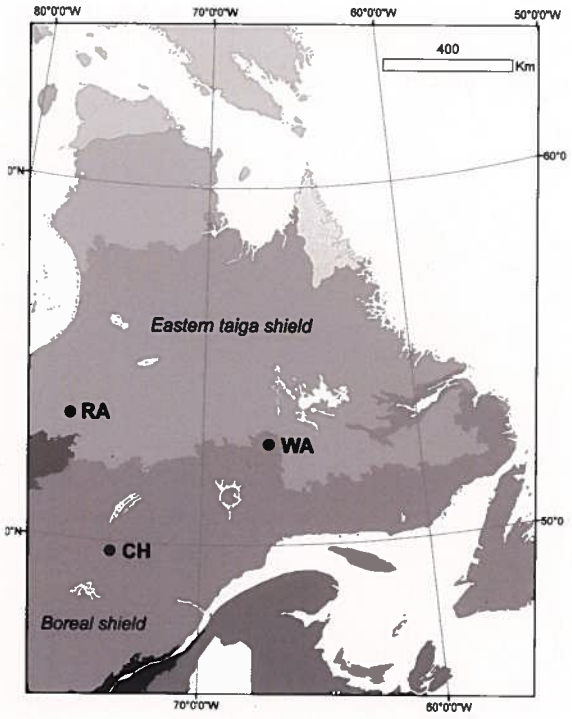


Figure 1 - Location of the three test sites in two northern ecozones of Eastern Canada: Chibougamau (CH) and Radisson (RA) and Wabush (WA).

2.2. Ground sample plots and forest maps

A total of 108 circular GSP (31 in CH, 49 in RA, and 28 in WA), each covering 400 m^2 and dominated by black spruce was established between 2002 and 2004 across the three test sites using National Forest Inventory (Gillis, 2001) procedures. The GSP were selected to represent a large range of the FA and understory types as per Leboeuf *et al.* (2007). Species composition of the GSP were such that: (i) black spruce represented more than 75% of the GSP basal area and (ii) any other species, such as trembling aspen (*Populus tremuloides* Michx.), white birch (*Betula papyrifera* Marsh.), *Alnus* spp., and *Salix* spp., represented less than 25% of the GSP basal area.

Several tree-level attributes were measured at each of the 108 GSP (Table 1). DBH was measured for all trees in the GSP with a DBH > 5 cm, and basal area was estimated from the DBH. In addition, tree-level attributes (height and volume) were estimated using published regression models where tree DBH was used as an independent variable (Table 2). Tree height was estimated from tree DBH using the Chapman-Richards model for black spruce as it produced the best results of the six models tested by Peng *et al.* (2001) for northern Ontario, Canada. In addition, the tree species and range of DBH at our test sites were similar to those from Peng *et al.* (2001). Tree stem volume was estimated using a general coniferous model, developed by Titus and Morgan (1988). For comparison purposes, the merchantable volume of the black spruce in the GSP was also estimated with a general coniferous model developed by Perron (1986) and used for black spruce in Quebec, Canada. The high correlation ($R^2 = 0.997$) and low bias (-0.007 m^3/ha) between tree volumes predicted by the two models confirmed the suitability of the general coniferous model for our test sites. The general coniferous model had the added benefit of being suitable for other coniferous species present in the GSP, for non-merchantable trees (DBH < 9 cm), and for areas outside of Quebec, and was also independent of local model attributes.

Tree-level forest attributes were further used to estimate plot-level FA (Table 2). For instance, plot-level volume was predicted by summing tree volumes of all merchantable trees (DBH > 9 cm) within the GSP. Plot-level basal area was estimated as the sum of the trunk circular areas at breast height (1.3 m) for all merchantable trees within the GSP. Crown closure was estimated visually on all GSP and validated on several GSP by two other techniques using the Plant Canopy Analyzer (LAI 2000® from LI-COR, Lincoln, Nebraska) and a densitometer (Beaudoin *et al.*, 2003). Plot-level height was estimated using Lorey's height as per Sharma *et al.* (2002) to improve objectivity and consistency of the estimation. Contrary to a qualitative mean height for the GSP, Lorey's height weights trees in direct proportion to their basal area (Shreuder *et al.*, 1993). Finally, the correlation among FA was assessed by graphing all plot-level values of stand volume by each of the other three FAs (Fig. 2).

A forest inventory stand map interpreted from 1:15 000 aerial color photography acquired in 1998 was available for the CH test site. The stand map contained stand boundary, species

composition, mean height class, crown closure class, stand age, and stand volume. Stand volume for all polygons was estimated from a volume model applied for this region (Perron, 1986).

Table 1 - Main forest attributes (FA) measured at the tree level for the 108 ground sample plots (GSP)

Inventory level	Forest attributs	Measure types	Attributes range
Tree level	Species	All trees with DBH > 0.05 m	Bs : 90 % ; Jp : 3 % ; Bf : 1 % and Ta : 1 %
	Diameter at breast height (DBH)	All trees with DBH > 0.05 m	0.3 to 39.2 cm (average: 8.75 cm)

Bs : Black spruce ; Jp : Jack pine ; Bf : Balsam fir ; Ta : Tamarack

Table 2 - Main forest attributes (FA) estimated at the plot level for the 108 ground sample plots (GSP)

Forest attributes	Estimation method	Attributes range
Volume	**Tree height:** $\mathbf{H_i} = 1.3+a(1-e^{-b*DBHi})^c$ (Peng et al, 2001) **Tree volume:** $\mathbf{TV_i} = (\mathbf{H_i} / 3)*\pi*(DBH_i/200)^2$ (Titus and Morgan, 1988) **Plot-level volume** $\mathbf{V_i} = \sum \mathbf{TV_i} *10\text{^}4/400$	0 to 200 m³/ha (average: 51 m³/ha)
Basal area	**Tree basal area :** $\mathbf{TBA_i} = \pi*(DBH_i/200)^2$ **Plot-level basal area** $BA_i = \sum \mathbf{TBA_i}$	0 to 37 m²/ha (average: 11 m²/ha)
Crown closure	Field visually assessment and measures	5 to 85 % (average: 33 %)
Height	Lorey's height: $\sum[\mathbf{TBA_i} * H_i] / \sum \mathbf{TBA_i}$	5 to 19 m (average: 10 m)

2.3. High spatial resolution satellite imagery (HSRSI)

QuickBird images were acquired with similar sun/sensor-viewing geometry for each test site and close to the peak of the growing season in 2002 or 2003. Table 3 summarizes the acquisition time, sun/sensor-viewing geometry, and coverage of each QB image. The panchromatic band (PAN) was used for this study following the recommendation of Leboeuf *et al.* (2007) on a similar study on stand biomass. The spatial, spectral, and radiometric resolutions were 0.6 m, 400–900 nm, and 16 bits, respectively. All images were geometrically referenced using at least

12 ground control points taken by a GPS with positional accuracies within 5 m. A first-order polynomial resulted in an average positional root mean square error (RMSE) between 10 m and 15 m. Relatively large average positional RMSE values were obtained compared with the spatial resolution of QB pixels because no digital elevation model or accurate reference maps were available for the test sites. However, this relatively large error relative to GSP location was mitigated by selecting GSP within homogeneous stands over a 1-ha area.

Table 3 - QuickBird image acquisition parameters for each test site

	CH	RA	WA
Date, local time	07/10/2003, 10:43	08/12/2003, 10:57	08/10/2002, 10:35
Location (lat/long) (°)	N 49.65 / W 74.27	N 53.64 / W 74.02	N 52.90 / W67.14
Sun zenithal / azimuth (°)	31.4 / 143	41.0 /152.9	39.0 / 157.7
Sensor zenithal / azimuth (°)	11.5 / 97.4	13.5 / 172.6	9.8 / 225.6
Coverage (km²)	147	93	107

3. Method

The SF method to map FA followed procedures developed by Leboeuf *et al.* (2007) for mapping aboveground forest biomass. The SF method as applied here involved three sets of procedures: (i) estimating SF from the QB PAN image, (ii) developing site-specific and global regression models with SF as the independent variable and each of the four selected FA as dependent variables, and (iii) mapping FA as a grid layer for each test site using the global regression model over forested area.

3.1. Estimating SF from QB images

As a first step, all pixels of the QB image were assigned a binary value of 0 or 1 for non-shaded and shaded pixels, respectively. This shadow bitmap was produced by applying a threshold value to the digital numbers of the QB PAN. Threshold values were defined visually by an analyst for each test site and followed two main considerations: (i) avoid overestimation of shaded areas and

(ii) ensure realistic proportions of shaded areas on varying understories, such as those dominated by lichen and moss. The impact of the threshold value on the estimated SF values was evaluated in Leboeuf *et al.* (2007). Surprisingly, they found that within a relatively large range of threshold values, the results were consistent which reduces the sensitivity of this parameter. Consequently, the exact threshold value was not a critical factor for developing predictive models based on SF. The SF was then estimated for each test site by overlaying a 30 m × 30 m reference square centered over each GSP on the shadow bitmap.

The size selected for the reference square was based on the work of Leboeuf *et al.* (2007), which demonstrated that the 900 m² reference square size generated results that were similar to those obtained with smaller or larger reference squares (R^2 varying less than 7%) and in addition, matched the spatial resolution of Landsat imagery, and reduced processing requirements. The SF was estimated as $\hat{SF} = A_{TS}/A_r$, where A_{TS} is the area of shaded pixels and A_r is the area of reference square: 900 m².

Shadow fraction values were normalized to a common geometry to account for differences in sensor viewing zenith (Z_v) and azimuth (A_v) angles, as well as sun zenith (Z_s) and azimuth (A_s) angles. The normalized SF values (SF_n) were estimated based on trigonometric formulas (Eq. 1) that modified the shadow area projected on the ground or on the crowns as viewed by the sensor. Normalization was performed as per Leboeuf *et al.* (2007) and took into account two elements: a correction factor that considered the shadow length projected in the plane of the sensor viewing geometry and a correction factor that considered the increasing shadow length with increasing sun zenithal angle.

$$\hat{SF}_n = \hat{SF} \times C_{ZV} \times C_{ZS} \qquad (1)$$

with

$$C_{ZV} = \frac{1}{\sin\theta_v}, \quad \text{where } \theta_v = \cos^{-1}(\sin Z_v \times \cos(A_v - A_s))$$

and

$$C_{ZS} = \frac{\tan Z_{s\_ref}}{\tan Z_s}$$

Topography was not corrected for in the normalization because precise digital elevation models were not available for our test sites.

3.2. Generating regression models between $\hat{SF}_n$ and FA_{GSP}

The estimated SF_n and FA were applied as independent and dependent variables, respectively, to build site-specific linear regression models using 70% of the GSP data set (GSP_{cal}) that were randomly selected:

$$FA_{GSP} = a + b \times \hat{SF}_n + \varepsilon, \quad (2)$$

where FA_{GSP} is one of the four FA available from GSP_{cal}, a and b are the intercept and slope of the regression model, and ε is the residual. The goodness of fit of the regression models was defined through the adjusted regression model coefficient (R^2_{adj}). Polynomial models were also tested to determine if they would improve the results.

Difference in variance of the data set (heteroscedasticity) was evaluated both by the Shapiro-Wilk test (Shapiro and Wilk, 1965) and by interpreting a graph of residuals. In the presence of heteroscedasticity, the regression models were weighted depending on the changing error variances of $\hat{SF}_n$ (Carroll and Ruppert, 1988). A Fisher test assessed differences among the site-specific models by evaluating the coincidence and parallelism hypothesis using a threshold of $\forall$=0.05 (Milton and Arnold, 2003). Coincidence assumes that both slopes and intercepts are the same for the regression models of the three sites. Parallelism assumes that the three regression models are parallel. For FA where the Fisher test was significant (i.e. where F>t), the data sets from the different test sites were merged to estimate global regression models between $\hat{SF}_n$ and FA using GSP_{cal}. R^2_{adj} was also estimated for all global regression models. Statistical tests were applied to evaluate if the intercept of the resulting model could be assumed to cross the origin.

Supplemental statistics were estimated to evaluate the accuracy of the global regression model. Thirty percent of the GSP remaining (GSP_{val}) were used to estimate two accuracy statistics: relative root mean square error ($RMSE_r$) and relative mean deviation, which provides an estimate of bias, hereafter annotated as $bias_r$.

$$RMSE_r = \sqrt{\frac{\sum_{i=1}^{i=N}\left[\left(FA_i - FA_{GSP\_i}\right)\right]^2}{\sum_{i=1}^{i=N}\left[FA_{GSP\_i}\right]^2}} \quad (3)$$

$$bias_r = \frac{\sum_{i=1}^{i=N}\left[FA_i - FA_{GSP\_i}\right]}{\sum_{i=1}^{i=N} FA_{GSP\_i}} \quad (4)$$

where the FA_{GSP_i} is the FA measured from the GSP_{val}, FA_i is the FA estimated from the regression model, and N is the number of observations. Weighted prediction intervals were also computed when the data set exhibited heteroscedasticity (Zorn *et al.*, 1997).

3.3. Producing and validating FA raster maps

Forest attribute raster maps with 30 m × 30 m square cells were generated by applying the global regression models for each FA to the SF rasters derived from the QB images. A first validation involved computing $RMSE_r$ and $bias_r$ using FA values of the GSP_{val} data set and the estimated values from the FA raster map cell that superimposed each GSP_{val}. A stand map was available from the provincial inventory for the CH site, which provided an alternative for map evaluation. Therefore, a second evaluation compared volumes available from the stand map and corresponding volumes estimates from the FA raster map generated using the global regression models. Stand volume for all polygons of the stand map was estimated by using growth and yield table (Perron, 1986) derived from regional data. In this table, stand volume is estimated according to stand composition and height, crown closure. The stand map was intersected with the FA raster map and areas affected by cloud or shadow were eliminated from the analysis. The total volume of all black spruce stands within the extent of the FA raster map was estimated using two independent processes: (i) summation of the volume of all black spruce stands from the stand

map and (ii) summation of the volume of all black spruce 30 m × 30 m pixels from the FA raster map. Cells that partially intersected polygons were weighted by the fraction of their area within the polygons. $RMSE_r$ and $bias_r$ were then estimated for all polygons dominated by black spruce (basal area of black spruce ≥75 %). Only volume was assessed among the four FA selected because basal area is not photointerpreted, and crown closure and stand height are particularly prone to error when using photointerpretation, and thus would not provide the desired validation of the SF method. Finally, volumes from the stand map were also compared with those from the raster map by volume ranges to determine whether there were biases for specific ranges.

3.4. Producing and validating FA polygon maps

Forest attribute polygon maps representing areas with homogeneous FA values were generated as a final step in the method to show that FA raster maps can be use to produce stand polygon maps. To do so, a segmentation algorithm, available in the eCognition® software (Definiens, 2002), was applied to the FA raster maps. In this case, FA values were used for segmentation instead of the more common use of spectral values in eCognition. The segmentation algorithm established object or polygon boundaries based on the homogeneity of FA values according to scale and shape factors specified by the operator. Forest attribute polygon maps were produced by including the four FA raster maps as input to the segmentation algorithm with an equal weight. Several iterations were performed to assess the segmentation factor that best delimited the stands (Baatz *et al.*, 2004; Benz *et al.*, 2004). The values for the segmentation parameters (scale and shape factors) were chosen by iterations such that mean polygon area and general polygon shape (patterns, areas and standard deviation) matched the forest inventory map covering the CH test site. $RMSE_r$ and $bias_r$ were estimated by comparing volume of the polygon map and the GSP_{val} data set.

4. Results

4.1. Regression models between $\hat{SF}_n$ and FA

Overall results from the individual sites or the global relationships led to statistical models with R^2_{adj} above 0.5 and often with values around 0.7 and 0.8. Lower R^2_{adj} values were obtained for the

site-specific regression models of the WA test site when compared with CH and RA test sites (Table 4). Higher R^2_{adj} values were obtained for volume and basal area and decreased gradually for crown closure and stand height. Polynomial regression models did not improve significantly the regression models for all FA ($P > 0.12$). Consequently only linear models were selected.

The Fisher test was significant (i.e. where F>t) when applied to the three sets of site-specific data points for the four FA (Table 4). This suggested that each of the four FA could be expressed as global linear regression models applicable to all test sites. Of the four FA, only volume and basal area exhibited heteroscedasticity, which required estimation of a regression model through a weighted analysis. Intercept was not significantly different than 0 for volume ($t = -0.073$; $p= 0.942$) and for basal area ($t = 0.449$; $p= 0.654$) but was significantly different for crown closure ($t = 12.67$; $p<0.01$ and for height ($t = 19.9$; $p<0.01$). For consistency, we decided not to force any of the models to cross the origin. The resulting global linear regression models gave relatively high R^2_{adj} values for volume (0.723), basal area (0.782), crown closure (0.634), and height (0.574). The $RMSE_r$ values for the four FA were 0.32 for volume, 0.29 for basal area, 0.32 for crown closure, and 0.17 for the height. $Bias_r$ values were also low for all four FA: 0.15 for volume, 0.14 for basal area, -0.04 for crown closure, and 0.05 for height. The weighted prediction interval for volume and basal area (Fig. 3) contained almost all GSP_{val} values, indicating that the global models were good predictors.

4.2. FA raster maps

The $RMSE_r$ and $bias_r$ obtained using the GSP_{val} data set and their corresponding values on the grid cells of each FA raster map were respectively 0.22 and 0.04 for volume, 0.34 and 0.15 for basal area, 0.35 and -0.02 for crown closure, and 0.14 and 0.09 for height. As an illustration of the results, Fig. 4a shows the volume mapped for black spruce stands derived from the application of the global regression model for the CH test site. Stand polygons taken from the polygonal stand maps were superimposed on the raster map to establish the area relevant for FA estimation (i.e., stands dominated by black spruce). The white polygons correspond to water bodies, clouds, and cloud shadows, or forest stands where black spruce was not dominant. The resulting patterns from the volume map corresponded well to the expected patterns, i.e., with volume values close to zero over regenerating and clearcut areas. In addition, total merchantable

volume estimated from $\hat{SF}_n$ (517 654 m³) of the CH test site was very similar to the one estimated from the traditional forest inventory (488 503 m³), with a difference of 5.6 %. The $RMSE_r$ and $bias_r$ estimated for volume when using the polygonal stand maps for reference were 0.42 and -0.08, respectively. Lastly, both maps provided a similar histogram of stand volume estimated by volume range (Fig. 5). This confirmed that no obvious bias resulted in a specific volume class from the volume map produced with the SF compared with volume estimated from a traditional polygon map.

4.3. FA Polygon maps

The FA polygon map showed similar patterns to the inventory stand map (Fig. 4) with mean polygon areas of 9 ha and a standard deviation of 15 ha. These values were similar to those of the corresponding area of the forest inventory map, which had 10 ha of mean polygon area and 15 ha of standard deviation. The scale and shape factor used to produce those polygons within the segmentation process were 15 and 8, respectively. The distribution of volumes in the FA polygon map is also very similar to the inventory stand map distributions as shown by Fig. 5. The $RMSE_r$ and $bias_r$ estimated for the stand volume of the CH test site using the GSP_{val} data set and their corresponding values on the stand polygon were 0.41 and 0.12, respectively. The $RMSE_r$ and $bias_r$ estimated for the resulting basal area was 0.38 and 0.11, respectively.

Table 4 - Coefficients (intercept a and slope b) for the three local (CH, RA, and WA) and global linear regression models with their R^2_{adj} estimated using GSP_{cal} for the four FA. Accuracy statistics ($RMSE_r$ and bias) estimated using GSP_{val}.

Volume

Data set	GSP_{cal}	a	b	R^2_{adj}	GSP_{val}	$RMSE_r$	$Bias_r$	Fisher test
Global	76	0.24	195.14	0.723	32	0.32	0.15	0.67 vs 2.49
CH	22	-5.16	186.22	0.783				
RA	35	-0.30	194.76	0.696				
WA	19	6.87	221.94	0.583				

Basal area

Data set	GSP_{cal}	a	b	R^2_{adj}	GSP_{val}	$RMSE_r$	$Bias_r$	Fisher test
Global	76	0.21	41.95	0.782	32	0.29	0.14	1.44 vs 2.49
CH	22	-0.89	40.32	0.843				
RA	35	0.27	43.93	0.763				
WA	19	2.49	38.99	0.527				

Crown closure

Data set	GSP_{cal}	a	b	R^2_{adj}	GSP_{val}	$RMSE_r$	$Bias_r$	Fisher test
Global	76	13.46	80.84	0.634	32	0.32	-0.04	1.49 vs 2.49
	22	14.86	73.77	0.553				
	35	9.05	111.94	0.766				
WA	19	18.71	50.85	0.069				

Height

Data set	GSP_{cal}	a	b	R^2_{adj}	GSP_{val}	$RMSE_r$	$Bias_r$	Fisher test
Global	76	6.23	14.17	0.574	32	0.17	0.05	0.40 vs 2.49
CH	22	4.97	16.27	0.707				
RA	35	6.08	13.46	0.391				
WA	19	6.66	17.13	0.273				

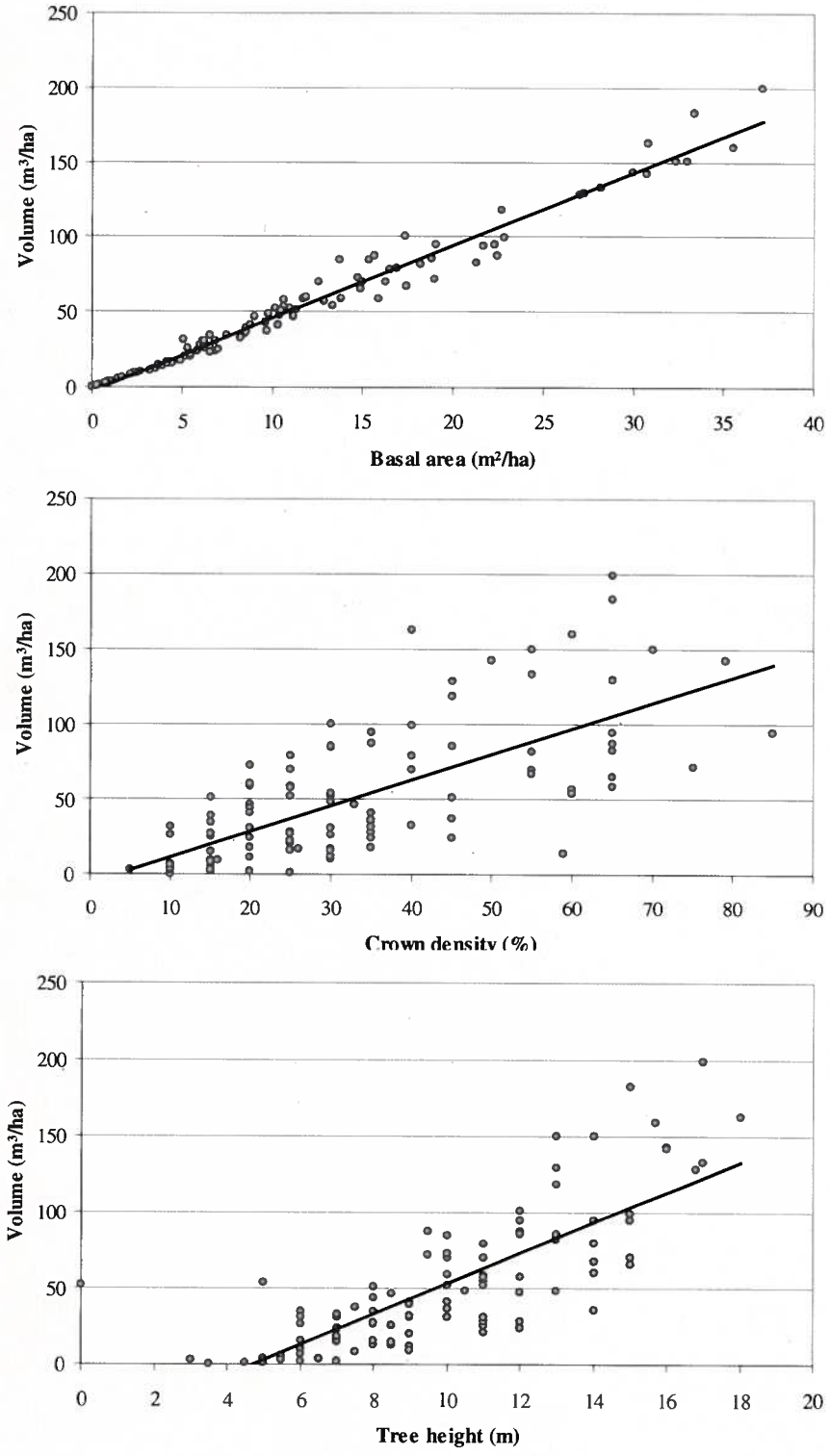
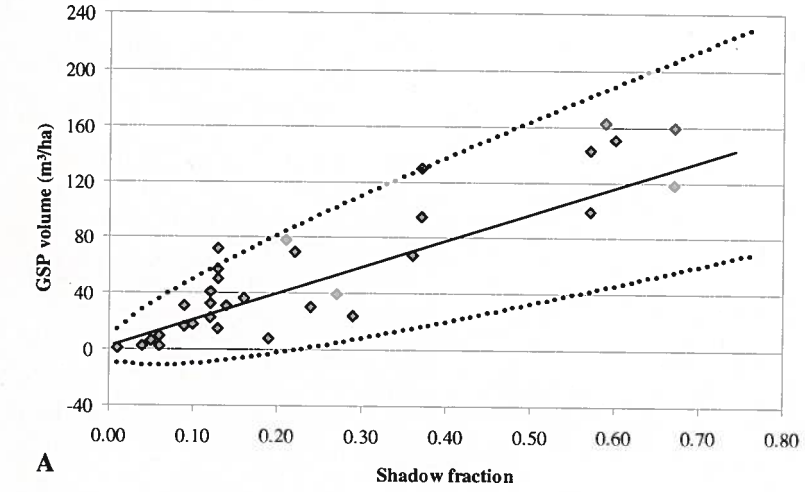
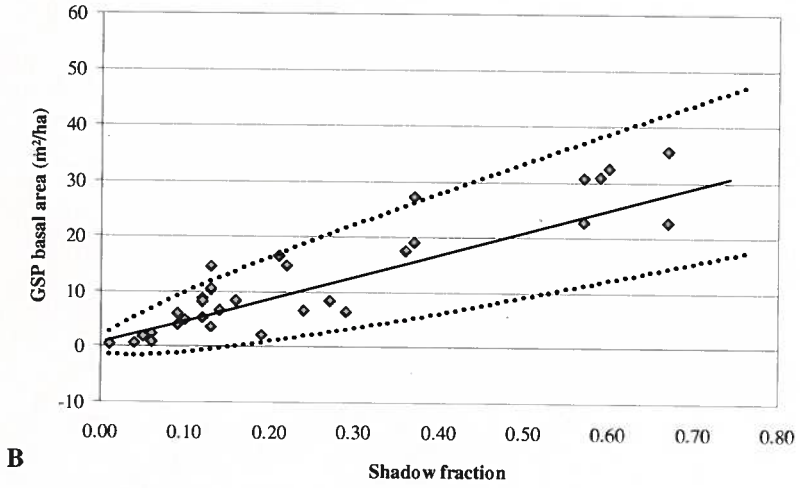


Figure 2 - Relationship between volume and each of the other forest attributes estimated at each GSP.

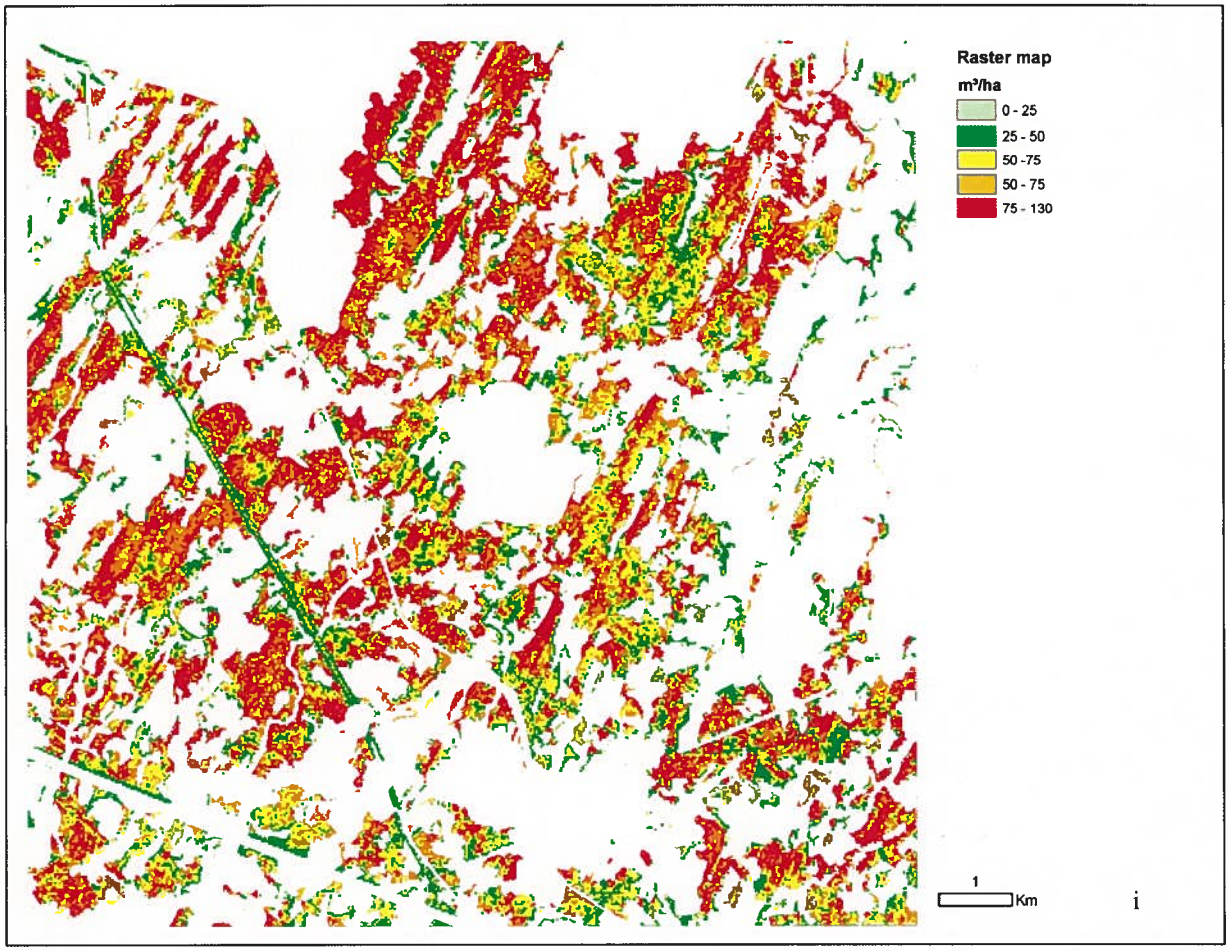


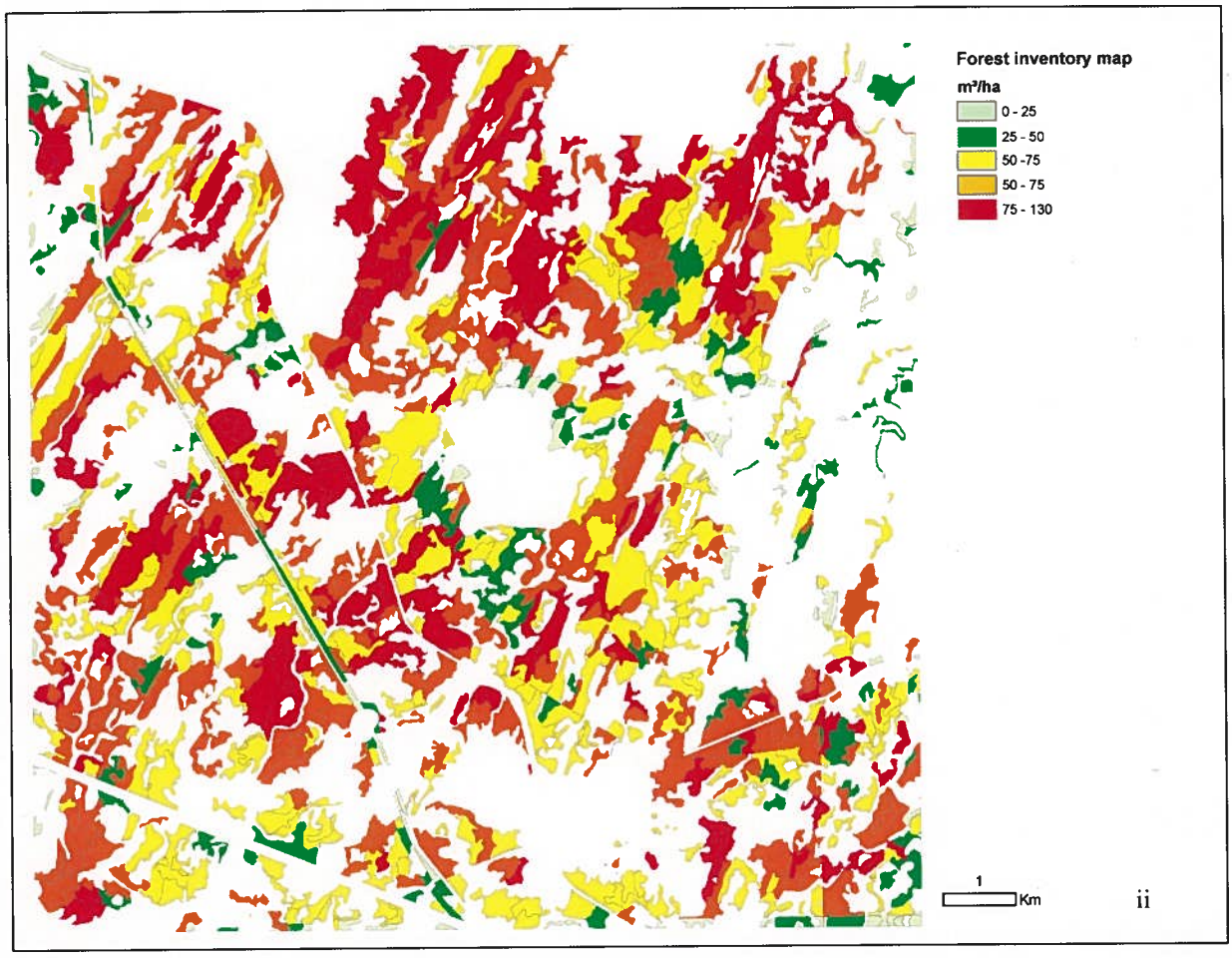
A



B

Figure 3 - All the GSP_{val} data points and the global linear regression model between SF_n and stand (A) volume or (B) basal area. Calibration line (—) and weighted prediction intervals (--).


Raster map
m³/ha
0 - 25
25 - 50
50 -75
50 - 75
75 - 130
1
Km


Forest inventory map
m³/ha
0 - 25
25 - 50
50 -75
50 - 75
75 - 130
1
Km
ii

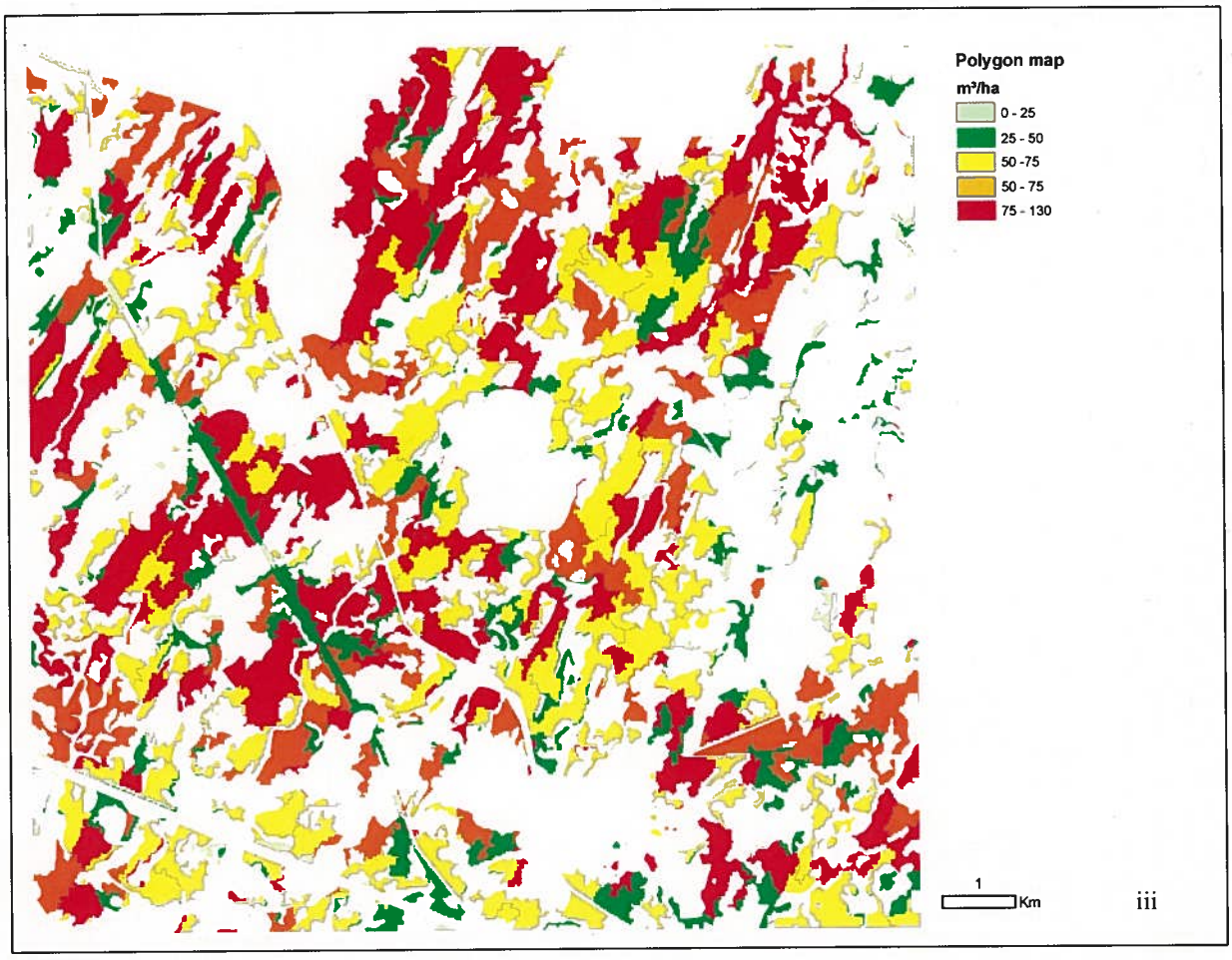


Figure 4 - Volume from (i) raster map, (ii) forest stand map and (iii) polygon map for black spruce stands on CH test site.

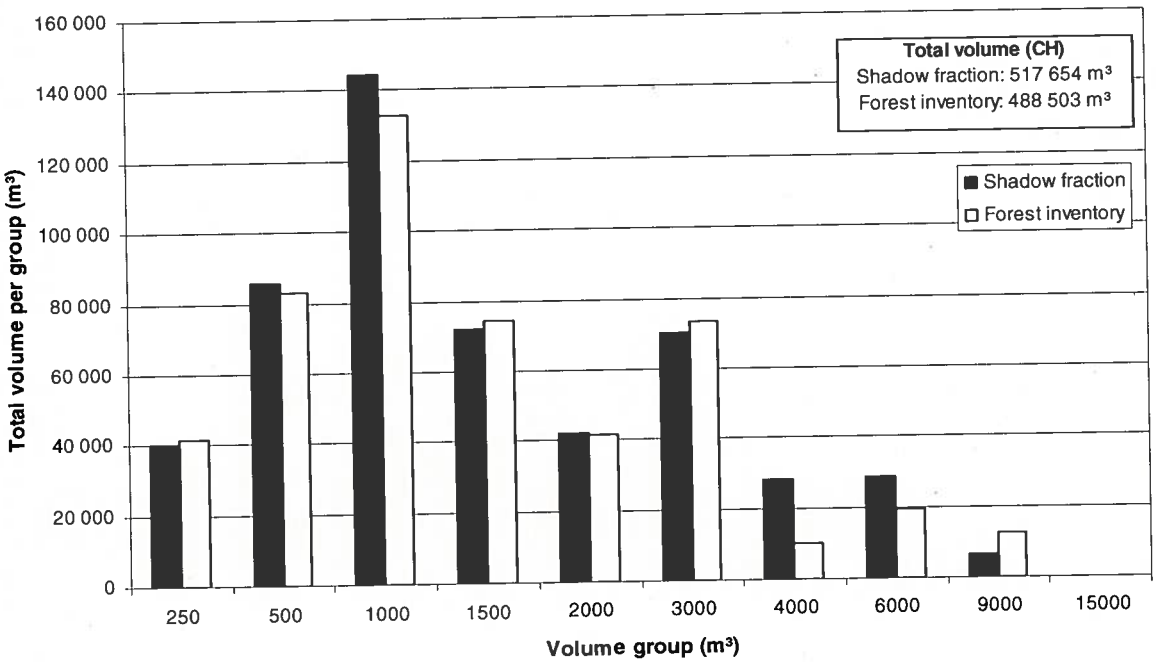


Figure 5 - Total volume calculated by volume ranges of 250 m^3/ha for a common area taken from a stand map and the shadow fraction excerpted from a satellite image. Values from shadow fraction were calculated from global volume equation in Table 4 and forest inventory values from the use of stand map attributes and a conversion table.

5. Discussion

Forest attribute mapping using the SF method developed in this study is a step toward more extensive forest inventory of large areas of northeastern forests in Canada dominated by black spruce stands. The three levels of evaluation demonstrated high potential for the SF method. First, error statistics (R^2_{adj} , $RMSE_r$, and $bias_r$) from the regression models demonstrated the relevance of a global model and its applicability to most black spruce stands and also to the four stand attributes, in addition to biomass (demonstrated in Leboeuf *et al.*, 2007). Weaker results from the WA site-specific regression models were explained partly by the more rugged relief, resulting in (i) weaker relationships between the FA and SF and (ii) higher co-registration errors

between GSP and the QB image. The fit between the predicted and observed values confirmed this trend. As the second evaluation showed, similar results were obtained comparing inventory stand maps made from photointerpretation with the FA raster maps produced by the SF method for the CH test site. Considering that both methods have a different set of uncertainties, the small difference of 1.2% between volume from the FA raster map and volume from the inventory stand map provided a convincing argument that results of the SF method were consistent at the landscape level. As the third evaluation showed, a small difference existed between mean merchantable volume generated by the SF method when compared with inventory stand map values, which also supported landscape-level consistency despite differences due to interpretative variability. This difference was explained by the difficulty in estimating crown closure and height classes using both the SF method and aerial photograph interpretation. In addition, SF captures crown size, crown density, and tree height information all together, which explains why it correlates well with volume or basal area, which integrate these attributes. Conversely, crown density and tree height were not as easily estimated from SF since crown density was estimated visually and tree height was estimated as Lorey's height (an approximation of stand height based on the height of all trees weighted by their DBH). Nevertheless, the applicability of the SF method for the spatial extents of a QB image gave results that were compatible with standard forest inventory for areas with few or no ground plots.

Although, we reached our main objectives of developing global regression models and producing raster and polygonal maps for the three test sites, our analysis raised the possibility of an alternative option for model development. For instance, a non-parametric technique could be used to further assess multivariate relationships between SF and FA (McRoberts *et al.* 2007). Fig. 2 shows how volume and basal area are strongly correlated. Stand biomass must also be strongly correlated with these two FA. These correlations, suggest that a multivariate analysis could help to develop FA models. While these new explorations were beyond the scope of our objectives, our results might be seen as stepping stone towards other options for model development.

The linear models proposed are suitable for the spatial extents of the respective QB images available, but they also open the possibility for generalization of the method over large areas. For

instance, many of the satellite images used in Google Earth are from QB. It would therefore be possible to transform QB images into FA maps where forest lands are compatible to the areas where the models were developed (Black spruce dominance). In addition the models proposed represent a step forward in spatially extending forest inventory methods. The general procedure to establish a model using the SF method is an interesting complement to earlier work by Peddle and Johnson (2000), Peddle *et al.* (2001), and Seed and King (2003) based on spectral mixture analysis. Several aspects of our study support the generalized application of the SF method to black spruce stands of northern boreal forests: results are consistent among the three test sites covering two ecozones, and the normalization process allows the use of QB images from different acquisition dates. Similarly, our study showed that strong correlation statistics resulted from global models for several FA, namely volume and basal area, and for biomass from Leboeuf *et al.* (2007). Results were also, to a lesser extent, satisfactory for crown density and stand height. Mapping forest stands usually requires ground plots for calibration or an interpretative process. The SF method removes this limitation through the availability of a global regression model. If GSP are available, they can be used to validate the global model. Otherwise, the model can be applied without GSP and provides a convenient alternative given the intensive resources required for GSP. Interpretation of aerial photographs suffers from subjective bias. The SF method appears compatible with the results of an experienced interpreter without the subjectivity. In addition, segmentation of the raster map produced by the SF method is an alternative map that may be closer to conventional stand maps used by foresters in operational situations. It produces stand polygons based on local homogeneity of SF values. Overall, the SF method is well adapted to mapping northern regions and compares favorably with existing forest inventories where GSP are not always available. It provides high local accuracy, either as raster- or vector-based maps, and allows easy update of FA polygon maps.

Despite the advantages of the SF method in mapping northern boreal forests, our study shows several limitations before it can be considered for an extensive inventory program. First, the regression models we developed used a range of FA values dictated by the GSP in our test sites. For example, stand volume of our GSP range from 0 to 202 m^3/ha. A complete assessment of the SF method before operational application should include GSP with FA values covering the full potential range for black spruce stands. An analysis of all permanent sample plots available for

black spruce stands from the government of Quebec suggests the following maximum values: 428 m^3/ha for volume, 56.4 m^2/ha for basal area, 100% for crown closure, and 23.5 m for height. Mean merchantable volume for medium productivity black spruce stands in the boreal forest of Canada is 160 m^3/ha (Bickerstaff and Hostikka, 1977), and black spruce stands dominate in that landscape. Without specific tests using these large values, we can expect that SF may reach a maximum from a given FA value assuming the SF will no longer be tied to an increase in the FA value due to saturation of SF linked to high crown closure with mutual shadowing. However, this hypothesis needs further testing. Meanwhile, our study provides regression models for the vast majority of cases in the boreal forest in northern Canada. Second, the regression models described in this study were developed only for black spruce stands. Compatibility of deciduous trees with the SF method remains to be demonstrated. Difficulties with deciduous stands when using a SF method were raised by Seed and King (2003) to explain their poor results between SF and LAI. In addition, it remains unknown if similar regression models will work on other coniferous stands and if they will be similar to our global regression models. Some results regarding biomass estimation across various test sites in Canada suggest this could be the case (Beaudoin *et al.*, 2007). Nevertheless, independent studies are required to test the SF method on other common coniferous stands in northern areas, such as balsam fir, jack pine, lodgepole pine (*Pinus contorta* (Dougl.)), etc. Third, the regression models developed in the present study can be applied only on QB PAN. Other HSRSI, such as IKONOS images, having similar PAN spatial resolution (1 m) to QB (0.6 m) must be tested to know if the SF method would generate accurate maps using such data. Because spatial resolutions are similar between these two images, regression models should generate similar results. The cost of IKONOS images is lower than QB, representing an operational advantage. Fourth, the SF values should be normalized according to sun–sensor–ground geometry as was done in the present study. We realize that the operator threshold applied to the image to create the shadow bitmap does apply a subjective normalization to a certain extent. However, the normalization procedure still requires development to be finalized, particularly to account for topographic effects. Fifth, intercept values (a) from linear regression model, as seen in Table 4, were not forced to pass through the origin (0, 0). This can introduce positive deviation for the lower FA values. However, at a landscape scale, such small FA values do not have an important impact and can be easily corrected by assigning null FA values to the smallest known FA value (regeneration, planting area, wet lands, etc.). Sixth, the SF

method applied only to HSRSI is too limiting to be able to map extended areas, therefore, another mapping method is required to generalize or expand the maps from HSRSI. Complete coverage of large areas is not realistic because of high cost or lack of image availability. Instead, a method like the K-Nearest Neighbors has the potential for such spatial generalization using HSRSI with another coarser satellite image such as those from Landsat TM, ETM or SPOT-5 (Beaudoin *et al.*, 2007; McRoberts *et al.*, 2007). Finally, the raster and/or polygon values obtained with the SF method are estimates, not observations as may be the case with ground inventory methods. Thus, large area estimates must incorporate not only the polygon-to-polygon variances but also the errors associated with estimates for individual polygons. Further work is needed to produce appropriate measures of uncertainty for large areas taking into account both sampling and model errors (McRoberts *et al.*, 2011). Despite the above limitations, the SF method tested in the present study can be applied to most black spruce stands in northeastern Canada.

6. Conclusion

The present study demonstrated through several validation processes that the SF method is expandable to several FA—namely, stand merchantable volume, basal area, crown closure, and height—for large black spruce-dominated areas of northeastern forests in Canada and that these FA maps could be used to generate FA polygon maps that are comparable to conventional forest inventory stand maps. Black spruce-specific regression models were first estimated between SF and FA over three test sites. These data sets were then used to develop accurate global regression models showing strong results for four key FA with R^2_{adj} from 0.574 to 0.782, $RMSE_r$ from 0.17 to 0.32, and $bias_r$ from -0.04 to 0.15. The study thus demonstrated that global regression models could be expanded to other FA beyond tree biomass, thus expanding the work of Leboeuf *et al.* (2007). Using this global regression model, FA raster maps were computed over the test site extents and also showed strong accuracy statistics with a small difference (5.6%) when compared with volume from polygonal stand maps. Our results showed that segmentation applied to the FA raster maps can produce FA polygon maps comparable to those currently used for forest management. The SF method provides a practical means to map four widely used FA, in raster or polygonal format, and can be applied over large black spruce-dominated areas of Canada's northeastern boreal forest using QB PAN imagery.

7. Acknowledgments

We acknowledge financial support by the Canadian Space Agency for field work and analysis. We also acknowledge funding from the National Science and Engineering Research Council of Canada, Natural Resources Canada (NRCan), and the Université de Sherbrooke toward salary for A. Leboeuf. We are grateful to M.-C. Lambert from the Ministère des Ressources Naturelles et de la Faune in Quebec for her advice and support regarding statistical elements and to P. Villemaire of NRCan for his support during field work and analysis. We also thank R. Thibault and C. Gagnon from Abitibi-Bowater Inc., for providing stand maps of the Chibougamau test site,. We thank the field crew from the Canadian Forest Service of NRCan for GSP measurements: J.-P. Bérubé, J. Boudreau, S. Côté, S. Dagnault, J. Donnelly, N. Laflamme, R. Morin, G. Simard, G. Strickland, and P. Villemaire. Finally, we appreciate comments made by two anonymous reviewers that resulted in significant improvements to the manuscript.

6. Article 2

Estimating stand attributes of boreal forest using digital aerial photography and a shadow fraction method

A. Leboeuf and R.A. Fournier

Centre d'Applications et de Recherches en Télédétection (CARTEL),

Department of Applied Geomatics, Université de Sherbrooke,

2500 boul. de l'Université, Sherbrooke, Québec, J1K 2R1, Canada

E-mail of corresponding author: antoine.leboeuf@usherbrooke.ca

Telephone: 819-821-8000 ext. 63209, Fax: 819-821-7944

Manuscript submitted to Canadian Journal of Remote Sensing in April 2013

ABSTRACT

The aim of this study was to develop an imputing method to map forest stand attributes (height and volume) using panchromatic aerial photographs and Landsat images. The method was tested on several sites in Québec, Canada, which were dominated by black spruce. The method involves four sets of procedures: (i) measuring the shadow fraction from panchromatic aerial photographs, (ii) generating regression models between the shadow fraction and forest attributes, (iii) mapping

locally the forest attributes as a grid layer (30 m × 30 m) and (iv) expand the forest attributes from the local maps to a large study area using an imputing approach. Regression models between the shadow fraction and stand attributes were calibrated with ground sample plots and a series of 73 aerial photographs. General results from linear relationships between the shadow fraction and two stand attributes showed linear shape but had low values for the goodness of fit (R^2_{adj} of 0.24 for height and 0.31 for volume). However, the results improved significantly (R^2_{adj} of 0.40 for height and 0.59 for volume) when the portion used of the aerial photograph was restricted to its centre (between 6.2 and 11.3 degrees). Such restricted area confines the use of aerial photographs to produce small local maps of stand attributes from the shadow fraction. These local maps were used as training dataset of an imputing approach, a KNN algorithm, aimed at mapping forest attributes over large area.

Keywords: Volume, stand parameter, forest inventory, shadow mapping, remote sensing, aerial photos, boreal.

1. Introduction

Forest inventory plays a key role in forest management and monitoring. In Canada, forest inventories are typically based on two datasets: (i) stand maps in polygon format resulting from the visual interpretation of aerial photographs, and (ii) ground sample plots (GSP) containing tree- and stand-level measurements such as diameter at breast height (DBH) of tree stems, tree height, and stand density from crown closure (Avery and Buckhart, 2001; Kangas and Maltamo, 2009). Forest inventories are expensive to implement in remote areas like the Canadian boreal forest because they are far from populated areas and the road infrastructure is sparse. There is increasing pressure to develop forest inventory procedures that improve the accuracy of forest attribute maps to support better forest management practices. For instance, several Canadian

provinces are considering an enhanced forest inventory where the minimum mapping unit will be around 2 ha and stand attributes such as tree height will be estimated as a continuous variable (DNR, 2011; Ministry of Forest of British Colombia, 2005; MRNF, 2011). These considerations may be realistic for forests in the southern part of Canada but they would be difficult to implement in northern regions. However, remote sensing and image processing methods have the potential to improve current estimates of forest attributes for northern regions within the limits of operational criteria and at a reasonable cost.

Methods to process digital images from satellite remote sensing provide new possibilities for improving forest inventory. For instance, LiDAR remote sensing has proven to be successful to map forest stands with great accuracy (Maltamo et al., 2004; Naesset, 2004; Parker, 2004). Unfortunately, this technology is still too expensive to map very large areas. The use of very high spatial resolution satellite images is another option in combination with object-oriented image processing (Leckie et al., 2003; Wulder et al., 2008). Here again, image acquisition and image processing are costly, both in terms of money and time. In a method proposed by Pellikka et al. (2000b), Greenberg et al. (2005), Leboeuf et al. (2007, 2011) and Wolter et al. (2012), the amount of shadow on a fixed grid is related to stand attributes in coniferous forests. Greenberg et al. (2005) reported reasonable accuracy (R^2 from 0.67 to 0.77) by linking shadowed vegetation areas to two tree structural parameters, the DBH and the crown area for mainly coniferous forest stands. Similarly, Leboeuf et al. (2007, 2011) used their "Shadow-fraction" (SF) method in models with relatively high predictive capability (R^2) for stand biomass (84%), volume (72%), basal area (78%), height (57%), and crown cover (63%), when using Quickbird satellite images and applied to areas dominated by coniferous stands. The SF-method has also been shown to work with balsam fir stands (*Abies balsamifera* (L.) Mill.) (Luther et al., 2012). The SF-method

appears to be well suited for mapping large areas of the eastern Canadian boreal forest, which is dominated by black spruce (*Picea mariana* (Mill.) B.S.P.) and balsam fir stands with various stand attributes up to 20 m height and 200 m^3/ha.

Applying the SF-method operationally to large areas of the Canadian boreal forest requires close attention to its limitations. This area is largely dominated by black spruce stands. In Quebec, the area north of the 52nd parallel covers more than 450,000 km^2, which is impossible to cover at a reasonable cost with very high spatial resolution satellite images. However, aerial photographs are routinely taken for forest inventory and are therefore available for large areas and for different periods of time over the last decades. In addition, the spatial resolution of digital aerial photographs is generally finer than the one from very high spatial resolution satellite images like QuickBird or GeoEye. Such fine spatial resolution facilitates the use of automatic image processing algorithms such as the SF-method.

Unfortunately, the radiometric patterns in a full frame aerial photographs present spectral bidirectional distortions called BRDF (bidirectional reflectance distribution function), resulting from reflectance characteristics of a given land cover type. BRDF leads to dissimilar reflectance values of a land cover type in various part of the image. This difference depends on the illumination and viewing (sun-target-sensor) geometry. Forest stands are indeed anisotropic scattering volumes with darker reflectance values when the viewing-target vector is opposed to the sun-target vector (forward viewing) and has a typical "hot-spot" (high peak of reflectance values) when the viewing-target vector is in a similar direction with the sun-target vector (backward viewing) (Pellikka et al., 2000a; Kane et al., 2008). The magnitude of this distortion depends on the land cover optical characteristics, atmospheric conditions, solar angle variables,

and the sensor view angle determined by lens focal length and sensor dimensions. Thus, aerial photographs taken with frame cameras with their large swat intensify the bidirectional effects (Pellikka et al., 2000a). To attenuated bidirectional reflectance distortion in the photographs caused by the, atmospheric conditions and vegetation effects, several models have been developed (Li and Strahler, 1992; Lillesand and Kiefer, 1994; Pellikka et al., 2000a; Soenen and Peddle, 2005). However, those models still need adjustments to account for practical issues.

Using the shadow fraction as an explanatory variable to estimate stand attributes do not use all the radiometric patterns in the image, which are strongly affected by the BRDF, but concentrates on the shadowing patterns. Shadow pattern differs for different location within the image due to the sun-target-sensor geometry and depends on forest structure and species (Pellikka et al., 2000a). A simple way to address the changes in tree shadow patterns due to the sun-target-sensor geometry is to normalize the tree shadow values according to a reference configuration (e.g. for a given view point such as nadir and a given solar direction such as the one at noon). This normalization procedure, or tree shadow normalization was used in Leboeuf et al. (2007). It uses a trigonometric model to correct the shadow fraction cast on the ground. This model was limited to account for sun and sensor angles and did not consider the effects of local topography and fine forest structure. However Pellikka et al. (2000b) suggested that topography has only a weak impact on the observed shadows of most stands except for very open stands. Tree shadow normalization should therefore be applied to aerial photographs if local patterns of shadow fraction are used to estimate stand attributes.

Mapping forest attributes at a regional scale may be done (i) using sample images covering partially the area at high spatial resolution and then (ii) apply the results to establish the statistics

between common area of the sample images and medium spatial resolution satellite images (MSRSI) such as Landsat-TM/ETM+, SPOT-5 or RapidEye images covering the whole area, and (iii) apply an imputation technique to produce a regional map (Crookson *et al.* 2002; McRobert and Tomppo, 2007; Tomppo *et al.* 2008). The MSRSI typically have a spatial resolution from 5 to 30 m, at least four spectral bands and are available at a reasonable cost or for free. Such approach to large area mapping has been widely suggested in the scientific literature (Cihlar, 2000; Wulder *et al.* 2012b). In addition, several projects have confirmed the applicability of remote sensing images using imputation techniques such as the k-nearest neighbour (KNN) technique for forest inventory (McRobert and Tomppo, 2007; Tomppo *et al.* 2008). This intuitive and versatile technique aim to estimate variables that are based on similarity of covariables between attributes to predict and observe samples. The case of the finish inventory is well adapted to the use of the KNN technique because of the numerous ground plots and a relatively homogeneous forest types (Tomppo *et al.* 2010). So far, several variants of the KNN were developed (Crookson *et al.* 2002; Ohmann and Gregory, 2002; Hudak *et al.* 2008; Bernier *et al.* 2010; Wilson *et al.* 2012) to increase its flexibility or powerfulness. The precision obtained from the imputation techniques is highly dependent on the number and quality of training samples (Franklin and Wulder, 2002; Tomppo *et al.* 2008). In the context of Canada's boreal forest, the number of available GSP is generally very low and they tend to be located in denser forest (density of 25% or more).

We propose, as the main goal of this study, to extend the SF-method to aerial photography for estimating and mapping two forest attributes (stand height and merchantable volume) over areas of the northern boreal forest in eastern Canada. Specifically, we wish (i) to assess how well the

SF-method works on aerial photographs and for what portion of the image, (ii) to assess if tree shadow normalization improves the results from the SF-method using the aerial photograph and (iii) to use the local maps of stand attributes from aerial photograph with an imputation technique to map large area with Landsat-TM images.

2. Study area and data

2.1. Study area

The study area is located in the Boreal Shield Ecozone, in northeastern Canada (Fig 1). It covers 120,000 km² between the 48th and 52rd parallels. This area is generally pretty flat but altitude varies from 0 to 988 m. A total of 73 aerial photographs were acquired in scattered locations throughout the study area. The number and location of the aerial photographs were selected to provide a representative sample for the range of stand height and merchantable volume and stand conditions that are typical of northeastern Canada. The forest stands of the study area were constituted of respectively of 3%, 10% and 87% of deciduous, mixed and coniferous stands. A majority (52%) of these stands were sparse (density < 40%), 31% were mid-dense (density between 40% and 60%) and 17% were dense (density > 60%).Generally, the forest stands in the study area covered by the aerial photographs were mainly composed of black spruce with a few other species such as jack pine (*Pinus banksiana* Lamb.) and balsam fir, which are typical of the region (Rowe, 1972). The understory of the black spruce stands was generally covered by lichens, mosses, and shrubs in various proportions.

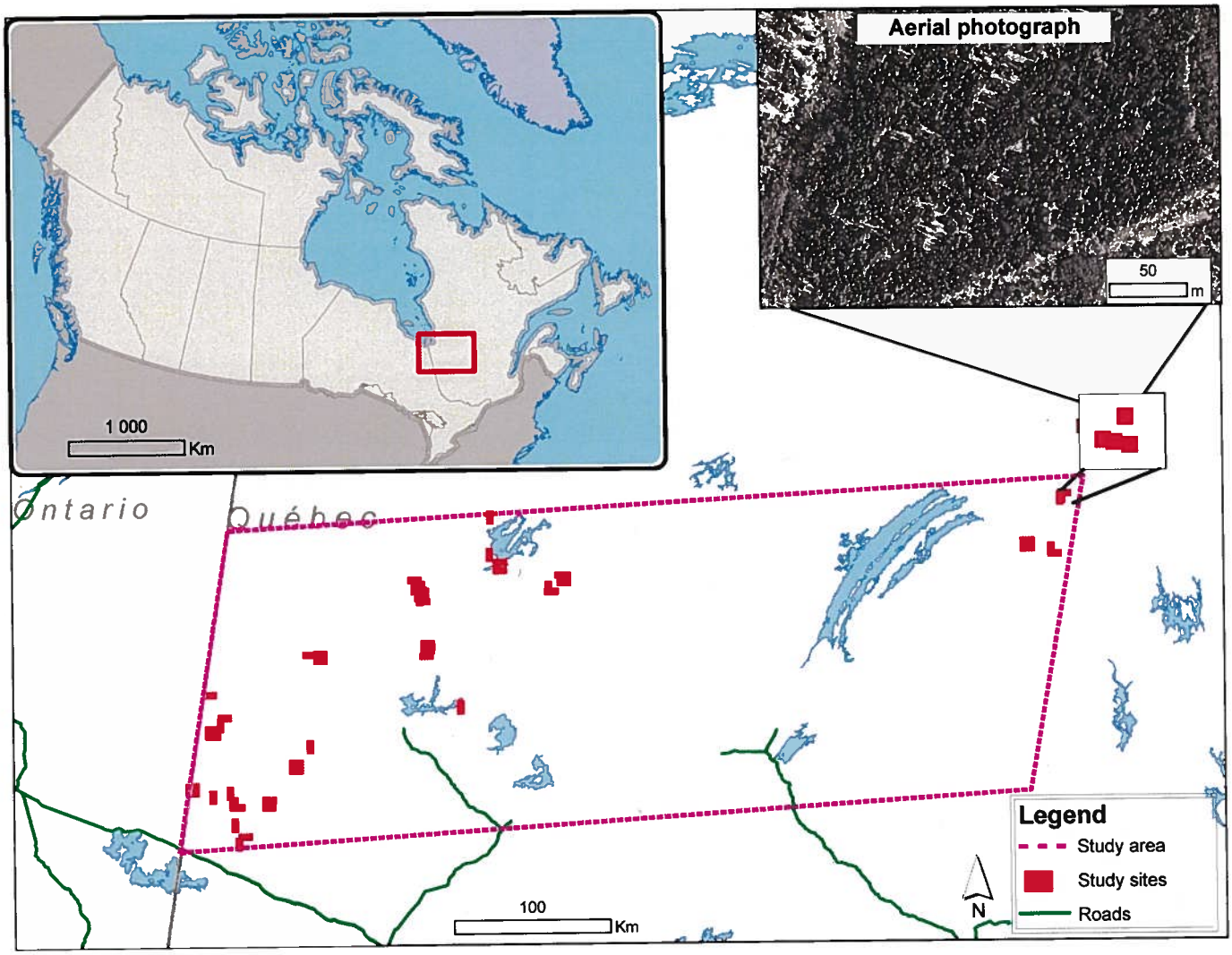


Figure 1. Location of the aerial photographs in the study area located in the western part of the province of Québec, Canada.

2.2. Ground sample plots and forest maps

The 142 circular ground sample plots (GSP) used in the study followed the measurements protocol of the Quebec Minisctry of Natural Ressources (MRNF, 2009). They were all located within the area covered by one of the 73 aerial photographs taken for a provincial inventory. Each

GSP covered an area of 400 m^2 (11.28 m radius), and the measurements for these plots were taken in 1996, 1997, 2000 or 2002. The GSP were selected to represent a wide range of stand attributes and understory types. The species composition of the GSP was largely dominated by black spruce such that: (i) black spruce represented more than 50% of the GSP basal area and (ii) any other species such as trembling aspen (*Populus tremuloides* Michx.), white birch (*Betula papyrifera* Marsh.), *Alnus* spp., and *Salix* spp. represented less than 25% of the GSP basal area. Table 1 summarizes the attributes range measured in the GSP.

Several tree-level attributes were measured at each of the GSP (Table 1). The DBH was measured for all trees in the GSP with a DBH > 9 cm. In addition, the stand height and volume were calculated using published regression models where tree DBH was used as an independent variable (Table 1). These models were developed using GSP data but for an ecological zone with a known range of potential vegetation, soil drainage and texture, disturbance regime and climatic data. Tree height was calculated from tree DBH using an existing model by Fortin et al. (2009) which is based on a database of 37,258 tree measurements (DBH and height). Stand height was estimated as the mean of the four trees with the highest DBH within the GSP. Tree stem volume (m^3/ha) was calculated using a methodology developed by Fortin et al. (2007) but adapted to our study area. A total of 3,912 sampled trees from 1,026 GSP were used to develop the volume model. Plot-level volume was calculated by summing the stem volume of all merchantable trees (DBH > 9 cm) within the GSP.

Table 1 - Forest attributes measured from ground sample plots (GSP).

Forest attributes	Calculation method	Attributes range
Altitude	**Measured on the field**	0 to 923 m (average: 429)
Slope	**Measured on the field**	0 to 59% (average: 8%)
DBH	**Measured on the field**	14.6 to 21.1 cm (average: 14.6)
Stand height	Tree height mean of the 4 highest DBH tree	7.8 to 22 m (average: 14.9 m)
Volume	**Plot-level volume** : $\sum$Tree volume$_i$ × 10,000/400 **Tree volume$_i$** : Fortin et al. (2007) adapted for the study area **Tree height$_i$** : Fortin et al. (2009) adapted for the study area	1.8 to 221 m³/ha (average: 88.4 m³/ha)

2.3. Aerial photographs

Aerial photographs were acquired in 2009 and 2010 with acquisition dates from July 10 to September 11 from 12:30 to 20:30 (UTC). The two cameras used were the Vexcel Ultracams D and X (Vexcel, 2013) with a focal distance of 100 mm for both cameras. The pixel size was respectively 9 µm × 9 µm and 7.2 µm × 7.2 µm resulting in 30 cm on the ground since the flight altitude was respectively 3,333 m and 4,200 m above ground. Flight lines followed an east-west transect. Each photograph covered an area on the ground of 3,450 m × 2,250 m and 4,329 m × 2,826 m for the Ultracam D and X. All photographs were orthorectified using a digital elevation model and high precision GPS data resulting in a X-Y positional accuracy within 5 m. The negative impact of this possible spatial displacement between the aerial photograph and the GSP was largely reduced by the fact that GSP were located in homogeneous and even-aged coniferous stands. Images were acquired in digital format using 8 bits with bandwiths of 390-530 nm, 470-

660 nm, 570-690 nm, and 670-940 for the blue, green, red and NIR bands of the Ultracam D respectively. The Ultracam X offered similar bandwidths but with noticeable differences: 400-600 nm, 480-660 nm, 580-720 nm, and 620-1000 respectively. Panchromatic images were produced from the linear average values of the digital data from the three visible bands: blue, green and red.

2.4. Landsat-TM images

The study area was completely covered by a mosaic of MSRSI composed of 16 Landsat-TM images. Although the images were acquired from June 23, 2008, to September 12, 2009, most of them were acquired during the growing season of 2009 from June to September. The mosaic was assembled using an histogram matching technique from Homer *et al.* (1997) implemented with the PCI Geomatics software (PCI Geomatics inc., 2003). This technique minimizes spectral differences between a reference satellite images and the "slave" images of the rest of the mosaic. This is done by looking at the distribution of brightness values (e.g., the shape of their histogram: maximum and spread) of all pairs of reference and slave images to bring them to a match. The zones with clouds or haze were masked out of the mosaic to avoid interference in the matching procedure. Three spectral bands were used: red (0.63 to 0.69 μm), which is sensitive to the presence of chlorophyll; near infrared (0.76 to 0.90 μm), which is sensitive to the structure of vegetation cover; and mid infrared (1.55 to 1.75 μm), which is also sensitive to the presence of chlorophyll and structure of vegetation cover (Gausman, 1984). The choice of these three bands was justified by their potential for forest class differentiation (Meng *et al.*, 2009). The spatial and spectral resolution of these images was 30 m and 8 bits, respectively. All 68 images were

orthorectified using a digital elevation model and hydrographic data, at a scale of 1:100,000, resulting in a positional accuracy of better than 30 m.

3. Methods

The SF-method was carried out following four sets of procedures: (1) the calculation of SF from aerial photography and the generation of a regression model with SF as the independent variable to predict stand height and volume as dependent variables and thus produce local maps from the model, (2) the evaluation of a tree-shadow normalization procedures, (3) the analysis of the most favourable area within the aerial photography for the application of the SF-method, and (4) the application of an imputation technique to map the study area for stand height and volume.

3.1. Calculation of SF and construction of a predictive model

The aim of the first set of procedures was to assign all pixels of a panchromatic image from a digital aerial photograph to a binary value of 0 for non-shaded status or 1 for shaded status. This shadow bitmap was produced by applying a threshold value to the digital data of the aerial photograph. Threshold values were defined visually by an analyst for each GSP within the photograph according to two main considerations: (i) avoid overestimates of shaded areas and (ii) ensure realistic proportions of shaded areas for various understories, such as those dominated by lichen and moss. Leboeuf et al. (2007) demonstrated that the threshold values selected had little effect on the resulting regressions since a wide range of threshold values resulted in small differences in the final model. In addition, experience has shown that the threshold value differs very little between analysts. The advantage of the manual thresholding in comparison of an automated approach is that it normalizes the SF values among each images considering the

radiometric difference between them. The radiometric difference can be caused by atmospheric conditions or radiometric difference of the cameras. Once the threshold value was selected, SF was calculated for each GSP by overlaying a 30 m reference square centred over each GSP and then using the shadow bitmap. The SF values were mapped as $SF = A_{TS}/A_r$ where A_{TS} is the area of shaded pixels and A_r is the area of the reference square in pixels. Leboeuf et al. (2007) reported very small differences in the results for reference sizes of 10, 20 and 30 m. Therefore, the size of the reference square was chosen to match the spatial resolution of Landsat imagery (900 m^2) for allowing the use of an imputation technique.

The SF values served as the independent variable to establish a statistical relationship to predict stand height and volume. To develop a regression model, 70% of the GSP (GSP_{cal}) was randomly selected to be used as a training or calibration dataset. Several model types (linear, polynomial, exponential) were evaluated using the goodness of fit, F-test, analysis of residuals and visual evaluation of the regressions. The goodness of fit of the regressions was calculated through the adjusted regression coefficient (R^2_{adj}). Supplemental statistics were calculated to evaluate the accuracy of the SF regression model. We used 30% of the remaining GSP (GSP_{val}) to calculate the relative Root Mean Square Error ($RMSE_r$) and the relative bias ($bias_r$) of the regression model:

$$RMSE_r = \sqrt{\frac{\sum_{i=1}^{i=N}\left[\left(FA_i - FA_{GSP\_i}\right)\right]^2}{\sum_{i=1}^{i=N}\left[FA_{GSP\_i}\right]^2}} \qquad (1)$$

$$Bias_r = \frac{\sum_{i=1}^{i=N} [FA_i - FA_{GSP\_i}]}{\sum_{i=1}^{i=N} FA_{GSP\_i}} \quad (2)$$

where FA_{GSP_i} are the measurements of the forest attribute (stand height or volume) taken from the GSP_{val}, FA_i is the attribute as calculated from the regression, and N is the number of observations. Two regression models were calculated to evaluate the effect of the time gap between the measures of the GSP and the aerial photograph: the first model with the older dataset from the GSP measured in 1997 and 1996 and the second model with the newer dataset from the GSP measured in 2000 and 2002. The impact of the species composition was also evaluated by ranking the GSP according to their proportion of black spruce. To do so, two groups of GSP were formed from the dataset: The first group included the 50 GSP with the highest proportion of black spruce (in terms of basal area) and the second group included the 50 GSP with the lowest proportion of black spruce.

A 30 m × 30 m raster layer was generated over the aerial photograph on which each cell had a SF value calculated from the shadow bitmap. Then, the SF value of each cell was converted in stand height and volume from the SF regression model. The raster map was validated by computing $RMSE_r$ and $Bias_r$ using the GSP_{val} dataset and their corresponding map values for height and volume because the map covered all GSP within the aerial photographs. Some elements like water bodies or cloud shadow had to be masked out from the raster map.

3.2. Tree-shadow normalization

Optical view angle variations on aerial photographs amplify bidirectional reflectance and affect the shaded areas due to the sun-target-sensor geometry. A tree shadow normalization was applied to the aerial photographs to rectify tree shadow fraction to a reference configuration. Normalization of the shaded area involved adjusting all SF values to a common geometry to account for differences in sensor viewing (zenith and azimuth) angles and differences in solar (zenith and azimuth) angles. The common geometry established for the normalization was a nadir viewing sensor (Zv=0), a flat terrain, black spruce trees represented by cones on posts, and solar disc illumination with Zs=40°. We considered tree shadow normalization based only on shadow length (main shadow dimension in the direction of sun illumination). Topography was ignored as study area had limited relief. The normalized SF values (SF_n) were calculated based on trigonometric formulas that modified the shadow area projected on the ground or on surrounding crowns as viewed by the sensor. Calculation of SF_n took into account two correction factors: (i) one that considered the shadow length projected in the plane of the sensor viewing geometry and (ii) one that considered the increasing shadow length with larger solar zenith angles. Consequently, SF_n was estimated based on trigonometric relationships including: (i) the solar configuration (sun azimuth (A_s) and zenith angles (Z_s) at the date and hour of the acquisition) and (ii) the viewing configuration (viewing azimuth (A_v) and zenith angles (Z_v) measured from the centre of the aerial photograph):

$$\hat{SF}_n = \hat{SF} \times C_{ZV} \times C_{ZS} \tag{3}$$

where C_{ZV} and C_{ZS} are the correction factors for the viewing and solar zenith angles:

$$C_{ZV} = \frac{1}{\sin \theta_V} \tag{4}$$

where

$$\theta_V = \cos^{-1}(\sin Z_V \times \cos(A_V - A_S)) \quad (5)$$

and

$$C_{ZS} = \frac{\tan Z_{s\_ref}}{\tan Z_S} \quad (6)$$

where $\tan Z_{s\_ref}$ is defined to 40 degrees.

3.3. The impact of sensor viewing geometry

We assessed what portion of the aerial photograph provided the most interesting results for the SF-method within the range of viewing conditions of the test sites. Then, two procedures were implemented to assess the impact of the viewing geometry. Firstly, the impact of the distance between the GSP and the centre of the aerial photograph was estimated on the SF regression models. Secondly, the impact of the azimuthal viewing angle between the GSP and the centre of the aerial photograph was also assessed on the SF regression models.

Firstly, we assessed the impact of the distance between the GSP and the centre of the aerial photograph on the SF regression models. The rationale for this procedure is based on the hypothesis that when the nadir angle of observation increased beyond a certain point the shadows were overestimated in the sun direction and underestimated in the opposite direction. We expect

that beyond a given angle there will be a reduced capability to predict forest attributes from the SF regression models. To test for the existence of such a threshold, we calculated the regression models between SF and the two forest attributes (stand height and volume) using a group of 35 GSP from the complete set of 142 GSP. A first trial used the 35 GSP closest to the north (0°). Then, we recalculated regression models and their R^2_{adj} by removing GSP with the smaller angle and replacing it by the next GSP with a position clockwise. This process was applied to new trials until we produced models using all 142 GSP, by gradually using the set of the 35 GSP closest to the nadir to the furthest. To provide comparable statistics, we always applied linear regression models. These procedures allowed a comparison of R^2 between trials to illustrate how distance between GSP and the centre of the photograph affects the ability to derive a model. We will thus be able to propose a distance threshold for the useful range for GSP selection within aerial photographs, implying that GSP beyond this range should not be used for model development.

Secondly, we also assessed the strength of the SF regression models according to the azimuthal viewing angles measured between the GSP and the centre of the aerial photograph. Following the previous hypothesis, the azimuthal viewing angle should affect the SF values. Logically, based on a sun azimuthal angle of 180 degrees (south), SF values should be overestimated in the southern angles and underestimated in the northern angles. Then, the GSP with angles that overestimated or underestimated SF could be eliminated from the dataset used in the regression models. To determine the impact of azimuthal viewing angles on the regression model, we calculated the regression models between SF and the two forest attributes (stand height and volume) using a group of 35 GSP. A first trial used the 35 GSP that were the closest to 0° (north). Then, we recalculated the regression models and their R^2 by removing the GSP with the smallest

angle and replacing it by the 36th GSP with the lowest angle. This process was successively applied to all 142 GSP. To provide comparable statistics, we always applied linear regression models. These procedures allowed comparison of R^2 between trials to illustrate how azimuthal viewing angles between GSP and the centre of the photograph affect the ability to derive a model. We will thus be able to suggest an appropriate range of angle values for GSP selection within aerial photographs, implying that GSP outside of this range of angles should not be used for model development.

3.4 Imputation technique to map stand height and volume

KNN techniques were chosen among several imputation techniques because of their proven potential to combine remote sensing imagery and GSP (Tomppo *et al.* 2010; Wilson *et al.* 2012). Moreover, such non-parametric techniques allow simultaneous estimation of multiple variables by preserving the covariance structure among the variables. Also, the KNN techniques are largely used in the scientific community and for national forest inventory of several countries (Tuominen *et al.* 2003; Tomppo and Halme, 2004, LeMay and Temesgen, 2005; LeMay *et al.* 2008). We trained the models using 1,200 randomly selected cells located within the center of the aerial photographs. We forced also the selection process to avoid the selection of two cells which would be closer that 150 m to avoid bias due to the spatial correlation. Using this dataset and the three Landsat-TM bands from the mosaic, the forest attribute value ($FAknn_i$) was calculated for each pixel of the MSRSI according to this equation:

$$FAknn_i = \frac{\sum_{k=1}^{k} W_k FA_{GSP\_k}}{\sum_{k=1}^{k} W_k} \quad for\ k \neq 1 \quad where\ W_k = \frac{1}{d_k^j}, \tag{7}$$

where $FAknn_i$ is the forest attribute value calculated from the KNN technique for cell i; k is the k^{th} nearest spectral neighbor; FA_{GSP_K} is the forest attribute value from the dataset; W_k is the weighting coefficient; d_k^j is the Euclidian distance to which is applied a power of j. The resulting $FAknn_i$ value.is in fact a mean value of the FA_{GSP_K} values, which have the nearest spectral value of the pixel evaluated. The parameter j = 0 generates a simple mean value, j = 1 generates a mean value weighted to the power 1, and j = 2 generates a mean value weighted to the power 2. For the non-parametric technique, two statistics were calculated to evaluate the results: relative Root Mean Square Error ($RMSE_r$) and bias (bias).

4. Results

4.1. Regressions between SF and forest attributes

Several regression models can be used to predict stand height and volume from SF values as the independent variable. To select the best model, different comparisons were tested. Table 2 shows the regression models that provided the lowest $RMSE_r$ and $Bias_r$ and highest R^2 for stand height and volume. The best results came from a linear regression with $RMSE_r$ and $Bias_r$ of 0.15 and 0.02 for stand height, and 0.39 and 0.12 for stand volume. A plot of these two linear models is given in Fig. 2 including measured values from ground plots. Other regression types (polynomial, exponential, logarithmic) were tested. The linear regression gave better results with the exception of the exponential regression that gave similar results. However, given the similar results, we

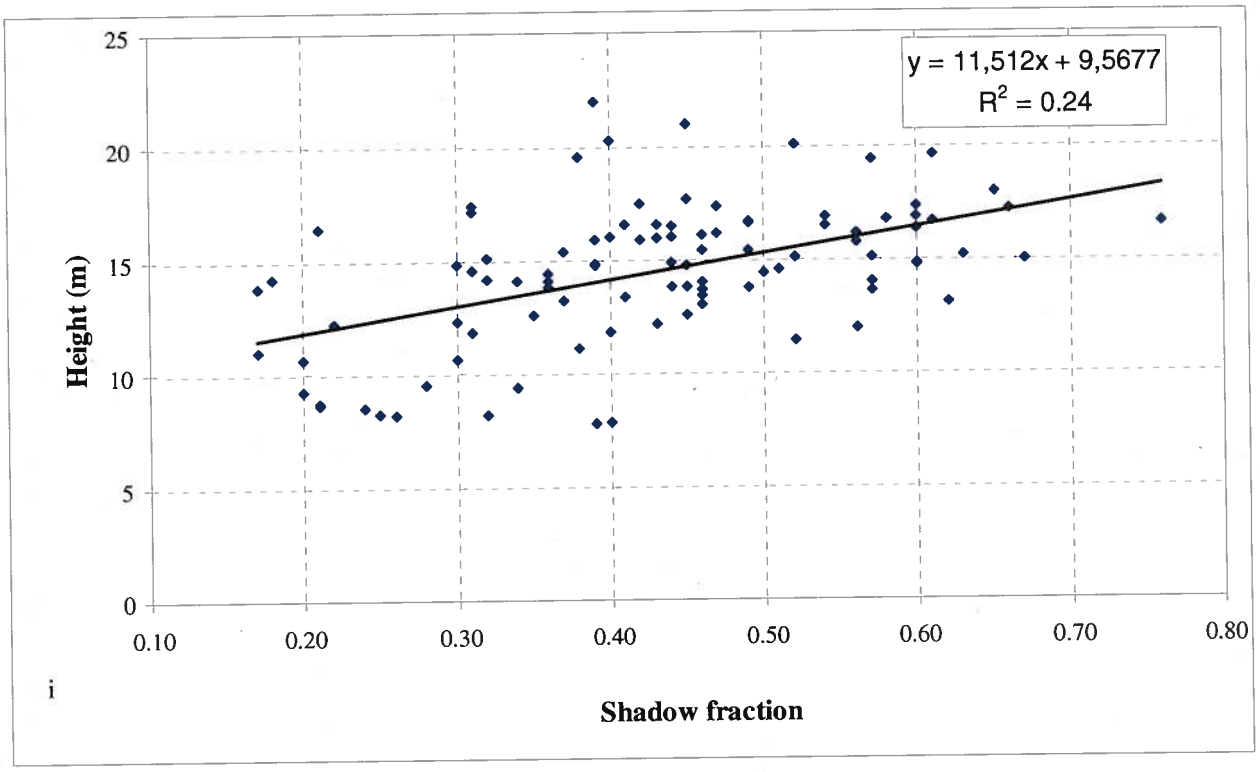
selected the linear regression because we believe that the curve is more representative of the expected increase of stand height or volume related to SF. An exponential increase of stand height or volume at high SF is not typical of field observations. On the contrary, height and volume values tend to saturate beyond a certain SF value. Also, we did not force the regressions to pass through the origin since there were no GSP with a stand height value less than 7 m. Consequently, we felt that it was acceptable as height and volume values were unavailable for the 0 – 7m range and we did not want to impose the shape of the relationship.

The linear models selected were used to produce a map of stand height and volume with cells of 30 m × 30 m. Fig. 3 shows the stand height map resulting from the application of the regression to one plot within an aerial photograph. The visual interpretation of the resulting patterns from the stand height map corresponded well to the expected patterns, i.e., stand height was close to zero for wetlands or openings in the stand canopy with a gradual increase towards denser forest stands. In addition to the general map, we produced an estimate of stand height and volume for a pixel of 30 m overlaid (centred) on all GSP_{val}. The resulting $RMSE_r$ and $Bias_r$ of these estimates using the validation dataset were 0.15 and 0.01 for stand height, and 0.61 and 0.23 for volume. The impact of time gap of GSP on regression models was noticeable with an increase of R^2 from 0.11 (GSP of 1996 and 1997) to 0.25 (GSP of 2000 and 2002) for stand height and from 0.20 (GSP of 1996 and 1997) to 0.49 for stand volume (GSP of 2000 and 2002). Also, the impact of species composition was important since the R^2 of 16.5% (stand height) and 34.2% (volume) for the 50 GSP with the higest proportion of black spruce dropped to a R^2 of 0.02% (stand height) and 18.9% (volume) for the 50 GSP with the lower proportion of black spruce. Other species

present in the GSP than black spruce included jack pine (56%), balsam fir (28%) and deciduous trees (12%).

Table 2 - Results from the best linear regression for stand height (m) and stand volume (m^3/ha).

Forest attribute	Regression	Equation		$RMSE_r$	$Bias_r$	R^2_{adj}
		Slope	Intercept			
Stand height	Linear P<0.001	11.51 P<0.001	9.57 P<0.001	0.15	0.02	0.23
Stand volume	Linear P<0.001	225.29 P<0.001	-12.49 P=041	0.39	0.12	0.31



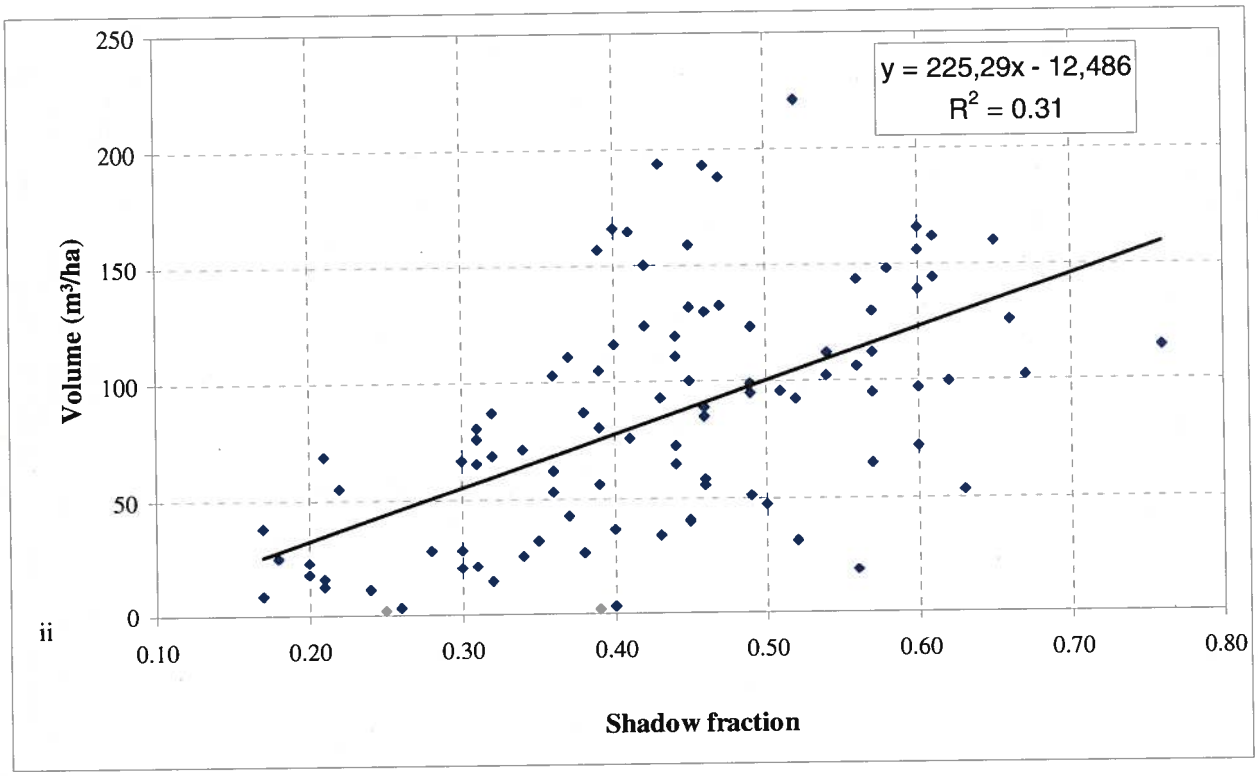


Figure 2. Linear regression using the shadow fraction of trees to predict (i) stand height and (ii) stand merchantable volume.

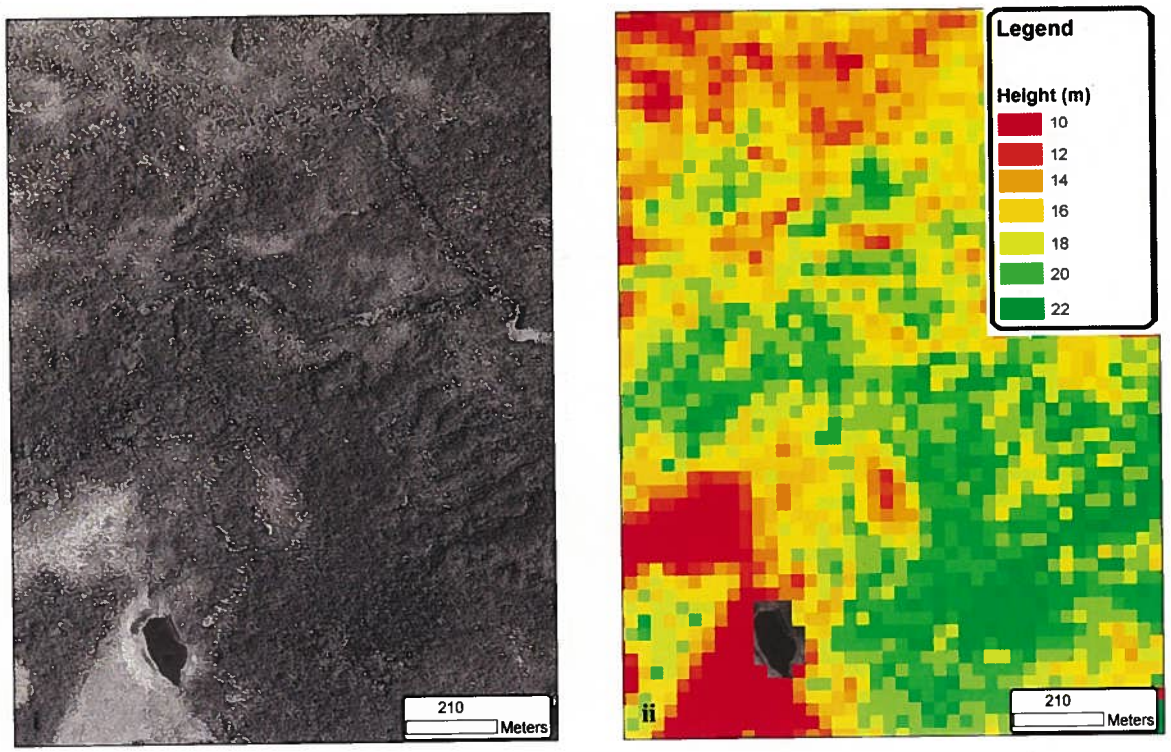


Figure 3. (i) A panchromatic image of the three bands of an aerial photography used to produce (ii) the stand height (m) raster map of 30 m of cell size from SF-method.

4.2. Impact of tree-shadow normalization

The SF_n values were calculated based on trigonometric formulas correcting the original SF values for sensor viewing (zenith and azimuth) angles, which were accounted for by C_{ZV}, and for solar (zenith and azimuth) angles, which were accounted for by C_{ZS}. The R^2_{adj} value for the linear model to predict the stand height decreased from 0.23 when using SF to 0.22 when using SF_n. Similarly, the R^2_{adj} value for the linear model to predict stand volume decreased from 0.31 when using SF to 0.22 when using SF_n. However, applying only the correction factor C_{ZV} had no impact on R^2_{adj} which remained at 0.23 and 0.31 for the height and volume predictions respectively. This

is easily explained since viewing angles vary within a limited range from 2 to 29 degrees. In contrast to C_{ZV}, the inclusion of C_{ZS}, the correction for solar zenith angles, had an important impact reducing the R^2_{adj} to such a low level that predictions were no longer reliable. The impact of these results is discussed later but overall we did not use SF_n for the rest of the analysis.

4.3. The impact of sensor viewing geometry

The first procedure assessed the impact of the distance between the GSP and the centre of the aerial photograph. The R^2_{adj} values were calculated for a group of 35 plots (25% of all the dataset) where the distance was defined at the centroid of the 35 GSP. The results are shown in Fig. 4 where R^2_{adj} values were higher for zenithal angles less than 14 degrees (e.g. 66 % of the center of the image surface) and at their peak ($R^2_{adj} > 0.30$) between zenithal angle of 8.5 and 14 degrees (or 600 to 1000 m from the centre of the aerial photograph in our case). Also, results gradually degraded beyond zenithal angle of 14 degrees (1,000 m) reaching a R^2_{adj} value of 0.16 at 29 degrees (2,216 m). Fig. 5 shows the regression model for the highest R^2_{adj} of 0.58 at 11.2 degrees (795 m).

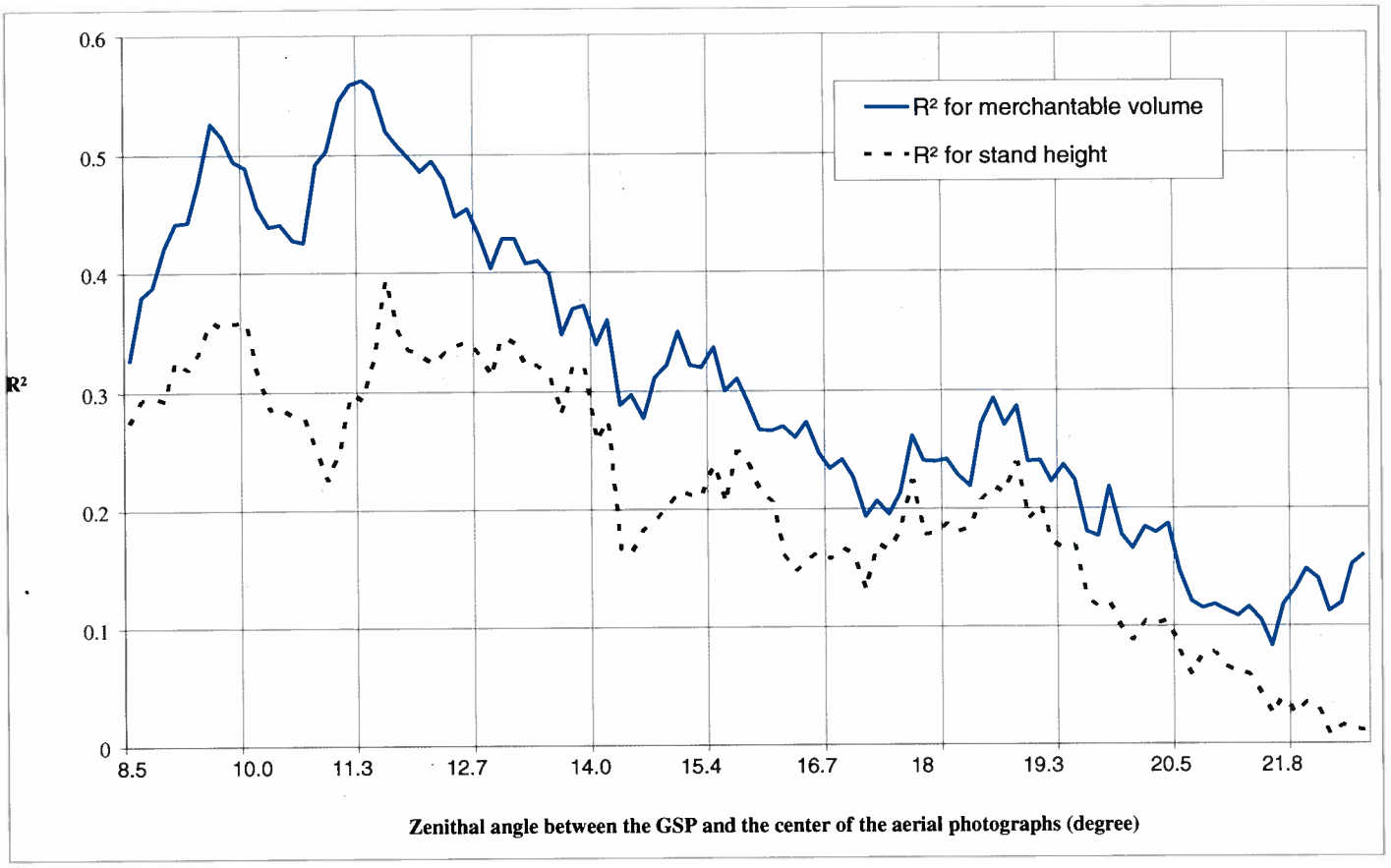
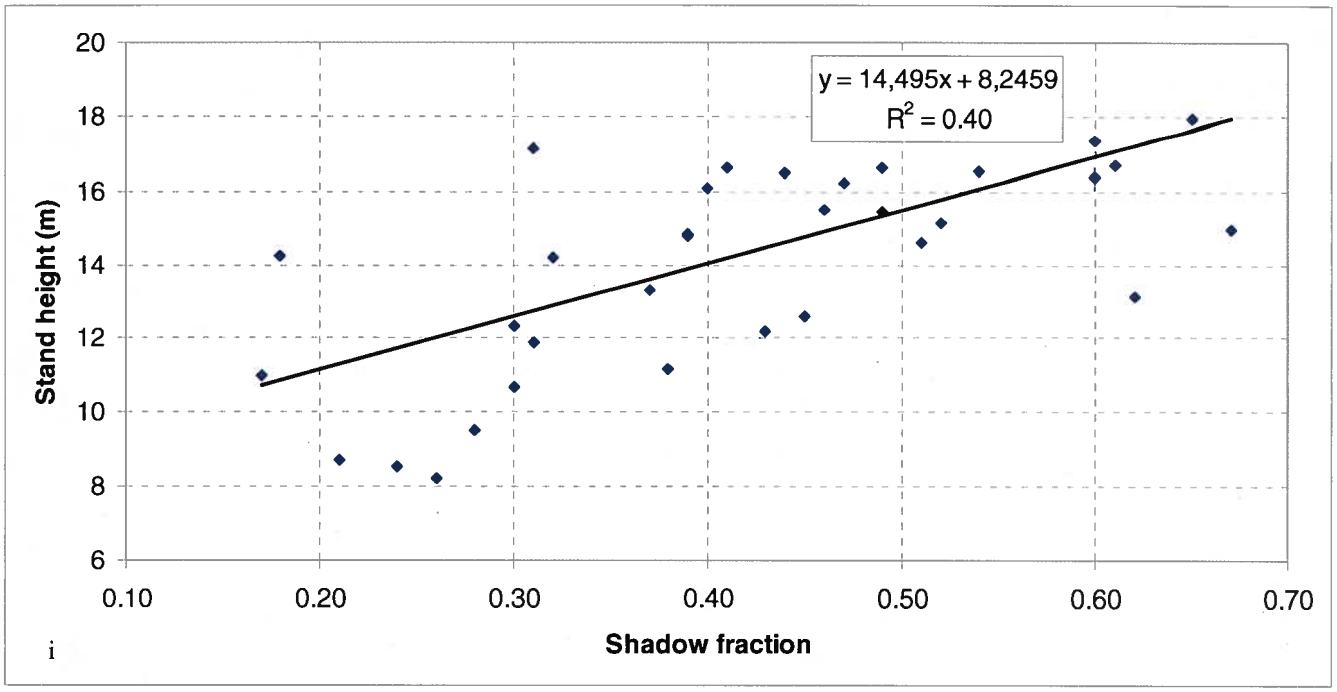


Figure 4. Coefficient of determination (R^2) for the linear regression between the shadow fraction of trees and stand merchantable volume using a cluster of 35 GSP located at several zenithal angles (degrees) from the center of the aerial photographs.



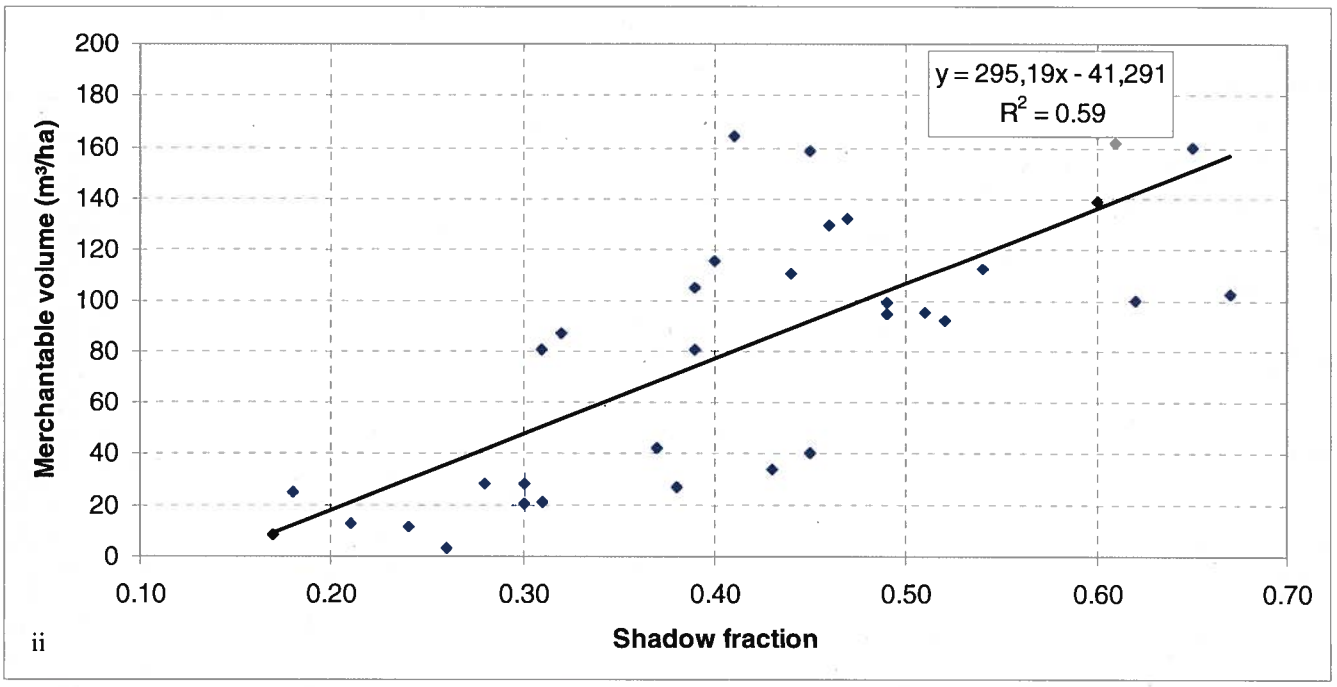


Figure 5. Linear regression between the shadow fraction of trees and (i) stand height or (ii) merchantable volume with the 35 GSP located at 11.2 degree (i.e. between 6.1 and 11.3 degree) of the center of the aerial photographs.

The second procedure assessed the impact of the azimuthal viewing angle between the GSP and the centre of the aerial photograph for the GSP centroid located between zenithal angle of 8.5 and 14 degrees. The R^2_{adj} values were also calculated for a group of 13 plots (25% of the GSP from 8.5 and 14 degrees of zenithal angle) where the azimuth angle was defined at the centroid of the 13 GSP_{cal}. The R^2_{adj} values were compared for all the azimuth angles between the centroid of the 13 GSP_{cal} and the centre of the aerial photograph. It is obvious that GSP_{cal} located at the centre of the aerial photograph should not be affected by the azimuthal viewing angles since the tree crowns are seen from the top and not from the side. Also, we demonstrated in the first procedure that results are affected by the distance to the centre of the photograph. Therefore, we compared the R^2_{adj} values for a series of azimuthal angles of the GSP_{cal} centroid located between zenithal angle of 8.5 and 14 degrees (Fig. 6). Fig. 7 shows that R^2_{adj} values were higher when the centroids of the GSP_{cal} were located towards the northeast of the centre of the aerial photograph. Conversely, the lowest R^2_{adj} values were obtained when the centroid was located towards the south and the north. The most favourable angles for the GSP_{cal} were between 28 and 129 degrees from the azimuth leading to R^2_{adj} of 0.42 for the stand height and 0.79 for merchantable volume (Fig. 7).

In summary, SF and both stand height and volume were correlated allowing us to obtain linear relationships linking them. They can therefore be used to estimate forest attributes for stands dominated by black spruce. However, since the use of normalized SF values was not suitable, two main limitations need to be taken into consideration to use the SF to map forest attributes with aerial photographs: (1) use the central portion of the aerial photograph with a zenithal angle lower than 14 degrees or from 1,000 m to the centre of the aerial photograph in our case and (ii)

when possible, avoid back and forward scattering zones of the aerial photograph (respectively the southern and the northern portions in our case). Fig. 8 illustrates to position of back and forward scattering zones according to sun and viewing angles.

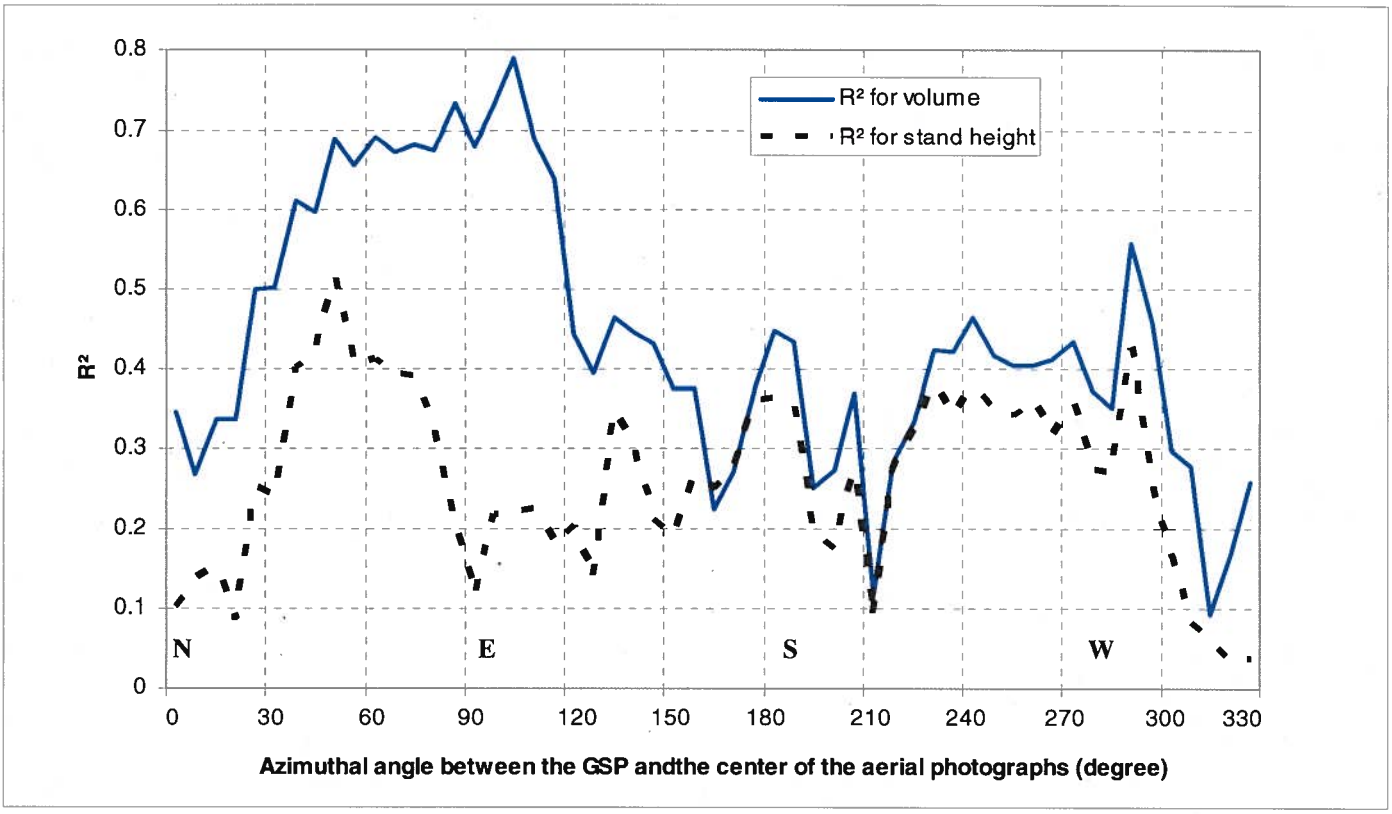
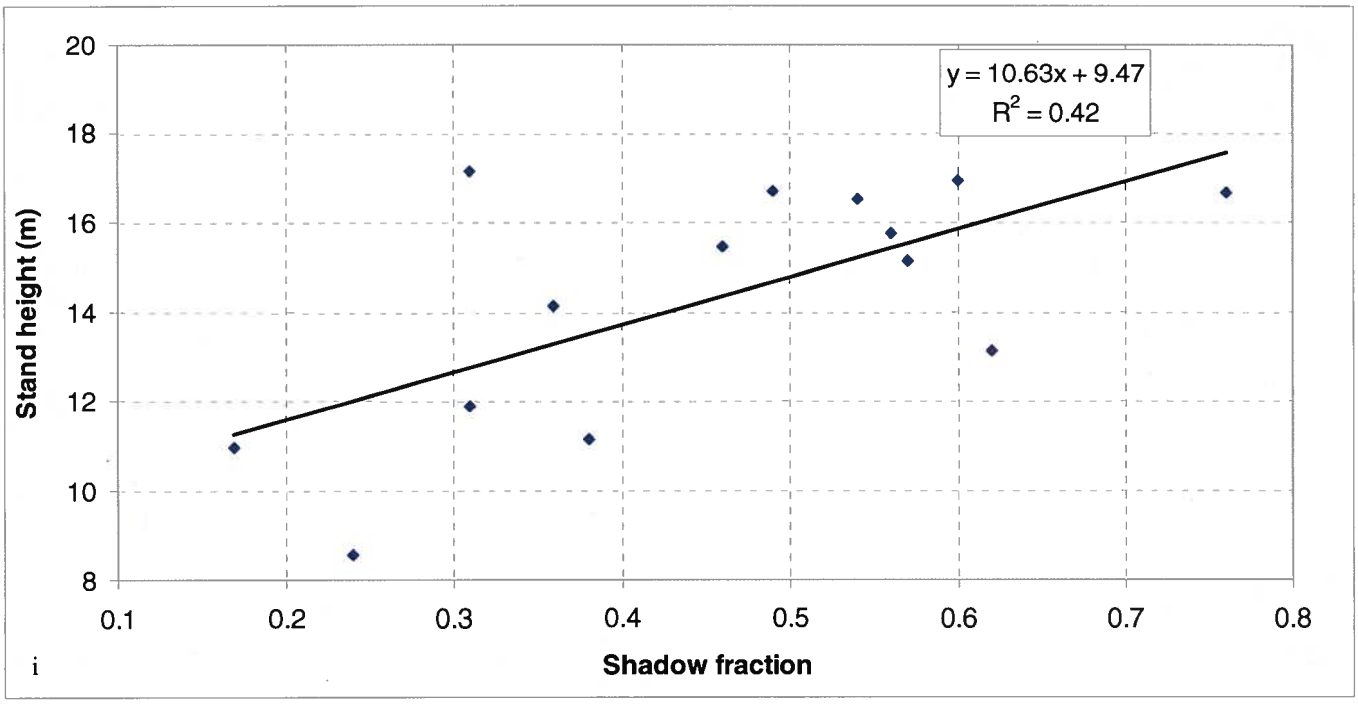
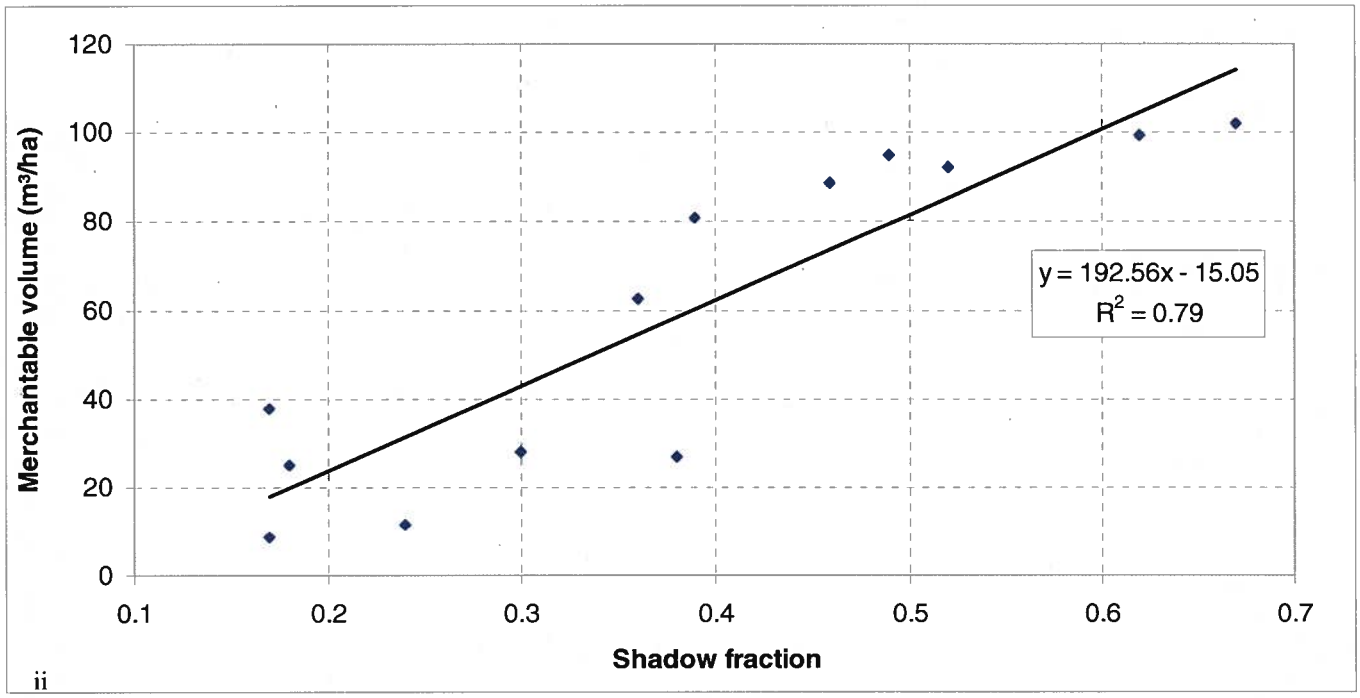


Figure 6. Coefficient of determination (R^2) for the linear regression between the shadow fraction of trees and (i) stand height (hashed line) or (ii) stand merchantable volume (full line). with a cluster of 13 GSP located at different azimuth angles. All GSP were between 8.5 to 14 degree to the center of the aerial photographs. Azimuth angles 0, 90, 180 and 270 represent north, east, south and west respectively.

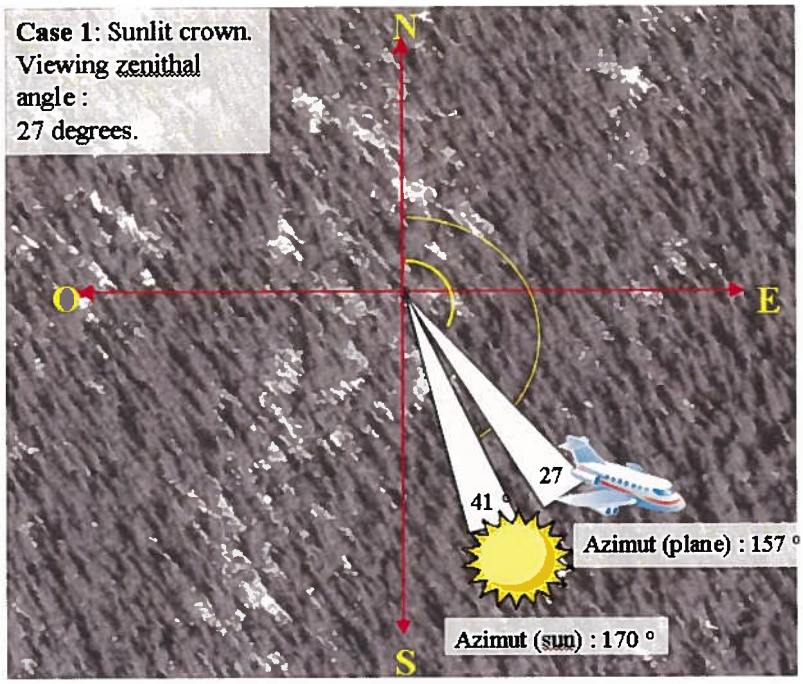


i

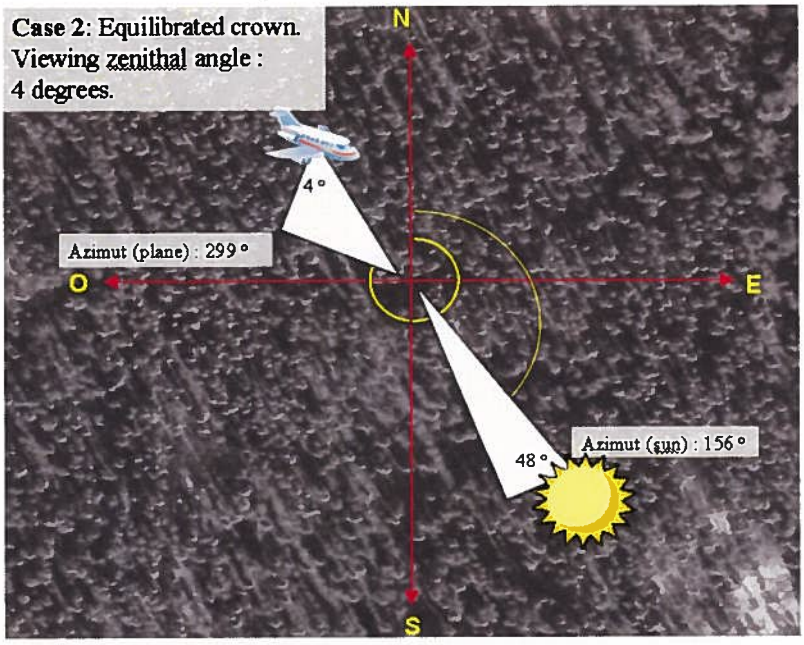


ii

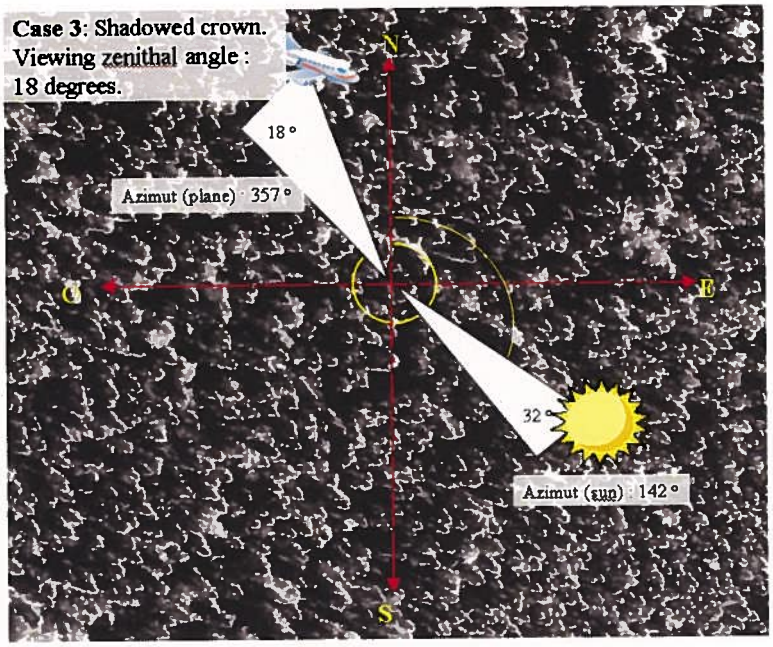
Figure 7. Linear regression between the shadow fraction of trees and (i) stand height or (ii) stand merchantable volume with 13 GSP_{ang}. The regression was applied for the GSP_{ang} located in the northern azimuth (between 9 – 64°) for the stand height model and for the GSP_{ang} located in the eastern azimuth (between 52 – 138°) for the stand volume model. All GSP_{ang} were also located between 600 and 1,000 m from the center of the aerial photograph.


Case 1: Sunlit crown.
Viewing zenithal angle :
27 degrees.
N
O
E
S
27
41 °
Azimut (plane) : 157 °
Azimut (sun) : 170 °

i


Case 2: Equilibrated crown.
Viewing zenithal angle :
4 degrees.
N
4 °
Azimut (plane) : 299 °
O
E
Azimut (sun) : 156 °
48 °
S

ii



iii

Figure 8. Examples of (i) back scattering that underestimated the SF, (ii) central zone that estimated correctly the SF and (iii) forward scattering that overestimated the SF.

4.4. The regional map generated from the imputing technique

A map of the merchantable volume for the the study area was produced using the KNN technique (Fig. 9). Expected ecological patterns were followed by this map with higher forest attributes values on richer and denser stands. Results also followed expected trends for area where interpretation is possible: e.g. drier sites or those that burnt few years ago show low stand volume. $RMSE_r$ and bias obtained from this method were respectively 15.8% and 0.04 m for stand height and 51% and -0.6 m^3/ha for stand volume.

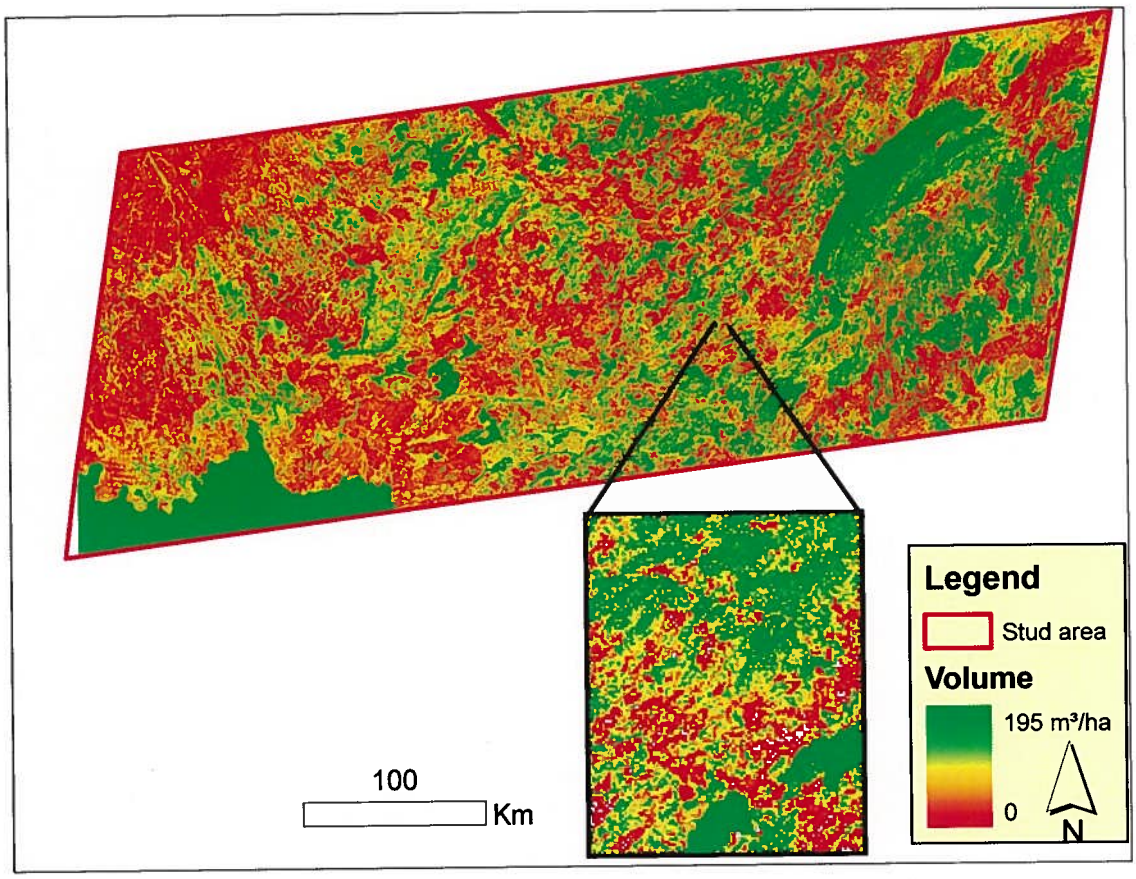


Figure 9. Volume (m^3/ha) map over the study area using a scaling up approach.

5. Discussion

The results from our study provide support for the expansion of the SF-method to aerial photographs. The aerial photographs present large geographic and temporal coverage. Different acquisitions projects (forest inventory, transportation, hydrographic mapping, etc.) make available aerial photographs at different spatial and spectral resolutions. Also, aerial photograph has been acquired since several decades. This temporal availability of the aerial photographs opens the potential for using the SF-method for the purpose of temporal monitoring of forest. Also, the SF-method was demonstrated to be applicable with high accuracy rate over northern

areas covered by black spruce and balsam fir stands (Luther et al., 2012). Results using aerial photographs were comparable to those from studies involving very high spatial resolution satellite imagery (Leboeuf et al., 2012) when we used only central part of aerial photographs for SF-method. Also, the fact that shadow fraction is relied on the shadowing patterns should minimise the radiometric effect between aerial photographs cameras. Thus, following methodological limitations of this paper, SF-methods should be applicable also on film based aerial photographs. For example, the R^2 for volume were of 0.72 from Leboeuf *et al.* (2012) compared to the value 0.79 obtained in this study for the central part of the photographs. The imputation technique using aerial photograph as training dataset paved the way for mapping forest attributes over large area of boreal shield. However, this can only be done with a better understanding of the limitations present when applying the method to aerial photographs. Therefore, it is useful to look at two sets of limitations: those related to the radiometry and those related to the available datasets. The former will guide us towards how the SF-method can be applied adequately to aerial photographs while the latter will help us understand our results.

The main limitations explored between the scene radiometry and the forest attributes included three aspects: (1) the tree shadow normalization, (2) the distance of the GSP from the centre of the photograph and (3) the azimuth angle of the GSP in the image. Our results showed that applying the tree shadow normalization decreased the resulting R^2_{adj} values from (0.27 to 0.08). This result is contrary to what was expected. We suggest that this is probably due to several aspects: (i) the trigonometry method for normalization was based on the shadow cast by a cone which was probably not appropriate to capture the complexity of the forest structure in the study area, (ii) the local topography was not taken into account in the normalization process, and (iii)

the viewing and sun angles may change greatly from one photograph to another. Our results and our interpretation of this limitation suggest that radiometric normalization for relief should be further examined before it can be used efficiently (Li and Strahler, 1992; Kane et al., 2008). This development, however, requires precise digital elevation model (beyond what is currently available) and a large number of GSP (that we did not have in the present study). Secondly, the results were clearly dependent on the distance between the GSP and the centre of the aerial photograph. More specifically, our results suggest using only a range of distances less than 14 degrees of zenithal angles (corresponding of 1,000 m) from the centre for satisfactory results. This concurs with Pellikka et al. (2000b) which suggested that shadow fraction measured for different part of aerial photographs were significantly altered for area located at zenith viewing angles larger than 10 degrees. Beyond that angle, the view of the tree shadows may be distorted: either reduced or enhanced according to the view direction and the crown density. In short, the empirical model derived was most accurate for a small range of the viewing angle (<14 degrees) close to nadir. Thirdly, results were strongly affected by the azimuth of the GSP relative to a line going from the centre of the photograph to the north (azimuth = 0 degrees). The best results were obtained for azimuth angles in the eastern quadrant (i.e., from 45 to 135 degrees; see Fig. 6). However, the GSP in the western quadrant maintained relatively high R^2_{adj} values. These two directions can also be considered for use in an operational setting if the accuracy requirements are sufficiently high for the purpose. The weaker results were consistently observed for GSP in the northern and southern directions. This can be explained by the fact that the mean sun azimuth value was 210 degrees (therefore generally towards the south at 180 degrees). This sun azimuth led to increase the BRDF effect in the axis of the sun resulting in darker pixel of shadow and then higher SF values for the GSP located in the southern part (forward scattering) of the aerial

photographs. The opposite effect can be observed in the northern part (back scattering) of the aerial photographs. Also, we observed an important discrepancy between stand height in R^2 values within 60 and 120 degrees. This discrepancy in the results can be explained by the fact that SF model of tree height is more fragile since the height range is restraint from 7.8 to 22 m. Thus, the sun-target-sensor geometry biased the SF value within these angles and reduces statistics more considerably than for SF models of volume. Our results related to scene radiometry show clearly the effects of the viewing sensor geometry toward the two dominant factors being (1) the position of the plot relative to the photograph's centre and (2) the azimuthal position of the plot.

Several factors, other than those mentioned so far, contributed to negatively affect the correlation coefficients of the regression models. Firstly, the study was done with databases that were not measured simultaneously, i.e., the aerial photographs were acquired in 2009 and 2010 and the GSP were measured between 1996 and 2007. Our result s showed clearly that using datasets collected at a time close to the time of acquisition of aerial images improved the results: R^2 values increased from 0.20 to 0.49 for predicted volume using older to more recent datasets respectively. Boreal forests in that area typically have a growth rate of 0.15 m and 1.04 m^3/ha per year for tree height and stand volume. These growth rates have been calculated for 18,394 GSP available for similar forest ecosystems. Thus, with such low growing rate, we expect the impact of temporal gap between acquisition of GSP and aerial images not to be related primarily with growth but more related with disturbances. Secondly, the regression models we calculated were limited by the range of forest attribute values present in the GSP within the 73 aerial photographs. For example, stand height values of our GSP ranged from 7.8 to 22 m. A complete assessment of the SF-method before operational application should include GSP with forest attribute values

covering the full height range for black spruce stands which spans from 0 up to 23.5 m in Quebec (MRNF, 2011). We can expect a saturation of the SF values for large stand volume (> 160 m^3/ha) or for tall and dense stands. This hypothesis is assumed from the fact that the SF would no longer be tied to an increase of the FA value at that point due to saturation of SF linked to high crown closure with mutual shadowing. However, this hypothesis needs further testing because we did not encounter that situation in our dataset. Meanwhile, our study provides regression models for the vast majority of cases in the boreal forest of northern Canada. Thirdly, the regression models described in this study were developed only for black spruce stands and the analysis on the composition of the GSP demonstrated that the other species seemed to reduced the R^2 values. Fortunately, other studies like Luther et al. (2012) and Wolter et al. (2012) successfully extended the SF-method using satellite images to other conifer species and different forest conditions. We can therefore expect that the SF-method can also be extended to other coniferous species using aerial photographs. Therefore, Seed and King (2003) obtained poor results between SF and LAI for mixed stand.

The limitations to apply the SF-method using aerial photographs demonstrated that only a portion of the photograph (close to nadir) is suitable to apply the model. Considering the large number of photographs taken during an inventory, this still allow producing local maps of forest attributes and apply them in an imputation technique to map large areas with the use of MSRSI. Therefore, in a National Forest Inventory perspective that process could be useful by using the central part of aerial photographs which overlaid the systematic grid of these programs. This systematic sampling of stand height and volume would be followed over the space by imputation and over the time by comparing the results over superimposed aerial photographs acquisitions over the time. The forest attribute maps produced by this present work demonstrated this possibility with

the production of the forest attribute map over the study area within the range of accuracy of similar works (Tomppo et al. 2008). Those maps showed fine local variation of these attributes and can give landscape statistics.

6. Conclusion

The present work demonstrated through several validation studies that the SF-method based on aerial photography can estimate two widely used forest attributes, stand height and volume, for black spruce dominated stands in the northeastern forests of Canada. Our results showed the importance of the solar and sensor viewing geometry for satisfactory results from aerial photographs. Two factors dominate: (1) selection of an area within a range of 8.5 to 14 degrees of zenithal angles from the centre of the aerial photograph and (2) selection of an area in the eastern or western quadrants of the aerial photograph. Within these limitations, predictive models had R^2_{adj} between 0.42 and 0.79 for stand height and volume. The SF-method tested in the present study provided a convenient model but must be applied, as much as possible, within the suggested distance range from the centre and for specific azimuth ranges. Such a model is applicable to the large majority of black spruce stands in northeastern Canada to estimate and map two forest attributes (stand height and merchantable volume). In addition, panchromatic aerial photography present both geographic and temporal availability, even for northern area, and the method works with few or no ground plots by applying the SF-method model previously calculated as done in this study. Raster maps obtained from the SF-method also showed promise for calculating intra-polygonal information such as horizontal stand structure, volume distribution, etc. Although several methods have already been developed to meet these requirements for forest inventory, they generally obtained weak results or results that could only

be applied locally. Other approaches were based on spectral mixture analysis (Peddle and Johnson, 2000; Peddle et al., 2001; Seed and King, 2003). Several aspects of our study support the generalized application of the SF-method to black spruce stands of northern boreal forests: (i) our results were consistent over the 73 aerial photographs, (ii) the SF-method is adapted to a large range of stand height and volume values, and (iii) it was applicable to aerial photographs acquired on different dates. Mapping forest stands usually requires ground plots for calibration or interpretation. The SF-method overcomes this limitation through using regression models. If GSP are available, they can be used to confirm the suitability of the model for a more extended area, as was done for other studies on the SF-method (Leboeuf et al., 2012; Luther et al., 2012; Wolter et al., 2012). Otherwise, the model can be applied without GSP and provides a convenient alternative because of the significant resources required for GSP measurements. The interpretation of aerial photographs suffers from subjective bias. The results from the SF-method are compatible with those from an experienced interpreter, but they are more consistent from one image to another without the subjectivity between interpreters. In addition, the SF-method can be applied to complement regular forest inventory using the same aerial photographs used in the photo-interpretation process. However, as demonstrated in this present study, only the central part of the aerial photograph is useful to derive a SF-model. This aspect limits the application of SF-method for mapping large area using aerial photographs. To overcome this limitation, we proposed applying an imputation technique which use the local maps produced by the use of aerial photographs as training dataset of a KNN algorithm aiming to regional map of forest attributes.

7. **Acknowledgments**

We are grateful to the Ministère des Ressources Naturelles in Quebec for the dataset and support for the project.

7. Article 3

Forest attribute estimation over large areas of Northern boreal forests using the shadow fraction method and the k-Nearest Neighbours

A. Leboeuf[1] and R.A. Fournier[1]

[1]Centre d'Applications et de Recherches en Télédétection (CARTEL), Université de Sherbrooke, 2500 boul. de l'Université, Sherbrooke, Québec, J1K 2R1, Canada

E-mail of corresponding author: antoine.leboeuf@usherbrooke.ca
Telephone: 819-821-8000 ext. 3209, Fax: 819-821-7944

Manuscript to be submitted to Journal of Environmental Monitoring and Management

ABSTRACT

Detailed mapping of the Canadian northern boreal forests is difficult because of the huge area to cover, the distance to urban centres and a generalized lack of road access. However, this ecosystem is highly dynamic and has a high impact on global climate and resources availability. A reduced road access and forest homogeneity in terms of tree species are compelling for the use of remote sensing techniques as an alternative. Our work aimed to develop an operating method to map forest attributes (merchantable volume and stand height) using satellite images and KNN techniques. The method was tested over a large area (470,000 km²) in Québec, Canada, largely dominated by black spruce (*Picea mariana* Miller). The method involved three sets of procedures: (i) for each test site, inferring forest attributes calculated from the shadow fraction of the panchromatic values of very high resolution satellite images, (ii) comparing several KNN approaches and, (iii) mapping the forest attributes as a grid layer (30 m × 30 m) for the whole area via Landsat images using the most appropriate method. To calibrate the regression model, available ground sample plots were randomly selected. These plots were acquired from 1996 to 2009. Relative RMSE values ranged from 0.18 to 0.58 for stand height and merchantable volume obtained by KNN techniques. The results showed that our method is a convenient and inexpensive way of mapping forest stands for large areas of Canada's northeastern boreal forest.

Keywords: Forest inventory , k-Nearest Neighbours, shadow fraction, remote sensing.

1. Introduction

The managed portion of Canada's eastern boreal forest, located roughly south of 51st parallel, is well known and mapped. As an example, the Ministère des Ressources Naturelles in Québec (MRNQ) produces, on a 10- to 15-year cycle, an ecoforest map at a scale of 1:20:000 (Robert and Robitaille, 2009b). However, all across Canada the boreal forest located north of the 51st parallel is not well documented. Only few vegetation mapping projects have been done in these remote areas including the Earth Observation for Sustainable Development (Wulder, 2000) and a project from the MRNQ (Robitaille *et al.* 2008). It is noticeable that these projects did not map stand height and merchantable volume. Only few ground sample plots (GSP) were available which is not statistically representative to calculate reliable models for merchantable volume. In addition, there is no GSP in Quebec with stand height of less than 7 m because it is assumed too juvenile for inclusion in the inventory. Other constraints included the large dispersion of the few available GSP, their measurements over a long period, and the diversity of the northern landscape. Regardless of these constraints, stand height and merchantable volume are essential for a basic knowledge of these forests and to support their sustainable management. Therefore the development of a mapping method of stand attributes requires increasing the number of GSP through some surrogate plot measures.

Several approaches can be used to map stand height and merchantable volume. First, interpretation of aerial photography is an operational tool traditionally used for forest management. Using this approach, stand height can be measured with satisfactory precision and merchantable volume can be derived using volume models. Remote sensing tools have also largely demonstrated their potential for calculating forest attributes (Gjertsen *et al.* 1999; Hyyppä *et al.* 2000). High spatial resolution satellite imagery (HSRSI), which is an alternative to aerial photography, is becoming more frequently used to map forest attributes such as stand height, merchantable volume, biomass, basal area and stand cover. With only a visual inspection of enhanced HSRSI acquired over boreal regions generally show a visual gradient of intensities where light-colored areas correspond to open stands and dark areas correspond to stands with

high crown density. This gradient of intensities is largely due to tree shadow. Shadows viewed from the HSRSI are composed of (1) the shadowed portion within the crown and (2) the shadow casted on the ground by crowns (Li and Strahler, 1985). Tree shadow fraction (SF) is the sum of all tree shadow areas divided by an area of fixed dimension (Seed and King, 2003). Several forest attribute mapping methods have been developed using HSRSI. Work by Peddle *et al.* (2001), Seed and King (2003), Greenberg *et al.* (2005), Leboeuf *et al.* (2007; 2012), Luther *et al.* (2012) and Wolter *et al.* (2012) offer good examples of the use of SF from HSRSI to calculate forest attributes such as stand height, volume, biomass, percentage of canopy cover, basal area and leaf area index. The SF-method seems adapted to provide mapped areas that can serve as surrogate GSP which in turn can be use to generalise the estimates from the local to the regional level. However, HSRSI are expensive and can difficultly cover large area or assess forest status at the regional scale.

Mapping forest attributes at a regional scale requires the use of medium spatial resolution satellite images (MSRSI) such as Landsat-TM/ETM+, SPOT-5 or RapidEye images. These images have a spatial resolution of from 5 to 30 m, several spectral bands and are available at a reasonable cost or for free. The use of a two- or three-stages approach for large area mapping has been widely suggested in the scientific literature (Cihlar, 2000; Wulder *et al.* 2012b). Within such approaches aiming to cover at different scales, several projects have confirmed the applicability of remote sensing images using imputation techniques such as the k-nearest neighbour (KNN) technique (McRobert and Tomppo, 2007; Tomppo *et al.* 2008). This intuitive and versatile technique aim to estimate variables that are based on similarity of covariables between attributes to predict and observe samples. It has been used with success for different applications such as the national forest inventory of Finland (Tomppo *et al.* 2010). The case of the finish inventory is well adapted to the use of the KNN technique because of the numerous ground plots and a relatively homogeneous forest types. So far, several variants of the KNN were developed (Crookson *et al.* 2002; Ohmann and Gregory, 2002; Hudak *et al.* 2008; Bernier *et al.* 2010; Wilson *et al.* 2012) to increase its flexibility or powerfulness. For instance, the KMSN (K Most similar neighbours) (Moeur and Stage, 1995) was developed to be more flexible by adding a weight to the distance between reference and target observations. The two imputation techniques most used up to now are the traditional approach to the KNN and the KMSN variant which are both well adapted to

serve in the last stage of a two- or three-stage approach to map forest attributes over large area. The precision obtained from the imputation techniques is highly dependent on the number and quality of training samples (Franklin and Wulder, 2002; Tomppo *et al.* 2008). In the context of Canada's boreal forest, the number of available GSP is generally very low and they tend to be located in denser forest (density of 25% or more). In addition, forests with a stand height of less than 7 m are poorly and sometimes not sampled. Information from the SF-method based on HSRSI can fill the sampling gap for stands of all categories. Remains to test how the KNN procedures and their variants can be effective at producing reliable maps when using these alternatives input datasets.

The main objective of this study was to map stand height and merchantable volume at a regional scale for a large area in the boreal forest of northern Quebec in Canada considering of the limitation in the training dataset (GSP). To address this main objective, we aim more specifically to fill gaps in the GSP database by developing GSP selection methods to increase the representativeness of this training dataset. We also aim to compensate anomalies and gaps of the GSP database by generating surrogate plots from the use of the SF-method and HSRSI. Spatially-explicit generalisation of the height and volume values was imputed from the extended plot dataset using the KNN and KMSN techniques.

2. Study area and data

2.1. Study area description

The study area is located in the Boreal Shield Ecozone (Fig. 1), in northeastern Canada. It covers 470,000 km² between the 49th and 53rd parallels. Nine test sites corresponded to existing HSRSI coverage. The location of the test sites was selected to provide a representative sample for the range of stand height and merchantable volume and stand conditions that are typical of northeastern Canadian forests. Forest species in the study area consisted mainly of black spruce for about 95% of the stands, with the presence of a few other species such as jack pine (*Pinus banksiana* Lamb.) and balsam fir (*Abies balsamea* (L.) Mill.), which is typical of the region (Rowe, 1972). The understory of black spruce stands consisted of lichen, moss, and shrubs in varying proportions.

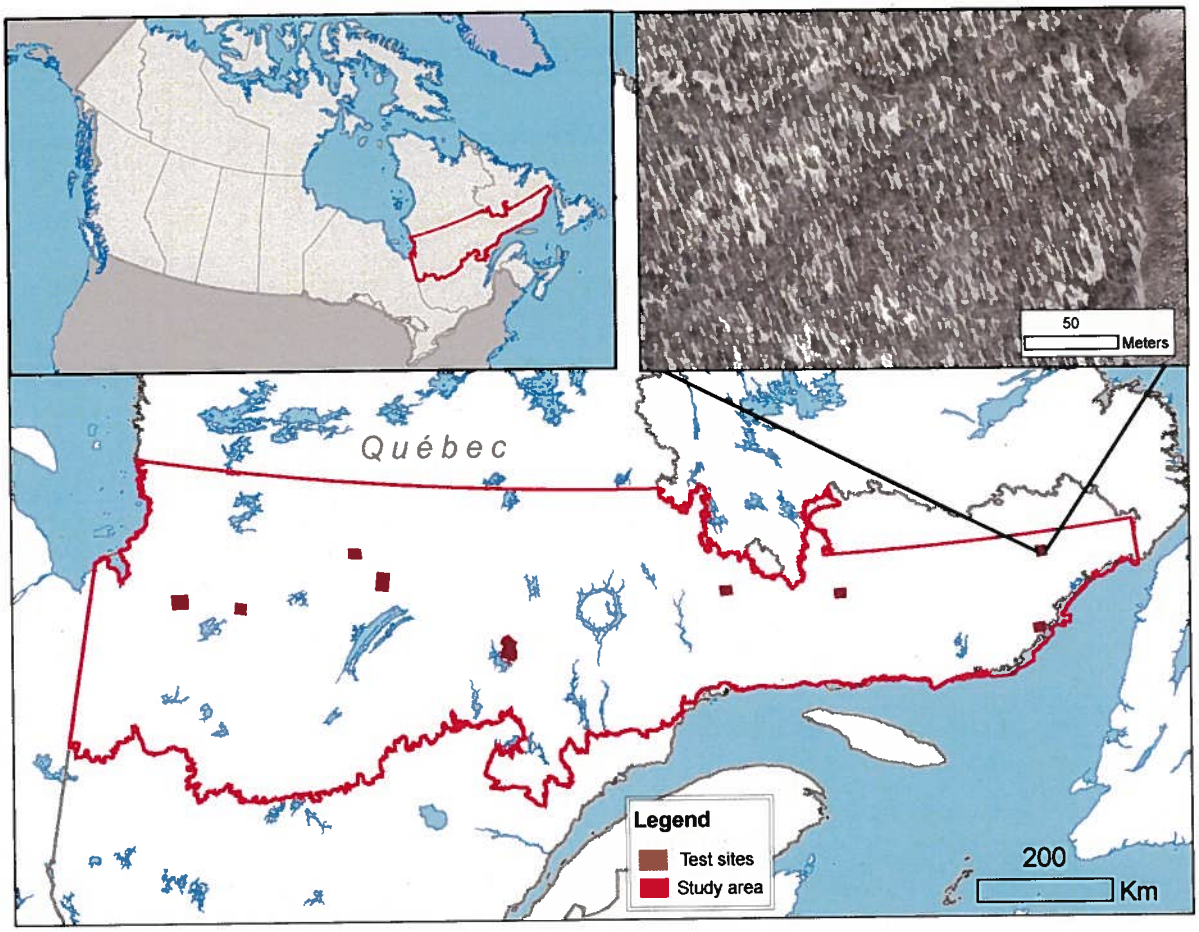


Figure 1 - Study area and test sites located in northeastern Canada.

2.2. Ground sample plots (GSP)

A total of 7,579 circular GSP each covering 400 m^2 (11.48 m radius) were established from 1996 to 2009 across the study area using procedures followed for the provincial forest inventory (MRNF, 2009b). The GSP were selected to represent the full range of stand height, merchantable volume and understory types for the black spruce stands. Species composition of the GSP were such that: (i) black spruce represented more than 50% of the GSP basal area and (ii) any other species such as trembling aspen (*Populus tremuloides* Michx.), white birch (*Betula papyrifera* Marsh.), *Alnus* spp., and *Salix* spp. represented less than 25% of the GSP basal area.

Several tree-level attributes were measured for each of the GSP (Table 1). DBH (diameter at breast height) was measured for all trees with a DBH > 9 cm in the GSP. In addition, both forest attributes (stand height and merchantable volume) were calculated using published regression

models where tree DBH was used as an independent variable (Table 1). Tree height was calculated from tree DBH using an existing model (Fortin *et al.* 2009) adapted to our study area using the 37,258 tree measurements (DBH and height). Stand tree height was estimated as the mean of the 4 trees with the highest DBH. Tree stem volume (m^3/ha) was calculated using a method developed by Fortin *et al.* (2007) also adapted to our dataset. A total of 3,912 sampled trees from 1,026 GSP were used to build the merchantable volume model. Finally, GSP merchantable volume was obtained by summing up the tree stem volumes of measured trees.

Table 1 - Forest attributes measured from GSP

Forest attributes	Calculation method	Attribute range
Merchantable volume	**Plot-level volume**: $\sum$Tree volume$_i$ *10,000/400 **Tree volume$_i$**: Fortin *et al.* (2007) adapted to the study area **Tree height$_i$**: Fortin *et al.* (2009) adapted to the study area	0.4 to 404 m^3/ha (average: 83.1 m^3/ha)
Stand height	Tree height mean of the 4 trees with highest DBH	5.9 to 30.9 m (average: 14.73 m)

2.3. HSRSI

Each of the 9 test sites was covered by HSRSI with a panchromatic band acquired from QuickBird or WorldView, both with a spectral window of 450-900 nm. The spatial resolution of one pixel was 0.6 m for QuickBird panchromatic images and 0.5 m for WorldView panchromatic images. The area covered by the images varied from 25 to 151 km^2 (Table 2), and the spectral resolution of all the images was 16 bits. All images were acquired from 2006 to 2009 during the growing season with: (i) clear skies (less than 10% cloud cover); (ii) viewing angles between 4.3 to 22.9 degrees; and (iii) solar zenith angles between 28.6 and 50 degrees. All images were orthorectified using a digital elevation model and high precision GPS data, resulting in a positional accuracy of better than 5 m. Then, unusable areas such as clouds, shaded hill faces, lake or ponds were manually masked out of the image.

Table 2 - Image acquisition parameters for each test site

	Date, UTC (dd/mm/yyyy, hh:mm)	Sun zenithal / azimuth (°)	Sensor zenithal / azimuth (°)	Coverage (km^2)
QuickBird	10/06/2006, 16:15	30.6/159.8	10.3/145	114
QuickBird	20/08/2006, 18:59	39.8/167.4	19.2/339.9	89
QuickBird	12/07/2008, 18:23	29.3/163.9	22.9/322.1	34
WorldView	11/07/2008, 16:10	31.6/150.4	14.5/132.5	151
WorldView	16/06/2009, 15:12	28.6/157.8	3.3/100.8	26
WorldView	17/09/2009, 15:28	50.0/166.9	18.6/49.2	25
WorldView	13/08/2009, 15:49	37.8/164.4	17.2/324.6	27
WorldView	28/06/2008, 16:12	30.0/158.2	19.4/292.9	36
WorldView	06/09/2009, 16:24	45.5/166.2	4.3/176.6	30

2.4. MSRSI

The study area was completely covered by a mosaic of MSRSI composed of 68 Landsat-TM images. Although the images were acquired from July 3, 2005, to October 18, 2009, most of the images were acquired during the growing season of 2009 from June to September. The mosaic was assembled using an histogram matching technique from Homer *et al.* (1997) implemented with the PCI Geomatics software (PCI Geomatics inc., 2003). This technique minimizes spectral differences between a reference satellite images and the "slave" images of the rest of the mosaic. This is done by looking at the distribution of brightness values (e.g., the shape of their histogram: maximum and spread) of all pairs of reference and slave images to bring them to a match. The zones with clouds or haze were masked out of the mosaic to avoid interference in the matching procedure. Three spectral bands were used: red (0.63 to 0.69 μm), which is sensitive to the presence of chlorophyll; near infrared (0.76 to 0.90 μm), which is sensitive to the structure of vegetation cover; and mid infrared (1.55 to 1.75 μm), which is sensitive to the presence of water in plants (Gausman, 1984). The choice of these three bands was justified by their potential for forest class differentiation (Meng *et al.*, 2009). The spatial and spectral resolution of these images was 30 m and 8 bits, respectively. All 68 images were orthorectified using a digital elevation model and hydrographic data, at a scale of 1:100,000, resulting in a positional accuracy of better than 30 m.

3. Method

Mapping the northern boreal forest of the study area was based on an approach in three steps using satellite remote sensing images. As the first step involve using measurements from GCP throughout the area to cover the full gradient of stand structure and conditions. However, considering the wide range of GSP and the variety of acquisition conditions we developed and tested different methods for selection of GSP. The GSP do not provide a complete coverage of all stand structure since they do not include any plot with stand height less than 7 m. Therefore the second step of the proposed approach involve generating surrogate or satellite plots (GSP_s) serving as a complement to the existing GSP database therefore increasing data representativeness. These GSP_s will be defined using the SF-method (Leboeuf *et al.* 2007, 2012; Luther *et al.* 2012) with the nine HSRSI. The test sites generated from these HSRSI were spread throughout the study area. The third and last step involved using the enhanced dataset (GSP + GSP_s) to train the mapping approach at the regional level, over the study area. This was accomplished with an imputation technique to map forest stand height and volume at a regional scale for the study area. The two candidate imputation techniques selected were the KNN (LeMay and Temesgen, 2005) and KMSN (Moeur and Stage, 1995).

3.1. Step 1: Selecting the representative GSP

Dataset representativeness can considerably affect KNN mapping results (Franklin and Wulder, 2002). Initial tests with the entire 7,579 GSP dataset showed that several GSP were weakly representative of the actual state of the forest. This low representativeness can be explained in large part by three factors: (i) forest disturbance (fire, insects, etc.) in the time gap between GSP measurement and image acquisition; (ii) a difference between GSP and image location; and (iii) site noise caused by minor flooding, for example, that changed the spectral values of pixels. These three factors can bias significantly the model because of the reduce correlation between image variables and the forest attribute measured. Two methods were tested to select the most representative GSP.

The first method of GSP selection was intended to eliminate GSP that were not representative of their forest attribute class. All GSP locations were known, therefore the spectral values (red, near

infrared and mean infrared) of the MSRSI pixel were linked with each GSP of the database. GSP were then classified into ten merchantable volume classes (0-5 m^3/ha; 5-15 m^3/ha, 15-25 m^3/ha, etc. and up to 395 to 405 m^3/ha) and the three mean spectral values (red, near infrared and mean infrared) were calculated for each merchantable volume classes. Merchantable volume was chosen as one of the two forest attributes since it integrates both stand height and DBH. The mean spectral value for each of the ten merchantable volume classes was calculated from inclusion of all the GSP corresponding to that class in the database. Then, the spectral distance between each GSP and the mean of its volume class was calculated and associated to the GSP. This spectral distance was calculated as the Euclidian sum of three standard deviations for each of the three selected bands. Several distance thresholds (0.5-, 1-, 1.5- and 2-time the standard deviation) between the spectral and the mean values were tested to eliminate the GSP with the highest spectral distance. The final threshold was selected to fulfill three conditions: (1) it had to make sure it would not exclude GSP belonging to the next height or volume class, i.e. the spectral values allowed in one class need to spread in their neighbouring classes, (2) more than 70% of all the GSP of a given class can be distributed in the center of the two neighbouring classes, and (3) produce the highest statistics over those selections that fulfils the two previous criteria. With such simple rules we know that it would be large enough to include a large variety of GSP considering the diversity of structure but it also limited GSP selection for those that may not be representative of that height or volume class for any of the bias factors. The remaining GSP from this method (GSP'_1) were used as training data to map forest attributes across the study area.

The second method of GSP selection was also intended to eliminate disturbed or unrepresentative GSP but following other rules than the first method. Unsupervised ISODATA classification was applied to each band of MSRSI to derive 40 spectral clusters. Some of these clusters were linked to five classes associated to disturbed sites: forest fire, insect defoliation, flooded area, windfall, and harvesting. These five classes were labelled based on training dataset established from polygons updates on the forest map available from MRN databases (MRN, 2011), where each disturbance was identified along with the date of occurrence. Once the disturbance classes were identified in the images, they were eliminated from the active area for potential GSP therefore removing all GSP on this masked area from future use in the imputation procedure. This disturbance mask removed the GSP from severely disturbed sites but several bias remains such as

co-registration errors, light disturbances or important changes of the stand structure. A supplemental procedure was thought to take in consideration these situation. The mean merchantable volume value and its standard deviation were calculated for each remaining clusters of the ISODATA procedure. Several thresholds values (0.5-, 1-, 1.5- and 2-times the standard deviation) were also tested to remove unsuitable GSP. The final threshold was selected to fulfill two conditions: (1) the threshold selects more than 70% of all the GSP of a given ISODATA class, and (2) produce the highest R^2, RMSE and bias among the thresholds values that fulfilled the previous criteria. Remaining GSP (GSP'_2), which were included within the range of the selected threshold, were used as training data to map forest attributes across the study area. The two methods of screening the GSP for disturbances were applied to assess which one provide the most suited results.

3.2. Step 2: GSP_S for stand height < 7m

The dataset covered all the range of stand volume (0.4 to 404 m^3/ha), however, no GSP covered a stand height below 5.9 m (Table 1) which created an important gap to fill in the database. Developing surrogate plots to fulfill this gap for stand with height < 7m is a way to overcome this important limitation for a complete dataset. As a result we used the models from the SF-method developed by Leboeuf *et al.* (2012) to map stand height and merchantable volume, over a series of test sites covered by HSRSI. These mapped areas were therefore useful to supply for an additional series of satellite plots (GSP_s) to compensate for the missing categories of stands in the original database, the stands with height < 7 m. Three steps were necessary to apply the SF-method: (i) calculating SF on a 30 m grid at each test site; (ii) applying the models published by Leboeuf *et al.* (2012) to obtain stand height and merchantable volume from the SF values; and (iii) selecting randomly a fix number of map cells to be defined as GSP_s in the enhanced database.

Firstly, the HSRSI was divided into a uniform grid with individual square cells with sides of 30 m. The size selected for the reference (900 m^2) square was based on the work of Leboeuf *et al.* (2007 and 2012) and it matched the spatial resolution of Landsat imagery. The SF was calculated for each of these cells from the percentage of shaded pixels over the whole area of the cell. Reformulating this statement we can say that SF was calculated as $SF = A_{TS}/A_r$ where A_{TS} is the

area of shaded pixels and A_r is the area of the reference square, 900 m². Therefore, all pixels of the HSRSI were assigned a binary value of 0 or 1 for non-shaded and shaded pixels, respectively. Defining a shaded pixel depends on a threshold value applied to the digital numbers of the HSRSI. The threshold values were defined visually by an analyst for each test site with two main considerations in mind: (i) to avoid overestimation of shaded areas and (ii) to ensure realistic proportions of shaded areas on varying understory types, such as those dominated by lichen and moss. Surprisingly, the work from Leboeuf *et al.* (2007) showed that the results are not overly sensitive to the threshold value which avoids subjective biases. From this first procedure we transformed the HSRSI into a 30 m grid with SF values.

Secondly, the SF from the 30 m grid could then be used as input of the models published by Leboeuf *et al.* (2012) to estimate stand height and merchantable volume. These models were however adapted by forcing them to the origin in order to detail the stands lower than 7 m. Also, SF values were normalized to a common geometry to account for differences in sensor viewing zenith and azimuth angles, as well as sun zenith and azimuth angles. The normalized SF values (SF_n) were calculated using trigonometric formulas that modified the shadow area projected on the ground or on the crowns as viewed by the sensor. Normalization was performed as per Leboeuf *et al.* (2007) and took into account two elements: a correction factor that considered the shadow length projected in the plane of the sensor viewing geometry and a correction factor that considered the increasing shadow length with increasing sun zenithal angle.

Lastly, a fix number of map grid cells were selected randomly from the HSRSI to serve as GSP_s in the enhanced database. Following the distribution of the 7,579 GSP, we observed that the lowest height classes 7, 8 and 9 m had respectively 18, 119 and 227 GSP, giving a mean of 121 GSP. Based on these statistics, we aimed to reach 100 GSP or GSP_s for each class less than 8 m. Thus, 779 GSP_s were selected randomly from all the available GSP_s from 0 to 7 m so that each 1 m class had 100 values from GSP or GSP_s. For example, 100 GSP_s were selected for each of the first height classes between 0 and 6 m, and 79 GSP_s from 6 to 7 m class representing a total of 779 GSP_s. Grid cells could easily be selected randomly because of their large number (373,770) and there was no attempt to assess spatial correlation. The resulting dataset comprising filtered

GSP and GSP_s covered the whole range of forest attribute values (stand height and merchantable volume) and was used as training dataset to the stage three for regional map of stand attribute.

3.3. Step 3: Regional map of stand attributes

Among the several imputations techniques, KNN were chosen because such non-parametric techniques allow simultaneous estimation of multiple variables by preserving the covariance structure among the variables (Bernier *et al.* 2010). Those techniques are largely used for different forest inventory contexts (Tuominen *et al.* 2003; Tomppo and Halme, 2004, LeMay and Temesgen, 2005; LeMay *et al.* 2008). The KNN models were trained using the extended dataset (GSP and GSP_s) of 8,358 plots or enhanced dataset (GSP'_1 or GSP'_2 ; and GSP_s) from selection methods described in section 3.1. Results obtained from all these datasets were compared in order to determine the impact of the selection methods. Mapping of the study area was based on well-documented techniques, KNN and k most similar neighbors (KMSN), that have largely demonstrated their effectiveness. Our analysis was geared towards defining what processing options would produce the most accurate maps of the study area. It was therefore necessary to compare results from two KNN techniques, KNN at a pixel scale (Equation 1) or KMSN at a polygon scale (Equation 2). It was also needed to assess the contribution of three ancillary mapping layers: spectral bands, physical data or climate data.

In the case of KNN technique, the mapping layers were selected based on the highest Pearson correlation values calculated between layers and GSP values. Thus, several elements, such as altitude, slope, drainage, precipitation, etc., were evaluated. Using the extended GSP dataset and the relevant mapping layers, the forest attribute value ($FAknn_i$) was calculated for each pixel of the MSRSI according with equation (1):

$$FAknn_i = \frac{\sum_{k=1}^{k} W_k FA_{GSP\_k}}{\sum_{k=1}^{k} W_k} \text{ for } k \neq 1 \text{ where } W_k = \frac{1}{d_k^j}, \qquad (1)$$

where $FAknn_i$ is the value of forest attribute (volume for example) calculated from the KNN technique for cell i; k is the k^{th} nearest spectral neighbor; FA_{GSP_K} is the forest attribute value from selected GSP and GSP_s; W_k is the weighting coefficient; d_k^j is the Euclidian distance to which is applied a power of j. The resulting $FAknn_i$ value is in fact a mean value of the FA_{GSP_K} values, which have the nearest spectral value of the pixel evaluated.

Application of the KMSN technique to map forest attribute was calibrated using only the extended dataset (GSP' + GSP_s) in order to compare the statistics obtained from the two non-parametric techniques (KNN or KMSN) applied respectively at a pixel- and polygon-scale. The two datasets obtained from the selection methods were not used when applying the KMSN technique since the choice of the selection method was done by KNN technique. The KMSN algorithm available allowed to apply the classification only at a polygon-scale (Moeur and Stage, 1995; Crookson *et al.* 2002). To implement the KMSN technique, we used the forest polygonal map from MRNQ on which we associated topographic and climatic data. For, example, we calculated for each polygon, the mean altitude, mean slope, mean aspect, mean temperature for the year, mean precipitation, etc. Also, we calculated a mean value of each spectral band of the MSRSI for each polygon. To do so, three methodological steps were necessary: (i) converting forest classes to numeric classes, for example density classes to numeric values; (ii) calculating the mean value within each polygon for image spectral bands, altitude, climatic map, etc., and (iii) testing several combinations from the available spatial layers. The forest attribute value ($FAkmsn_i$) was calculated for each forest polygon using this equation (2):

$$FAkmsn_i = reference\ observation \quad j\ with\ d_{ij}^2 \quad minimum \qquad (2)$$

$$d_{ij}^2 = (X_i - X_j)W(X_i - X_j)\ for\ all\ j = 1,...,n,$$

where $FAkmsn_i$ is the forest attribute value of the nearest neighbor of the i^{th} target observation and the j^{th} observation among n reference observations; j is the reference value (among all the reference values) that minimizes the Euclidian weighted distance; X_i is the vector of the

normalized X-variables for the i^{th} target observation; X_j is the vector of the normalized X-variables for the j^{th} target observation, and W is a weight matrix.

For the two non-parametric techniques, three statistics were calculated to evaluate the results: R^2, relative Root Mean Square Error ($RMSE_r$), and relative bias ($bias_r$).

$$RMSE_r = \sqrt{\frac{\sum_{i=1}^{i=N}\left[\left(FA_i - FA_{GSP\_i}\right)\right]^2}{\sum_{i=1}^{i=N}\left[FA_{GSP\_i}\right]^2}}, \qquad (2)$$

$$Bias = \sum_{i=1}^{i=N}\left[FA_i - FA_{GSP\_i}\right], \qquad (3)$$

where the FA_{GSP_i} is the Forest Attribute or FA (height or volume) measured from the GSP, FA_i is the FA calculated from KNN technique ($FAknn_i$ or $FAkmsn_i$), and N is the number of observations. The samples were selected using the "leave one out" approach.

4. Results and their interpretation

4.1. Step 1: Selecting the representative GSP

Two GSP selection methods were tested: the first method was intended to eliminate GSP that were not representative of their forest attribute class and the second method was intended to eliminate disturbed or unrepresentative GSP. For each selection method, statistics obtained using several threshold were compared based on fraction of the standard deviation values (0.5-, 1-, 1.5- and 2- times the standard deviation). For the two selection methods (1^{st} and 2^{nd} method), only the 0.5 and 1 time the standard deviation fulfilled the second condition established that more than 70% of all the GSP of a given class can be distributed in ISODATA classes. Also, in the two selection methods, statistics (R^2, RMSE and bias) were more favourable for 1 time of the standard deviation.

Table 3 presents the statistics (mean, range, standard deviation) of stand height and merchantable volume before and after the application of the two selection methods. The first selection method reduced the stand height mean values from 14.7 to 11.6 m, increased the merchantable volume mean values from 83.1 to 111.4 m³/ha and reduced the number of GSP available for the stage three process from 7,579 to 6,578 GSP. The second selection method maintained the statistics. The stand height mean values were stable with 14.7 and 14.6 m. The merchantable volume mean values decreased also slightly from 83.1 to 76 m³/ha and the number of GSP was reduced from 7,579 to 4,577 GSP. For all the dataset, the date of measurement was not modified using any selection methods with respectively a minimum, mean and maximum values of 1996, 2001 and 2009.

Table 3 - Statistics for the available GSP and GSP_s

Dataset (Number of plots)	Stand height (m) : min / mean / max	Merchantable volume (m³/ha): min / mean / max
GSP (7,579)	5.9 / 14.7 / 30.9	0.4 / 83.1 / 404.4
First selection method (GSP'_1) (6,578)	8.1 / 11.6 / 17.0	2.2 / 111.4 / 319.5
Second selection method (GSP'_2) (4,577)	6.6 / 14.6 / 25.7	0.4 / 76.0 / 339.3
GSP + GSP_s (8,358)	0 / 13.7 / 30.9	0.24 / 75.9 / 404.4

4.2. Step 2: GSP_S for stand height < 7m

Remaining GSP, which were included within the optimal threshold, were used as training data and were added to the 779 randomly selected GSP_s. Table 4 presents the minimum, mean and maximum values of stand height and merchantable volume for the GSP_s. These statistics demonstrate that the selected GSP_s respected the distribution of forest attribute values. Table 3 presents also that the resulting dataset which comprise GSP and GSP_s did not affect the stand height and merchantable volume variability. The size of the dataset was increased from 7,579 to 8,358 with the input of the GSP_s. The stand height and merchantable volume mean values decreased respectively from 14.7 to 13.7 m and from 83.1 to 75.9 m³/ha when we added the GSP_s to the training dataset.

Table 4 - Statistics for the 779 randomly selected GSP_s

Stand height class	Number of cells	Stand height (m) : min / mean / max	Merchantable volume (m^3/ha): min / mean / max
0 to 0.5 m	100	0.0 / 0.0 / 0.4	0.2 / 0.3 / 1.6
0.51 to 1.5 m	100	0.8 / 1.0 / 1.4	0.4 / 1.0 / 2.2
1.51 to 2.5 m	100	1.7 / 2.1 / 2.4	2.4 / 2.7 / 3.0
2.51 to 3.5 m	100	2.6 / 3.1 / 3.5	3.2 / 3.7 / 4.1
3.51 to 4.5 m	100	3.6 / 4.1 / 4.5	4.3 / 5.1 / 5.7
4.51 to 5.5 m	100	4.6 / 5.0 / 5.4	5.9 / 6.8 / 7.9
5.51 to 6.5 m	97	5.5 / 6.0 / 0.4	8.0 / 9.6 / 11.0
6.51 to 7.5 m	82	6.5 / 6.8 / 0.4	11.2 / 12.2 / 13.1

4.3. Step 3: Regional map of stand attributes

Table 5 summarizes the statistics obtained for pixel-scale KNN according to the two methods for GSP screening defined in section 3.1 (GSP'_1 + GSP_s or GSP'_2 + GSP_s) and the use of the different spatial layers. Pearson correlation results between available layers and GSP suggested seven relevant layers: Landsat red, near infrared and mid infrared bands, slope, degree-day, temperature in Celcius, and radiation during the growing season (MJ/m^2). We first compared two sets of layers: (1) three spectral bands (R, NIR, MIR) only, and (2) the three spectral bands with physical and climatological layers (R, NIR, MIR, SL, DD, T, R). We assessed variable contribution according to the dataset used. Using the GSP + GSP_s similar statistics were observed for stand height (R^2: 38 to 48; $RMSE_r$: 24 to 27; $Bias_r$: -0.09 to -0.06 m) and for merchantable volume (R^2: 37 to 36; $RMSE_r$: 58 in both cases; $Bias_r$: -1.15 to -1.08 m^3/ha) respectively for the set of layers (1) and (2). The selection method of the training dataset (1st and 2nd method) had a greater impact on the statistics than the input of the physical and climatological layers. For example, R^2 increased from 38% to 63% (1st method) for stand height and from 37% to 47% (2nd method) for volume. Finally, comparison between all selection methods and all layer combinations indicated that the 1st method associated with spectral bands and physical and climatological layers presented most favourable statistics for stand height with 63%, 21% and 0 m^3/ha for R^2, $RMSE_r$ and $Bias_r$, respectively. The 2nd method associated with spectral bands presented most favourable statistics for merchantable volume with 47%, 37% and -0.36 m^3/ha for R^2, $RMSE_r$ and $Bias_r$, respectively. The parameters k=5 and t=2 were determined by comparing

the $RMSE_r$ produced for k and t ranging respectively from 3 to 30 and t from 0.5 to 2 using GSP + GSP_s dataset.

Table 6 describes the statistics for polygon-scale KMSN. The results obtained are comparable but the best results were obtained using the pixel-scale KNN technique presented in Table 5. For stand height, R^2, $RMSE_r$ and $Bias_r$ were 53%, 18% and -0.01 m, respectively. For volume, R^2, $RMSE_r$ and $Bias_r$ were 45%, 50% and -0.05 m^3/ha, respectively. KMSN parameters, k=20 and t=0 were also determined by comparing the $RMSE_r$ produced for k and t ranging respectively from 3 to 30 and t from 0.5 to 2 using GSP + GSP_s dataset.

Finally, the map in Figure 2 shows merchantable volume of the study area using the pixel-scale KNN technique. We observed that expected ecological patterns are followed by this map with higher forest attributes values on richer and denser stands. Also, it follows the volume tendencied that we expect by interpreting the satellite images. Drier sites or those that burnt few year ago show low forest stand volume values.

Table 5 - Error statistics (R^2, $RMSE_r$ and $bias_r$) obtained using the pixel-scale KNN technique with various input datasets.

Forest Attribute (FA)	Selection method	Base layers applied to the KNN	k	t	R^2 (%)	RMSE (%)	Bias
Stand height (m)	GSP + GSP_s	R, NIR, MIR	5	2	38	27	-0.09 m
		R, NIR, MIR, SL, DD, T, R	5	2	48	24	-0.06 m
	GSP'_1 + GSP_s	R, NIR, MIR	5	2	48	27	-0.03 m
		R, NIR, MIR, SL, DD, T, R	**5**	**2**	**63**	**21**	**0 m**
	GSP'_2 + GSP_s	R, NIR, MIR	5	2	46	28	-0.03 m
		R, NIR, MIR, SL, DD, T, R	5	2	59	24	-0.05 m
Merchan-table volume (m^3/ha)	GSP + GSP_s	R, NIR, MIR	5	2	37	58	-1.15 m^3/ha
		R, NIR, MIR, SL, DD, T, R	5	2	36	58	-1.08 m^3/ha
	GSP'_1 + GSP_s	R, NIR, MIR	5	2	41	55	-0.38 m^3/ha
		R, NIR, MIR, SL, DD, T, R	5	2	49	50	-0.42 m^3/ha
	GSP'_2 + GSP_s	R, NIR, MIR	**5**	**2**	**47**	**37**	**-0.36 m^3/ha**
		R, NIR, MIR, SL, DD, T, R	5	2	45	38	-0.74 m^3/ha

N.B.: GSP + GSP_s (8,358 plots); GSP'_1 + GSP_s: Euclidian method with 1 time the standard deviation (7,357 plots); GSP'_2 + GSP_s: ISODATA filtering (5,356 plots); R: Landsat red band; NIR: Landsat near infrared band; MIR: Landsat mid infrared band; SL: slope in degrees; DD: degree-day; T: temperature in Celcius; R: radiation during the growing season (MJ/m^2).

Table 6 - Statistics (R^2, $RMSE_r$ and $bias_r$) obtained using the polygon-scale KMSN method following different methodological approaches.

Forest Attribute (FA)	Selection method	Base layers applied to the KNN	k	t	R^2 (%)	RMSE (%)	Bias
Stand height (m)	GSP + GSP_s	R, NIR, MIR, Lat, Long, Alt, SL, TP, NP, MT, CP, DP, SLcos, SLsin	20	0	53	18	-0.01 m
Merchan -table volume (m^3/ha)	GSP + GSP_s	R, NIR, MIR, Lat, Long, Alt, SL, TP, NP, MT, CP, DP, SLcos, SLsin	20	0	45	50	-0.04 m^3/ha

N.B.: GSP + GSP_s (8358 plots); R: Landsat red band; NIR: Landsat near infrared band; MIR: Landsat mid infrared band; Lat: Latitude; Long: Longitude; Alt: Altitude in metres; SL: slope in degrees; TP: total precipitation, NP: net precipitation; MT: mean temperature in Celcius; CP: coniferous proportion; DP: density percentage; SLcos: Slope * cos(aspect); SLsin: Slope * sin(aspect).

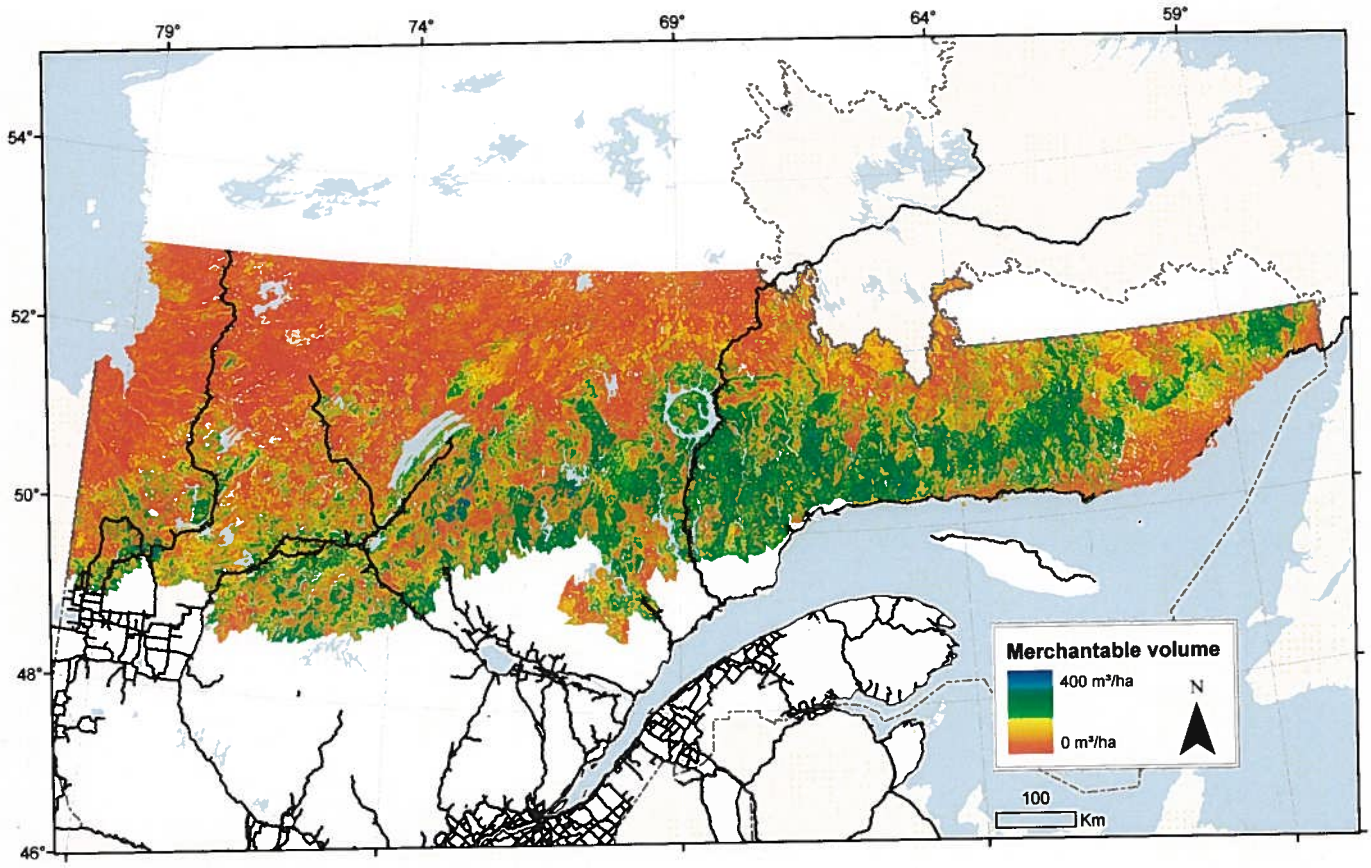


Figure 2 - Merchantable volume over the study area using the pixel-scale KNN technique.

5. Discussion

Two main challenges had to be considered to produce a forest attribute map of a huge boreal forest area in Quebec. First, the lack of recent and representative training datasets (GSP) influenced significantly the accuracy of the results (Franklin and Wulder, 2002). Second, the work on a large area and its remoteness inhibited the use of traditional forest inventory techniques (photo-interpretation associates with field works). To overcome these two challenges, we had to generate a reliable and representative dataset and chose the most suitable remote sensing based techniques to map forest attributes over the large area.

To overcome the first challenge, related to the representativeness of the dataset, we had first to answer the issue of time interval between GSP measurements and acquisition time of remote

sensing images. Also, associated with the first challenge, we had to answer the lack of representativeness of the dataset since measured GSP covered only stand height from 5.9 m to 30.9 m. The temporal gap was answered by developing and comparing two GSP selection methods. Application of these methods modified lightly the minimum, maximum and mean forest attribute values and then maintained the values distribution. The representativeness issue was answered by the application of the SF-method from Leboeuf *et al.* (2012) over 9 test sites. GSP_s selected from forest attribute raster map filled the gap of lower forest attributes and increase largely the representativeness of the dataset. Again, the data distribution among GSP_s was maintained. At this point, the enhanced dataset produced (GSP' + GSP_s) contained few disturbed or biased GSP compared with the original dataset (GSP). This enhanced dataset (GSP' + GSP_s) also respects the initial distribution of GSP and fills the gap of lower values (less than 7 m) with the input of GSP_s.

The second challenge related to the most convenient forest attribute mapping process was overcome by evaluating (i) the GSP selection method, (ii) the non parametric techniques (KNN or KMSN), (iii) the basis mapping layers and (iv) technical parameters (k and t). The results of the different iterations tested indicated the best options for mapping stand height and merchantable volume across the study area. First, for stand height, results indicated that spectral bands associated with physical and climatological layers (SL, DD, T and R), as base layers, improved statistical results. The dataset selection method also had a significant impact on the results as shown by the 1st method (GSP'1 + GSPs) with best results of R^2=63%, RMSE=21% and a bias=0 m. The favorable impacts of physical and climatological layers are probably due to these layers being closely correlated to stand height and thus influencing site productivity. The estimated height values cover a range from 0 to 30.9 m which is encouraging since such variability was expected. Second, for merchantable volume, the best option was obtained by using only spectral layers (R, NIR and MIR) and the 2nd selection (GSP'2 + GSPs) method resulting in R^2=47%, RMSE=37% and a bias=-0.36 m. For volume, the addition of physical and climatological layers produced very little impact. Also, the better statistics produced by the 2nd selection method (GSP'2 + GSPs) can be explained by the fact that it excluded disturbed plots from the dataset. In many cases, stand height in partially disturbed plots was not affected by the elimination of some trees of the plot but merchantable volume was. Finally, based on GSP' +

GSPs dataset, statistics obtained from KMSN were slightly improved but comparable to those from KNN. However the KMSN technique impute an attribute value to the polygon, which is a limit since no intra-polygonal analysis is then possible. Thus, the pixel-scale KNN technique presents several advantages. The first advantage is that the technique is independent of the photo-interpreted map. Therefore, mistakes by or variations between interpreters are not reflected on the merchantable volume or stand height map. The second advantage is that the pixel-scale KNN technique can also detail stand structure by way of a raster forest attribute map and intra-polygonal analysis can be used to calculate, for example, the standard deviation of merchantable volume within the polygon. This information can be very useful for forest managers and ecologists to, first, determine the uncertainty level of the area in terms of merchantable volume and, second, determine the fragility rate of the site in terms of forest structure. The final map of stand height and merchantable volume produced, as presented in Figure 2, gives useful information at a landscape scale about the forest attributes. Also, forest attribute gradients followed expected localization and ecological patterns. In summary, the methods developed in this study to map forest attributes enable the use of old and not well distributed dataset and is a step forward toward a more extensive forest inventory of large areas of Canada's northeastern forest using remote sensing techniques. Also, statistics obtained were comparable when compared with similar works done for National Forest Inventory (Rosengren *et al.*, 1999; Katila, 2006; Tomppo *et al.*, 2008).

Although a sizeable area can be covered efficiently, some limitations inherent to the methods were observed. First, deciduous stands did not cover a large proportion of the study area and the shadow fraction method did not consider such stands (Leboeuf *et al.* 2007). This situation results in underrepresentation of deciduous stands in training datasets and can introduce bias on the resulting forest attribute maps. A second limitation is that the quality of the resulting map depends on the quality and homogeneity of spectral values of the MSRSI mosaic and thus depends on the mosaic assembly work. A last limitation due to averaging nature of non-parametric techniques is the introduction of biais in the extreme forest attribute values (McRoberts *et al.* 2002). For example, the low volume stand were generally overestimated and the high volume stand are generally underestimated. In spite of the above limitations, we demonstrated through the results obtained that the method tested in this study can be applied over

extensive areas of northeastern Canada with a high precision rate at a fraction of the cost of traditional methods which require field measurements, aerial photograph acquisition and interpretation.

6. Conclusion

This study demonstrated, using several validation processes, an efficient method for mapping forest attributes at landscape level using GSP and HSRSI. The proposed approach allows to use GSP measured more than 10 years ago and generate surrogate GSP_s from HSRSI and SF-models. Also, two main approaches, pixel-scale KNN and polygon-scale KMSN were tested and produced similar results. However, since pixel-scale KNN details the intra-polygon pattern, this approach was selected. This method now provides a practical and cost effective means of mapping two widely used forest attributes in raster format and can be applied over sizeable areas of Canada's northeastern boreal forest.

7. Acknowledgments

We are grateful to André Robitaille, Robert Jobidon, Jean-Gabriel Élie and Serge Gagné from Québec's Ministère des Ressources Naturelles for the dataset and support in the project. We would also like to thank Luc Guindon and André Beaudoin of the Canadian Forest Service for their involvement in the project and making the KNN algorithms available.

8. Discussion

Les méthodes développées à travers ces trois articles scientifiques ont démontré leur fort potentiel pour la cartographie d'attributs forestiers en milieu de forêt boréale. Ces méthodes sont supportées par une approche statistique rigoureuse mais sont néanmoins contraintes à des limites méthodologiques pour leur application à différentes situations géographiques ou différents jeux de données. Pour mettre en lumière les forces et limites de ces méthodes, chacun des trois projets de recherche (sections 5, 6, 7) sera détaillé pour soulever les forces, les faiblesses et les améliorations qui pourraient être réalisées. Cette description permettra de connaître différentes avenues de recherche qui pourraient être réalisées afin de poursuivre les avancées scientifiques de ces axes de recherche.

8.1. L'extension de la méthode par fraction d'ombre à quatre attributs forestiers

La section 5 (Leboeuf *et al.*, 2012) détaille l'utilisation de la fraction d'ombre par images satellite à très haute résolution spatiale pour effectuer la cartographie d'attributs forestiers. Les résultats obtenus ont démontré l'efficacité de la méthode pour les milieux boréaux dominés par l'épinette noire. À l'instar de ce projet, Greenberg *et al.* (2005), Leboeuf *et al.* (2007) et Luther *et al.* (2012) présentent le potentiel de cette méthode pour son application à d'autres attributs forestiers comme le diamètre à hauteur de poitrine et la biomasse forestière ainsi qu'à d'autres types de peuplements forestiers comme les sapinières. Par exemple, les modèles développés à travers la section 5, principalement pour l'épinette noire, ont généré des résultats statistiques assez élevés avec un R^2 entre 0,57 et 0,78 ainsi que des RMSE relatifs entre 0,17 et 0,32. Ces modèles permettent donc de cartographier quatre attributs forestiers largement utilisés en foresterie : la hauteur de peuplement, la densité, la surface terrière et le volume marchand. Les modèles de Luther *et al.* (2012) appliqués à Terre-Neuve et au Labrador pour le sapin baumier sont également appuyés par des résultats statistiques significatifs avec des R^2 se situant entre 0,15 et 0,81, des RMSE relatifs se situant entre 0,06 et 0,43. Ces modèles, associés au sapin baumier, (Luther *et al.*, 2012) complètent donc ceux d'épinette noire (Section 5 : Leboeuf *et al.*, 2012). Enfin, Greenberg *et al.* (2005) ont aussi démontré que la méthode s'applique pour le calcul des

diamètres à hauteur de poitrine, à la densité de couvert et à la biomasse du peuplement, et ce, pour des territoires forestier résineux de la Californie.

Néanmoins, jusqu'à maintenant en forêt boréale, la méthode de fraction d'ombre s'applique aux essences résineuses d'épinettes noires, d'épinettes blanches et de sapin baumier. La méthode n'a donc pas été démontrée sur des peuplements de pin gris, essence commune en milieu boréal. Notre compréhension du fonctionnement de la fraction d'ombre nous amène à penser que son application sur des peuplements forestiers de pin gris devrait être plus difficile vu la forme de cime plus arrondie et translucide chez cette essence par rapport aux épinettes et aux sapins. En effet, le feuillage du pin gris est plus clairsemé et laisse donc passer davantage de lumière que les épinettes et le sapin baumier. Dans ce même sens, la végétation de sous-bois comme le lichen peut apporter du bruit à la valeur spectrale de l'image. Les cimes plus coniques des épinettes et sapins semblent permettre aux capteurs optiques de bien détecter l'ombre au sol et les branches ombragées le long de l'arbre. Aussi, contrairement au pin gris, les cimes coniques sont généralement assez garnies en terme de feuillage ce qui minimise le retour spectral du sous-bois. De plus, certains éléments restent à tester afin de bien cerner les limites de l'application de la méthode. Premièrement, plusieurs nouveaux capteurs ont fait leur apparition au cours des dernières années (WorldView 2, GeoEye, etc.) Il serait intéressant de voir si l'approche fonctionne bien avec les différentes caractéristiques inhérentes à chacun des capteurs (résolution spatiale, spectrale, etc.). Leboeuf *et al.* (2007) ont démontré son application sur des images QuickBird à 0,6 m et Ikonos à 1 m. Néanmoins, la démonstration reste à faire pour des capteurs de résolution plus grossière de 1 m. Il est à prévoir que les résultats pour de tels capteurs devrait se dégrader rapidement puisque l'ombrage deviendra plus difficile à classifier. En effet, le pixel sera bruité par des cimes illuminées et par la végétation de sous-bois.

Deuxièmement, la méthode de fraction d'ombre a démontré son efficacité pour les peuplements résineux constitués d'épinettes ou de sapins baumiers présentant les caractéristiques suivantes : hauteur plus petite que 19 m, densité de 5 à 85 %, volume entre 2 et 200 m^3/ha, etc. Les peuplements ayant une valeur d'attributs forestiers plus élevés restent à tester afin de cerner la fenêtre d'application de la méthode. Dans ce cas-ci, nous supposons que le modèle devrait se

saturer après un certain niveau d'attribut forestier car logiquement un couvert trop fermé présente moins d'ombre et de cimes ombragées qu'un peuplement plus ouvert. Toutefois, la fenêtre d'application démontrée jusqu'à maintenant permet la cartographie de la très grande majorité de la forêt boréale où seulement 0,35% des placettes des domaines bioclimatiques de la pessière à mousse et de la pessière à lichen (Saucier *et al.*, 1998) ont des attributs forestiers plus élevés que ceux déjà testés.

Troisièmement, les outils de normalisation de l'ombre développés par Leboeuf *et al.* (2007) ne tiennent pas en compte l'effet de la topographie sur la fraction d'ombre. Néanmoins, l'effet de la topographie est visible sur les images où les fortes pentes qui sont exposées au nord accroissent les surfaces d'ombre tandis que les fortes pentes exposées au sud réduisent les surfaces d'ombre. Une approche théorique comme celle de Leboeuf *et al.* (2007) pourrait être améliorée en tenant compte du facteur de la topographie. Un modèle de normalisation empirique pourrait également être développé afin d'ajuster les effets des angles de vue du satellite et des angles solaires. Un tel modèle nécessiterait néanmoins un grand nombre de points de contrôle ou une superposition de plusieurs images satellitaires afin d'obtenir pour un même site plusieurs angles de vue différents.

Quatrièmement, les angles de vue testés jusqu'à maintenant varient entre 7,2 et 13,5 degrés (Luther *et al.*, 2012 ou Leboeuf *et al.*, 2012). L'angle maximal au-delà duquel le modèle de fraction d'ombre se dégraderait pourrait également être défini. Il est néanmoins à noter que l'approche de fraction d'ombre est appliquée aux images satellitaires, ce qui réduit les effets d'angles sur l'image puisque ces images sont acquises à très haute altitude en comparaison aux photographies aériennes ce qui réduit donc les déformations de l'ombre en bordure de l'image.

Cinquièmement, les modèles testés jusqu'à maintenant pour lier la fraction d'ombre à des attributs forestiers étaient linéaires ou polynomiaux. Cependant, la comparaison entre les modèles linéaires ou polynomiaux et une approche non paramétrique comme KNN ou KMSN n'a pas été faite jusqu'à maintenant. L'avantage de cette approche est que plusieurs couches d'information peuvent être utilisées pour effectuer la cartographie des attributs forestiers et que plusieurs attributs forestiers pourraient être testés en même temps. Contrairement aux modèles linéaires

développés à travers cette section 5, le désavantage des approches non paramétriques est que le résultat est difficilement applicable à d'autres secteurs. Le travail de cartographie doit donc être refait en entier pour chaque nouveau secteur à cartographier.

Sixièmement, l'utilisation d'images satellitaires d'hiver pour l'application de la méthode de fraction d'ombre pourrait accroître les résultats statistiques des modèles de régression. La végétation de sous-bois varie grandement entre différents milieux forestiers ce qui influence le retour spectral au satellite de l'ombre au sol et de l'ombrage des branches. Par exemple, un couvert de lichen va faire pâlir l'ombre et l'ombrage retourné au satellite par rapport à une végétation de sous-bois de mousse. L'utilisation d'une image ayant un couvert de neige au sol permettrait d'uniformiser le retour spectral de tous les types de sous-bois. Néanmoins, avec notre expérience avec ce type d'image, il serait préférable que la date d'acquisition de celle-ci soit le plus tard possible avant la fonte de la neige afin d'avoir un angle zénithal solaire le plus petit possible et aussi éliminer le problème potentiel de cimes enneigées. Ceci permettra de minimiser le biais causé par les pentes fortes et les pentes exposées au nord.

Septièmement, jusqu'à maintenant les méthodes de fraction d'ombre impliquent un seuillage manuel basé sur une interprétation de l'image. Cette opération pourrait être automatisée afin de stabiliser le seuillage entre les différentes images satellitaires. Deux approches sont possibles. La première nécessite une bonne quantité de placettes de calibration couvrant l'image. Ces placettes permettraient de connaître quel est le seuil qui maximise les statistiques (R^2, RMSE, etc.). Une seconde, basée sur les valeurs spectrales de l'image, pourrait aussi être développée. Un filtre de détection des hautes variances spectrales « edges » pourrait être appliqué. Ainsi, un seuillage automatisé serait appliqué sur l'image panchromatique en partant de la valeur numérique de 0 en montant afin de minimiser les zones d'ombres sous ces « edges ». L'avantage d'une telle approche est qu'elle devient indépendante des placettes de calibration terrain qui sont parfois difficiles ou impossibles à obtenir.

Malgré les limites soulevées et les voies de recherche pour améliorer la méthode, celle-ci est tout de même applicable aux forêts dominées par les épinettes et le sapin. Or, ces deux essences

représentent près de 75% du territoire forestier québécois et la grande majorité du territoire de la forêt boréale. Les zones d'application sont donc très grandes. De plus, l'approche a démontré qu'elle reste stable sur une grande fenêtre d'attributs forestiers (ex. 0 à 200 m³/ha). Néanmoins, une application de la méthode sur de grands territoires, nécessiterait l'acquisition d'une grande quantité d'images satellitaires dont le coût reste élevé. C'est pourquoi la méthode de fraction d'ombre a été testée également sur des photographies aériennes, qui couvrent déjà la presque totalité du territoire par différents programmes d'acquisitions. De plus, les photographies aériennes apportent des avantages méthodologiques et opérationnels puisqu'elles permettent une restitution 3D et que les attributs forestiers calculés par fraction d'ombre peuvent aisément être associés aux cartes forestières photo-interprétées.

8.2. Application de la méthode de fraction d'ombre à la photographie aérienne

La section 6 qui décrit l'approche de fraction d'ombre associée aux photographies aériennes démontre que cette méthode demeure applicable mais en tenant compte des limites soulevées à travers ce chapitre. Le fait que les résultats ont été démontrés significatifs pour la cartographie d'attributs forestiers ouvre la voie à son utilisation à de larges étendues vue la très grande disponibilité de ces images sur le territoire boréal. Le chapitre cerne néanmoins quelques limites méthodologiques qui doivent être prises en considération pour une application sur de grands territoires. Une fenêtre d'application est décrite à travers ce projet de recherche (distance par rapport au centre et angles optimaux), ce qui permet l'utilisation immédiate de la méthode. Par contre, différents travaux pourraient être réalisés pour faire avancer l'opérationnalité de ce projet.

Premièrement, pour minimiser l'effet de la bordure sur les valeurs de fraction d'ombre, un modèle pourrait être conçu à partir d'une superposition de plusieurs photographies aériennes acquises à partir de lignes de vol décalées les unes par rapport aux autres. Cette façon de faire permettrait de concevoir une normalisation de la fraction d'ombre en comparant les valeurs d'une cellule localisée à différentes distances sur d'autres photographies aériennes voisines. La même démarche pourrait ainsi être réalisée sur la base de l'angle azimutal et zénithal. Cette normalisation, si elle s'avère efficace, permettrait de faire une cartographie complète pour l'ensemble d'un grand territoire des attributs forestiers car les biais causés par les bordures de

photographies aériennes seraient réduits. Dans un même ordre d'idée, un modèle trigonométrique de normalisation de la fraction d'ombre comme celui proposé par Leboeuf *et al.* (2007) pourrait être ajusté aux photographies aériennes mais en y ajoutant des variables de topographie. Par contre, comme démontré à travers la section 6, les variables d'angles de la prise de vue et d'angles solaires n'affectaient pas les résultats statistiques de façon notable. Ainsi, une révision de la base du modèle (autre fondement que trigonométrique) serait peut-être pertinente. Enfin, un modèle empirique de normalisation de la fraction d'ombre pourrait être conçu sur la base de placettes terrain. En effet, si un grand nombre de placettes terrain étaient disponibles, une correction des valeurs de fraction d'ombre pourrait être dérivée d'un modèle de régression. Concrètement, il s'agirait de sélectionner les valeurs d'un attribut forestier à l'intérieur d'une fenêtre donnée, par exemple, entre 100 et 110 m^3/ha et de lier la distance du centre de la photographie aérienne sur la fraction d'ombre pour ces placettes sélectionnées. Le même exercice pourrait, encore là, être fait pour les angles azimutaux et zénithaux. Un tel exercice pourrait donc se faire sur des sites de recherche comme la Forêt Montmorency (forêt de recherche de l'Université Laval au nord de Québec) où il y existe plusieurs centaines de placettes terrain.

Deuxièmement, comme il a été démontré dans la section 6, une zone d'application est optimale entre 0 et 1 000 m du centre de la photographie aérienne. Pour permettre de maintenir une zone optimale continue entre les photographies aériennes, il faudrait réduire la distance entre les lignes de vol afin d'augmenter le recouvrement à 40% alors qu'il est généralement établis à 30%. Dans ce même ordre d'idée, il serait intéressant de tester différents types de caméras comme les caméras à balayage (pushbroom) d'ADS qui sont largement utilisées pour des projets de photo-interprétation comme en Ontario. Ces caméras devraient logiquement réduire les effets de bordure à l'intérieur de la ligne de vol bien que ces effets devraient demeurer présents entre les lignes de vol si le recouvrement entre ces lignes demeure à 30%.

Troisièmement, comme ce chapitre se base sur les photographies aériennes, plusieurs avantages technologiques intrinsèques aux photographies aériennes peuvent améliorer les résultats statistiques. En effet, plusieurs méthodes de corrélations de photographies aériennes existent (Gruber et Schneider, 2007; St-Onge et al, 2008). Ces méthodes permettent d'obtenir un modèle

numérique d'élévation de canopée où chaque pixel de ce modèle représente une hauteur absolue de la canopée. Différentes statistiques, telles que la variance et la texture de la canopée, peuvent s'ajouter à la fraction d'ombre pour la mesure d'attributs forestiers.

Quatrièmement, comme mentionné précédemment, des méthodes non paramétriques comme KNN permettraient d'utiliser plusieurs couches d'information comme la fraction d'ombre, les valeurs spectrales des photographies aériennes et des variables de modèle numérique d'élévation de canopée. L'avantage de ces méthodes est qu'elles peuvent gérer plusieurs couches d'informations. De plus, certaines de ces méthodes permettent de donner un poids supplémentaire aux couches qui apportent davantage à la précision de la cartographie d'attributs forestiers.

Bien que certaines limites méthodologiques aient été soulevées, les sections 5 et 6, en lien avec la méthode de fraction d'ombre, ont démontré que ces méthodes peuvent être utilisées pour cartographier les attributs forestiers à une échelle locale. Cependant, certains éléments contraignent l'application de ces méthodes à un grand territoire comme le coût des ISTHRS ou le biais inter-photographies aériennes. Face à ces contraintes, des méthodes de cartographies à de grands territoires devaient être développées afin de répondre aux objectifs de départ.

8.3. Méthode de cartographie régionale adaptée à la forêt boréale québécoise

Dans le cas de la cartographie d'attributs forestiers à de grands territoires, les contraintes étaient reliées (i) à la faible représentativité de la donnée d'entraînement (PET mesurées de 7 m et plus) (ii) à l'écart temporel entre la mesure des PET et l'acquisition d'image et (iii) au grand territoire à couvrir. Par l'apport des PET_s, la première contrainte a pu être surmontée. La deuxième contrainte a pu être répondue par des méthodes de sélection qui ont été développée à la section 7. Le fait d'avoir comme données d'entraînement des placettes datant de plus de dix ans, dans certains cas, réduisait les résultats statistiques. C'est donc pour résoudre ces problématiques que nous avons développé des méthodes de sélection de placettes non perturbées entre la date de mesure de la placette et la date d'acquisition de l'image. Ces méthodes se sont aussi avérées efficaces puisqu'elles ont permis d'accroître les résultats statistiques au maximum de 15% pour le R^2 et de réduire l'erreur RMSE au maximum de 21%. La troisième contrainte a été surmontée par

l'utilisation des méthodes de cartographie non paramétriques. Ces dernières, permettent d'utiliser un grand nombre d'observation (couches numériques dans notre cas) et de prédire plusieurs attributs en même temps. Le projet de recherche décrit dans la section 7 utilise d'ailleurs ces forces pour couvrir avec une grande précision un territoire de 470 000 km² de la forêt boréale. Ces méthodes ouvrent donc la voie à plusieurs applications pratiques et scientifiques comme la sélection de placettes d'inventaire qui datent de plusieurs années pour les calculs de volumes marchands à l'échelle de grands territoires. Bien que ces méthodes ont démontré leur fort potentiel pour le territoire d'étude, celles-ci pourraient être poussées plus loin afin d'accroître leur efficacité et leur spectre d'application.

Premièrement, nous avons utilisés trois bandes de l'image Landsat, soient les bandes 3, 4 et 5. Or, il serait intéressant de tester l'apport des bandes 1, 2 et 7 sur les résultats statistiques. Pour utiliser ces autres bandes, il serait nécessaire d'utiliser des techniques de normalisation automatique des images comme celle proposée par Olthof *et al.* (2005). Un problème qui pourrait néanmoins survenir avec de telles techniques est que la résolution spectrale pourrait être dégradée et affecter négativement les résultats statistiques. Un autre élément à considérer avec l'apport de ces trois bandes est que les bandes 1 et 2 sont très corrélées avec la bande 3 et la bande 1 est régulièrement bruitée. Néanmoins, Jensen (2005) décrit le lien entre la bande 2 et la santé de la végétation.

Deuxièmement, différents travaux de cartographie ont récemment été réalisés à l'aide d'approches très performantes telle que Random Forest. L'algorithme a été d'abord proposé par Breiman, (2001) ainsi que par Cutler *et al.* (2007). Il effectue un apprentissage sur une multitude d'arbres de décision entraînés sur des sous-ensembles de données. Entre autres avantages, l'approche permet de gérer une base de données considérable, elle peut utiliser des milliers de variables en intrant et elle donne un estimé de quelles variables sont importantes dans la classification. Donc, puisque celle-ci permet de déterminer les couches de base (ex. bandes spectrales, couches thématiques, etc.) qui fournissent le plus grand apport sur la précision des résultats, il serait intéressant de tester une telle approche. Par exemple, Random Forest pourrait nous permettre de déterminer les quatre couches les plus pertinentes parmi celles établies au

départ. Par la suite, ces quatre couches pourraient être utilisées et appliquées à d'autres territoires similaires sans avoir à générer à nouveau les dix couches.

Troisièmement, comme mentionné à travers la section 7, les placettes terrain ont été mesurées en 1996 ou plus tard, ce qui réduit la correspondance entre les données mesurées et l'image satellitaire. Bien que la méthode de sélection des placettes proposée dans cette section réduise cet effet, un décalage d'attribut forestier existe néanmoins. En effet, entre 1996 et 2009, les arbres de certains sites plus riches ou provenant de peuplements forestiers plus jeunes, peuvent croître de plusieurs centimètres. En utilisant des placettes terrain récentes, les résultats statistiques de la carte résultante seraient très probablement plus élevés. Dans un même ordre d'idée, les méthodes non paramétriques utilisées (KNN et le KMSN) dans cette section ne permettent pas d'ajouter du poids à certains éléments moins représentés dans le jeu d'entraînement. Ainsi, les peuplements forestiers feuillus sont peu échantillonnés puisqu'ils ne couvrent que de petites surfaces. De plus, la signature spectrale de ces peuplements varie de façon importante selon le site, l'espèce dominante et la densité du couvert. Ce faible taux d'échantillonnage réduit donc la précision du résultat de ces peuplements forestiers feuillus. Il est à noter que ce biais affecte peu la carte dans l'ensemble puisque ces milieux ne couvrent pas de grandes superficies. En fait, les feuillus couvrent moins de 2 % de la superficie des domaines de la pessière à mousse et de la pessière à lichen. Néanmoins, afin de pallier cette faiblesse, un échantillonnage aléatoire stratifié selon les types de peuplements pourrait être réalisé. Cette façon de faire permettrait de tenir compte de la variabilité des peuplements forestiers plus rares.

Quatrièmement, de nouveaux capteurs ont récemment fait leur apparition et sont disponibles sur de grands territoires comme les images RapidEye. Ces dernières ont une résolution spatiale de 6,5 m et cinq bandes spectrales (bleu, vert, rouge, « red-edge » et infrarouge). Cette plus grande résolution spatiale permettrait d'ajouter des statistiques spatiales comme par exemple de la variance ou de la texture à l'intérieur des cellules d'analyses, soient de 30 mètres par 30 mètres. Cette information supplémentaire, devrait permettre de discriminer les sites homogènes de ceux hétérogènes en termes de structure de peuplement et de ce fait, influencer les résultats statistiques de la carte d'attributs forestiers résultante.

En dépit de toutes les limitations, améliorations et avenues de recherches possibles, ces travaux de recherche permettent de cartographier les peuplements forestiers dominés par l'épinette noire à partir d'ISTHRS ou de photographies aériennes. Ces méthodes permettent donc l'application des modèles à de grands territoires puisque l'épinette noire est l'essence la plus commune de la forêt boréale de l'est du Canada. Le fait que la méthode fonctionne également sur des photographies aériennes facilite grandement l'application de l'approche étant donné que la majeure partie du territoire canadien est couvert par au moins une série de photographies aériennes. Enfin, la méthode de cartographie des attributs forestiers à l'ensemble du territoire d'étude innove avec le développement de méthodes de sélection de placettes d'inventaires. Ces méthodes permettent donc d'utiliser de vieux jeux de données pour l'entraînement de méthodes de cartographies. Enfin, les cartes d'un des deux attributs forestiers (hauteur de peuplement) produites à l'ensemble du territoire d'étude ont été utilisés comme intrant par le Comité Scientifique visant à établir une limite nordique des forêts attribuables pour un aménagement forestier durable (MRN, 2012).

9. Conclusions

Nos travaux sur la cartographie forestière nordique par méthode de fraction d'ombre et de KNN ou KMSN ont permis de pousser plus loin les méthodes actuelles d'inventaire forestier nordique sur trois aspects principaux: (i) la cartographie locale d'attributs forestiers par fraction d'ombre à partir d'ISTHRS, (ii) la cartographie locale d'attributs forestiers par fraction d'ombre à partir de photographies aériennes et (iii) le développement d'une méthode complète pour la cartographie d'attributs forestiers adaptée au contexte du nord du Québec par l'apport de PET_s, de méthodes de sélection des PET et d'adaptation des méthodes non paramétriques. Jusqu'à maintenant, l'utilisation des relations statistiques entre la fraction d'ombre et certains attributs forestiers a été développée et vérifiée par Peddle *et al.* (2001), par Seed et King (2003) et Leboeuf *et al.* (2007). La méthode actuelle innove avec la démonstration que ces approches appliquées aux ISTHRS peuvent être utilisées pour le calcul et la cartographie d'attributs forestiers pertinents à la foresterie (hauteur de peuplement, densité, surface terrière et volume marchand). Celle-ci va plus loin en démontrant l'application de la méthode de la fraction d'ombre aux photographies

aériennes. Dans ce sens, nos travaux ont permis de spécifier différentes contraintes à l'application de cette approche à la photographie aérienne. Les cartes ainsi produites à l'aide de la méthode de fraction d'ombre permettent un calcul précis d'attributs forestiers qui peuvent accroître le nombre de données d'entraînements liées à la carte forestière régionale et ainsi permettre de couvrir l'ensemble de l'éventail de valeurs d'attributs forestiers, 0 à 7 m dans notre cas. Finalement, les nouvelles approches de sélection des PET développées à travers ces travaux démontrent le fort impact positif sur les résultats de la cartographie régionale et peuvent donc être appliquées pour l'utilisation de vieilles PET et ce pour différentes problématiques.

La nouvelle méthode de cartographie forestière proposée à l'échelle régionale répond donc aux enjeux cruciaux auxquels font face ceux qui effectuent des inventaires forestiers dans un milieu nordique ; faire beaucoup avec peu ! D'abord les méthodes KNN et KMSN proposées innovent puisqu'elles permettent une cartographie sur un immense territoire et en utilisant peu de PET et des portions de cartes d'attributs forestiers calculées à l'échelle locale. Cette approche réduit la dépendance à un grand nombre de PET qui sont rares et coûteuses étant donné l'immense surface à couvrir et le réseau routier limité. La méthode choisie s'applique donc favorablement au milieu nordique principalement à cause de la dominance et l'homogénéité des forêts d'épinette noire, du fort gradient de densité de ces milieux et de la rareté des milieux plus complexes à la cartographie tels que les milieux feuillus et mixtes. En somme, puisque la méthode requiert peu de ressources et permet de cartographier différents milieux présents dans le gradient nord-sud du site de cartographie, elle répond à l'enjeu principal de faire beaucoup avec un minimum de moyens.

10. Références bibliographiques

Avery, T.E. et Burkhart, H.E. (2001) Forest Measurements, Fifth ed. McGraw-Hill, New York.

Baatz, M., Benz, U., Dehghani, S., Heynan, M., Höltje, A., Hofmann, P., Lingenfelder, I., Mimler, M., Sohlabach, M., Weber, M. et Willhauck, G. (2004) eCognition User Guide 4. Definiens Imaging, Munich, Germany.

Beaudoin, A., Guindon, L., Lambert, M.C., Ung, C.H., Simard, G., Luther, J.E., et Fournier, R. (2003) A method for scaling up biomass of Canadian subarctic forests from tree to landscape levels using ground plots, QuickBird and Landsat data. Proceedings of the 25th Canadian Remote Sensing Symposium & 11th Congress of the Association québécoise de télédétection, 14–16 October 2003, Montreal, Quebec.

Beaudoin, A., Guindon, L., Hall, R.J., Luther, J.E. et Fournier, R.A. (2007) A strategy to map biomass of Canadian northern boreal forests using multi-sensor/multi-resolution imagery: approach, methods and preliminary results. Proceedings of Forestsat 2007, 5–7 November 2007, Montpellier, France.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. et Heynen, M. (2004) Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 58, p. 239–258.

Bernier P.Y., Daigle, G., Rivest, L.-P., Ung, C.H., Labbé, F., Bergeron, C. et Patry A. (2010) From plots to landscape: A k-NN-based method for estimating stand-level merchantable volume in the Province of Québec, Canada. Forestry Chronicle. Vol. 86, N° 4, p. 461-468.

Bickerstaff, A. et Hostikka, S.A. (1977) Growth of forests in Canada. Part 1: an annotated bibliography. Natural Resources Canada, Canadian Forest Service Inf. Report FMR-X-98.

Breiman, L. (2001) Random Forests. *Machine Learning*. Vol. 45, N°. 1, p. 5-32.

Carroll, R.J. et Ruppert, D. (1988) Transformation and Weighting in Regression. Chapman & Hall, New York.

Cihlar, J. (2000) Land cover mapping of large areas from satellites: status and research priorities. International Journal of Remote Sensing. Vol. 21, p. 1093-1114.

Crookson, N.L, Moeur, M. et Renner, D. (2002) User guide to the most similar neighbor imputation program. USDA, Forest Service. 40 pp.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson J. et Lawler. J. J. (2007) Random forests for classification in ecology. Ecology, Vol. 88, p. 2783-2792.

Definiens Inc. (2002) eCognition, version 3.1. Definiens Inc., Boulder, Colorado.

DNR (2011) Ontario Forest Resources Inventory. Photo Interpretation Specifications. 92 pp.

Fortin, M., DeBlois, J., Bernier. S. et Blais, G. (2007) Mise au point d'un tarif de cubage général pour les forêts québécoises : une approche pour mieux évaluer l'incertitude associée aux prévisions. Forestry Chronicle, Vol. 83, N° 5, p. 754-765.

Fortin, M., Bernier, S., Saucier, J.-P. et Labbé, F. (2009) Une relation hauteur-diamètre tenant compte de l'influence de la station et du climat pour 20 espèces commerciales du Québec. Mémoire de recherche forestière n°153, Direction de la recherche forestière.

Franklin, S. et Wulder, M. (2002) Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. Progress in Physical Geography, Vol. 26, p. 173–205.

Gausman, H.W. (1984) Evaluation of factors causing reflectance differences between sun and shade leaves. Remote Sensing of Environment. Vol. 15, p.177–181

Gillis, M.D. (2001) Canada's national forest inventory (responding to current information needs). Environment Monitoring and Assessment, Vol. 67, p. 121–129.

Gjertsen, A.K., Tomppo, E. et Tomter, S. (1999) National forest inventory in Norway: Using sample plots, digital maps, and satellite images. In: Proc. of IGARSS, Hamburg, Germany, p. 729-731.

Gougeon, F.A. (1995) A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. Canadian Journal of Remote Sensing. Vol. 21, p. 274–284.

Gougeon, F.A. et Leckie, D.G. (2003) Forest information extraction from high spatial resolution images using an individual tree crown approach. Natural Resources Canada, Canadian Forest Service - Pacific Forestry Centre Inf. Rep BC-X-396.

Greenberg, J. A., Dobrowski, S. Z. et Ustin, S. L. (2005) Shadow allometry: Estimating tree structural parameters using hyperspatial image analysis. Remote Sensing of Environment, Vol. 97, p. 15-25.

Gruber, M. et Schneider, S. (2007) Digital surface models from UltraCam-X images, Photogrammetric Image Analysis, Munich, Germany, 36 (3/W49B), p.47-52.

Hart, P.E. (1967) Nearest neighbor pattern classification. IEEE Transactions on Information, Vol. 13. p. 21–27.

Homer, C.G., Ramsey, R.D., Edwards, T.C. et Falconer, A. (1997) Landscape cover type modeling using a multi-scene thematic mapper mosaic. Photogrammetric Engineering and Remote Sensing. Vol. 63, p. 59-67.

Hudak, A. T., Crookston, N. L., Evans, J. S., Hall, D. E. et Falkowski, M.J. (2008) Nearest neighbor imputation of species-level, plot-scale forest structure attributes from LiDAR data. Remote Sensing of Environment, Vol. 112, p. 2232-2245.

Hyyppä, J., Hyyppä, H., Inkinen, M., Engdahl, M., Linko, S. et Zhu, Y.H. (2000) Accuracy comparison of various remote sensing data sources in the retrieval of forest stand attributes. Remote Sensing of Environment, Vol. 128, p.109–120.

Jensen, J.R. (2005) Introductory digital image processing. A remote sensing perspective. Third edition. Pearson prentice hall, 526 p.

Kane, V., Gillespie, A., McGaughey, R., Lutz, J., Ceder, K. et Franklin, J. (2008) Interpretation and topographic compensation of conifer canopy self-shadowing. Remote Sensing of Environment, Vol. 112, p. 3820-3832.

Kangas A. et Maltamo M. (Ed). (2009) Forest inventory: Methodology and applications, Series on Managing Forest Ecosystems, Springer, Vol. 10, 362 p.

Katila, M. (2006). Empirical errors of small area estimates from the multisource National Forest Inventory in Eastern Finland. Silva Fennica, Vol. 40, N°. 4, p. 729–742.

Ker, M.F. (1974) Metric yield tables for the major forest cover types of Newfoundland. Natural Resources Canada, Canadian Forest Service – Atlantic Forestry Centre, Inf. Rep. M-X-141.

Leboeuf, A., Beaudoin, A., Fournier, R.A., Guindon, L., Luther, J.E. et Lambert, M.-C. (2007) A shadow fraction method for mapping biomass of northern boreal black spruce forests using QuickBird imagery. Remote Sensing of Environment, Vol. 110, p. 488-500.

Leboeuf, A., Fournier, R.A., Luther, J.E., Beaudoin, A., and Guindon, L. (2012) Forest attribute estimation of northern Canadian forests using QuickBird imagery and a shadow fraction method. Forest Ecology and Management, Vol. 266, p. 66-74.

Leckie, D.G., Gougeon, F.A., Walsworth, N. et Paradine. D. (2003) Stand delineation and composition estimation using semi-automated individual tree crown analysis. Remote Sensing of Environment, Vol. 85, p. 355–369.

LeMay, V. et Temesgen, H. (2005) Comparison of nearest neighbor methods for estimating basal area and stems per hectare using auxiliary variables. Forest Science, Vol. 51, p. 109–119.

LeMay, V., Maedel, J. et Coops, N.C. (2008) Estimating stand structural details using nearest neighbor analyses to link ground data, forest cover maps, and Landsat imagery. Remote Sensing of Environment, Vol. 112, p. 2578–2591.

Lévesque, J. et King, D. (2003) Spatial analysis of radiometric fractions from high-resolution multispectral imagery for modelling individual tree crown and forest canopy structure and health. Remote Sensing of Environment, Vol. 84, p. 589–602.

Létourneau, J.P., Bard, A. et Lambert, J. (2003) Normes de cartographie écoforestière. Troisième inventaire écoforestier. Ministère des Ressources naturelles, de la Faune et des Parcs du Québec. Direction des Inventaires Forestiers.

Li, X. et Strahler, A.H. (1985) Geometric-optical modeling of a conifer forest canopy. IEEE Transactions on Geoscience and Remote Sensing, Vol. 23, p. 705–721.

Li, X., et Strahler, A.H. (1992) Geometric-optical bidirectional reflectance modeling of the discrete crown vegetation canopy: Effect of crown shape and mutual shadowing. IEEE Transactions on Geoscience and Remote Sensing, Vol. 30, p. 276–292.

Lillesand, T.M., and Kiefer. R.W. (1994) Remote Sensing and Photo Interpretation, 3rd. ed. John Wiley& Sons: New York. 750 p.

Lowe, J.J., Power, K. et Gray, S.L. (1994) Canada's forest inventory 1991. Natural Resources Canada, Canadian Forest Service – Petawawa National Forest Institute Inf. Rep. PI-X-115.

Luther, J.E., Fournier, R.A, Houle, M., Leboeuf, A. et Piercy D.E. (2012) Application of shadow fraction models for estimating attributes of northern boreal forests. Canadian Journal of Forest Research. Vol. 4, p. 1750-1757.

Maltamo, M., Eerikäinen, K., Pitkänen, J., Hyyppä, J. et Vehmas, M. (2004) Estimation of timber volume and stem density based on scanning laser altimetry and expected tree size distribution functions. Remote Sensing of Environment, Vol. 90, p. 319-330.

McRoberts, R.E., Nelson, M.D. et Wendt, D.G. (2002) Stratified estimation of forest area using satellite imagery, inventory data, and the *k*-Nearest Neighbors technique. Remote Sensing of Environment, Vol. 82, p. 457–468.

McRoberts, R.E., Tomppo, E.O., Finley, A.O. et Heikkinen, J. (2007) Estimating areal means and variances of forest attributes using the k-nearest neighbors technique and satellite imagery. Remote Sensing of Environment, Vol. 111, p. 466–480.

McRoberts, R.E. et Tomppo, E. (2007) Remote Sensing Support for national forest inventories. Remote Sensing of Environment, Vol. 110, p. 412-419.

McRoberts, R.E. (2011) Satellite image-based map: Scientific inference or pretty pictures? Remote Sensing of Environment, Vol. 115, p. 715–724.

Meng, Q., Cieszewski, C. et Madden, M. (2009) ISPRS Journal of Photogrammetry and Remote Sensing. Vol 64, N°.. 1, p. 27-36.

Milton, J.S. et Arnold, J.C. (2003) Introduction to probability and statistics: Principles and applications for engineering and the computing sciences. 4th edition.

Ministry of Forest of British Colombia (2005) British Columbia Mapping Standards For use in results submissions. 24 p.

Moeur, M. et Stage, A.R. (1995) Most similar neighbor: an improved sampling inference procedure for natural resource planning. Forest Science, Vol. 41: p. 337-359.

Mora, B., Wulder, M.A. et White, J.C. (2010) Identifying leading species using tree crown metrics derived from very high spatial resolution imagery in a boreal forest environment. Canadian Journal of Remote Sensing, Vol. 36, p. 332-344.

MRN (1998) Les zones de végétation et les domaines bioclimatiques du Québec. Bibliothèque nationale du Québec.

MRNF (2009a) Norme de stratification écoforestière du Nord. Ministère des Ressources naturelles et de la faune du Québec. Direction des Inventaires Forestiers, 33 p.

MRNF (2009b) Norme d'inventaire forestier. Placettes-échantillons permanentes. Ministère des Ressources naturelles et de la faune du Québec. Direction des Inventaires Forestiers, 290 p.

MRNF (2011) Norme d'inventaire forestier. Normes de cartographie écoforestière. Quatrième inventaire écoforestier. Ministère des Ressources naturelles, de la Faune du Québec. Direction des Inventaires Forestiers. 33 p.

MRN (2012) Limite nordique des forêts attribuables pour un aménagement forestier durable. *In* Documents électroniques, Québec, http://www.mrn.gouv.qc.ca/forets/amenagement/amenagement-limite-nordique.jsp, 1 p.

Naesset, E. (2004) Practical large-scale forest stand inventory using a small-footprint airborne scanning laser. Scandinavian Journal of Forest Research, Vol. 19, N°. 2, p. 164-179.

Ohmann, J. L. et Gregory, M. J. (2002) Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, U.S.A. Canadian Journal of Forest Research, Vol. 32, p. 725–741.

Olthof, I., Butson, C., Fernandes, R., Fraser, R., Latifovic, R. et Orazietti, J. (2005) Landsat ETM+ mosaic of northern Canada. Journal canadien de télédétection. Vol. 31. p. 412-419.

Ozdemir, I. (2008) Estimating stem volume by tree crown area and tree shadow area extracted from pan-sharpened QuickBird imagery in open Crimean juniper forests. International Journal of Remote Sensing, Vol. 29, p. 5643-5655.

Parker R.C. et Evans, D. L. (2004) An application of LiDAR in a double-sample forest inventory. Western Journal of Applied Forestry, Vol. 19, No. 2, p. 95–101.

PCI Geomatics inc. (2003) Geomatica Focus, version 9.1.

Peddle, D.R. et Johnson, R.L. (2000) Spectral mixture analysis of airborne remote sensing imagery for improved prediction of leaf area index in mountainous terrain, Kananaskis Alberta. Canadian Journal of Remote Sensing, Vol. 26, p. 177–188.

Peddle, D. R., Brunke, S. P. et Hall, F. G. (2001) A comparison of spectral mixture analysis and ten vegetation indices for estimating boreal forest biophysical information from airborne data. Canadian Journal of Remote Sensing, Vol. 27, p. 627-635.

Pekkarinen, A. (2002) Image segment-based spectral features in the estimation of timber volume. Remote Sensing of Environment, Vol. 82, p. 349–359.

Pellikka, P., King, D.J. et Leblanc, S.G. (2000a) Quantification and reduction of bidirectional effects in aerial CIR imagery of deciduous forest using two reference land surface types. Remote Sensing Reviews, Vol. 19, p. 259–291.

Pellikka, P., Seed, E.D. et King, D.J. (2000b) Modelling deciduous forest ice storm damage using aerial CIR imagery and hemispheric photography. Canadian Journal of Remote Sensing, Vol. 26, p. 394–405.

Peng, C., Zhang, L. et Liu, J. (2001) Developing and validating nonlinear height–diameter models for major tree species of Ontario's boreal forest. Northern Journal of Applied Forestry, Vol. 18, p. 87–94.

Perron, J.Y. (1986) Standard volume tables: gross merchantable volume. Gouvernement du Québec, Ministère de l'Énergie et des Ressources, Québec.

Puttonen, E., Litkey, P. et Hyyppä, J. (2010) Individual Tree Species Classification by Illuminated—Shaded Area Separation. Remote Sensing, Vol. 2, p. 19-35.

Robitaille, A., Leboeuf, A., Létourneau, J.-P., Saucier, J.-P. et Vaillancourt, E. (2008) Integrated Ecoforest Mapping of the Northern Portion of the Continuous Boreal Forest, Québec, Canada. Workshop Proceedings: Circum Boreal Vegetation Mapping. 3-6 November, Helsinki, Finland, p.180-183.

Robitaille, A., Leboeuf, A., Létourneau, J.P., Saucier, J.P. et Vaillancourt, É. (2009) Extensive Ecoforest Map of Northern Continuous Boreal Forest, Québec, Canada. IUFRO – the International Union of Forest Research Organizations, 19-22 mai. Québec, Canada.

Robert, D. et Robitaille, A. (2009a) Cartographie forestière. In Manuel de foresterie, Ordre des ingénieurs forestiers du Québec, Éditions Multi Mondes, p. 507-540.

Robert, D. et Robitaille, A. (2009b) Integrated Ecoforest Mapping of Southern Québec Forests. Extended forest inventory and monitoring over space and time, IUFRO – the International Union of Forest Research Organizations, May 19-22th, Québec City, Canada.

Rosengren, M., Tomppo, E., Pereira, J. M., Nilsson, M., Aalto, P., Hagner, O., Katila, M., Malmberg, U., Paul, J., Tome, M., et Willen, E. (1999) FMERSII final report—Forest monitoring in Europe with remote sensing (biomass and wood volume mapping). Final report XP-FMERSII-12, Swedish Space Corporation (SSC), Solna, Sweden.

Rowe, J.S. (1972) Forest Regions of Canada. Can. For. Serv., Depart. Environ., Publication No. 1300.

Saucier, J.-P, Grondin, P., Robitaille, A. et Bergeron, J.-F. (1998) Les zones de végétation et les domaines bioclimatiques du Québec. Gouvernement du Québec. Ministère des Ressources naturelles.

Seed, E.D. et King, D.J. (2003) Shadow brightness and shadow fraction relations with effective leaf area index: importance of canopy closure and view angle in mixedwood boreal forest. Canadian Journal of Remote Sensing, Vol. 29, p. 324-335.

Shapiro, S.S., Wilk, M.B. (1965) An analysis of variance test for normality (complete samples). Biometrika, Vol. 52, p. 591–611.

Sharma, M., Amateis, R.L. et Burkhart, H.E. (2002) Top height definition and its effect on site index determination in thinned and unthinned loblolly pine plantations. Forest Ecology and Management, Vol. 168, p. 163–175.

Shreuder, H.T., Gregoire, T.G. et Wood, G.B. (1993) Sampling methods for multiresource forest inventory. J. Wiley and Sons, New York.

Soenen, S.A. et Peddle, D.R. (2005) SCS+C: A modified sun-canopy-sensor topographic correction in forested terrain. *IEEE transactions on Geoscience and Remote Sensing*, Vol. 43, p. 2148-2159.

St-Onge, Vega, B.C., Fournier, R.A. et Hu. Y. (2008) Mapping canopy height using a combination of digital stereo-photogrammetry and lidar, International Journal of Remote Sensing, Vol. 29, p. 3343-3364.

Titus, S.J. et Morgan, D.J. (1988) Theory and methodology of single tree volume estimation. Natural Resources Canada, Canadian Forest Service – Petawawa National Forest Institute Inf. Rep. PI-X-85. 8 p.

Tomppo, E. et M. Halme. (2004) Using coarse scale forest variables as ancillary information and weighting of variables in the k-NN estimation: a genetic algorithm approach. Remote Sensing of Environment, Vol. 92, p. 1–20.

Tomppo, E., H Olsson, G Ståhl, M Nilsson, O Hagner, et Katila, M. (2008) Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment, Vol. 112, p. 1982–1999.

Tomppo, E., Schadauer, K., McRoberts, R.E., Gschwantner, T., Gabler, K. et Ståhl, G. (2010) History of NFIs. In: National Forest Inventories. Eds: Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R., Springer, p. 1-2.

Tuominen, S., Fish S. et Poso. S. (2003) Combining remote sensing, data from earlier inventories, and geostatistical interpolation in multisource forest inventory. Canadian Journal of Forest Research, Vol. 33, p. 624–634.

Vexcel. 2013. Vexcel Corporation. A Microsoft® compagny.

Wolter, P.T., Berkley, E.A., Peckham, S.D., Singh, A. et Townsend, P.A. (2012) Exploiting tree shadows on snow for estimating forest basal area using Landsat data. Remote Sensing of Environment, Vol. 121, p. 69–79.

Wilson, B.T., Lister, A.J., et Riemann, R.I. (2012) A nearest-neighbor imputation approach to mapping tree species over large areas using forest inventory plots and moderate resolution raster data. Forest Ecology and Management, Vol. 271, p. 182-198.

Wulder, M. (2000) Large area remote sensing land cover mapping projects. Canadian Forest Service, Natural Resources Canada. URL: http://www.pfc.forestry.ca/eosd/cover/largearea_e.html.

Wulder, M., Niemann, K.O. et Goodenough, D.G. (2000) Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. Remote Sensing of Environment, Vol. 73, p. 449–476.

Wulder, M., White, J.C., Hay, G.J., et Castilla, G. (2008) Towards automated segmentation of forest inventory polygons on high spatial resolution satellite imagery. The Forestry Chronicle, Vol. 4, p. 221-230.

Wulder, M.A., White, J.C., Coggins, S., Ortlepp, S.M., Coops, N.C., Heath, J., Mora, B. (2012a) Digital high spatial resolution aerial imagery to support forest health monitoring: the mountain pine beetle context. Journal of Applied Remote Sensing, Vol. 6, p. 1-10.

Wulder, M.A., White, J.C., Nelson, R.F., Naesset, E., Orka, H.O., Coops, N.C. Hiker, T., Bater, C.W., et Gobakken, T. (2012b) Lidar sampling for large-area forest characterisation: a review. Remote Sensing of Environment. Vol. 121, p. 196-209.

Zorn, M.E., Gibbons, R.D. et Sonzogni, W.C. (1997) Weighted least-square approach limits of detection and quantification variability as a function of concentration. Analytical Chemistry, Vol. 69, p.3069-3075.

11. Listes d'acronymes

$AFknn_i$: valeur d'attributs forestiers calculés par knn
$AFkmsn_i$: valeur d'attributs forestiers calculés par kmsn
DHP : diamètre à hauteur de poitrine
ISTHRS : images satellitaires à très haute résolution spatiale
kNN: k Nearest Neighbor
KMSN: k Most Similar Neighbors
LAI : surface foliaire
MRNQ : Ministère des Ressources naturelles du Québec
PET: placette d'échantillonnage terrain
PET_{CAL} : PET de calibration (70% du jeu de données)
PET_s : placette satellitaire
PET_{VAL} : PET de validation (30% du jeu de données)
RMSE : erreur quadratique moyenne (RMSE)
SCF: Service Canadien des Forêts

Article 1, 2 et 3 :
A_s : azimuth angles
A_v : azimuth angles
CH: Chibougamau
DBH: diameter at breast height
FA: forest attributes
FA_{GSP} : one of the four FA available from GSPcal,
GSP: ground sample plots
GSP'1 : GSP from first selection method
GSP'2: GSP from second selection method
GSP_{cal} : 70% of the GSP data set
GSP_{val} : Thirty percent of the GSP remaining
HSRSI: High spatial resolution satellite imagery
PAN: The panchromatic band
QB: QuickBird image.
RA: Radisson
R^2_{adj} : adjusted regression model coefficient
RMSE : root mean square error
$RMSE_r$: relative root mean square error
SF: tree shadow fraction
SF_n : normalized SF values
SF_n: normalized SF values
WA: Wabush
Z_v: viewing zenith angles
Z_s : sun zenith angles