

Résumé d'article scientifique

Regression modeling and mapping of coniferous forest basal area and tree density from discrete-return lidar and multispectral data

Maxime Lacroix

23 novembre 2018

- 1 Introduction
- 2 Collecte des données
- 3 Modélisation
- 4 Principaux constats
- 5 Conclusion et discussion

Introduction

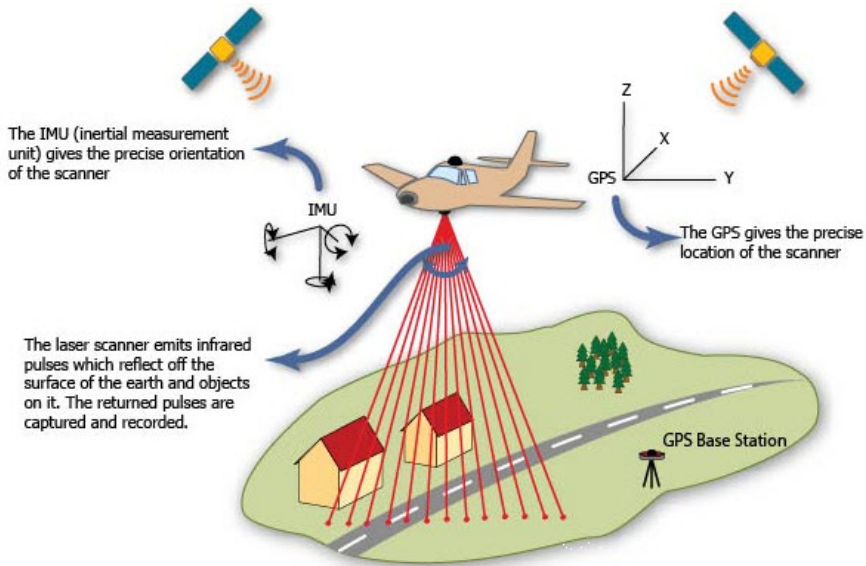
Contexte

- Article publié par Hudak et al. en 2006
- *Canadian Journal of Remote Sensing*
- Contexte de modélisation forestière

Objectif

- Modéliser deux variables réponses :
 - Surface terrière (*basal area*)
 - Densité de l'arbre (*tree density*)
- À l'aide de variables provenant de deux méthodes de détection :
 - Le lidar (*discrete-return light detection and ranging*)
 - Multispectral satellite data (*ALI*)
- Pour résumer, on tente de prédire ces deux variables en fonction des données qu'on est capable de capter.

Compréhension du lidar

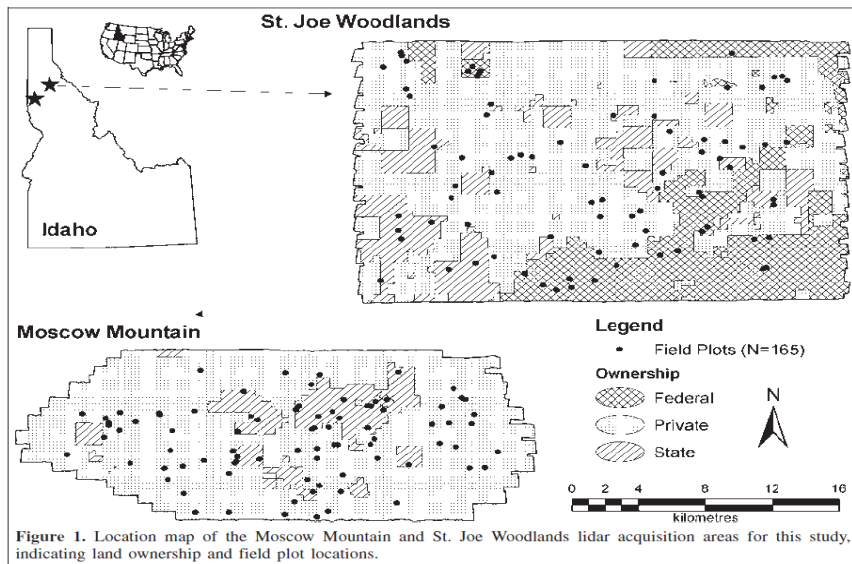


Collecte des données

Plan d'échantillonnage

- On s'intéresse à deux forêts au nord de l'Idaho, aux États-Unis
 - St-Joe Woodlands
 - Moscow Mountain
- Plan d'échantillonnage en deux étapes
 - Dans chaque forêt, on crée 9 strates basées sur :
 - 3 classes d'élévation
 - 3 classes d'insolation solaire
 - Dans chaque strate : 9 classes de *leaf area index*
 - Sélection aléatoire des points pour chacun des regroupements d'élévation, insolation et LAI.

Résumé du plan d'échantillonnage



Collecte des données

- Mesure de la surface terrière et la densité de l'arbre sur site
- Mesures ALI faites en octobre 2004
- Mesures lidar faites en Juillet, Août et Septembre 2003

Modélisation

Variables disponibles

- En tout, 26 variables sont utilisées pour prédire
- 2 variables à prédire (BA et TD), quelques transformations :
- On pose 0 si arbre trop petit
- On utilise $\ln(\text{BA}+1)$ et $\ln(\text{TD}+1)$

Liste des variables

Table 1. Predictor variables used for multiple linear regression modeling of natural-logarithm-transformed basal area and tree density.

Predictor variable	Description
Geographic	
Easting	UTM (Zone 11) easting at plot center
Northing	UTM (Zone 11) northing at plot center
Elevation	Elevation (m) above mean sea level at plot center
Advanced land imager (ALI)	
B1 _{mean}	Mean of 30 m ALI band 1 pixels intersecting plot
B2 _{mean}	Mean of 30 m ALI band 2 pixels intersecting plot
B3 _{mean}	Mean of 30 m ALI band 3 pixels intersecting plot
B4 _{mean}	Mean of 30 m ALI band 4 pixels intersecting plot
B5 _{mean}	Mean of 30 m ALI band 5 pixels intersecting plot
B6 _{mean}	Mean of 30 m ALI band 6 pixels intersecting plot
B7 _{mean}	Mean of 30 m ALI band 7 pixels intersecting plot
B8 _{mean}	Mean of 30 m ALI band 8 pixels intersecting plot
B9 _{mean}	Mean of 30 m ALI band 9 pixels intersecting plot
PAN _{mean}	Mean of 10 m PAN band pixels intersecting plot
PAN _{std}	Standard deviation of 10 m PAN band pixels intersecting plot
Lidar	
Intensity	
INT _{mean}	Mean of 2 m intensity pixels intersecting plot
INT _{std}	Standard deviation of 2 m intensity pixels intersecting plot
INT _{min}	Minimum of 2 m intensity pixels intersecting plot
INT _{max}	Maximum of 2 m intensity pixels intersecting plot
Height	
HT _{mean}	Mean of 6 m height pixels intersecting plot
HT _{std}	Standard deviation of 6 m height pixels intersecting plot
HT _{min}	Minimum of 6 m height pixels intersecting plot
HT _{max}	Maximum of 6 m height pixels intersecting plot
Canopy cover	
CC _{mean}	Mean of 6 m canopy cover pixels intersecting plot
CC _{std}	Standard deviation of 6 m canopy cover pixels intersecting plot
CC _{min}	Minimum of 6 m canopy cover pixels intersecting plot
CC _{max}	Maximum of 6 m canopy cover pixels intersecting plot

Modèle et sélection de variables

- Régression linéaire (utilisation de `lm`)
- Deux étapes de sélection de variable :
 - Méthode algorithmique `stepAIC` pour déterminer le nombre de paramètres dans le modèle
 - `leaps::regsubsets()`, qui permet de trouver le meilleur modèle selon le nombre de paramètres.
- Méthode de validation croisée pour tenter de prédire le BA et le TD

Principaux constats

Comparaison des deux méthodes

Table 2. Multiple linear regression models for predicting natural-logarithm-transformed basal area and tree density from geographic, ALI, or lidar variable groups.

Method	Selected variables	R^2	Adjusted R^2	Residual SE	AIC
Basal area					
Geographic	Easting, northing, elevation	0.0946	0.0777	31.3400	1140.80
ALI	B1 _{mean} , B2 _{mean} , PAN _{mean} , PAN _{std}	0.5599	0.5489	0.8540	-47.15
Lidar					
All lidar	INT _{mean} , Ht _{mean} , Ht _{std} , Ht _{min} , CC _{mean} , CC _{std}	0.8941	0.8901	0.4216	-278.17
Intensity	INT _{mean} , INT _{std} , INT _{min}	0.7779	0.7738	0.6048	-161.98
Height	HT _{mean} , HT _{max}	0.7958	0.7960	0.5744	-180.01
Canopy cover	CC _{mean} , CC _{std} , CC _{min}	0.7058	0.7003	0.6962	-115.57
Tree density					
Geographic	Easting, northing, elevation	0.0871	0.0700	368.1000	1953.69
ALI	B1 _{mean} , B2 _{mean} , B7 _{mean} , B8 _{mean} , B9 _{mean} , PAN _{mean} , PAN _{std}	0.6568	0.6415	1.0390	20.32
Lidar					
All lidar	INT _{mean} , INT _{std} , INT _{min} , CC _{mean} , CC _{max}	0.8698	0.8657	0.6358	-143.56
Intensity	INT _{mean} , INT _{std} , INT _{min}	0.7779	0.7737	0.8252	-59.45
Height	HT _{mean} , HT _{max}	0.4962	0.4900	1.2390	73.67
Canopy cover	CC _{mean} , CC _{std} , CC _{min} , CC _{max}	0.8354	0.8313	0.7126	-106.87

Note: Variable groups were best subsets selected based on Mallows (1973) Cp statistic.

Présentation des coefficients du modèle BA

Table 4. Parameters, coefficients, and statistics for the model used to map basal area (natural logarithm transformed).

Parameter	Estimate	SE	<i>t</i> value	Pr(> <i>t</i>)	Significance
Intercept	-4.05×10^1	2.04×10^1	-1.981	4.9407×10^{-2}	*
Easting	-1.23×10^{-5}	3.80×10^{-6}	-3.244	1.4480×10^{-3}	**
Northing	9.67×10^{-6}	4.32×10^{-6}	2.237	2.6750×10^{-2}	*
Elevation	1.04×10^{-3}	1.81×10^{-4}	5.713	5.6700×10^{-8}	***
PAN _{mean}	-9.18×10^{-4}	3.68×10^{-4}	-2.495	1.3675×10^{-2}	*
INT _{mean}	-2.39×10^{-2}	4.99×10^{-3}	-4.794	3.8600×10^{-6}	***
HT _{mean}	3.56×10^{-2}	1.02×10^{-2}	3.492	6.2800×10^{-4}	***
HT _{std}	7.22×10^{-2}	1.75×10^{-2}	4.126	6.0500×10^{-5}	***
HT _{min}	2.22×10^{-2}	9.46×10^{-3}	2.342	2.0454×10^{-2}	*
CC _{mean}	1.74×10^{-2}	5.21×10^{-3}	3.342	1.0470×10^{-3}	**
CC _{std}	4.90×10^{-2}	1.56×10^{-2}	3.144	2.0060×10^{-3}	**
CC _{min}	8.45×10^{-3}	4.98×10^{-3}	1.695	9.2107×10^{-2}	.
CC _{max}	-1.46×10^{-2}	6.06×10^{-3}	-2.407	1.7288×10^{-2}	*

Note: Regression sum of squares = 244.138 at 12 degrees of freedom (df); error sum of squares = 21.049 at 152 df; mean square error = 0.1385; residual standard error = 0.3721; multiple $R^2 = 0.9206$; adjusted $R^2 = 0.9144$; *F* statistic = 146.9 on 12 and 152 df ($p < 2.20 \times 10^{-16}$). Significance levels are as follows: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$; ., $p < 0.1$.

Présentation des coefficients du modèle TD

Table 5. Parameters, coefficients, and statistics for the model used to map tree density (natural logarithm transformed).

Parameter	Estimate	SE	<i>t</i> value	Pr(> <i>t</i>)	Significance
Intercept	-7.15×10^1	3.15×10^1	-2.266	2.4850×10^{-2}	*
Easting	-1.42×10^{-5}	6.00×10^{-6}	-2.360	1.9550×10^{-2}	*
Northing	1.63×10^{-5}	6.63×10^{-6}	2.462	1.4940×10^{-2}	*
Elevation	5.62×10^{-4}	2.83×10^{-4}	1.987	4.8700×10^{-2}	*
B5 _{mean}	1.24×10^{-3}	4.31×10^{-4}	2.874	4.6200×10^{-3}	**
B7 _{mean}	-3.57×10^{-3}	1.45×10^{-3}	-2.461	1.5000×10^{-2}	*
B9 _{mean}	2.51×10^{-2}	9.63×10^{-3}	2.607	1.0030×10^{-2}	*
PAN _{mean}	-1.88×10^{-3}	6.72×10^{-4}	-2.797	5.8200×10^{-3}	**
INT _{mean}	-3.54×10^{-2}	4.88×10^{-3}	-7.261	1.7800×10^{-11}	***
INT _{std}	4.12×10^{-2}	1.37×10^{-2}	3.003	3.1200×10^{-3}	**
CC _{max}	3.49×10^{-2}	3.88×10^{-3}	8.992	8.3600×10^{-16}	***

Note: Regression sum of squares = 436.288 at 10 degrees of freedom (df); error sum of squares = 57.279 at 154 df; mean square error = 0.3719; residual standard error = 0.6099; multiple R^2 = 0.8839; adjusted R^2 = 0.8764; *F* statistic = 117.3 on 10 and 154 df ($p < 2.20 \times 10^{-16}$). Significance levels are as follows: ***, $p < 0.001$; **, $p < 0.01$; *, $p < 0.05$.

Prédictions avec validation croisée

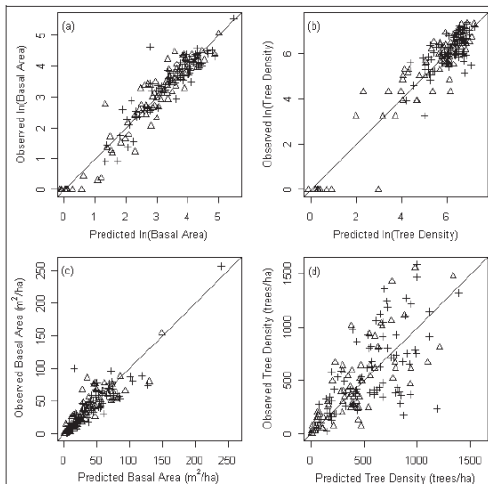


Figure 2. Scatterplots of cross-validation predictions versus observations ($N = 165$) for natural-logarithm-transformed basal area (a) and tree density (b) and natural-scale basal area (c) and tree density (d). The two highest values in (c) are the two old-growth plots. The lines indicate 1:1 relationships. Δ , Moscow Mountain plots; +, St. Joe Woodlands plots.

Conclusion et discussion

Discussion

- Ajout pertinent d'un effet aléatoire forêt ?
- Mention d'utilisation potentielle d'un GLM
- Vérification de la multicollinéarité ?
- Pertinence de la transformation \ln

Conclusion

- Le lidar et le multispectral satellite data permettent de bien prédire le BA et le TD ($R^2 \approx 0.9$)
- Serait intéressant de refaire l'exercice avec des modèles un peu plus complexes
- La situation a-t-elle changée depuis 2006 ? Avons-nous accès à de nouvelles technologies qui seraient meilleures?

Sources

- Image sur le lidar : <https://www.elprocus.com/lidar-light-detection-and-ranging-working-application/>
- Autres images et tableaux proviennent directement de l'article