

Incorporation de contraintes spatiales et opérationnelles dans des plans d'échantillonnage pour des inventaires forestiers

Audrey-Anne Vallée¹

en collaboration avec

Bastien Ferland-Raymond²,

Louis-Paul Rivest³,

Yves Tillé¹ .

¹Université de Neuchâtel

²Ministère des forêts, de la faune et des parcs du Québec

³Université Laval

Journées de Méthodes Avancées pour l'Analyse de Sondages Complexes

8 novembre 2016

Nantes

Table des matières

Contexte de l'inventaire forestier

Région d'intérêt

Objectif

Problèmes

Solution

Probabilités d'inclusion

Plan de sondage pour inventaire forestier

Illustration numérique



Figure: Région d'intérêt dans l'inventaire régulier du Québec: au Sud du 52ième parallèle.



Épinette noire, pin gris et sapin baumier.





Peuplier faux-tremble, bouleau à papier et érable.



Objectif

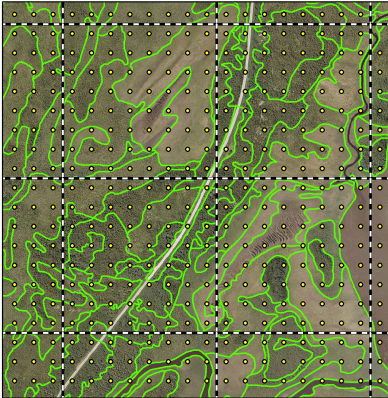


Figure: Ensemble de polygones, tuiles et placettes.

Photo aérienne prise sur tout le territoire étudié.

Objectif

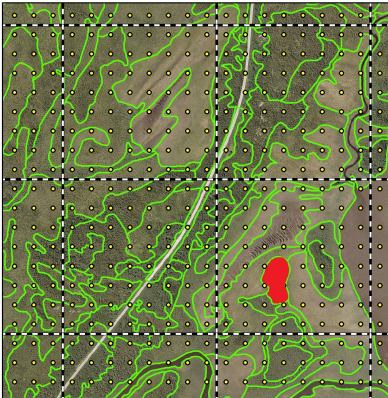


Figure: Ensemble de polygones, tuiles et placettes.

Polygones.

- Information auxiliaire: Espèces d'arbre.
- H ensembles de polygones (types de forêt).

Objectif

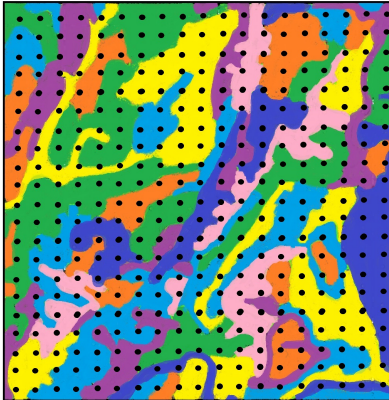


Figure: Ensemble de polygones, tuiles et placettes, avec représentation des types.

$h = 1$ Conifères 0-10 ans.

$h = 2$ Conifères 10-30 ans.

$h \dots$

$h = H - 1$ Feuillus 0-10 ans.

$h = H$ Feuillus 10-30 ans.

Objectif

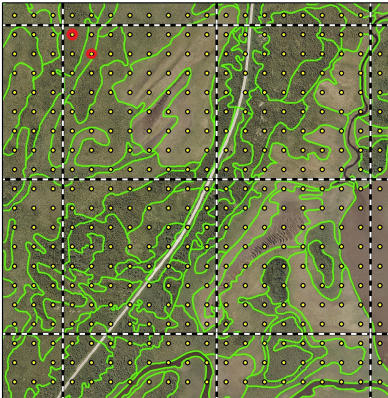


Figure: Ensemble de polygones, tuiles et placettes.

Placettes.

- ▶ Unité d'échantillonnage.
- ▶ Information auxiliaire:
Des polygones: Espèces d'arbre,
Landsat 5: Bandes de couleurs
spectrales.

Objectif

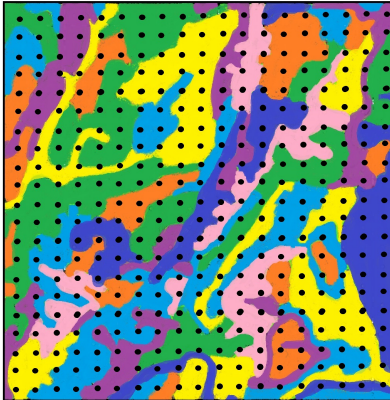


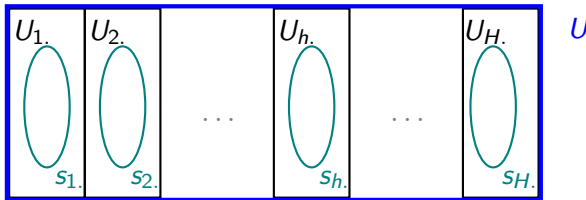
Figure: Ensemble de polygones, tuiles et placettes, avec représentation des types.

- ▷ Type 1
 N_1 . placettes dans la population
 n_1 . placettes dans l'échantillon s
- ▷ ...
- ▷ Type H
 N_H . placettes dans la population
 n_H . placettes dans l'échantillon s

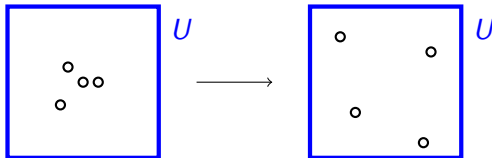
Objectif

On aimerait un échantillon

- Stratifié par type de forêt: n_h placettes de type h .



- Spatialement étalé.



Problèmes

- ▶ Les placettes échantillonnées sont étalées sans contraintes.
- ▶ Un enquêteur peut se retrouver à visiter une placette par jour.
- ▶ Cela entraîne beaucoup de déplacements.
- Enquête très longue.
- Enquête très coûteuse.

Solution

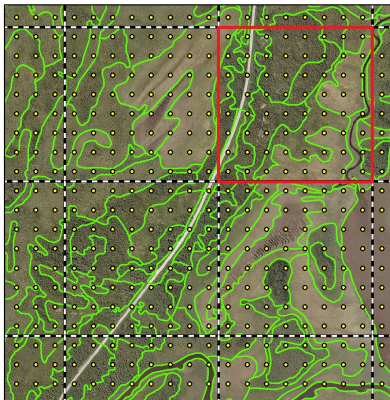


Figure: Ensemble de polygones, tuiles et placettes.

Tuiles.

- ▶ Grille de carrés de 1 km^2 ;
- ▶ M tuiles au premier niveau;
- ▶ ≤ 64 placettes par tuile.

Solution

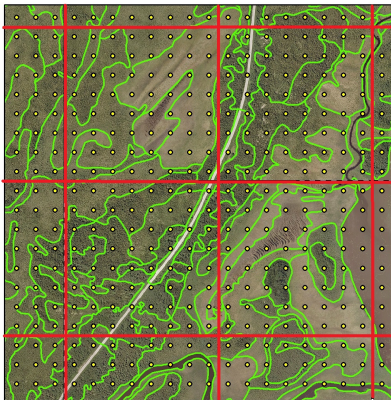


Figure: Ensemble de polygones, tuiles et placettes. Plan à deux degrés.

- ▷ Premier degré: m tuiles échantillonnées;
- ▷ Deuxième degré: $n_i = 4$ placettes échantillonnées par tuile.
- ▷ Qu'en est-il des types?

Solution

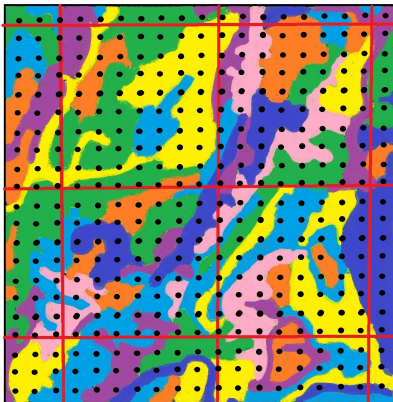


Figure: Example of H types.

- ▷ Premier degré: m tuiles échantillonnées;
- ▷ Deuxième degré: $n_{.i} = 4$ placettes échantillonnées par tuile.
- ▷ Défi: respecter les tailles d'échantillon par type dans un plan à deux degré:
 n_1 . plots of type 1
...
 n_H . plots of type H

Problématique

Table: Plans de sondage.

Plan de sondage	Information auxiliaire utilisée	
	Tuiles	Types
Stratifié sur les types		✓
Deux-degrés	✓	
Deux-degrés, comme stratifié sur les types	✓	✓

Exigences

Exigences

- (i) Plan de sondage à deux degrés;
- (ii) Sélection de $n_i = 4$ placettes dans chaque tuile;
- (iii) $\pi_k = n_{h.}/N_{h.}$, si la placette k est de type h .
 $\Rightarrow$ L'espérance du nombre de placettes du type h est $n_{h.}$

π_k : probabilité de sélectionner la placette k dans l'échantillon final.

π_{1i} : probabilité de sélectionner la tuile i au premier degré.

$\pi_{k|i}$: probabilité de sélectionner la placette k au deuxième degré, si la tuile i a été tirée (k est dans i).

Exigences

Exigences

- (i) Plan de sondage à deux degrés;
- (ii) Sélection de $n_i = 4$ placettes dans chaque tuile;
- (iii) $\pi_k = n_h./N_{h.}$, si la placette k est de type h .
 $\Rightarrow$ L'espérance du nombre de placettes du type h est n_h .

On veut définir π_{1i} et $\pi_{k|i}$ tel que $\pi_k = \pi_{1i}\pi_{k|i} = n_h./N_{h.}$.

Exigences

Exigences

- (i) Plan de sondage à deux degrés;
- (ii) Sélection de $n_{.i} = 4$ placettes dans chaque tuile;
- (iii) $\pi_k = n_{h.}/N_{h.}$, si la placette k est de type h .
 $\Rightarrow$ L'espérance du nombre de placettes du type h est $n_{h.}$

Probabilités d'inclusion

$$\pi_{1i} = \frac{1}{4} \sum_{h=1}^H \frac{n_{h.}}{N_{h.}} N_{hi},$$

$$\pi_{k|i} = \frac{4 \times n_{h.}/N_{h.}}{\sum_{h=1}^H n_{h.}/N_{h.} \times N_{hi}},$$

pour k dans la tuile i et $i = 1, \dots, M$.

Table des matières

Contexte de l'inventaire forestier

Plan de sondage pour inventaire forestier

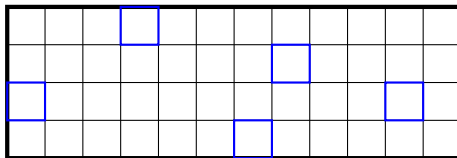
Premier degré

Second degré

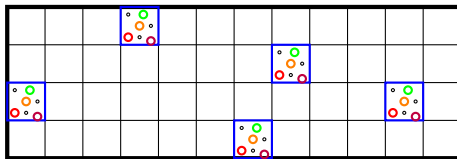
Échantillon étalé spatialement

Illustration numérique

1. Premier degré: Sélectionner un échantillon de tuiles assurant une distribution des types adéquate pour le degré 2.



2. Second degré: Sélectionner 4 placettes par tuiles choisies en 1. tel que globalement, on obtient n_h placettes de type h .



Premier degré - Échantillonnage équilibré

- Soit U_T la population de tuile.
Il faut un échantillon de m tuiles.



S'il n'est pas choisi avec soins, rien n'assure qu'il y aura suffisamment de placettes de chaque type pour le second degré.

- **Solution:** Échantillon équilibré
 - assure la sélection de m tuiles,
 - assure une bonne distribution du type dans l'échantillon de tuiles,
 - assure une cohérence dans l'estimation sur des variables auxiliaires.

Échantillonnage équilibré

Soit le vecteur de p variables auxiliaires $\mathbf{x}_i$ pour $i \in U_T$.
Un plan de sondage est équilibré sur les variables $\mathbf{x}_i$ si

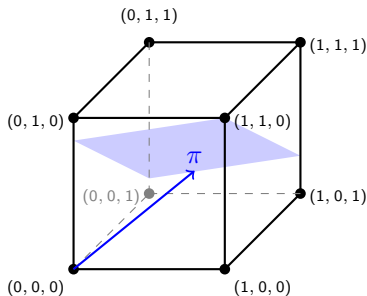
$$\hat{\mathbf{X}} = \sum_{i \in s_T} \frac{\mathbf{x}_i}{\pi_i} = \sum_{i \in U_T} \mathbf{x}_i = \mathbf{X}.$$

La méthode du cube (Deville et Tillé, 2004):

- ▶ Sélectionne aléatoirement un échantillon (approximativement) équilibré;
- ▶ Assure la cohérence dans l'estimation de $\mathbf{X}$;
- ▶ Assure une taille d'échantillon fixe si $\sum_{i \in U_T} \pi_i$ est un entier.

Échantillonnage équilibré

La méthode du cube de Deville et Tillé (2004):
Séquence de vecteurs aléatoires mettant à jour le vecteur de probabilités d'inclusion $\pi = (\pi_1, \dots, \pi_k, \dots, \pi_N)'$ jusqu'à ce que l'échantillon $\mathbf{s} = (0, 1, 0, \dots, 0, 1)'$ soit sélectionné.



Échantillonnage équilibré

La méthode du cube de Deville et Tillé (2004):

Deux phases:

- ▶ Phase de vol:
 - ▶ Séquence de vecteurs débutant avec π qui respecte toujours les contraintes d'équilibrage.
 - ▶ À chaque étape, une composante du vecteur est changée à 0 ou 1 de façon aléatoire.
 - ▶ Fin de la phase de vol lorsqu'on ne peut plus continuer en respectant les contraintes.
 - ▶ Il reste au maximum p unités avec une composante $\notin \{0, 1\}$.
- ▶ Phase d'atterrissage: Choisir un échantillon qui est aussi équilibré que possible.

Premier degré de l'inventaire

Vecteur d'équilibrage

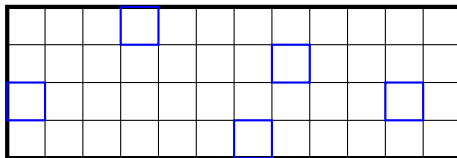
$$\mathbf{x}_{1i} = \left(\pi_{1i}, N_{1i} \frac{n_{1.}}{N_{1.}}, \dots, N_{hi} \frac{n_{h.}}{N_{h.}}, \dots, N_{Hi} \frac{n_{H.}}{N_{H.}}, \mathbf{z}_{1i}^T \right)^T,$$

pour $i = 1, \dots, M$.

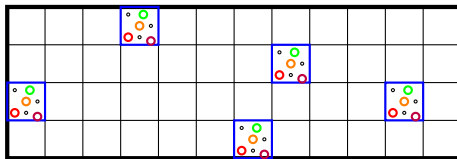
$$\sum_{i \in s_T} \frac{\mathbf{x}_i}{\pi_i} = \sum_{i \in U_T} \mathbf{x}_i.$$

- Fixe le nombre de tuiles échantillonnées à $\sum_{i \in U_T} \pi_{1i} = m$, où m est le nombre de tuiles requis;
- Assure un nombre de placettes fixe pour chaque type;
- Équilibre sur les totaux des bandes de couleur, où $\mathbf{z}_{1i}$ contient les totaux des bandes de couleur dans la tuile i .

1. Premier degré: Sélectionner un échantillon de tuiles assurant une distribution des types adéquate pour le degré 2.



2. Second degré: Sélectionner 4 placettes par tuiles choisies en 1. tel que globalement, on obtient n_h placettes de type h .



Second degré - Échantillon stratifié et équilibré

- ▶ Soit U_i l'ensemble de placettes dans la tuile i .
Il faut un échantillon de 4 placettes par tuile.
- La population de placettes étant donné le premier degré est divisée en m strates (tuiles échantillonnées) U_i .



S'il n'est pas choisi avec soins, rien n'assure que globalement, on obtiendra n_h placettes de type h .

- ▶ **Solution:** Échantillon équilibré pour population hautement stratifiée
 - assure la sélection de 4 placettes par tuile (strate),
 - assure la sélection de n_h placettes de type h dans l'échantillon global,
 - assure la cohérence dans l'estimation sur des variables auxiliaires.

Second degré - Échantillon stratifié et équilibré

Un plan stratifié est équilibré sur les variables $\mathbf{x}_k$ si

$$\sum_{i \in s_T} \sum_{k \in s_{.i}} \frac{\mathbf{x}_k}{\pi_k} = \sum_{i \in s_T} \sum_{k \in U_{.i}} \mathbf{x}_k, \quad \text{pour } i = 1, \dots, m.$$

Plan équilibré pour population hautement stratifiée
(Hasler et Tillé, 2014) :

- ▶ Sélectionne aléatoirement un échantillon équilibré stratifié;
- ▶ Assure la cohérence dans l'estimation des totaux des variables x entre les strates;
- ▶ Assure la taille d'échantillon fixe dans les strates;

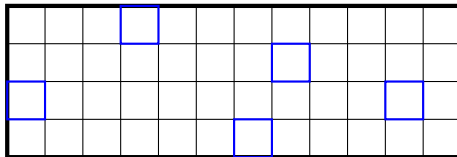
Second degré de l'inventaire forestier

Vecteur d'équilibrage

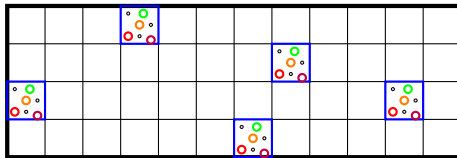
$$\mathbf{x}_{2k} = \left(\mathbb{1}_{k \in U_1} \pi_{k|i}, \dots, \mathbb{1}_{k \in U_H} \pi_{k|i}, \mathbf{z}_k^T \right)^T, \quad \text{pour } k \in U_i.$$

- ▶ Dans les tuiles: assure la sélection de 4 placettes dans la tuile i ;
- ▶ Entre les tuiles: assure la sélection de n_h placettes de type h ;
- ▶ Équilibre sur les totaux des bandes de couleur, où $\mathbf{z}_k$ contient la bande de couleur Landsat de la placette k .

1. Premier degré: Sélectionner un échantillon de tuiles assurant une distribution des types adéquate pour le degré 2.



2. Second degré: Sélectionner 4 placettes par tuiles choisies en 1. tel que globalement, on obtient n_h placettes de type h .



Étalement de l'échantillon

Échantillon doublement équilibré (Grafström et Tillé, 2013):

But: Éviter la sélection de tuiles voisines et de placettes voisines dans les tuiles tout en respectant les contraintes d'équilibrage.

Motivation: Les unités proches peuvent être semblables à cause d'une corrélation spatiale.

- ▶ Échantillon équilibré (Méthode du Cube);
- ▶ Échantillon étalé spatialement (Méthode du Pivot Local).

Échantillon doublement équilibré

Méthode du pivot local (Grafström, Lundström et Schelin, 2012)

Séquence de vecteurs aléatoires mettant à jour π jusqu'à l'obtention d'un échantillon s .

- ▶ Sélection de deux unités voisines k et ℓ .
- ▶ Si $\pi_k + \pi_\ell < 1$, une unité sera rejetée, l'autre recevra la probabilité $\pi_k + \pi_\ell$ d'être choisie.
- ▶ Si $\pi_k + \pi_\ell > 1$, une unité sera choisie dans l'échantillon, l'autre reçoit $\pi_k + \pi_\ell - 1$.
- ▶ Répéter jusqu'à l'obtention de s .

→ Donne des petites probabilités d'inclusion d'ordre 2 aux unités voisines.

Étalement de l'échantillon

Échantillon doublement équilibré (Grafström et Tillé, 2013):

- ▶ Choisir $p + 1$ unités voisines.
- ▶ Lancer la phase de vol de la Méthode du Cube dans les $p + 1$ unités.
→ Au moins une unité est rejetée ou sélectionnée dans s .
- ▶ Recommencer jusqu'à ce qu'il reste moins de $p + 1$ unités.
- ▶ Lancer la phase d'atterrissage de la Méthode du Cube.

Sélection de $p + 1$ unités voisines

- ▶ Choisir une unité avec probabilités égales. Choisir ses p plus proches voisins.
- ▶ Calculer la position moyenne des $p + 1$ unités.
- ▶ Choisir les $p + 1$ unités les plus proches de cette position moyenne.
- ▶ Recommencer jusqu'à stabilité (somme des distances carrées).

Table des matières

Contexte de l'inventaire forestier

Plan de sondage pour inventaire forestier

Illustration numérique

Illustration numérique

Retour au monde réel

Contexte

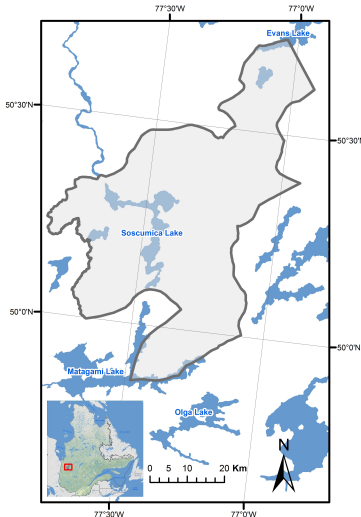


Figure: Région de l'inventaire Québécois étudié dans une étude par simulations

- ▶ 60'913 placettes;
- ▶ 2'522 tuiles;
- ▶ $H = 15$ types;
- ▶ $n = 300$ placettes;
- ▶ $m = 75$ tuiles;
- ▶ Estimation de la *hauteur* moyenne.

Plans de sondage étudiés

Table: Plans de sondage dans l'étude par simulations.

Plan de sondage	Information auxiliaire utilisée		
	Tuiles	Types	Couleurs
Équilibré (stratifié) sur les types		✓	
Deux-degrés non équilibré	✓		
Deux-degrés, équilibré sur couleurs et types	✓	✓	✓

- ▶ Avec et sans étalement spatial;
- ▶ Estimateur de la *hauteur* moyenne:

$$\hat{\bar{Y}} = \frac{\sum_{k \in S} y_k / \pi_k}{\sum_{k \in S} 1 / \pi_k}.$$

Resultats

Table: RRMSE de $\hat{\bar{Y}}$.

Plan de sondage	Sans étaler	En étalant
Équilibré (stratifié) sur les types	0.016	0.015
Deux-degrés non équilibré	0.027	0.024
Deux-degrés, équilibré sur les couleurs et les types	0.018	0.017

$$RRMSE = \frac{\sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\bar{Y}}^{(r)} - \bar{Y})^2}}{\bar{Y}}$$

- ▶ Respecte les contraintes (deux degrés, 4 placettes par tuiles);
- ▶ Propriétés d'un plan stratifié sur les types;
- ▶ Résultats presque aussi bons qu'avec un échantillon stratifié.

Retour au monde réel

Hypothèse pour le plan de sondage précédent:

les placettes présentes dans la population initiale sont toutes observables.

Réalité pour janvier 2017:

le tri des placettes observables/non-observables n'est pas fait avant l'enquête. Il ne peut être réalisé que sur une partie de la population.

Retour au monde réel

Exigences

1. Tirer un échantillon de tuiles 3 fois plus grand que supposé.
2. Trier les placettes en identifiant observable/non-observable.
Cela entraîne l'élimination possible de tuiles (si moins de 4 placettes).
3. Tirer un échantillon de m tuiles.
4. Tirer un échantillon de 4 placettes par tuiles.

- Deville, J.-C. et Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, **91**, 893–912.
- Grafström, A., Lundström, N. L. P. et Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, **68**, 514–520.
- Grafström, A. et Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, **14**, 120–131.
- Hasler, C. et Tillé, Y. (2014). Fast balanced sampling for highly stratified population. *Computational Statistics and Data Analysis*, **74**, 81–94.