

INCORPORATION DE CONTRAINTES SPATIALES ET OPÉRATIONNELLES DANS DES PLANS D'ÉCHANTILLONNAGE POUR DES INVENTAIRES FORESTIERS

Audrey-Anne Vallée ¹, Bastien Ferland-Raymond ², Louis-Paul Rivest ³ & Yves Tillé ¹

¹ *Institut de Statistique, Université de Neuchâtel, Bellevaux 51, 2000 Neuchâtel, Suisse. (audrey-anne.vallee@unine.ch, yves.tille@unine.ch)*

² *Direction des inventaires forestiers, Ministère des forêts, de la faune et des parcs du Québec, Québec, Canada. (bastien.ferland-raymond@mffp.gouv.qc.ca)*

³ *Département de mathématiques et de statistique, Université Laval, 1045 rue de la médecine, Québec, Canada. (louis-paul.rivest@mat.ulaval.ca)*

Résumé. Un but des inventaires forestiers est l'étude de l'évolution temporelle de la forêt et des écosystèmes instables. Dans la province du Québec, la forêt est examinée par l'entremise d'inventaires réguliers. La région d'intérêt couvre un grand territoire qui n'est pas accessible facilement et qui est composé de plusieurs types d'arbres. Les exigences opérationnelles par rapport aux coûts de la collecte, à la répartition et au type des placettes sélectionnées sont difficiles à gérer. Dans plusieurs inventaires, les placettes sont sélectionnées à l'aide d'un plan de sondage systématique à deux dimensions. L'information auxiliaire rendue disponible par les différentes techniques d'imagerie n'est utilisée qu'à l'étape de l'estimation. Le but de cette présentation est de montrer que les techniques d'échantillonnage modernes peuvent être utilisées pour sélectionner un échantillon et pour améliorer la planification des inventaires forestiers, et ce, malgré les nombreuses exigences. En effet, notre plan de sondage inclut l'échantillonnage équilibré, l'échantillonnage équilibré hautement stratifié et l'étalement de l'échantillon à travers un plan à deux degrés. L'impact de ces techniques sur la satisfaction des exigences opérationnelles et sur la précision des estimations est étudié à l'aide des données d'un inventaire forestier québécois.

Mots-clés. Échantillonnage hautement stratifié, échantillonnage équilibré spatialement, inventaire forestier, probabilités inégales.

Abstract. The goals of forest inventories include understanding the forest temporal evolution and monitoring fragile ecosystems. In the province of Quebec, the forest is examined through regular inventories. The area of interest covers a large territory which is not easily accessible and which has various types of tree species. Operational requirements for the spreading and the type of the sampled plots and for the cost of the sampling are difficult to manage. In many inventories, the sampled plots are selected using two dimensional systematic sampling. The rich auxiliary information available through various types of imaging techniques is used only at the estimation stage of the survey. The goal

of this presentation is to show that modern and advanced sampling techniques can be used to select a sample and to improve the planning of the forest inventories, even if there are many requirements. Indeed, our sampling design includes balanced sampling, highly stratified balanced sampling and sample spreading through a two stage sample. The impact of these novel techniques on the satisfaction of the operational requirements and on the precision of survey estimates is investigated using field data from a Quebec inventory.

Keywords. Forest inventory, balanced sampling, stratified sampling, spatially balanced sampling, unequal probability sampling.

1 Inventaire forestier

Les inventaires forestiers sont des sondages collectant des données sur la composition de la forêt. Les espèces d'arbres, la hauteur et le diamètre des arbres, la santé et la qualité de la forêt sont des aspects étudiés. Ces études peuvent avoir pour but de surveiller l'évolution et l'écologie des forêts. Dans plusieurs inventaires forestiers, les unités d'échantillonnage sont sélectionnées avec un plan de sondage systématique à deux dimensions. Par conséquent, il peut être difficile de respecter certaines contraintes opérationnelles et l'information auxiliaire fournie par les techniques d'imagerie n'est pas utilisée pour construire le plan de sondage. Dans cette présentation, des contraintes opérationnelles et de l'information auxiliaire sont pris en compte pour construire le plan de sondage d'un inventaire forestier dans la province du Québec.

La région d'intérêt de l'inventaire forestier couvre un vaste territoire d'environ 600 000 km². La majorité de la forêt étudiée est située au Nord de la région habitée du Québec, elle est donc difficilement accessible. De plus, la forêt est très variée étant donné qu'elle couvre plusieurs zones écologiques. D'autre part, deux à trois ans avant l'inventaire forestier, le territoire est photographié et divisé en polygones homogènes par des interprètes. Pour chaque polygone, certaines variables sont disponibles telles la hauteur des arbres et les espèces d'arbres sur ce polygone. Avec cette information, H types de polygones sont déterminés. Ils serviront à déterminer des tailles d'échantillons par type étant donné que certains types de polygone sont plus hétérogènes que d'autres et que certains sont plus rares que d'autres.

Les unités échantillonnées sont des placettes circulaires de 400 m² centrées en des points sélectionnés aléatoirement sur le territoire. Les placettes échantillonnées sont visitées par une équipe collectant des mesures telles le volume et la hauteur des arbres. Au moment de l'élaboration du plan de sondage, l'information donnée pour les polygones est disponible comme variable auxiliaire. D'autres informations auxiliaires sont fournies par imagerie spectrale. Deux contraintes sont à respecter lors de l'élaboration du plan de sondage. La première contrainte est établie afin de respecter un budget et un temps d'enquête. Il est connu qu'une équipe d'enquêteurs est capable d'étudier quatre placettes

dans la même journée de travail si ces placettes sont dans un carré de 1 km². Une grille est donc tracée sur le territoire afin d'obtenir un plan à deux degrés. Premièrement, un échantillon de carrés doit être choisi, puis quatre placettes par carré choisi doivent être sélectionnées. La deuxième contrainte concerne les types de polygones. Des tailles échantillonnales de placettes sont déterminées par type de polygones afin de surreprésenter certains polygones. Ainsi, le plan de sondage doit avoir une caractéristique d'un plan stratifié. Il doit pouvoir assurer un nombre fixe de placettes pour chaque type.

2 Plan de sondage

Dans cette présentation, un plan de sondage est élaboré pour respecter les deux contraintes opérationnelles et pour intégrer l'information auxiliaire disponible pour l'inventaire forestier. Ce plan de sondage est un plan à deux degrés. Au premier degré, des carrés de 1 km² sont sélectionnés. L'échantillonnage équilibré (Deville et Tillé, 2004) est utilisé pour obtenir une bonne répartition des types parmi les carrés choisis. Des variables d'équilibrage sont construites afin d'assurer la taille échantillonnale fixe de carrés choisis, afin d'avoir assez de placettes de chaque type dans ces carrés et afin d'utiliser l'information auxiliaire fournie par imagerie spectrale. Au second degré, quatre placettes sont choisies dans chaque carré sélectionné. Pour respecter les tailles échantillonnales fixes par type de polygones, l'échantillonnage équilibré pour populations hautement stratifiées (Hasler et Tillé, 2014) est utilisé. Des variables d'équilibrage sont construites afin d'assurer la sélection de quatre placettes par carré et des tailles échantillonnales fixes par type et afin d'utiliser l'information auxiliaire fournie par imagerie spectrale. De plus, pour éviter la sélection d'unités voisines, il serait intéressant d'étaler spatialement l'échantillon. Les unités voisines étant semblables, un maximum d'information peut être collecté si l'échantillon est bien étalé. Pour étaler l'échantillon en respectant des contraintes d'équilibrage, l'échantillonnage doublement équilibré est utilisé (Grafström et Tillé, 2013) aux deux degrés.

Le plan de sondage proposé permet donc de respecter les contraintes opérationnelles et d'utiliser de l'information auxiliaire. Une étude par simulation permet d'évaluer l'impact des techniques de sondage suggérées sur la précision des estimations et sur la satisfaction des contraintes. Cette étude par simulations est réalisée avec des données d'un inventaire forestier québécois. La région étudiée couvre 2 522 carrés et 60 913 placettes. Il y a $H = 15$ types de polygones et la hauteur moyenne \bar{Y} doit être estimée par un estimateur $\hat{\bar{Y}}$. Un échantillon de 300 placettes, donc de 75 carrés est choisi. Trois plans de sondages sont comparés. Le premier plan de sondage, est celui qui serait utilisé s'il n'y avait pas de contraintes de budget et de temps d'enquête, pour obtenir des tailles échantillonnales fixes par type: le plan de sondage stratifié par type de polygones. Le deuxième plan est celui qui résulte de l'utilisation des carrés : le plan de sondage à deux degrés. Le troisième plan est celui proposé respectant les contraintes de budget, de temps et de tailles échantillonnales

Table 1: RREQM de \widehat{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

par type: un plan à deux degrés avec des contraintes d'équilibrage. Ces trois plans sont répétés deux fois, une fois sans étaler spatialement et une fois en étalant. Suite aux simulations, la racine de l'erreur quadratique moyenne relative (RREQM) de l'estimateur de la moyenne est calculée dans chaque cas,

$$RREQM = \frac{\sqrt{\frac{1}{R} \sum_{r=1}^R (\widehat{Y}^{(r)} - \bar{Y})^2}}{\bar{Y}},$$

où R est le nombre de simulations. La Table 1 contient les résultats de l'étude. La valeur du RREQM pour le plan à deux degrés (0.027) est nettement plus élevée que celle du plan stratifié (0.016). Cependant, en incluant des contraintes d'équilibrage, le résultat s'améliore (0.018) et ressemble à celui du plan stratifié. Le gain de l'étalement dans les plans équilibrés est faible, il est un peu plus grand dans le cas du plan à deux degrés (0.027 à 0.024). Ceci peut s'expliquer par l'importance du gain de l'équilibrage.

Le plan de sondage proposé permet de respecter les contraintes budgétaires et temporelles avec un plan à deux degrés et la sélection de quatre placettes par carrés. Il permet aussi d'avoir la propriété d'un plan stratifié et de choisir un nombre fixe de placettes par type de polygones. En ajoutant l'étalement spatial de l'échantillon, la précision des estimations avec le plan proposé est presque celle du plan souhaité mais irréaliste, le plan stratifié.

Bibliographie

- [1] Deville, J.-C. et Tillé, Y. (2004), Efficient balanced sampling: The cube method, *Biometrika*, 91, 893–912.
- [2] Hasler, C. et Tillé, Y. (2014), Fast balanced sampling for highly stratified population, *Computational Statistics and Data Analysis*, 74, 81–94.
- [3] Grafström, A. et Tillé, Y. (2013), Doubly balanced spatial sampling with spreading and restitution of auxiliary totals, *Environmetrics*, 14, 120–131.