Journées de méthodologie statistique de l'Insee 2022

Une généralisation de la méthode de partage des poids dans le cas où la base de sondage est continue

Philippe BRION (*), Olivier BOURIAUD (**), Guillaume CHAUVET (***)

(*) Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France
(**) IGN, Laboratoire de l'Inventaire Forestier (***) Univ Rennes, ENSAI, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

philippe.brion55@gmail.com

Mots-clés. : Inventaire forestier, plan de sondage en population continue.

Domaines. Échantillonnage

Résumé

La définition de l'unité statistique utilisée dans les enquêtes statistiques est une question difficile : les différents "univers" enquêtés n'ont pas nécessairement une base de sondage directement utilisable, et il arrive que l'on utilise des unités à échantillonner d'une nature différente de celle des unités observées. La production d'estimations statistiques pose alors des problèmes méthodologiques complexes, qui peuvent être traités en utilisant la méthode dite du partage des poids, formalisée par [3]. Cette méthode est basée sur les liens existant entre les deux populations : population échantillonnée et population observée. Cependant, les deux populations considérées dans cette approche sont des populations discrètes. Pour certains domaines d'étude, en particulier liés à des aspects environnementaux, la population échantillonnée est une population continue : c'est par exemple le cas des inventaires forestiers pour lesquels, fréquemment, les arbres enquêtés sont ceux situés sur des placettes dont les centres sont des points tirés de manière aléatoire dans une zone donnée. La production d'estimations statistiques à partir de l'échantillon d'arbres enquêtés pose alors des difficultés de méthode, ainsi que les calculs de variance associés. L'objet de ce papier est de procéder à une généralisation de la méthode de partage des poids au cas continu (population échantillonnée) - discret (population enquêtée), à partir de la formalisation proposée par [2] sur l'extension de l'estimateur de Horvitz-Thompson au tirage de points réalisé dans un univers continu.

Abstract

The issue of the definition of the statistical unit is very pregnant in the domain of sample surveys. Not all the populations surveyed have a readily available sampling frame. For some populations, the sampled units are distinct from the observation units. Producing estimations on the population of interest raises complex questions, which can be addressed by using the weight

share method of [3]. However, the two populations considered in this approach are discrete. In some fields of study, the sampled population is continuous : this is for example the case of forest inventories for which, frequently, the trees surveyed are those located on plots of which the centers are points randomly drawn in a given area. The production of statistical estimates from the sample of trees surveyed then poses methodological difficulties, as do the associated variance calculations. The purpose of this paper is to generalize the weight-share method to the continuous (sampled population) - discrete (surveyed population) case, from the extension proposed by [2] of the Horvitz-Thompson estimator for drawing points carried out in a continuous universe.

1 Rappels concernant le cas où la population échantillonnée est discrète

1.1 Notations

Nous nous intéressons à une population finie, discrète U^A de taille N^A pour laquelle une base de sondage est disponible. Nous considérons une variable d'intérêt y^A prenant la valeur y_i^A pour une unité $i \in U^A$, et nous voulons estimer le total sur la population U^A

$$\tau_y^A = \sum_{i \in U^A} y_i^A. \tag{1}$$

Un échantillon S^A est tiré dans U^A selon un plan de sondage $p^A(\cdot)$, avec π_i^A la probabilité de l'unité i d'appartenir à l'échantillon. L'estimateur de Horvitz-Thompson (HT) vaut

$$\hat{\tau}_y^A = \sum_{i \in S^A} d_i^A y_i^A, \tag{2}$$

avec $d_i^A = 1/\pi_i^A$ le poids de sondage de l'unité i. Cet estimateur est un estimateur sans biais de τ_y^A, sous la condition que tous les π_i^A sont > 0.

La variance de $\hat{\tau}_y^A$ vaut

$$V\left(\hat{\tau}_y^A\right) = \sum_{i,j \in U^A} \frac{y_i^A}{\pi_i^A} \frac{y_j^A}{\pi_j^A} \left(\pi_{ij}^A - \pi_i^A \pi_j^A\right), \tag{3}$$

π_{ij}^A étant la probabilité que les unités i and j soient sélectionnées simultanément dans S^A. Si tous les π_{ij}^A sont strictement positifs, cette variance est estimée sans biais par

$$\hat{V}\left(\hat{\tau}_y^A\right) = \sum_{i,j \in S^A} \frac{y_i^A}{\pi_i^A} \frac{y_j^A}{\pi_j^A} \left(\frac{\pi_{ij}^A - \pi_i^A \pi_j^A}{\pi_{ij}^A}\right). \tag{4}$$

1.2 La méthode du partage des poids

Supposons que nous nous intéressons à une autre population U^B, avec une variable d'intérêt y^B prenant la valeur y_k^B pour l'unité $k \in U^B$. Nous voulons estimer le total de cette population

$$\tau_y^B = \sum_{k \in U^B} y_k^B. \tag{5}$$

Nous nous situons dans un cas où aucune base de sondage n'est disponible pour U^B, mais où cette population est liée à la population U^A. Plus précisément, le lien entre les unités de U^A et

U^B peut être représenté par les variables indicatrices

$$L_{ik}^{AB} = \begin{cases} 1 & \text{si les unités } i \in U^A \text{ et } k \in U^B \text{ sont liées,} \\ 0 & \text{sinon.} \end{cases} \tag{6}$$

L'ensemble des *ancêtres* d'une unité $k \in U^B$ est $Anc_k = \{i \in U^A;\ L_{ik}^{AB} = 1\}$. L'ensemble des *descendants* d'une unité $i \in U^A$ est $Des_i = \{k \in U^B;\ L_{ik}^{AB} = 1\}$. Pour toute unité $k \in U^B$,

$$N_{+k}^{AB} = \sum_{i \in U^A} L_{ik}^{AB} \tag{7}$$

est le nombre total d'ancêtres. Il est nécessaire que toute unité $k \in U^B$ soit liée à au moins une unité de U^A ; ce qui veut dire que $N_{+k}^{AB} > 0$ pour toute unité $k \in U^B$. Plus généralement, l'indicateur de lien L_{ik}^{AB} peut être remplacé par une constante positive θ_{ik}^{AB} [3], et ces constantes θ_{ik}^{AB} peuvent être utilisées pour chercher un estimateur optimal en minimisant la variance, cf [5].

Un échantillon S^B est obtenu dans U^B en enquêtant l'ensemble des descendants des unités i échantillonnées dans S^A. De manière formelle, nous avons

$$S^B = \bigcup_{i \in S^A} Des_i. \tag{8}$$

Pour obtenir un estimateur de τ_y^B, la méthode de partage des poids ([3]) part du principe de dualité entre les populations U^A et U^B, basé sur la fonction de lien présentée en (6). Le total τ_y^B peut être écrit comme

$$\tau_y^B = \sum_{i \in U^A} y_i^A \quad \text{avec} \quad y_i^A = \sum_{k \in U^B} \frac{L_{ik}^{AB}}{N_{+k}^{AB}} y_k^B, \tag{9}$$

voir le Résultat 2 dans [3]. L'équation (9) utilise le fait que la variable y_k^B peut être distribuée sur les unités de U^A pour obtenir une variable synthétique y_i^A. Ceci étant réalisé en partageant chaque valeur y_k^B de manière égale entre les ancêtres dans Anc_k.

A partir de l'équation (9), le total τ_y^B peut être estimé sans biais en utilisant l'estimateur HT sur l'échantillon S^A, pour la variable synthétique y_i^A. Cet estimateur HT peut être ré-écrit comme

$$\begin{aligned} \hat{\tau}_y^B &= \sum_{i \in S^A} d_i^A y_i^A = \sum_{k \in S^B} w_k^B y_k^B, \qquad (10) \\ \text{avec } w_k^B &= \frac{1}{N_{+k}^{AB}} \sum_{i \in S^A} L_{ik}^{AB} d_i^A, \end{aligned}$$

voir le Résultat 3 dans [3]. Chaque unité $k \in S^B$ se voit affecter la somme des poids des unités échantillonnées $i \in S^A$ qui sont liées à k, somme divisée par le nombre de liens N_{+k}^{AB}. Les poids d_i^A des unités $i \in S^A$ sont donc partagés entre les unités $k \in S^B$, d'où le nom de la méthode. Il est important de noter que les poids w_k^B peuvent être calculés seulement si le nombre d'ancêtres N_{+k}^{AB} est connu pour toute unité $k \in S^B$. Cette information doit donc être collectée lors de l'enquête.

A partir de l'équation (10), la méthode du partage des poids permet d'attribuer à chaque unité $k \in S^B$ un poids w_k^B utilisable pour toute variable d'intérêt y_k^B et tel que l'estimateur $\hat{\tau}_y^B$ est sans biais. Ceci constitue une propriété très forte. En revanche, l'estimateur HT relatif à l'échantillon S^B ne peut pas être calculé. La probabilité d'inclusion d'une unité k dans l'échantillon S^B vaut

$$Pr(k \in S^B) = \sum_{\substack{s^A \subset U^A \\ s^A \cap Anc_k \neq \emptyset}} p^A(s^A). \tag{11}$$

Calculer les valeurs de ces probabilités d'inclusion nécessiterait de connaître à la fois le plan de sondage $p^A(\cdot)$ et les liens complets entre les deux populations, ce qui est en général impossible.

Puisque $\hat{\tau}_y^B$ peut être écrit comme un estimateur HT sur l'échantillon S^A, la variance de $\hat{\tau}_y^B$ est donnée par l'équation (3), où y_i^A est la variable synthétique définie en (9), et un estimateur de cette variance est donné par (4). Notons que la variable y_i^A peut être calculée pour toute unité $i \in S^A$, puisqu'on a supposé que toutes les unités de U^B liées aux unités de S^A sont enquêtées, voir l'équation (8).

2 Cas où la population échantillonnée est continue

2.1 Notations

Nous rappelons d'abord le cadre de référence relatif au sondage et aux estimateurs associés pour un univers continu $\mathcal{U}^A$ qui a été présenté par [2]. Ce cadre de référence a été proposé pour répondre à des problématiques environnementales (échantillonnage dans un territoire géographique supposant le tirage de points dans un plan), et a conduit à une généralisation de l'estimateur de Horvitz Thompson existant pour des populations discrètes, voir section 1.1.

Supposons que l'univers $\mathcal{U}^A$ est inclus dans $\mathbf{R}^q$, avec $q \geq 1$. Nous considérons une fonction intégrable au sens de Lebesgue $y^A : \mathcal{U}^A \to \mathbf{R}$, et nous voulons estimer le total de cette fonction sur $\mathcal{U}^A$.

$$\tau_y^A \quad = \quad \int_{\mathcal{U}^A} y^A(x)dx. \tag{12}$$

Il faut remarquer que y^A est ici une densité. Par exemple, si l'on veut estimer le volume total de neige sur un lac gelé, $y^A(x)$ sera la hauteur de neige au point x.

Un échantillon $S^A = \{s_1^A, \ldots, s_{n^A}^A\}$ de n^A points s_i^A est sélectionné aléatoirement dans $\mathcal{U}^A$. Nous supposons l'existence d'une fonction de densité de probabilité jointe (PDF)

$$f(s_1^A, \ldots, s_{n^A}^A) \tag{13}$$

pour les unités échantillonnées, ainsi que l'existence de fonctions de densité marginale et jointe

$$f_i(s_i^A) \quad \text{and} \quad f_{ij}(s_i^A, s_j^A) \tag{14}$$

de s_i^A et s_j^A pour $i \neq j$. Par exemple, si l'échantillon S^A résulte de n^A tirages indépendants, réalisés de manière uniforme dans $\mathcal{U}^A$, nous avons

$$f_i(s_i^A) \quad = \quad \frac{1(s_i^A \in \mathcal{U}^A)}{M^A} \tag{15}$$

et $f(s_1^A, \ldots, s_{n^A}^A) = \prod_{i=1}^{n^A} f_i(s_i^A)$, avec $M^A = \int_{\mathcal{U}^A} dx$ la mesure globale de l'univers, $1(\cdot)$ étant la fonction indicatrice.

Supposons que la densité de probabilité est absolument continue relativement à la mesure de Lebesgue. Comme l'ont remarqué [2] et [8], on trouve des plans de sondage pour lesquels cette hypothèse n'est pas vérifiée. Pour tout point $x \in \mathcal{U}^A$, la densité d'inclusion est définie par

$$\pi^A(x) \quad = \quad \sum_{i=1}^{n^A} f_i(x). \tag{16}$$

Ce qui peut être vu comme une mesure locale du nombre de points échantillonnés par unité de mesure. Nous avons en particulier $\int_{\mathcal{U}^A} \pi^A(x)dx = n^A$, ce qui est une propriété classique des plans de sondage de taille fixe n^A. De la même manière, la densité d'inclusion jointe est définie par

$$\pi^A(x, x') \quad = \quad \sum_{i=1}^{n^A} \sum_{\substack{j=1 \\ j \neq i}}^{n^A} f_{ij}(x, x') \tag{17}$$

pour $x, x' \in \mathcal{U}^A$.

L'estimateur HT de τ_y^A vaut

$$\hat{\tau}_y^A \quad = \quad \sum_{s \in S^A} d^A(s) y^A(s), \tag{18}$$

avec $d^A(x) = 1/\pi^A(x)$ le poids de sondage du point x. Cet estimateur est sans biais pour τ_y^A, sous condition que $\pi^A(x) > 0$ presque partout, voir le Théorème 1 dans [2].

Si la fonction $y^A(\cdot)$ est bornée et $\int_{\mathcal{U}^A} \{1/\pi^A(x)\} dx < \infty$, la variance de $\hat{\tau}_y^A$ est donnée par la formule de Horvitz-Thompson

$$\begin{aligned} V\left(\hat{\tau}_y^A\right) \quad &= \quad \int_{x \in \mathcal{U}^A} \frac{\{y^A(x)\}^2}{\pi^A(x)} dx \\ &+ \quad \int_{x \in \mathcal{U}^A} \int_{x' \in \mathcal{U}^A} \left\{\pi^A(x, x') - \pi^A(x)\pi^A(x')\right\} \frac{y^A(x)}{\pi^A(x)} \frac{y^A(x')}{\pi^A(x')} dx dx', \end{aligned} \tag{19}$$

ou de manière équivalente par celle de Sen-Yates-Grundy

$$V\left(\hat{\tau}_y^A\right) = \frac{1}{2} \int_{x \in \mathcal{U}^A} \int_{x' \in \mathcal{U}^A} \left\{\pi^A(x)\pi^A(x') - \pi^A(x, x')\right\} \left\{\frac{y^A(x)}{\pi^A(x)} - \frac{y^A(x')}{\pi^A(x')}\right\}^2 dx dx', \tag{20}$$

voir la Section 2 de [2]. Les estimateurs de variance associés sont respectivement

$$\begin{aligned} \hat{V}_{HT}\left(\hat{\tau}_y^A\right) \quad &= \quad \sum_{s \in S^A} \left\{\frac{y^A(s)}{\pi^A(s)}\right\}^2 \\ &+ \quad \sum_{s \in S^A} \sum_{\substack{s' \in S^A \\ s' \neq s}} \left\{\frac{\pi^A(s, s') - \pi^A(s)\pi^A(s')}{\pi^A(s, s')}\right\} \frac{y^A(s)}{\pi^A(s)} \frac{y^A(s')}{\pi^A(s')} \end{aligned} \tag{21}$$

et

$$\hat{V}_{YG}\left(\hat{\tau}_y^A\right) = \frac{1}{2} \sum_{s \in S^A} \sum_{\substack{s' \in S^A \\ s' \neq s}} \left\{\frac{\pi^A(s)\pi^A(s') - \pi^A(s, s')}{\pi^A(s, s')}\right\} \left\{\frac{y^A(s)}{\pi^A(s)} - \frac{y^A(s')}{\pi^A(s')}\right\}^2. \tag{22}$$

Si de plus à la fois $\pi^A(x) > 0$ et $\pi^A(x, x') > 0$ presque partout sur $\mathcal{U}^A$, alors les deux estimateurs de variance sont sans biais, voir le Théorème 2 de [2]. Il faut noter que la condition sur $\pi^A(x, x')$ peut ne pas être respectée, c'est par exemple le cas quand on utilise un sondage systématique.

2.2 La méthode de partage des poids : le cas continu - discret

Nous supposons que nous nous intéressons à la population U^B et à l'estimation du total τ_y^B donné dans l'équation (5). Les unités de U^B ne sont pas échantillonnées de façon directe, mais un univers continu $\mathcal{U}^A$ lié à U^B est utilisé pour l'échantillonnage. C'est par exemple le cas si nous nous intéressons à une population U^B d'arbres situés dans un territoire $\mathcal{U}^A$. Un échantillon de points est tiré dans $\mathcal{U}^A$ à partir d'un plan de sondage sur un univers continu, et des placettes circulaires ayant pour centre ces points sont ensuite dessinées, en utilisant un rayon pré-défini. L'ensemble des arbres situés dans ces placettes devront alors faire l'objet de l'enquête. Un exemple utilisant cette méthode est détaillé dans la section suivante.

Les liens entre les unités des populations $\mathcal{U}^A$ et U^B sont représentés par la fonction indicatrice

$$L_k^{AB}(x) = \begin{cases} 1 & \text{si } x \in \mathcal{U}^A \text{ et } k \in U^B \text{ sont liés,} \\ 0 & \text{sinon.} \end{cases} \tag{23}$$

Nous gardons la même terminologie que dans la section 1.2, and notons Anc_k pour l'*ensemble des ancêtres* d'un $k \in U^B$, et $Des(x)$ pour l'*ensemble des descendants* d'un $x \in \mathcal{U}^A$. Pour tout $k \in U^B$,

$$M_{+k}^{AB} = \int_{x \in \mathcal{U}^A} L_k^{AB}(x) dx \tag{24}$$

est la mesure de l'ensemble des ancêtres de l'unité k. Comme pour le cas discret, nous supposons que $M_{+k}^{AB} > 0$ pour tout $k \in U^B$.

Un échantillon S^B est obtenu dans U^B en enquêtant tous les descendants des points tirés dans S^A. Nous pouvons écrire

$$S^B = \bigcup_{s \in S^A} Des(s). \tag{25}$$

Pour obtenir un estimateur de τ_y^B, nous établissons à nouveau un principe de dualité entre les populations $\mathcal{U}^A$ et U^B. Ceci est résumé dans la proposition 1. Ce principe de dualité est semblable à celui obtenu dans le cas discret : chaque valeur y_k^B est partagée de manière égale sur les points appartenant à l'ensemble des ancêtres de k. La fonction synthétique $y^A(x)$ peut être interprétée comme une mesure locale de densité de la variable y^B par unité de mesure. Cette approche a déjà été étudiée dans le cadre des inventaires forestiers, voir par exemple la Section 4 de [6] et le chapitre 10 de [4]. Ce point est développé dans la section 3.

Proposition 1. *Le total τ_y^B peut se réécrire*

$$\begin{aligned} \tau_y^B &= \int_{x \in \mathcal{U}^A} y^A(x) dx \qquad (26) \\ \textit{avec } y^A(x) &= \sum_{k \in U^B} \frac{L_k^{AB}(x) y_k^B}{M_{+k}^{AB}}. \end{aligned}$$

La proposition 1 permet d'écrire τ_y^B comme une intégrale sur l'univers $\mathcal{U}^A$, et donc de pouvoir utiliser l'estimateur généralisé de Horvitz Thompson présenté dans (18). Cet estimateur peut également être écrit comme une somme pondérée sur l'échantillon S^B. Ceci est résumé dans la proposition 2.

Proposition 2. *Le total τ_y^B peut être estimé dans biais par*

$$\begin{aligned} \hat{\tau}_y^B &= \sum_{k \in S^B} w_k^B y_k^B \qquad (27) \\ \textit{avec } w_k^B &= \frac{1}{M_{+k}^{AB}} \sum_{s \in S^A} L_k^{AB}(s) d^A(s). \end{aligned}$$

La méthode de partage des poids fournit donc une solution pour l'estimation de τ_y^B, par l'intermédiaire d'un estimateur pondéré calculé sur l'échantillon S^B, où les poids w_k^B sont donnés dans l'équation (27). Chaque unité $k \in S^B$ se voit affecter la somme des poids des points échantillonnés $s \in S^A$ qui sont liés à k, divisée par M_{+k}^{AB}, la mesure de l'ensemble des ancêtres de l'unité k. Le principe est donc le même que celui utilisé dans le cas de populations discrètes. Il est important de remarquer que, pour toute unité $k \in S^B$, il est nécessaire de connaitre la mesure M_{+k}^{AB} de l'ensemble de ses ancêtres.

Puisque $\hat{\tau}_y^B$ peut être écrit comme l'estimateur HT sur l'échantillon S^A, la variance est obtenue à partir de l'équation (19) ou l'équation (20), en utilisant la variable synthétique $y^A(x)$ donnée dans l'équation (26). Un estimateur de variance peut alors être proposé, à partir de l'équation (21) ou l'équation (22).

3 Discussion

Il existe différentes raisons pour lesquelles une population n'est pas directement "atteignable" via une base de sondage. Quand il existe une population discrète pour laquelle une base de sondage est disponible et que cette population est liée à la population qu'on veut enquêter, la méthode classique de partage de poids [3] permet d'utiliser des échantillons probabilistes, et de produire des estimateurs non biaisés, ainsi que des estimateurs de variance. Nous avons montré dans cet article que cette approche peut être généralisée quand la population d'intérêt est liée à une population continue, par l'intermédiaire d'une fonction synthétique sur cette population continue qui peut être interprétée comme une mesure locale de densité.

Cette approche a déjà été utilisée dans le domaine des inventaires forestiers par [6] et [4], par exemple. Dans la Section 4.2 de [6], Mandallaz étudie le cas où un échantillon d'arbres est obtenu à partir d'un échantillon de points à l'intérieur d'une zone donnée, points qui servent à dessiner des placettes circulaires centrées sur ceux et servant à déterminer les arbres à enquêter. La variable synthétique présentée dans l'équation (4.5) est appelée "densité locale". De manière plus générale, le chapitre 10 de [4] introduit l'approche dite de l'intégration Monte Carlo, et appellent la variable synthétique "attribute density". Plusieurs domaines d'applications sont suggérés pour l'utilisation de cette approche dans le cadre des inventaires forestiers (par exemple la méthode relascopique, ou l'échantillonnage à l'aire de transects).

De notre point de vue, replacer cette méthode dans le cadre de la méthode de partage des poids présente différents avantages. D'abord, cela permet de traiter facilement les effets de bord, à savoir les cas où certaines unités ont un ensemble d'ancêtres qui déborde de la zone d'étude. Si l'on utilise la surface de l'ensemble d'ancêtres M_{+k}^{AB} dans les poids de sondage (voir équation 27), on est assuré d'avoir des estimateurs sans biais, alors que les méthodes de correction proposées dans ces papiers peuvent s'avérer complexes, voir par exemple la section 10.7 de [4], ou [7]. L'approche par le partage des poids permet également de se référer à la population qui a été échantillonnée. Ceci afin de pouvoir s'appuyer sur l'estimateur de Horvitz Thompson, et de calculer ensuite un estimateur de variance de manière rigoureuse. De plus, la méthode peut être appliquée à n'importe quel plan de sondage, grâce à la théorie présentée dans l'article de [2].

Cette méthode n'est sans doute pas réservée aux seuls inventaires forestiers. On peut citer l'exemple de l'enquête sur les pratiques culturales menée par le service statistique du Ministère de l'agriculture français, jusqu'en 2006, avec un échantillon de parcelles déterminées à partir de l'échantillon de points de l'enquête TER-UTI ([1]). L'extrapolation des parcelles se faisait dans cette enquête en considérant que les parcelles sont tirées avec des probabilités inégales selon leur taille, ce qui « retrouve » la manière dont la méthode du partage des poids proposerait des poids de sondage, aux éventuels effets de bord près. Utiliser comme cadre de référence la méthode de partage des poids permettrait, dans ce cas, d'avoir une formalisation plus rigoureuse de la démarche utilisée. D'autres domaines d'investigation liés aux problématiques environnementales peuvent certainement trouver également un bénéfice dans l'application de la méthode.

Références

[1] Chapelle-Barry, C. Enquête sur les pratiques culturales en 2006. *Agreste Chiffres et Données* (2008).

[2] Cordy, C. B. An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters 18*, 5 (1993), 353–362.

[3] Deville, J., and Lavallée, P. Indirect sampling : The foundations of the generalized weight share method. *Survey Methodology 32*, 2 (2006), 165.

[4] Gregoire, T. G., and Valentine, H. T. *Sampling strategies for natural resources and the environment.* CRC Press, 2007.

[5] Lavallée, P. *Indirect sampling.* Springer Science & Business Media, 2009.

[6] Mandallaz, D. *Sampling techniques for forest inventories.* CRC Press, 2007.

[7] Roesch, F. A., Green, E. J., and Scott, C. T. An alternative view of forest sampling. *Survey Methodology. December 1993 Vol. 19, No. 2, pp. 199-204 Statistics Canada* (1993).

[8] Stevens, D. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics : The official journal of the International Environmetrics Society 8*, 3 (1997), 167–195.