

Plans d'échantillonnage pour les Évaluations forestières nationales

Ronald E. McRoberts¹, Erkki O. Tomppo² et Raymond L. Czaplewski³

CE CHAPITRE ABORDE LES THÈMES SUIVANTS:

- Comment définir une population pour un échantillonnage.
- Comment sélectionner la taille et la forme d'une placette pour la configurer.
- Comment différencier les plans d'échantillonnage aléatoire simple, systématique, par strates et par grappes.
- Méthodes de construction d'un plan d'échantillonnage.
- Estimation des variances et des moyennes relatives à une population.
- Estimation des erreurs d'échantillonnage.
- Considérations particulières relatives aux inventaires de forêts tropicales.

Résumé

Les Évaluations forestières nationales (EFN) doivent être conduites à partir d'attributs forestiers estimés de façon précise et défendable sur un plan scientifique. Cette rubrique aborde le concept statistique du plan d'échantillonnage d'un inventaire forestier, notamment le processus utilisé pour définir la population à échantillonner et le choix d'un échantillon répondant aux exigences de précision de l'EFN. Toute équipe chargée de mettre au point un inventaire forestier national doit inclure un statisticien expérimenté. Cependant, dans le cas où un tel expert ne serait pas disponible, cette rubrique fournit des directives et des recommandations qui permettront de créer un plan d'échantillonnage relativement simple, qui réduit les risques tout en améliorant les chances de succès.

1. Introduction

Le plan d'échantillonnage qui soutient le programme technique d'une Évaluation

forestière nationale (EFN) requiert un plan théorique qui devra être mis en œuvre sur le terrain (consulter Réalisation d'une EFN). La compréhension des principes élémentaires liés aux statistiques et aux méthodes d'estimation représente l'une des composantes du processus global Gestion des informations et enregistrement des données d'une Évaluation forestière nationale.

OBJECTIFS

Le but est d'évaluer l'état des forêts de l'ensemble d'un pays au moyen de données collectées à partir d'un échantillonnage de placettes relevées sur le terrain. Les objectifs de base d'une EFN sont au nombre de quatre: (1) obtenir au niveau national des estimations de la surface totale des forêts, subdivisée en catégories principales de types et d'état des forêts, nombre et répartition des arbres par catégories d'espèce et de taille, volume de produits ligneux par type d'arbre, produits forestiers non ligneux, estimation du changement de ces attributs forestiers et indicateurs de biodiversité; (2) obtenir des

1 Northern Research Station, U.S. Forest Service, 1992 Folwell Avenue, Saint Paul, Minnesota 55108 USA

2 The Finnish Forest Research Institute, PO Box 18. FI-01301 Vantaa, Finland

3 Rocky Mountain Research Station, U.S. Forest Service, 2150 Center Avenue, Building A, Fort Collins, Colorado 80526 USA

estimations suffisamment précises pour les régions géographiques sélectionnées, par exemple au niveau national, sous-national, provincial, régional et communal ; (3) collecter des informations en quantité et en diversité suffisantes pour satisfaire aux exigences internationales de rapport et (4) parvenir à un compromis acceptable entre le budget prévu d'une part, et la précision et la résolution géographiques des estimations d'autre part. Consulter Variables généralement évaluées dans les Inventaires forestiers nationaux.

HYPOTHÈSES ET SIMPLIFICATION DES CONTRAINTES

Plusieurs hypothèses soutiennent la discussion qui suit. En premier lieu, nous supposons que des statisticiens expérimentés dans le domaine des inventaires des ressources naturelles et de l'analyse des données ne sont pas disponibles. Ensuite, nous supposons que des données existantes sont disponibles, sous la forme de cartes décrivant des caractéristiques telles que régions écologiques, couverture des terres, sols, élévation, frontières politiques et administratives, et réseaux de transport. Enfin, nous supposons que des modèles sont disponibles pour la prévision d'attributs tels que les volumes de produits ligneux obtenus à partir des mesures de base des arbres. Même après avoir posé ces hypothèses, envisager toutes les possibilités de plan d'échantillonnage pour une EFN dépasse largement le cadre de cette rubrique. Par conséquent, nous avons adopté trois contraintes qui en définissent les limites. La première est de limiter le débat à des plans relativement simples, polyvalents, qui peuvent être utilisés de manière fiable avec des compétences uniquement locales. La deuxième est de nous limiter à des plans dotés d'une grande flexibilité, afin de réduire les risques d'imprécision et de perte de crédibilité. La troisième est d'aborder des plans incluant les mêmes probabilités d'échantillonnage ou, en cas de plans stratifiés, les mêmes probabilités d'échantillonnage au sein de chaque strate.

POURQUOI UTILISER UN ÉCHANTILLONNAGE?

La description la plus précise d'une population est obtenue à partir des mesures précises de chaque membre de la population, en d'autres termes un recensement. Généralement, un recensement est impossible pour des raisons de budget et de logistique. Imaginons une tentative de mesurer chaque arbre d'une forêt de 1 million d'hectares. Un échantillon mesure une portion de la population et, en foresterie, il s'agit habituellement d'une très petite portion. Les estimations basées sur des données provenant d'un échantillon mesuré sont ensuite extrapolées à la population tout entière, dont la plus grande partie n'a pas été mesurée.

Il s'agit donc de "deviner" ou "estimer" l'état d'une population après avoir échantillonné quelques membres de cette population. Si l'échantillon est représentatif de l'ensemble de la population, l'estimation sera exacte et et moins sujette à des déviations par rapport aux valeurs réelles de population. Dans le cas contraire, l'estimation sera inexacte et trompeuse ; il ne sera pas évident qu'elle est inexacte, et son inexactitude sera ignorée car l'état véritable de l'ensemble de la population sera ignoré. Le mieux qu'il est possible de faire est d'augmenter les chances de mesurer un échantillon représentatif. Pour y parvenir, des règles défendables du point de vue scientifique doivent être appliquées pour sélectionner l'échantillon, augmenter le nombre d'unités d'échantillonnage observées ou mesurées, et diminuer les erreurs de mesure de chaque échantillon (voir Qualité des données). S'il n'est pas difficile de produire des données, il est beaucoup plus problématique de produire des données exactes, avec une fiabilité reconnue, qui serviront à prendre des décisions de grande importance.

DÉFINITION D'UNE POPULATION

Une estimation scientifiquement défendable des attributs d'une population est fondée sur un ensemble officiel de théories mathématiques qui doivent être respectées si elles doivent

servir à défendre l'exactitude d'une estimation basée sur un échantillon. Les sélections rigoureuses du cadre d'échantillonnage, de la configuration des placettes de relevé et du plan d'échantillonnage sont des étapes cruciales du processus qui ne peuvent pas être accomplies indépendamment l'une de l'autre. Chaque décision concernant l'une a un impact sur les autres. La théorie mathématique commence par une définition précise de la population dont les attributs seront évalués. Par exemple, pour une commune de 5 millions d'hectares dont 1 million sont en forêt, la population statistique peut être décrite de plusieurs façons, différentes mais logiques :

- Milliers d'hectares de polygones forestiers et non forestiers
- Dizaines de millions de placettes potentielles d'échantillonnage de 0,1 hectare
- Dix millions de pixels télédétektés de 30 x 30 m
- Milliards d'arbres
- Nombre infini de points

Pour plus de détails, consultez la rubrique consacrée aux Unités d'observation.

Il n'existe pas une seule et unique bonne définition de la population des inventaires forestiers. Le problème clé pour l'application d'un échantillonnage forestier est de définir précisément les frontières géographiques de la population ciblée, par exemple l'ensemble des terres, à la fois forestières et non forestières, à l'intérieur d'un pays qui sont situées en dehors des frontières géopolitiques des zones urbaines. Il n'est pas inhabituel de découvrir que certaines portions d'une population cible ne sont pas échantillonnables. Il peut s'agir, par exemple, de zones éloignées, inaccessibles ou dangereuses d'accès. Ces zones doivent être identifiées avec précision sous une forme cartographique, même si les frontières réelles ne sont pas évidentes, et exclues de la population étudiée. Une estimation scientifiquement défendable doit être limitée uniquement à la population échantillonnée.

CHOIX D'UN CADRE D'ÉCHANTILLONNAGE

Faisons une distinction claire entre ces trois expressions : cadre d'échantillonnage, plan d'échantillonnage et configuration de placette. Le cadre d'échantillonnage se réfère à l'ensemble de toutes les unités d'échantillonnage possibles ; le plan d'échantillonnage se réfère à la sélection d'un sous-ensemble d'unités d'échantillonnage qui représentent la population ; la configuration de placette se réfère à la taille, à la forme et aux composantes d'une placette relevées sur le terrain.

Un cadre d'échantillonnage présente quelques avantages s'il considère une forêt comme une population infinie de points. Une approche de l'échantillonnage en tant que cadre est d'appliquer la fameuse méthode Bitterlich, qui se révèle efficace pour estimer les variables corrélées à la taille des arbres. D'autres configurations de relevé de placette par point mesurent une région donnée et imputent ses attributs à un point. Lorsqu'il se trouve près d'une frontière ou de la bordure d'un peuplement, un point est plus facilement attribué à un côté ou à l'autre, alors que les placettes de configuration différente risquent de chevaucher largement ces limites. Nous recommandons que la population d'une forêt soit considérée comme étant un ensemble infini de points et que les mesures physiques dans une région donnée soient utilisées pour décrire l'état d'un point échantillon.

CHOIX D'UNE CONFIGURATION DE PLACETTE

La configuration d'une placette décrit la taille et la forme de la placette étudiée et détermine les variables à mesurer à chaque emplacement d'une placette échantillon. Les options d'une configuration de placette incluent les placettes à surface variable, les placettes à surface fixe, les sous-divisions de placettes en sous-placettes, et les placettes en grappe, qui toutes requièrent un relevé de taille et de forme. La méthode Bitterlich appliquée au relevé de placettes à surface variable est particulièrement efficace pour

obtenir une estimation précise des attributs d'une forêt associés à la taille des arbres. Le relevé de placettes à surface fixe, qui n'est pas nécessairement optimal pour un attribut de forêt particulier, est un excellent compromis si l'échantillonnage doit produire l'estimation d'un grand nombre d'attributs différents et tend à être plus compatible avec des données existantes. L'échantillonnage par grappes réduit les déplacements entre placettes tout en fournissant un nombre suffisant de relevés. La taille et la forme optimales d'une placette peuvent être déterminées par une simulation d'échantillonnage et des informations existantes, mais les placettes circulaires sont souvent les plus utilisées dans les inventaires forestiers.

Les problématiques du choix d'une configuration de placette sont abordées sous les rubriques Unités d'observation et Optimisation des configurations de placette.

MESURE DES PLACETTES ÉCHANTILLONS

Sous la rubrique Observations et mesures des Évaluations forestières nationales sont résumés les principaux facteurs à prendre en considération pour la mesure des placettes échantillons. Pour obtenir des informations plus détaillées, consultez la référence en ligne *Statistical Techniques for Sampling and Monitoring Natural Resources* (Schreuder et al., 2004). Sous la présente rubrique, nous étudierons les deux aspects de cette étape que sont l'utilisation de données de télédétection pour mesurer les placettes, et les placettes temporaires comparées aux placettes permanentes.

Tout d'abord, les données de télédétection générées par les satellites de résolution normale et les photographies aériennes à haute altitude (échelles 1:24 000 à 1:60 000) fournissent des mesures peu coûteuses pour des indicateurs généraux de l'état des forêts, principalement de l'évolution des surfaces forestières. Toutefois, il est impossible d'obtenir la plupart des mesures détaillées de l'état des forêts avec ce type de détection (voir Télédétection et Évaluations forestières nationales). Des mesures plus

détaillées de l'état des forêts sont fournies par les photographies aériennes à basse altitude et des détecteurs tels que le Lidar. Toutes ces méthodes de détection sont généralement coûteuses et leur champ de vision est encore trop étroit. Elles ne permettent pas, actuellement, de couvrir un pays entier de frontière à frontière. Néanmoins, elles pourraient théoriquement servir à mesurer quelques emplacements d'échantillonnage dans le cadre d'une étude nationale. Par exemple, il pourrait être rentable de mesurer d'abord une placette par une méthode de télédétection pour déterminer si elle possède un couvert forestier accessible ou des surfaces forestières exploitables. Sinon, il n'est pas garanti que les équipes de terrains pourront la visiter.

Ensuite, l'estimation des changements et des tendances des forêts à l'échelon national est une part importante d'une EFN. Si les emplacements des placettes échantillons sont bien documentés, ces mêmes placettes peuvent être remesurées au fil du temps pour obtenir des estimations plus précises de changements tels que croissance des arbres, mortalité, abattage, régénération ainsi que les changements d'état général de la forêt et de catégories d'utilisation des terres. Pour plus de détails, consultez les rubriques Observations permanentes ou temporaires, Observations et mesures et Évaluation des changements. Le remesurage des placettes accroît l'efficacité de l'estimation et contribue à une meilleure compréhension des composantes du changement. Toutefois, si des placettes permanentes sont utilisées, leurs emplacements doivent être très exactement documentés. Cela peut consister à mettre en place une balise au sol au centre ou au coin d'une placette, puis à documenter soigneusement comment la trouver depuis un lieu de départ approprié, peut-être distant de plusieurs kilomètres. La balise doit se trouver hors du champ de vision normal afin que la placette reste représentative des milliers d'hectares qui ne seront jamais mesurés. Une placette échantillon ne sera plus représentative si elle est traitée d'une manière spéciale, par

exemple si elle est protégée de l'abattage ou d'autres perturbations. Une balise trop visible dans le sol peut influencer sur le traitement normal de l'emplacement par d'autres personnes.

Bien que le remesurage des mêmes arbres procure des estimations du changement extrêmement précises, cette approche est plus coûteuse puisque les mêmes arbres et les mêmes centres de placette doivent être relocalisés au moment de chaque mesure. Une alternative pour l'estimation du changement des placettes temporaires est d'évaluer la croissance des arbres par les perçages de croissance et l'estimation brute des changements de surface et de volume d'une forêt en comparant des estimations indépendantes, obtenues par la mesure de différents ensembles de placettes temporaires à différentes époques. Cependant, l'abattage, la mortalité et la régénération sont difficiles à estimer à partir des données de placettes temporaires. Par conséquent et dans la mesure du possible, nous recommandons l'utilisation de placettes permanentes ou d'une combinaison de placettes temporaires et permanentes (p. ex. Ranney et al., 1987).

2. Plan d'échantillonnage

Il existe deux approches générales de l'échantillonnage : échantillonnage subjectif ou intentionnel, et échantillonnage aléatoire. L'échantillonnage subjectif tente d'appliquer un jugement professionnel à la sélection d'unités d'échantillonnage considérées comme représentatives de la population tout entière. Ces unités conviennent souvent aux mesures réalisées avec un budget limité. Bien que les données collectées de cette façon décrivent avec précision l'état des sites échantillonnés, elles peuvent ne pas caractériser la population dans sa totalité. Les partisans de l'échantillonnage subjectif font confiance aux compétences des experts pour sélectionner un échantillon représentatif et affirment que cette approche convient parfaitement à des fins pratiques. Dans certaines situations simples, cela peut être vrai. Mais que se passe-t-il si l'utilisateur des données n'accorde pas la

même confiance aux experts ? Des données coûteuses peuvent perdre toute valeur si le plan d'échantillonnage n'est pas défendable d'un point de vue scientifique. De même, les sites d'échantillonnage les plus pratiques se trouvent souvent à proximité de routes, lesquelles sont fréquemment associées à une uniformité de la morphographie, de l'occupation des sols, des antécédents de gestion foncière et des paysages. De tels sites sont-ils vraiment représentatifs de la population tout entière ? La réponse est matière à débat. Il est beaucoup plus facile de discréditer la véracité des populations estimées à partir d'un échantillon subjectif que de prouver le contraire.

L'échantillonnage aléatoire remplace les jugements subjectifs par des règles objectives fondées sur des probabilités connues de sélection de chaque membre d'une population. Par exemple, supposons qu'une forêt de 1 million d'hectares consiste en une population de placettes de 10 x 10 m. La population serait alors composée de 100 millions de placettes. Si l'une de ces placettes était choisie de façon aléatoire, sa probabilité d'être sélectionnée serait de $1/100\,000\,000$. Si un seul échantillon aléatoire de 1000 placettes était sélectionné pour estimer l'état de la population du million d'hectares, chaque membre de cette population aurait alors une probabilité de sélection de $1000/100\,000\,000 = 1/100\,000$ approximativement, et chaque placette mesurée au sein de l'échantillon pourrait être considérée comme représentant les 99 999 autres placettes non mesurées. La leçon importante de cet échantillonnage aléatoire est qu'il constitue une méthode objective, avec des règles précises et un fondement mathématique, applicable à l'estimation des attributs d'une population basée sur un échantillon. La probabilité qu'un expert sélectionne l'une quelconque des placettes échantillons potentielles n'est pas connue, et les règles mathématiques d'un échantillonnage subjectif sont inapplicables d'un point de vue scientifique. Par conséquent, nous recommandons d'utiliser un échantillonnage aléatoire plutôt que subjectif, et de plus, un

échantillonnage aléatoire égalitaire dans lequel les unités d'échantillonnage potentielles ont les mêmes probabilités de sélection.

SÉLECTION D'UN PLAN D'ÉCHANTILLONNAGE ALÉATOIRE

Une bonne part des difficultés associées au choix d'un plan d'échantillonnage est due à deux facteurs : tout d'abord, les unités d'échantillonnage sont réparties dans un espace et leurs observations peuvent être spatialement corrélées. Ensuite, les plans d'échantillonnage diffèrent par leur coût. La corrélation spatiale de certaines observations de variables d'intérêt influe fortement sur le choix d'un plan d'échantillonnage. Des facteurs écologiques, climatiques et pédologique, ainsi que les pratiques de gestion forestière, font que les observations de parcelles proches entre elles sont, en moyenne, plus semblables que les observations de parcelles éloignées les unes des autres. Il en résulte que, au sens strict du terme, la conception d'un plan d'échantillonnage totalement optimal est une mission impossible car les nombreuses variables mesurées et dérivées d'une EFN varient très différemment dans l'espace. Des plans d'échantillonnage optimisés étant différents en raison des différences de variables, leur optimisation peut donc requérir en premier lieu une diminution de l'erreur standard d'une seule variable importante, par exemple le volume ligneux, ou bien une pondération des erreurs standard d'un petit nombre de variables. Une solution partielle est de minimiser les effets de la corrélation spatiale en établissant des emplacements d'échantillonnage les plus éloignés les uns des autres que possible. Cela s'ajoute aussi au fait que des observations de placettes échantillons qui dévient beaucoup les unes des autres permettent de cumuler plus d'informations sur l'échantillon. En matière d'échantillonnage forestier, cela suggère souvent des plans d'échantillonnage hexagonaux. Les principaux coûts d'un échantillonnage sont représentés par les déplacements pour se rendre à l'unité d'échantillonnage et en revenir, et par la

mesure de l'unité elle-même. Ces coûts, en retour, dépendent de la structure du paysage et des forêts, des mesures à prendre, des conditions topographiques et économiques, et des moyens de transport.

Un point de départ courant dans le choix d'un plan d'échantillonnage est de connaître la limite supérieure acceptable des erreurs standard de l'estimation et les limites budgétaires à ne pas dépasser. Optimiser le plan d'échantillonnage, en fonction du cadre d'échantillonnage et de la configuration des placettes, implique de sélectionner une procédure de répartition spatiale des emplacements des unités d'échantillonnage de telle sorte que les erreurs standard soient minimisées tout en restant dans les limites budgétaires autorisées. Parfois, il n'est pas possible d'y parvenir et rechercher un compromis peut s'avérer nécessaire.

ÉCHANTILLONNAGE ALÉATOIRE SIMPLE

Un échantillonnage aléatoire simple place des placettes échantillons au hasard au sein de la population échantillonnée (Figure 1a). Par hasard, il existe des grappes spatiales et des vides dans la répartition de la placette ; malgré tout, cet échantillon reste une probabilité valide. Dans un échantillon aléatoire, les coordonnées géographiques de chaque placette échantillon peuvent être choisies avec un générateur de nombre aléatoire si les coordonnées autorisées sont limitées à la population étudiée. Sinon, la sécurité, la difficulté de la mesure des placettes ou les trajets vers et depuis les emplacements des placettes ne seraient pas pris en compte. À probabilité égale, c'est le plan d'échantillonnage le moins risqué, mais c'est aussi généralement le moins efficace, en ce qui concerne à la fois le coût et la précision des estimations, en partie à cause de la corrélation spatiale entre les observations.

ÉCHANTILLONNAGE SYSTÉMATIQUE

Un échantillonnage systématique utilise une grille ou une matrice fixe pour assigner des placettes selon un schéma régulier (Figure 1b). L'intérêt d'un échantillonnage systématique

est qu'il augmente la distance moyenne entre les placettes et donc diminue la corrélation spatiale entre les observations et accroît l'efficacité statistique. De plus, un échantillon systématique qui apparaît clairement comme représentatif d'une façon ou d'une autre, peut être très convaincant aux yeux des décideurs qui ne possèdent aucune expérience de l'échantillonnage. Les échantillons systématiques peuvent être basés sur des grilles rectangulaires ou des matrices hexagonales. Par exemple, une placette échantillon peut être placée aux intersections d'une grille de 2 km x 2 km. Un nombre aléatoire permet de choisir le point de départ et l'orientation de cette grille, mais aucun autre nombre aléatoire n'est requis. Ce plan d'échantillonnage est courant en foresterie. Le risque le plus important est que, par un pur hasard, l'orientation de la grille coïncide avec ou soit parallèle à des éléments naturels ou d'origine

humaine tels que des routes ou des monceaux de graviers laissés par la fonte des glaciers. Dans le cas de grandes zones géographiques, il est conseillé d'éviter d'aligner les lignes de quadrillage sur les lignes de longitude. Pour les latitudes élevées, la nature convergente des lignes de quadrillage nord-sud peut entraîner un rapprochement des emplacements des placettes échantillons plus marqué dans les latitudes hautes que dans les latitudes basses. Les plans d'échantillonnage basés sur des matrices hexagonales amoindrissent ce problème (White et al., 1992).

Les plans d'échantillonnage systématiques non alignés combinent les caractéristiques des deux plans aléatoires et systématiques. Dans ces plans, une placette échantillon est assignée à un emplacement choisi au hasard dans chaque cellule de la grille ou de la matrice (Figure 1c).

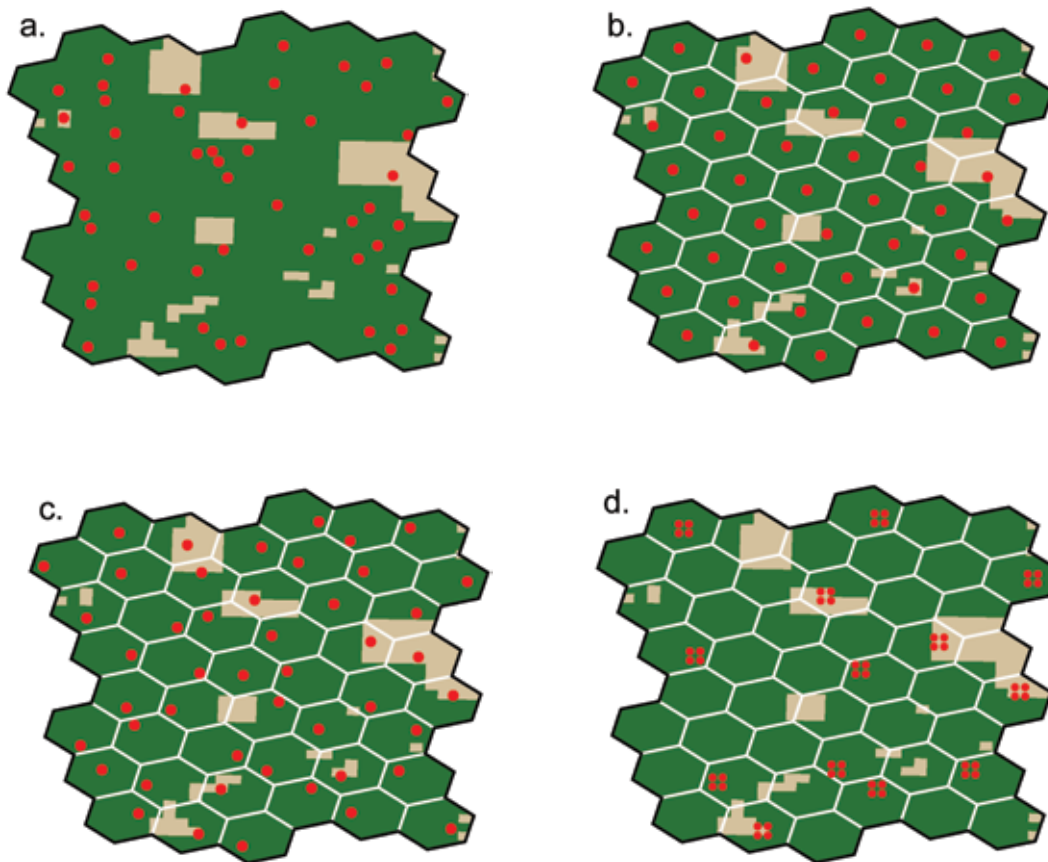


Figure 1. (a) plan d'échantillonnage aléatoire simple, (b) plan d'échantillonnage systématique aligné, (c) plan d'échantillonnage systématique non aligné, (d) plan d'échantillonnage systématique non aligné, en grappe, avec le même nombre de placettes mais groupées en grappes.

ÉCHANTILLONNAGE EN GRAPPE

Pour des raisons pratiques, comme augmenter l'efficacité budgétaire et réduire les déplacements de l'équipe de terrain, les placettes échantillons peuvent être organisées en grappes, conduisant alors à un échantillonnage en grappe systématique et un échantillonnage en grappe systématique stratifié. Dans l'échantillonnage en grappe systématique, les grappes sont réparties dans la population à l'aide de grilles ou de polygones comme les hexagones.

Plusieurs questions se posent alors, lors de la planification d'un plan d'échantillonnage basé sur l'utilisation de grappes : (1) quel est l'espacement entre les grappes ? (2) quelle est la forme de la grappe ? (3) quel est le nombre de placettes par grappe ? et (4) quelle est la configuration de la placette échantillon ? Pour répondre à ces questions, des informations préliminaires sur la répartition spatiale et la corrélation des variables d'intérêt sont éventuellement nécessaires. La corrélation, en tant que fonction de distance entre les placettes sur le terrain estimée à l'aide de variogrammes, peut servir à comparer l'efficacité des différents plans d'échantillonnage.

2.1 ÉCHANTILLONNAGE STRATIFIÉ

L'échantillonnage stratifié exige d'abord de diviser la population en sous-populations sans chevauchement appelées strates dont l'ensemble représente la population tout entière, puis de délimiter un échantillon indépendant dans chaque strate. Si chaque strate comprend un échantillon aléatoire simple, la procédure entière porte le nom d'échantillonnage aléatoire stratifié. De nombreuses raisons peuvent justifier un échantillonnage stratifié (Cochran, 1977; Schreuder et al., 1993). La première est qu'une stratification permet d'augmenter la précision de l'estimation de la population. Pour comprendre le potentiel du gain de précision qui peut être atteint grâce à une stratification, quelques remarques et formules sont nécessaires. Avec un échantillonnage aléatoire simple (SRS), la moyenne de

l'estimation de la population est

$$\bar{y}_{SRS} = \frac{1}{n} \sum_{i=1}^n y_i \quad [1]$$

et l'estimation de la variance de la moyenne est

$$Var(\bar{y}_{SRS}) = \frac{s^2}{n} \quad [2]$$

où n est la taille de l'échantillon, y_i est une observation, et

$$s^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2 \quad [3]$$

est l'estimation de la variance de la population. L'étude Cochran (1977) fournit la formule de base d'une estimation stratifiée. Ignorant les facteurs de correction de la population finie et les erreurs d'estimation de la strate pondérée, l'estimateur objectif de la moyenne et de la variance de la population est

$$\bar{y}_{Str} = \sum_{h=1}^L W_h \bar{y}_h \quad [4]$$

et

$$Var(\bar{y}_{Str}) = \sum_{h=1}^L W_h^2 \frac{s_h^2}{n_h} \quad [5]$$

où

$$\bar{y}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} y_{jh} \quad [6]$$

$$s_h^2 = \frac{1}{n_h - 1} \sum_{j=1}^{n_h} (y_{jh} - \bar{y}_h)^2$$

sont les moyennes et les variances au sein de la strate, respectivement ; $h=1, 2, \dots, L$ décrit la strate ; j décrit les observations au sein de la strate ; n_h décrit le nombre d'observations au sein de la h e strate avec $n_1+n_2+\dots+n_L=n$; et W_h est la strate pondérée représentant la

proportion de la population dans la strate. Les effets de la stratification et de l'estimation stratifiée sur la précision sont souvent évalués à partir d'une efficacité relative RE, définie comme

$$R = \frac{Var(\bar{y}_{SRS})}{Var(\bar{y}_{Str})} \quad [8]$$

où $RE > 1$ indique un effet bénéfique. L'efficacité relative peut être interprétée comme l'augmentation de la taille globale de l'échantillon qui serait nécessaire pour atteindre la même précision qu'une estimation basée sur un échantillonnage aléatoire simple, comme c'est le cas lorsque la stratification et l'estimation stratifiée sont utilisées. D'un point de vue quantitatif, les gains de précision sont obtenus lorsque les variances des moyennes estimées d'une strate sont nettement inférieures à la variance de la moyenne estimée

globale (c.-à-d., $\frac{s_h^2}{n_h} < \frac{s^2}{n}$) et/ou lorsque des

strates avec un grand rapport $\frac{s_h^2}{n_h}$

représentent de petites proportions de la population (c.-à-d., lorsque Wh est faible). D'un point de vue qualitatif, les gains de précision sont obtenus lorsque des populations hétérogènes sont divisées en sous-populations plus homogènes. Cela signifie généralement que les strates ont en substance des moyennes différentes, des variances différentes ou les deux.

La deuxième justification de la stratification est qu'elle peut contribuer à éviter une déviation, en fonction de l'estimateur sélectionné. Par exemple, les équipes de terrain d'une EFN ont généralement un accès garanti aux placettes situées sur des terres du domaine foncier public. Par contre, si une autorisation des propriétaires fonciers privés est requise pour mesurer des placettes échantillons sur leurs terres, inévitablement certains propriétaires refuseront de donner leur accord. Dans des

cas extrêmes, le rapport placettes publiques-placettes privées dans la portion accessible de l'échantillon peut être considérablement moindre que le même rapport concernant les propriétés foncières publiques et privées dans la population. Si la répartition des espèces et/ou les pratiques de gestion des propriétés foncières publiques et privées diffèrent de façon significative, l'estimation peut se trouver biaisée. La solution consiste à stratifier les terres par propriété foncière, ce qui aboutit à des estimations indépendantes pour les deux strates (McRoberts, 2003).

Une troisième raison justifiant la stratification est qu'elle permet d'englober différents protocoles d'échantillonnage ou procédures d'estimation pour différentes sous-populations. Par exemple, une partie importante du budget d'échantillonnage peut être consacrée aux frais de déplacement vers et depuis les emplacements des placettes. Si des données de télédétection peuvent servir à déterminer que certaines placettes sont situées sur des terres non forestières, les frais de déplacement seront considérablement réduits en évitant l'envoi d'équipes de terrain vers ces emplacements. Selon la technique de mesure utilisée, il peut être nécessaire d'utiliser un estimateur différent pour ces strates.

L'estimation stratifiée présente le plus d'avantages lorsque la population est stratifiée et la taille d'une strate déterminée préalablement à l'échantillonnage. Le processus de détermination de la taille d'une strate ou, de manière équivalente, le processus d'attribution d'échantillons à des strates, peuvent être réalisés de différentes façons et en fonction de plusieurs objectifs différents. Fréquemment, des échantillons sont attribués aux strates proportionnellement à l'un de leurs attributs. Une approche facile à mettre en œuvre est d'attribuer des placettes échantillons à des strates proportionnellement à leur taille. Si un échantillonnage aléatoire ou systématique est utilisé au sein des strates, cette approche conduit alors à des échantillons de probabilité égale entre strates, ce qui peut simplifier l'estimation. Toutefois, avec cette approche,

les variances des moyennes de strates peuvent être très différentes. Si l'on souhaite obtenir des estimations des moyennes de strates de précision comparable, les échantillons doivent alors être attribués proportionnellement aux variances des strates. Un inconvénient potentiel de cette approche est que de bonnes estimations des variances sont requises avant que les échantillons soient attribués aux strates. Enfin, il est possible que les estimations des moyennes de certaines strates soient plus importantes que d'autres. Dans ce cas, l'attribution des échantillons peut être proportionnelle à une évaluation subjective de l'importance des strates.

Il arrive parfois que les objectifs de l'échantillonnage interdisent l'utilisation d'un échantillonnage aléatoire stratifié. Par exemple, un plan d'échantillonnage systématique peut être utilisé comme un moyen d'optimiser la précision des estimations pour plusieurs variables simultanément. Même si la stratification ne présente pas de grand intérêt pour une variable particulière, les effets positifs de l'augmentation de la précision et de la réduction des déviations éventuelles peuvent encore justifier le choix d'une stratification post-échantillonnage et d'une estimation stratifiée. Par conséquent, même si un échantillonnage stratifié n'est pas utilisé, nous recommandons d'envisager une stratification post-échantillonnage et une estimation stratifiée car elles peuvent apporter des gains importants de précision à peu de frais ou d'efforts supplémentaires.

La quasi-totalité des sources de données peut servir à créer des strates aussi longtemps que les deux tâches suivantes peuvent être réalisées de manière constante. La première, la pondération des strates, calculée en tant que proportion de la population représentée par chaque strate, doit être déterminée. La seconde est que chaque placette soit attribuée à une strate et une seule. La disponibilité croissante de diverses couches de données numériques thématiques ouvre de vastes possibilités de sources de données qui peuvent servir à la création de strates. En outre, la disponibilité

croissante de systèmes d'informations géographiques (SIG) simplifie grandement la réalisation de ces deux tâches. Un choix populaire de données de stratification est celui des classifications de couvertures des terres à partir desquelles des catégories de couvert forestier et non forestier peuvent être définies et utilisées comme strates (McRoberts, 2002). L'utilisation d'un SIG possédant ce type de couche de données simplifie grandement la réalisation des deux tâches de stratification. Au sein du SIG, chaque unité cartographique de la classification de couverture des terres est attribué à une strate sur la base de la catégorie attribuée à l'unité cartographique. Le calcul de la pondération des strates revient alors simplement à pouvoir utiliser la fonctionnalité du SIG pour déterminer la surface totale de toutes les unités cartographiques attribuées à une même strate et à la diviser par la surface totale de la population échantillonnée. Une placette est attribuée à la strate de l'unité cartographique contenant son centre. D'autres choix de couches de données numériques utilisables pour créer des strates incluent notamment les cartes pédologiques, les cartes de divisions climatiques, les régions écologiques, les frontières administratives, les cartes foncières, et les unités de gestion foncière.

2.2 Estimateurs de rapport et estimateurs de l'erreur selon Matérn

Bien que les estimateurs statistiques soient abordés sous une autre rubrique ou puissent être obtenus en ligne dans l'ouvrage *Statistical Techniques for Sampling and Monitoring Natural Resources* (Schreuder et al., 2004), nous noterons ici l'importance du choix d'estimateurs qui soient cohérents avec le plan d'échantillonnage afin d'aboutir à des estimations valides des variances. Avec des plans d'échantillonnage systématiques et fondés sur des grappes, il est particulièrement important que les estimateurs prennent correctement en compte la possible

corrélation spatiale entre les observations. En raison de leur adéquation avec les plans d'échantillonnage qui doivent incorporer la corrélation spatiale, nous publions ici un bref aperçu des estimateurs Matérn (Matérn, 1960).

Les estimations des inventaires forestiers étant fréquemment des valeurs moyennes ou totales de surface ou de volume, les variables dérivées pertinentes d'un inventaire forestier sont souvent exprimées par la formule

$$M = \frac{X}{Y} \quad [9]$$

où X et Y sont des prévisions de variables aléatoires, x et y. Par exemple, considérons l'estimation de la surface moyenne d'une forêt par strate d'utilisation foncière, pour des placettes échantillons qui peuvent croiser plusieurs strates, toutes comprises dans la catégorie des terres forestières. La méthode pour rendre compte de ce phénomène, particulièrement utile pour l'échantillonnage par points, est d'utiliser les informations fournies uniquement par le point central. Soit $x_i=1$ lorsque le point central de la placette appartient à la strate en question et $x_i=0$ s'il ne l'est pas, et soit $y_i=1$ lorsque le point central est sur une terre forestière, et $y_i=0$ s'il ne l'est pas. Par suite, l'estimateur de rapport de la surface moyenne est

$$m = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} = \frac{\bar{X}}{\bar{Y}} \quad [10]$$

où n est le nombre d'unités d'échantillonnage. Soit E(.) le descripteur de la prévision statistique ; par suite,

$$E(m) \approx \frac{\overline{Ex}}{\overline{Y}} = \frac{\overline{X}}{\overline{Y}} = M \quad [11]$$

signifie que m n'est pas biaisé par approximation lorsque n est élevé.

L'estimation des erreurs standard est compliquée par la corrélation spatiale qui peut émerger des changements/tendances des variables et de l'échantillonnage systématique ou en grappe. Matérn (1947, 1960) a suggéré de prendre la variance

des erreurs, $E(m - M)^2$, comme mesure de la fiabilité de l'estimateur, et proposé également un estimateur de la variance. Soit i le descripteur des placettes de terrain ; soit r le descripteur des grappes des placettes de terrain ; et considérons les résidus des

grappes $z_r = x_r - m_r$, où $x_r = \sum_{i \in r} x_i$ et

$y_r = \sum_{i \in r} y_i$. Nous présumons que les résidus représentent un processus stochastique stationnaire au second ordre (stationnarité hebdomadaire). La variance du processus peut être estimée au moyen de formules

$$T = \sum_r \sum_s c_{rs} z_r z_s, \quad \text{quadratiques, où } c_{rs} = c_{sr},$$

$$\sum_r \sum_s c_{rs} = 0 \quad \text{et} \quad \sum_r c_{rr} = 1,$$

où r et s se réfèrent aux grappes des placettes de terrain. Les estimateurs de cette formule ne sont pas biaisés si le processus z n'est pas spatialement corrélé et sont conservatoires si le processus est corrélé positivement (Matérn, 1960). Cette approche a été utilisée pour les inventaires de la Suède et de la Finlande (cf. Ranney, 1981, voir aussi Tomppo et al., 1997 et Heikkinen, 2006) et appliquée par strate d'échantillonnage comme suit. Au sein de chaque strate, le groupe g de grappes de quatre placettes de terrain (r1, r2, r3, r4) est composé de façon que chaque grappe appartienne à quatre groupes différents (Figure 2). La déviance de la moyenne de grappe, \bar{y}_r , à partir de la moyenne de strate \bar{y} , est calculée pour chaque grappe r. Notez que $z_r = (\bar{y}_r - \bar{y})n_r$, où n_r est le nombre de points d'échantillon pertinents dans la grappe r (dans cet exemple,

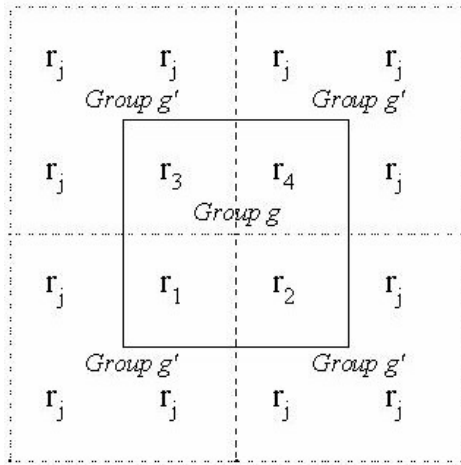


Figure 2. Groupes de grappes et grappes de placettes échantillons.

nr=4). Les pondérations $c_{r_1} = c_{r_4} = -c_{r_2} = -c_{r_3} = 1/2$ sont souvent utilisées. Il est alors possible d'exprimer les formules quadratiques sous la forme $T_g = (z_{r_1} - z_{r_2} - z_{r_3} + z_{r_4})^2 / 4$, et les estimateurs d'erreur standard résultants pour chaque strate sont

$$s = \sqrt{\frac{k \sum_g T_g}{\sum_i y_i}} \quad [12]$$

où g décrit un groupe de grappes dans la strate, i décrit les placettes de la strate, et k est le nombre de grappes dans chaque groupe de grappes (dans cet exemple, k =1). Les estimateurs d'erreur standard pour la surface totale étudiée sont obtenus en combinant les estimateurs spécifiques de la strate avec la formule courante de l'échantillonnage stratifié (Eqs. [4] et [5]). Cette procédure est pertinente pour les strates possédant un grand nombre de placettes de terrain, de préférence plusieurs centaines au minimum.

3. Taille de l'échantillon

La détermination de la taille de l'échantillon est l'une des étapes les plus importantes de la conception d'un plan d'échantillonnage. Si l'échantillon est trop petit, l'incertitude sera grande ; s'il est trop grand, les coûts

seront inutilement élevés. Il est possible de quantifier la confiance attendue dans les futures estimations découlant d'un échantillon dont la probabilité est valide. Plus le nombre de placettes échantillons augmente, la variance de l'erreur d'estimation diminue, la précision de l'estimation augmente, et plus l'estimation est digne de confiance. Généralement, la valeur exacte de l'estimation est connue mais l'état réel de la forêt ne l'est pas. Avec des échantillons de probabilité, la probabilité qu'une estimation se trouve à une distance prédéfinie de la valeur réelle peut être déterminée. Ce sont les rôles de ce que l'on nomme "l'intervalle de confiance", une plage de proportions estimées qui est susceptible d'inclure une proportion réelle mais inconnue de forêt, et "le coefficient de confiance", la probabilité que des intervalles de confiance similaires, bâtis sur différents échantillons, contiendront la proportion réelle de forêt.

Le cas le plus simple est d'estimer les proportions sur un échantillon aléatoire simple ; par exemple, estimer la proportion d'un pays qui est forestière. Supposons qu'une EFN couvre une population échantillonnée de 5 millions d'hectares et que, dans un échantillon aléatoire simple comptant n=1000 placettes, 400 sont couvertes de forêts. La proportion estimée des forêts est de 40 %, mais quel est le niveau de confiance que l'on peut accorder à cette estimation ? Supposons qu'un coefficient de confiance de 80 % est acceptable. Cela signifie que pour 80 placettes échantillons, le pourcentage réel mais inconnu de forêt est situé dans l'intervalle de confiance. D'après les tables et les figures disponibles (Czaplewski, 2003), avec n=1000, et une estimation de 40 % de forêt, l'intervalle de confiance est situé entre 38,0 et 42,0 %. Autre exemple, supposons qu'un type de forêt très rare existe dans la population, mais que la quantité exacte n'est pas connue. Cependant, aucune forêt de ce type n'a été observée dans l'échantillon aléatoire simple de n=1000 placettes, et le pourcentage estimé du pays portant ce type rare de forêt est de 0 %. Avec le même coefficient de confiance

de 80 %, l'intervalle de confiance de cette estimation est situé entre 0, et 0,2 %. Par suite, l'estimation de la surface de ce type de forêt très rare dans l'ensemble du pays de 5 millions d'hectares est de 0 à 10 000 hectares. Enfin, cet exemple d'une commune de 100 000 hectares pour laquelle la mesure d'un échantillon de $n=20$ sur 1000 placettes a montré que 18 sont couvertes de forêts. L'estimation du couvert forestier de cette commune est donc de 90 %, avec un intervalle de confiance de 75,5 à 97,3 %, soit 75 500 à 97 300 hectares. D'autres calculs de tailles d'échantillons sont possibles avec des "calculateurs de taille d'échantillon" interactifs qui sont disponibles sur Internet. Ces exemples montrent que des estimations précises à l'échelon national pour les types courants de couvert forestier sont possibles avec relativement peu de parcelles échantillons. Toutefois, des tailles d'échantillon plus amples sont souvent nécessaires si l'EFN requiert l'estimation de types de forêt rares ou de petites parties d'un pays. C'est la taille de l'échantillon qui détermine la précision des estimations d'une EFN, et non pas la taille de la population échantillonnée tout entière.

Déterminer la taille de l'échantillon voulu requiert une estimation de l'écart type des différences entre les valeurs individuelles au niveau de la placette et leurs moyennes. Cet écart type peut être évalué par une étude ou un inventaire-pilote qui mesure un petit échantillon de placettes forestières afin de déterminer leur variabilité. Par exemple, supposons que l'inventaire pilote inclut 60 placettes, et que le volume ligneux est mesuré sur chaque placette. Ensuite, supposons que le volume moyen est $\bar{x}=100$ m³/ha, la variance entre placettes est $s^2=2\ 500$ m⁶/ha², et l'écart type est $s=50$ m³/ha. Si les observations des placettes-pilotes sont normalement réparties, environ 1/6e des placettes auront $(100-50)=50$ m³/ha ou moins, et un autre 1/6e des 60 placettes auront $100+50=150$ m³/ha ou plus. Supposons que la précision requise pour l'EFN est d'estimer un volume ligneux moyen à l'hectare avec une "tolérance" de $\pm 5\%$ ou une "différence admissible maximale"

($D_{\max}=0,05$) avec un coefficient de confiance de 66 %. La taille d'échantillon requise n est de 100 placettes échantillons approximativement.

$$n = \left(\frac{\hat{s}}{\hat{x} D_{\max}} \right)^2 = \left(\frac{50}{(100 \times 0,05)} \right)^2 = 100 \quad [15]$$

Si cette précision requise de l'EFN est applicable à l'ensemble du pays, alors 100 placettes échantillons est un nombre suffisant. Si elle est applicable à chacune de 10 unités sous-nationales, alors les placettes échantillons nécessaires sont au nombre de 1000 au total. La taille des échantillons augmente fortement dès lors que la tolérance acceptable devient plus faible. Dans cet exemple, une tolérance de $\pm 1\%$ exigerait que la taille de l'échantillon passe de $n=100$ à $n=2500$ placettes échantillons (Eq. 15). La taille requise de l'échantillon augmente si les coefficients de confiance augmentent. Par exemple, il faut quatre fois plus de placettes échantillons pour améliorer la précision d'un coefficient de confiance de 66 % au niveau de 95 %. Des calculs plus exacts et détaillés des tailles d'échantillons requises sont possibles avec les "calculateurs de taille d'échantillon" qui sont disponibles sur Internet.

4. Comparaison des plans d'échantillonnage

Un moyen efficace de comparer des plans d'échantillonnage est de procéder à une simulation si un modèle de zone forestière est disponible. Le modèle peut provenir de précédents inventaires ou d'une estimation de variables d'intérêt basée sur des images satellite. Un exemple d'erreurs standard obtenues avec un plan d'échantillonnage pour l'estimation du volume moyen d'un peuplement final est illustré à la Figure 3. Le site testé se trouve au nord de la Finlande et possède une surface foncière de 6,47 millions d'hectares, une surface forestière de 4,19 millions d'hectares, et un volume moyen en zone forestière de 52,7 m³/ha.

Une carte forestière de frontière à frontière,

à l'échelle du pixel, a été produite à partir des données de terrain de l'inventaire précédent, d'images satellite et de données cartographiques numériques (Tomppo & Halme, 2004; Tomppo, 2004). Des images satellite de différentes résolutions constituent une source d'informations qui s'ajoute aux cartes existantes. De plus, un inventaire-pilote peut servir à collecter des informations pour planifier le plan d'échantillonnage final. Des sous-zones représentatives peuvent être sélectionnées dans la population où des inventaires-pilotes peuvent être effectués. Ces inventaires-pilotes doivent néanmoins être reconnus et acceptés comme loin d'être optimaux. Par ailleurs, de nouveaux plans d'échantillonnage peuvent être créés à partir des informations des inventaires précédents, comme c'est le cas dans les pays où des inventaires forestiers ont été réalisés depuis les années 20 et 30 du siècle précédent (p. ex. Ilvessalo, 1927).

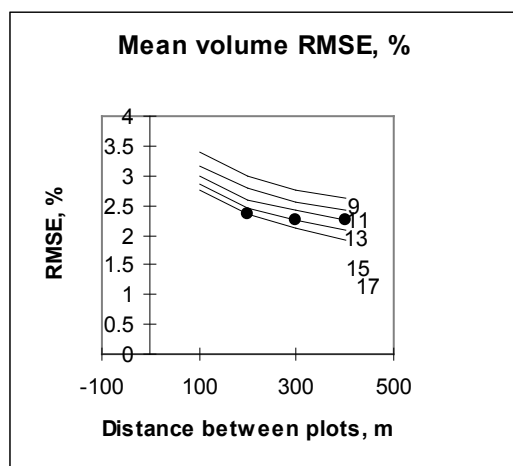


Figure 1. Erreurs standard basées sur des simulations d'échantillonnage avec différentes distances entre placettes de terrain et un nombre de placettes par grappe allant de 9 à 17. La distance entre les grappes est de 10 km. (Figure extraite de Tomppo, et al. 2001, par Helena Henttonen)

5. Facteurs à considérer pour l'échantillonnage des forêts tropicales inventoriées

Au cours des dernières années, les préoccupations liées au changement climatique et aux mesures destinées à en atténuer les effets ont motivé un intense intérêt pour les inventaires forestiers dans les pays tropicaux afin d'évaluer les stocks de carbone forestiers et leur évolution. De tels inventaires, souvent présentés comme des systèmes de Supervision, rapport et vérification (SRV) lorsqu'ils sont réalisés dans le cadre du mécanisme REDD (programme de réduction des émissions liées à la déforestation et à la dégradation des forêts), sont similaires aux IFN (Inventaires forestiers nationaux) bien que la portée d'un SRV puisse être restreinte aux variables liées à la biomasse et sa population d'intérêt, aux terrains dont les émissions de gaz à effet de serre sont d'origine humaine. Cependant, en raison des similarités entre SRV et IFN, les pays tropicaux en voie de développement planifient leurs IFN de telle sorte qu'ils puissent servir également de SRV, ou bien planifient leurs SRV de façon à pouvoir en étendre facilement la portée à celle d'un IFN complet. Les directives développées ci-dessous concernent donc les SRV comme les IFN.

Par définition, un programme de supervision porte principalement sur les changements et les tendances. En outre, les IFN réalisés ces dernières années ont accru leurs observations des changements et des tendances. Par conséquent, le choix de la configuration des placettes, des plans d'échantillonnage, et peut-être des schémas de stratification, est conduit au moins partiellement par la volonté d'évaluer les changements.

CONFIGURATION DES PLACETTES

Le choix d'une configuration de placette est fondé sur plusieurs principes élémentaires dont beaucoup sont les mêmes qu'il s'agisse d'inventaire de région boréale, tempérée

ou tropicale, et dont certains sont aussi différents. Il est bien connu qu'une estimation précise des changements est plus difficile à obtenir que celle des conditions actuelles, en particulier lorsque les changements affectent uniquement une petite surface. La précision des estimations des changements peut être améliorée en remesurant les mêmes placettes à des périodes successives. De plus, la zone étudiée peut être stratifiée afin de réduire la variance en utilisant une variable associée à la probabilité du changement. L'accent mis sur l'estimation des changements dans les inventaires tropicaux plaide en faveur d'une proportion relativement importante de placettes permanentes ce qui, à son tour, plaide en faveur du marquage ou de la détermination de l'emplacement des arbres, de façon à les relocaliser facilement lors des inventaires successifs. Bien que l'établissement et la mesure d'une placette temporaire soient moins coûteux que ceux d'une placette permanente, établir et mesurer différentes placettes temporaires en deux occasions n'est pas nécessairement moins coûteux que d'établir, mesurer et remesurer une seule placette permanente.

Une forme spécifique de placette ne fait pas l'unanimité, mais les placettes circulaires sont souvent préférées car elles requièrent uniquement un point de contrôle unique, leur centre. Les placettes rectangulaires requièrent quatre points de contrôle, un à chaque coin. De plus, pour une surface donnée, une placette circulaire possède un périmètre plus petit, ce qui signifie que moins de décisions sont à prendre concernant des arbres limitrophes. De la même façon, déterminer les coordonnées d'arbres individuels, ce qui peut être nécessaire pour évaluer leur évolution, peut être plus facile dans une placette circulaire qui a un seul point de contrôle que dans une placette rectangulaire qui en a quatre. Par contre, si la densité des arbres est exceptionnellement forte, des placettes rectangulaires longues et étroites peuvent représenter une alternative plus réaliste.

Par souci d'efficacité logistique, les

programmes d'inventaire et de supervision comportent généralement des configurations de placettes en grappe. Configurer des placettes en grappe peut même être encore plus essentiel pour les programmes de pays tropicaux, en raison de problèmes d'accès prévisibles. La taille d'une placette individuelle et le nombre de placettes dans une grappe dépendent de plusieurs facteurs importants, généralement liés à des considérations logistiques, budgétaires et de précision (Tomppo et al., 2010a, 2011; Scott, 1993). En premier lieu, les placettes doivent être de petite taille et peu nombreuses au sein des grappes pour que l'équipe de terrain puisse mesurer une grappe entière en une journée. La part la plus importante du budget de mesure d'une placette dans les forêts boréales et tempérées est celle consacrée aux trajets vers et depuis l'emplacement de la placette ; cette part est susceptible d'être encore plus importante pour les forêts tropicales qui, dans beaucoup de régions, sont éloignées et peu accessibles. Le fait que les équipes de terrain ne soient pas obligées de retourner plusieurs fois de suite au même emplacement permet d'améliorer l'efficacité. Deuxième facteur important, les caractéristiques dimensionnelles des placettes, telles que le rayon des placettes circulaires ou la longueur des placettes rectangulaires, doivent être mesurées sur un plan horizontal et non pas sur un terrain irrégulier. Des mesures sur un plan horizontal étant plus difficiles à relever dans les grandes placettes, en particulier en terrain vallonné ou montagneux, des placettes de petite taille sont donc, là encore, préférables. Troisième facteur important, l'établissement de placettes permanentes plutôt que temporaires pour faciliter l'estimation des changements requiert généralement le marquage des arbres individuels ou la détermination de leur coordonnées. Cette approche est plus difficile dans les placettes de grande taille des forêts tropicales denses car un plus grand nombre d'arbres seront situés entre les arbres repérés et les points de contrôle. Un argument en faveur des placettes de grande taille dans les

inventaires de forêts tropicales est que ces dernières sont généralement plus diversifiées que les forêts boréales et tempérées, ce qui signifie que la surface totale inventoriée à chaque emplacement d'échantillonnage doit être plus étendue pour capturer une plus large diversité. Cette surface plus étendue peut aussi être obtenue en augmentant le nombre de petites placettes dans une même grappe. Cette approche est plus économique si les corrélations spatiales entre les observations des variables d'intérêt sont élevées mais diminuent à mesure que la distance augmente.

L'utilisation de sous-placettes de petite taille pour mesurer les arbres de petit diamètre permet également d'améliorer l'efficacité de l'échantillonnage. Pour les placettes circulaires, les sous-placettes sont généralement imbriquées, en d'autres termes elles forment des cercles concentriques, toutes ayant le même centre. Les tailles particulières des sous-placettes et les seuils de diamètre correspondant aux sous-placettes doivent être fondés sur le nombre supposé d'arbres présents dans les sous-placettes, les similarités prévisibles des arbres et la durée du trajet entre les sous-placettes d'une même placette, ou les placettes d'une même grappe.

Enfin, la nature distante et quasi inaccessible de nombreuses forêts tropicales signifie que les inventaires devront peut-être s'appuyer sur une combinaison de parcelles et de données de télédétection. Lors du choix d'une configuration des placettes, les facteurs dont dépend la télédétection devront donc être pris en considération. En guise d'exemple, une placette doit être suffisamment grande pour constituer un échantillon adéquat des arbres d'un élément au sol correspondant à un élément télédétekté (p. ex., pixel d'image satellite, empreinte Lidar) contenant le centre de l'élément au sol. De plus, l'alignement souhaité de différentes placettes d'une même grappe avec différents éléments télédétektés peut requérir que les distances entre les placettes soient au moins aussi grandes que les dimensions des éléments télédétektés.

PLAN D'ÉCHANTILLONNAGE

Le choix d'un plan d'échantillonnage pour un inventaire de forêts tropicales conduit à prendre également en considération plusieurs principes. Le plus important est généralement celui de l'équilibre spatial, selon lequel de grandes régions géographiques de la population étudiée ne doivent pas rester exclues de l'échantillonnage. L'équilibre spatial est souvent obtenu en incorporant une composante systématique dans le plan d'échantillonnage. Il peut s'agir d'un réseau de lignes de quadrillage perpendiculaires ou d'un morcellement de la population en polygones réguliers. Les plans alignés spatialement établissent des placettes aux intersections de la grille ou au centre des polygones tandis que les plans non alignés établissent des placettes de façon aléatoire dans les rectangles formés par les lignes quadrillées ou dans les polygones réguliers.

La prise en compte de caractéristiques de télédétection influe également sur le choix d'un plan d'échantillonnage. Par exemple, les forêts tropicales sont souvent caractérisées par une couverture nuageuse quasi-permanente. Par conséquent, il peut s'avérer difficile d'obtenir des images dénuées de nuages prises par des satellites comme Landsat ou SPOT. Les données Lidar, actuellement acquises depuis des plates-formes aériennes qui utilisent des techniques laser, sont souvent proposées comme alternative. L'avantage des impulsions laser est qu'elles pénètrent la canopée des forêts et produisent des informations utiles pour estimer le volume, la biomasse et le contenu en carbone des arbres. Si les placettes sont situées aux intersections des lignes de quadrillage, l'acquisition de données Lidar depuis des plates-formes aériennes en bandes est facilitée par l'utilisation de lignes de vol rectilignes.

Enfin, l'établissement de réseaux quadrillés et d'un morcellement de polygones offre la possibilité d'utiliser des projections de surfaces égales. Sinon, les placettes situées à une grande distance de l'équateur représenteront moins de surface de la population étudiée que

les placettes situées plus près de l'équateur. Bien qu'il existe des schémas de pondération applicables aux projections de surfaces inégales, ils sont souvent complexes et fastidieux.

Comme mentionné plus haut, la part la plus importante du budget de mesure d'une placette est celle consacrée aux trajets vers et depuis l'emplacement de la placette ; cette part peut être encore plus importante dans le cas de forêts tropicales situées dans des régions lointaines et inaccessibles (Tomppo et al., 2010a, 2011). Pour ces raisons, les placettes doivent être établies en grappe plutôt qu'isolées. Plusieurs approches de l'échantillonnage par grappe sont bien connues. L'une est de configurer une placette en plusieurs sous-placettes selon une répartition régulière et à proximité les unes des autres (McRoberts et al., 2005). Avec une telle approche, il est possible d'agréger les données de toutes les sous-placettes et de les attribuer au centre de la placette. Une autre approche est d'établir des placettes en grappes de forme rectangulaire, semi-rectangulaire ou autre forme géométrique (Tomppo, 2006). Une troisième approche est de choisir un échantillonnage par grappe en deux étapes. Cela consiste à sélectionner aléatoirement des unités d'échantillonnage principales telles que des polygones rectangles, puis d'établir plusieurs unités d'échantillonnage secondaires sous forme de placettes au sein des polygones sélectionnés, à des emplacements également aléatoires. Lors de l'utilisation d'un échantillonnage par grappe, la corrélation spatiale entre les observations des placettes d'une même grappe doit être prise en considération. Si les distances entre des paires de placettes sont inférieures à la plage de corrélation spatiale, les observations tendront à être similaires et l'échantillonnage risque de devenir moins efficace.

STRATIFICATION

Les approches de stratification de l'échantillonnage sont appliquées pour plusieurs raisons, mais la principale est de varier les intensités d'échantillonnage pour

respecter des critères sélectionnés. Par exemple, pour un SRV qui met l'accent sur les régions géographiques sujettes à des émissions de carbone d'origine humaine, des intensités d'échantillonnage moindres sont acceptables pour des régions lointaines et inaccessibles qui sont donc moins susceptibles d'être développées ou exploitées. De plus, les coûts liés à des intensités d'échantillonnage élevées dans ces régions éloignées risquent d'être prohibitifs. L'échantillonnage, même avec des intensités moindres, doit néanmoins être effectué dans ces régions en vertu de l'équilibre spatial à atteindre.

Plusieurs principes soutiennent également les approches de stratification de l'échantillonnage. Le premier est que des strates avec des frontières stables sont généralement préférables. Sinon, les changements de frontières des strates avec différentes intensités d'échantillonnage conduisent à des probabilités d'inclusion différentes d'échantillonnage et compliquent l'estimation. De plus, une estimation stratifiée requiert qu'une placette soit attribuée à une seule et unique strate. Si la strate à laquelle est attribuée une placette change entre les mesures, il devient difficile de savoir à quelle strate doivent être attribuées les observations de changement de la placette. Des strates définies par la topographie, les zones climatiques, les biomes ou les frontières politique peuvent donc être préférables à des strates définies par des attributs forestiers tels que la densité ou peut-être le type de forêt.

L'échantillonnage stratifié est le plus souvent mis en œuvre en appliquant l'un des trois schémas d'attribution de placette. Avec une attribution égale, le même nombre de placettes est attribué à toutes les strates quelle que soit leur taille. Ce schéma est choisi si l'objectif est d'obtenir des estimations par strate individuelle. Avec une attribution optimale, les intensités d'échantillonnage sélectionnées pour les strates sont basées sur des critères d'optimisation tels que le coût des mesures et/ou la variation, au sein d'une strate, des observations des variables d'intérêt telles que le volume ou la biomasse, ou leurs

changements éventuels. Des intensités plus élevées d'échantillonnage sont sélectionnées selon que les strates ont des variations plus importantes et/ou des coûts de mesure plus bas. Avec une attribution proportionnelle, les intensités d'échantillonnage sélectionnées sont proportionnelles à la taille des strates. L'étude Cochran (1977) fournit une discussion complète concernant des stratégies alternatives. Pour les pays tropicaux possédant de grandes régions éloignées et quasi inaccessibles, une certaine forme d'attribution optimale sera généralement nécessaire pour atténuer le coût excessif associé à un échantillonnage de ces régions. Une attribution proportionnelle et optimale peut facilement être mise en œuvre en utilisant des plans d'échantillonnage basés sur des réseaux de lignes de quadrillage perpendiculaires. Dans le cas d'une attribution proportionnelle, des placettes ou des grappes de placettes sont établies aux intersections du quadrillage, sans tenir compte de la strate associée à l'intersection. Dans le cas d'une attribution optimale, il est possible d'augmenter ou de diminuer les intensités d'échantillonnage de différentes strates en sélectionnant les intersections du quadrillage où des placettes sont établies. Par exemple, si l'intensité de l'échantillonnage doit être divisée par quatre, les placettes peuvent être établies aux intersections d'une ligne de quadrillage sur deux dans chaque direction.

ÉTUDE DE CAS – TANZANIE

Comme plan d'échantillonnage pour la Tanzanie, Tomppo et al. (2010a) ont utilisé un échantillonnage en deux phases pour la stratification et une attribution optimale de placettes aux strates. L'échantillonnage de la première phase consistait en une évaluation au niveau du bureau d'une grille dense de placettes de terrain afin de définir des catégories de volume et de budget. Sur la base de ces évaluations, des strates ont été créées en utilisant le volume moyen prévisible du peuplement final au niveau des grappes et le coût estimé pour mesurer une grappe de placettes. Les catégories de volume étaient

basées sur les prévisions de volume découlant de l'imagerie par satellite, des observations de placettes de terrain en dehors de la Tanzanie, et de modèles éprouvés dont les prévisions étaient étalonnées à partir d'estimations de volume aréal pour la Tanzanie. La méthode Neyman d'attribution (Cochran, 1977) a été appliquée à la sélection des catégories de volume de façon à maximiser la précision de l'estimation globale des volumes en présumant une taille fixe d'échantillon. Les prévisions budgétaires étaient basées sur les analyses SIG et l'opinion d'un expert local sur le nombre de jours (un, deux, plus de deux) nécessaires pour mesurer une grappe de placettes. Le choix des intervalles de catégorie, qui influe sur le gain éventuel s'agissant de la stratification, requiert une plus large investigation. L'échantillonnage de la seconde phase consistait en mesures sur le terrain de placettes où les intensités d'échantillonnage au sein des strates étaient sélectionnées selon le mode d'attribution optimale (Cochran, 1977). Dans une attribution optimale, les intensités d'échantillonnage sont proportionnelles à la quantité $\sigma_h / \sqrt{c_h}$ où σ_h est l'écart type, dans la strate, des observations de la variable d'intérêt (volume moyen de peuplement final) et c_h , le coût moyen en termes de temps de mesure d'une grappe de placettes dans la strate h . Des informations détaillées concernant le plan d'échantillonnage sont disponibles dans l'étude Tomppo et al. (2010a).

Sous les tropiques, l'utilisation des cartes de végétation existantes pour délimiter les terres en forestières et non forestières peut être tentante. Toutefois, si des grappes de placettes ne sont pas établies sur les terres non forestières de la même façon que sur les terres forestières qui ont été délimitées, des erreurs de carte pourraient conduire à des déviations car des terres forestières faussement délimitées comme non forestières ne seraient pas échantillonnées. Les coûts associés à l'échantillonnage des terres non forestières peuvent toujours être diminués en attribuant à ces terres des intensités moindres

d'échantillonnage. Par ailleurs, la mesure sur le terrain de grappes de placettes entièrement situées en dehors des forêts et dépourvues de peuplement final peut souvent être évitée en évaluant ces grappes grâce aux informations d'occupation des sols, obtenues d'autres sources fiables, comme ce qui a été proposé concernant le Brésil (Tomppo, 2009).

Le manque de voies de circulation, autres que les rivières, représente un défi particulier propre aux inventaires de forêt tropicale telle que le biome amazonien. Par exemple, les routes peuvent être praticables seulement une partie de l'année, approximativement six mois en Amazonie. S'y ajoute le fait que certaines forêts peuvent être réservées à des fins de préservation de la nature ou au seul usage de peuples indigènes. La stratification basée sur des variables pertinentes, telles que la probabilité de changements et les coûts des mesures, parle en faveur d'une adhésion à de sains principes de rentabilité et de statistique des inventaires.

6. Récapitulatif

La conception d'un plan d'échantillonnage approprié pour une EFN, un IFN ou un SRV, est une étape cruciale pour conduire à des estimations suffisamment précises et défendables d'un point de vue scientifique. L'une des premières étapes de ce processus est de définir la population cible et de choisir un cadre d'échantillonnage. Nous recommandons d'opter pour un cadre d'échantillonnage appliqué à une population infinie, dans lequel les observations et les mesures d'une zone supportant une placette de terrain sont attribuées au point central de la placette. Les inventaires étant généralement supposés fournir des estimations du changement, nous recommandons que le plan d'échantillonnage inclue un minimum de placettes permanentes. L'étape suivante consiste à répartir des placettes de terrain à travers la population à échantillonner. À partir des informations présentées, nous avons passé en revue plusieurs plans d'échantillonnage

connus : échantillonnage aléatoire simple, échantillonnage stratifié, échantillonnage systématique et échantillonnage par grappes. Si le plan d'échantillonnage inclut une composante systématique, nous recommandons la prudence lors de l'utilisation de grilles rectangulaires pour les populations cibles dont les composantes nord-sud sont de grande amplitude. Bien que le choix d'un plan d'échantillonnage particulier dépende de nombreuses considérations, nous recommandons que, si un échantillonnage stratifié n'est pas choisi, une stratification post-échantillonnage et une estimation stratifiée soient prises en considération. Enfin, des informations complémentaires sur ces thèmes et sur la problématique de plans d'échantillonnage plus complexes sont disponibles dans la documentation référencée.

7. Références et ressources techniques

Cochran, W. G. (1977). *Sampling techniques*. (3rd ed.). New York: Wiley.

FAO, 2001. *Global Forest Resources Assessment 2000. Main report.* FAO Forestry Paper 140. FAO Rome. 479 p.

Cunia, T. 1978. On the objectives and methodology of national forest inventories. In: *Forest resource inventory. Joint Meeting of IUFRO Groups S4.02 and S4.04.* 18-26 June 1978. Institutul de Cercetari si Amenageri Silvice., XI-XXIX. Bucharest.

Czaplewski, Raymond L. 2003. Chapter 5: Accuracy assessment of maps of forest condition: statistical design and methodological considerations, pp. 115-140. *Remote Sensing of Forest Environments: Concepts and Case Studies.* (Michael A. Wulder and Steven E. Franklin, Eds.) Kluwer Academic Publishers, Boston. 515p

Heikkinen, J. 2006. Assessment of uncertainty in spatially systematic sampling. In: Kangas, A. & Maltamo, M. (eds.). *Forest inventory. Methodology and applications. Managing Forest Ecosystems. Vol 10.* Springer, Dordrecht. p. 155-176.

Ivessalo, Y. 1927. The forests of Suomi Finland. Results of the general survey of the forests of the country carried out during the years 1921--1924. In *Finnish with English summary. Communicationes ex Instituto Quaestionum Forestalium Finlandie* 11. Summary in

English p. 321-395.

Loetsch, F., & Haller, K. E. 1973. Forest Inventory. Volume I. BLV Verlagsgesellschaft mbH, München. 439 pp.

Loetsch, F., Zöhrer, F. & Haller, K. E. 1973. Forest Inventory. Volume II. BLV Verlagsgesellschaft mbH, München. 469 pp.

Matérn, B. 1960. Spatial variation. Meddelanden från statens skogsforskningsinstitut. 49(5). 144 pp. Also appeared as Lecture Notes in Statistics 36. Springer-Verlag. 1986.

McRoberts, R.E. 2003. Compensating for missing plot observations in forest inventory estimation. Canadian Journal of Forest Research 33:1990-1997.

McRoberts, R.E., Wendt, D.G., Nelson, M.D., and Hansen, M.H. 2002. Using a land cover classification based on satellite imagery to improve the precision of forest inventory area estimates. Remote Sensing of Environment 81:36-44.

McRoberts, R.E., Bechtold, W.A.B., Patterson, P.L., Scott, C.T., & Reams, G.A. (2005). The enhanced Forest Inventory and Analysis program of the USDA Forest Service: historical perspective and announcement of statistical documentation. Journal of Forestry 103: 304-308.

Ranneby, B. 1981. Medelfelsformer till skattningar baserade på material från den 5:e riksskogstaxeringen. Abstract: Estimation of standard errors in systematic sampling. Swedish University of Agricultural Sciences, Section of Forest Biometry. Report 21. 19 s. (in Swedish).

Ranneby, B., Cruse, Th., Häggglund, B., Jonasson, H., Swärd, J. 1987. Designing a new national forest survey for Sweden. Studia Forestalia Suecica 177. 29 p.

Schreuder, H.T., Ernst, R., and Ramirez-Maldonado, H. 2004. Statistical techniques for sampling and monitoring natural resources. Gen. Tech. Rep. RMRS-GTR-126. USDA Forest Service, Rocky Mountain Research Station, Ft. Collins, CO. Available online at: <http://www.treearch.fs.fed.us/pubs/viewpub.jsp?index=6287>.

Schreuder, H. T., Gregoire, T. G., Wood, G. B. 1993. Sampling Methods for Multiresource Forest Inventory. John Wiley & Sons, Inc, New York. 446 p. Scott, C.T. 1993. Optimal design of a plot cluster for monitoring. In: K. Rennolls, and G. Gertner (Eds.). Proceedings of the IUFRO S.4.11 Conference. The optimal design of forest experiments and forest surveys. pp. 233-242. Available by request from the author: ctscott@fs.fed.us.

Tomppo, E. 2004. Resource assessment: Forest

resources. In: Burley, J., Evans, J. & Youngquist, J.A. (eds.). Encyclopedia of forest sciences. Elsevier, p. 965-973.

Tomppo, E. 2006. The Finnish National Forest Inventory. In: Kangas, A., & Maltamo, M. (Eds.). Forest inventory, methodology and applications. Springer. 179-194.

Tomppo, E. 2009. Establishing Methodological Basis and Building Partnership for Brazil's National Forest Inventory. A mission report for the Food and Agriculture Organization of the United Nations (FAO), Division FOMRD.

Tomppo E., & R.L. Czaplewski. 2002. Potential for a remote-sensing-aided forest resource survey for the whole globe. Unasylva 53(210):16-18.

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

Tomppo, E., Czaplewski, R. L., & Makisara, K. 2002. The role of remote sensing in global forest assessment: A remote sensing background paper for Kotka IV expert consultation. Forest Resources Assessment (FRA) Working Paper 61. Food and Agriculture Organization of the United Nations, Forest Resources Division, Forest Resources Assessment Programme, Rome, Italy.

Tomppo, E., Henttonen, H. & Tuomainen, T. 2001. Valtakunnan metsien 8. inventoinnin menetelmä ja tulokset metsäkeskuksittain Pohjois-Suomessa 1992-94 sekä tulokset Etelä-Suomessa 1986-92 ja koko maassa 1986-94. Metsätieteen aikakauskirja 1B/2001: 99-248. (in Finnish).

Tomppo, E., Gschwantner, T., Lawrence, M. & McRoberts, R.E. (eds.) 2010b. National Forest Inventories - Pathways for common reporting. Springer, 612 p. ISBN 978-90-481-3232-4

Tomppo, E., Heikkinen, J., Henttonen, H.M., Ihalainen, A., Katila, M., Mäkelä, H., Tuomainen, T. & Vainikainen, N. 2011. Designing and conducting a forest inventory - case: 9th National Forest Inventory of Finland. Springer, Managing Forest Ecosystems 21. 270 p. ISBN 978-94-007-1651-3

Tomppo, E., Varjo, J., Korhonen, K., Ahola, A., Ihalainen, A., Heikkinen, J., Hirvelä, H., Mikkilä, H., Mikkola, E., Salminen, S. & Tuomainen, T. 1997. Country report for Finland. In: Study on European Forestry Information and Communication Systems. Reports on forestry inventory and survey systems. Vol. 1. European Commission, p. 145-226.

Tomppo, E. Katila, M., Mäkisara, K., Peräsaari, J., Malimbwi, R., Chamuya, N., Otieno, J. Dalsgaard, S., & Leppänen, M. 2010a. A Report to the Food and Agriculture Organization of the United Nations

(FAO) in support of Sampling Study for National Forestry Resources Monitoring and Assessment (NAFORMA) in Tanzania. Available at <http://dl.dropbox.com/u/6700344/NAFORMA%20Sampling%20Study.pdf>

White, D, Kimerling, J., and Overton, S.W. 1992. Cartographic and geometrics components of a global sampling design for environmental monitoring. *Cartography and Geographic Information Systems* 19:5-22.