



HAL
open science

Comparaison d'estimateurs de la ressource forestière tropicale basés sur plan de sondage et sur modèle compte tenu d'une post-stratification

Hélène Dessard

► To cite this version:

Hélène Dessard. Comparaison d'estimateurs de la ressource forestière tropicale basés sur plan de sondage et sur modèle compte tenu d'une post-stratification. *Annals of Forest Science*, 1999, 56 (8), pp.651-665. <hal-00883307>

HAL Id: hal-00883307

<https://hal.science/hal-00883307v1>

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Comparaison d'estimateurs de la ressource forestière tropicale basés sur plan de sondage et sur modèle compte tenu d'une post-stratification

Hélène Dessard*

Cirad-forêt, Campus international de Baillarguet, BP 5035, 34 032 Montpellier Cedex 1, France

(Reçu le 17 décembre 1998 ; accepté le 16 juin 1999)

Abstract – Comparison of design-based and model-based estimates for tropical forestry resource with post-stratification.

Forest resource management planning provides today the guide line for harvesting: the forest is cut in high yield units from which forest managers successively extract quantities of timber. Management mapping needs local assessment of the resource over the whole forest from a forest inventory. Prediction of the resource is made by two methods of kriging: one is ordinary kriging and the other one, named stratified kriging, takes into account an auxiliary qualitative variable. Results obtained from these technics are compared with those yielded by classical sampling. Thanks to an exhaustively sampled area, one can judge more objectively the suitability of each technique. If the total is similar for the different estimators, variances obtained by kriging are smaller. The difference between ordinary and stratified kriging concerns the distribution of the estimation. Taking into account the stratification was not very efficient and it might be better to use survey estimators. © 1999 Inra/Éditions scientifiques et médicales Elsevier SAS.

management planning / forest inventory / kriging / post-stratification

Résumé – L'exploitation forestière est aujourd'hui guidée par un plan d'aménagement, c'est-à-dire par un découpage en grandes unités de la forêt d'où l'exploitant extrait successivement une certaine quantité de bois. L'établissement de ce plan passe en premier lieu par l'estimation locale de la ressource sur l'ensemble de la forêt, effectuée à partir d'un inventaire forestier. La prévision de la ressource est réalisée par deux méthodes de krigeage : l'un est le krigeage ordinaire et l'autre, appelé krigeage stratifié tient compte d'une variable auxiliaire qualitative. Les résultats obtenus par ces diverses procédures sont comparés à ceux obtenus par les méthodes usuelles relevant de la théorie des sondages. Il est également possible de juger de façon plus objective la pertinence de chaque technique en comparant les estimations des totaux à ceux d'une région échantillonnée exhaustivement. Si le total estimé ne diffère guère pour chacune des méthodes les variances sont plus faibles pour les procédures de krigeage. La distribution des estimations par krigeage stratifié est plus homogène que par krigeage ordinaire pour un total équivalent. La prise en compte de la stratification n'a finalement pas permis d'améliorer les estimations et nous recommandons plutôt d'utiliser les estimateurs relevant des plans de sondage. © 1999 Inra/Éditions scientifiques et médicales Elsevier SAS.

plan d'aménagement / inventaire forestier / krigeage / post-stratification

1. Introduction

Dans le cadre d'une gestion durable de la forêt tropicale, les inventaires forestiers servent à établir un plan

d'aménagement. Au delà de la simple estimation de la ressource forestière et de sa précision, la gestion des forêts s'appuie de plus en plus [11] sur le découpage de la forêt en régions ou unités, exploitées les unes après

* Correspondance et tirés à part
E-mail : dessard@cirad.fr

les autres pendant un certain temps. Le temps de rotation total, c'est-à-dire la période au bout de laquelle l'exploitant revient à la première unité exploitée doit permettre la reconstitution d'un certain taux de la ressource. Lorsque le parcellaire est établi à quantité de ressource égale, en l'occurrence à volume exploitable égal par unité, il est beaucoup plus simple d'effectuer le découpage en unité si l'on possède déjà une cartographie de l'estimation sur l'ensemble de la forêt. Or la construction du parcellaire est à l'heure actuelle très empirique puisque les outils permettant son élaboration, voire son optimisation « automatique », compte-tenu de diverses contraintes, quelles soient de nature géographiques ou statistiques, n'existent pas encore.

Classiquement, les inventaires en forêts tropicales sont réalisés selon un plan d'échantillonnage systématique en transects continus. Ces transects constituent des unités primaires et sont découpés en unités secondaires, dans ce cas des placettes rectangulaires contiguës, repérées par leur centre géométrique et échantillonnées exhaustivement. Le traitement statistique devrait en principe relever de la procédure de *single stage cluster sampling* [4] pour estimer la ressource sur l'ensemble de la forêt. En revanche, la délimitation des unités d'exploitation (UE) nécessite une estimation locale sur de petites sous-unités dont l'agrégation permet la constitution des UE. Cette estimation locale pourrait être effectuée par exemple par la méthode des estimateurs à noyaux [13] : il faut cependant choisir la taille du voisinage, c'est-à-dire un nombre « raisonnable » de placettes. Le calcul du coefficient de variation donnera par la suite une idée de la précision de l'estimation. Cette méthode constitue une première approche du problème mais reste néanmoins soumise à l'arbitraire du voisinage. Traditionnellement, la méthode employée, tout aussi fastidieuse mais plus simple, utilise la théorie des sondages [5] et consiste pour chacune des UE délimitées a priori, à appliquer les formules de calcul du total (ou de la moyenne) et de sa variance à un échantillon de placettes considéré comme aléatoire simple. L'opération est répétée pour divers découpages de la forêt en UE jusqu'à l'obtention d'UE de volumes approximativement égaux [11].

Une cartographie de l'estimation de la ressource s'impose donc comme aide à la constitution du parcellaire, à condition que la méthode d'estimation soit la plus objective possible, autrement dit qu'un critère d'erreur soit minimisé.

L'analyse statistique usuellement employée néglige en général les corrélations entre les unités secondaires. Cette pratique est justifiée lorsque la variable d'intérêt est distribuée selon une loi uniforme dans la forêt. En revanche, si l'on suppose que la variable mesurée est structurée spatialement, une méthode naturelle d'estima-

tion spatiale qui prenne en compte les corrélations est le krigeage [1].

On dispose également d'une variable auxiliaire correspondant au type de strate forestière déterminée par photo-interprétation (carte de la stratification du massif de Sangha-Mbaéré). On considère que cette information supplémentaire doit a priori améliorer les estimations locales puisque d'une part elle est échantillonnée exhaustivement sur le domaine et d'autre part elle est sensée être corrélée avec la variable d'étude.

L'objectif de ce travail est de comparer deux approches de l'estimation de la ressource : l'une, basée sur plan de sondage et l'autre basée sur un modèle par la technique du krigeage, compte-tenu ou non d'une variable auxiliaire de nature qualitative.

Pour cela, toutes les estimations sont effectuées au sein d'un bloc du parcellaire contenant quatre transects tronqués. Ce nombre restreint d'unités primaires ne permet pas d'estimer correctement la variance de l'estimateur du total de la ressource sur le bloc par *single stage cluster* [4]. On explicitera donc les hypothèses émises pour travailler à partir de plans de sondage aléatoire stratifié et post-stratifié.

On aimerait donc apporter des éléments de réponse à deux questions :

- i) quelle est la « meilleure » technique d'estimation au sens de l'erreur quadratique moyenne : celle fondée sur un modèle, comme dans le cas du krigeage ou celle qui se fonde sur un plan de sondage, même inadapté a priori ?
- ii) la stratification proposée est-elle pertinente, autrement dit permet-elle réellement d'améliorer la précision des estimateurs de la ressource ? On pourrait aussi reformuler la question dans le sens suivant : la variable retenue explique-t-elle tout ou en partie la stratification ?

2. Matériel

2.1. L'inventaire d'aménagement

L'inventaire est réalisé selon un plan de sondage systématique : des layons de 25 m de largeur espacés de 1 000 m les uns des autres traversent la forêt dans sa totalité si bien que la longueur de chaque layon est déterminée par les lisières de la forêt. Chaque layon est découpé en parcelles de 100 m de longueur. Pour chaque parcelle, tous les arbres de diamètre supérieur à 40 cm sont identifiés et classés dans une classe diamétrique. La surface de la forêt est de 105 521 hectares (ha), le taux de sondage avoisine donc 2,5 %.

2.2. Une variable auxiliaire : la stratification du couvert

Une stratification du couvert forestier a été établie par [10] à partir de l'analyse de photographies aériennes : dix strates ont été déterminées (tableau I).

Les strates 1, 2, 3 représentent un pourcentage de fermeture du couvert végétal, 4, 5, 6, 7, 9 caractérisent l'état de la forêt (zones de recrues, forêt inondable, etc), tandis que les strates 10 et 11 correspondent à des formations homogènes.

2.3. L'inventaire d'exploitation

Une parcelle de 1 500 (3 km x 5 km) hectares environ située au centre de la forêt (la carte de la stratification situe l'inventaire) a été échantillonnée exhaustivement, mais à la différence de l'inventaire d'aménagement, les comptages des arbres de diamètre supérieur à 120 cm sont tous regroupés au sein d'une même classe. Les comptages sont effectués au sein de quadrats de 25 m x 25 m de surface. Cet inventaire servira de jeu de valida-

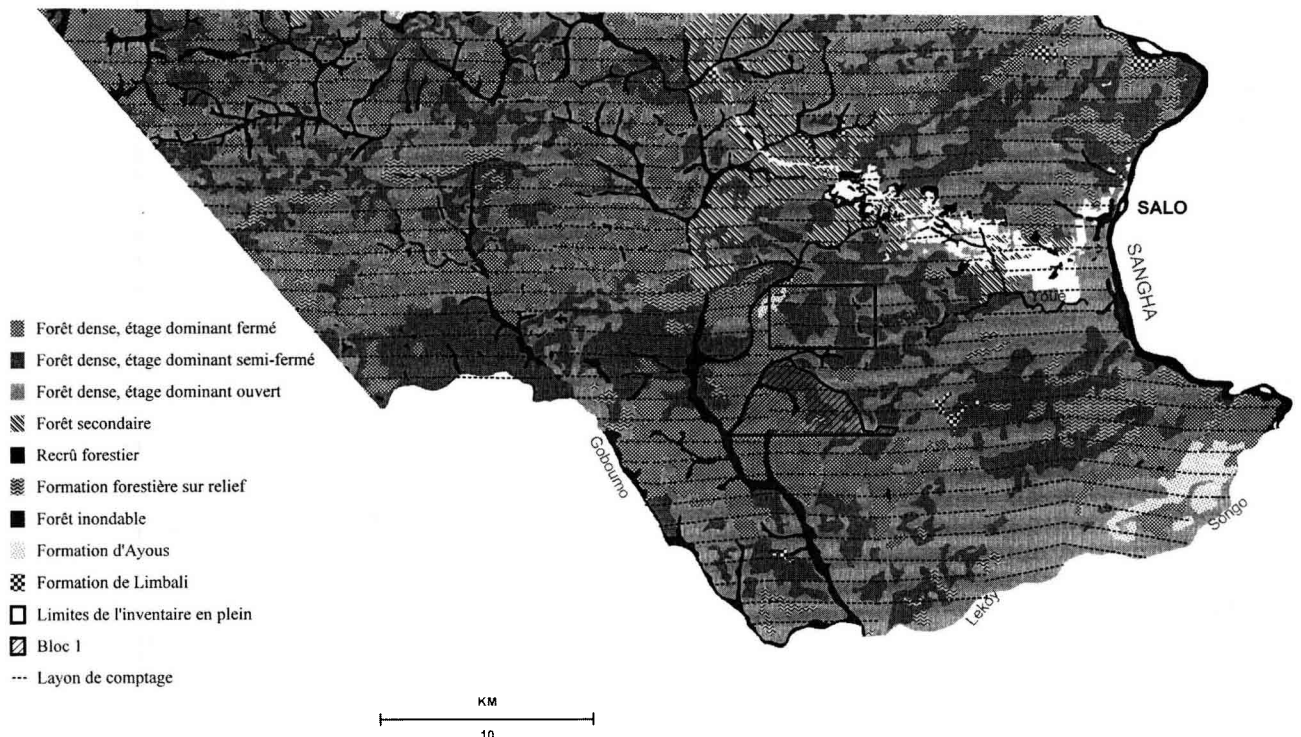
Tableau I. Type de strates.

Identifiant	Type	Surface total (ha)
1	Forêt dense D > 60 %	21 171
2	Forêt dense 20 % < D < 60 %	26 683
3	Forêt dense D < 20 %	37 780
4	Forêt secondaire	4 398
5	Recrû forestier	508
6	Formation forestière sur relief	4 522
7	Forêt inondable	7 353
9	Jachères	1 699
10	Ayous	1 032
11	Limballi	375

tion des techniques utilisées après transformation des variables.

L'ensemble de l'étude portera sur une essence commercialisée prioritaire : le Sapelli (*Entandrophragma cylindricum*). Cette essence est à peu près présente sur l'ensemble de la forêt à raison de 0,5 à 1 tige par hectare pour des diamètres supérieurs à 60 cm. Elle représente 70 % des espèces exploitées en République Centrafricaine.

STRATIFICATION FORESTIERE DU MASSIF DE SANGHA



La variable interpolée est la surface terrière, c'est-à-dire la surface de la section de tous les arbres à hauteur de 1,30 mètre par unité d'aire, en principe un hectare. Cette surface sera en fait ramenée à l'unité de base, c'est-à-dire la placette de 0,25 hectare. On a choisi la surface terrière plutôt que le volume, usuellement estimé, afin de ne pas utiliser les tarifs de cubage qui ajoutent un arbitraire aux données brutes.

3. Méthodes

3.1. Analyses exploratoires

On examine, dans un premier temps, la distribution de la surface terrière à l'aide d'un histogramme sur l'ensemble de la forêt, puis à l'intérieur des strates par des box-plots. Ces descriptions seront suivies d'une analyse de variance à un facteur à dix modalités en plan déséquilibré de façon à détecter un effet de la stratification.

Dans un deuxième temps, on examine les données en tenant compte de leur spatialisation au moyen du *h-scatterplot*. Introduite par [12], cette analyse consiste à représenter le graphe des valeurs observées de $Z(x+h)$ contre $Z(x)$. La dispersion du nuage de points autour de la bissectrice reflète la variabilité des mesures séparées d'une distance h . La forme du nuage permet de détecter aussi bien des points qui présentent des valeurs particulières qu'une dérive, autrement dit, un écart à la stationnarité.

3.2. Interpolation par krigeage

L'interpolation est d'abord effectuée par krigeage ordinaire [7], sans tenir compte de la stratification puis nous utilisons un système de krigeage qui incorpore une variable auxiliaire qualitative. Ce sont surtout les méthodes de cokrigeage [15], qui tiennent compte d'une ou plusieurs variables auxiliaires quantitatives, qui ont fait l'objet de développements théoriques. En revanche, le cas des variables exogènes qualitatives n'est que peu traité [6] et c'est pourquoi nous proposons la méthode suivante.

3.2.1. Krigeage stratifié

Modèle

On fait l'hypothèse que la variable $Z(x)$ se décompose en une moyenne $m(x)$ constante dans chaque strate au point x et un résidu $Y(x)$ qui vérifie l'hypothèse stationnaire [15] d'ordre 2 :

$$Z(x) = m(x) + Y(x) \text{ où } m(x) = \sum_{j=1}^J m_j \mathbb{1}_{S_j}(x),$$

Avec

$$E[Y(x)] = 0 \text{ et } E[Y(x)Y(x+h)] = C(h) \quad \forall x, h, j \in W.$$

Le covariogramme $C(h)$ est donc évalué indifféremment pour toutes les paires de points x et $x+h$, que x et $x+h$ appartiennent à la même strate ou non. La part de variabilité spatiale éventuellement due à la stratification est donc incorporée dans le covariogramme.

Système de krigeage

On construit de manière classique l'estimateur du krigeage $Z^*(x_0)$ au point x_0 , qui est linéaire, sans biais et dont la variance d'erreur $\sigma_E^2 = \text{Var}(Z^*(x_0) - Z(x_0))$ est minimale. On note $Z^*(x_0) = \sum_{\alpha} \lambda_{\alpha} Z(x_{\alpha})$. La condition de non biais $E[Z^*(x_0) - Z(x_0)] = 0$ se développe comme suit :

$$\begin{aligned} & E \left[Z^*(x_0) - Z(x_0) \right] \\ &= E \left[\sum_{\alpha} \lambda_{\alpha} Y(x_{\alpha}) - Y(x_0) \right] \\ &+ \sum_{j=1}^J m_j \left(\sum_{\alpha} \lambda_{\alpha} \mathbb{1}_{S_j}(x_{\alpha}) - \mathbb{1}_{S_j}(x_0) \right) \end{aligned} \quad (1)$$

Le premier terme de (1) est toujours nul en raison des hypothèses sur Y . En revanche, pour que le second terme soit toujours nul pour tous m_j et x_{α} il suffit que :

$$\sum_{\alpha} \lambda_{\alpha} \mathbb{1}_{S_j}(x_{\alpha}) = 1 \text{ si } x_0 \in S_j$$

$$\text{et } \sum_{\alpha} \lambda_{\alpha} \mathbb{1}_{S_j}(x_{\alpha}) = 0 \text{ si } x_0 \notin S$$

Il faut noter que comme $\sum_{j=1}^J \mathbb{1}_{S_j}(x) = 1$, l'ensemble des pondérateurs vérifie toujours $\sum_{\alpha} \lambda_{\alpha} = 1$. Ce système de pondération permet donc d'accorder une importance plus grande aux points appartenant à la même strate que x_0 (ils ont un poids total égal à 1) et plus faible dans les autres strates (le poids total y est nul), tout en faisant intervenir dans le système de krigeage. En notant $C_{\alpha\beta} = E[Y(x_{\alpha})Y(x_{\beta})]$, la variance d'erreur s'écrit :

$$\sigma_E^2 = \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C_{\alpha\beta} - 2 \sum_{\alpha} \lambda_{\alpha} C_{\alpha 0}$$

La minimisation de σ_E^2 est obtenue classiquement en dérivant l'expression ci-dessus par rapport aux λ_{α} sous

les J contraintes (3). À cet effet, on introduit les multiplicateurs de Lagrange $2\mu_j, j = 1, \dots, J$. On aboutit alors au système :

$$\begin{pmatrix} \mathbf{C} & \mathbf{F} \\ \mathbf{F}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Lambda} \\ \boldsymbol{\mu} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_0 \\ \mathbf{F}_0 \end{pmatrix}. \quad (2)$$

Dans le système (4), \mathbf{C} est la matrice $n \times n$ des covariances $C_{\alpha\beta}$, \mathbf{C}_0 est le vecteur de dimension n des covariances $C_{\alpha 0}$, $\boldsymbol{\Lambda}$ est le vecteur de dimension n des pondérateurs λ_α , $\boldsymbol{\mu}$ est le vecteur de taille J des pondérateurs de Lagrange, et $\mathbf{0}$ est une matrice $J \times J$ dont tous les éléments sont nuls. \mathbf{F} est la matrice $n \times J$ de l'indicatrice des strates au point x_α : $F_{\alpha i} = 1$ si $x_\alpha \in S_i$ et $F_{\alpha i} = 0$ si $x_\alpha \notin S_i$. De la même façon, \mathbf{F}_0 est le vecteur de taille J d'appartenance aux strates.

On vérifie aisément que la variance d'estimation vaut

$$\sigma_E^2 = \sigma^2 - \sum_{\alpha} \lambda_{\alpha} C_{\alpha 0} - \mu_{j_0}$$

où $\sigma^2 = \text{Var}(Y(x_0))$, j_0 est la strate au point 0 et μ_{j_0} le multiplicateur de Lagrange associé à la strate au point 0.

3.2.2. Procédure d'estimation par krigeage

Ajustement d'un modèle de variogramme

Il est plus usuel de donner les expressions pour le variogramme mais on retrouve la covariance par la formule $C(h) = C(0) - \gamma(h)$. Le variogramme retenu est un modèle sphérique de portée a , de seuil $C_1 = C(0)$, comportant un effet de pépité $\gamma(0)$, soit :

$$\gamma(h) = \begin{cases} \gamma(0) + C_1 \left[\frac{3}{2} \frac{h}{a} + \frac{1}{2} \frac{h^3}{a^3} \right] & 0 \leq h \leq a \\ \gamma(0) + C_1 & h > a \end{cases}$$

Il est plus usuel de donner les expressions pour le variogramme mais on retrouve $C(h) = C_1 - \gamma(h)$. Le variogramme est estimé par :

$$\hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} (Z(x_i) - Z(x_j))^2$$

où $N(h)$ représente l'ensemble de tous les points i et j tels que $d(x_i, x_j) = h$ (d pour la distance euclidienne) et $|N(h)|$ correspond au nombre de paires qu'il contient.

Le variogramme est ajusté par la méthode des moindres carrés [3]. Le paramètre $\gamma(0)$ appelé couramment « effet de pépité » correspond à une discontinuité à l'origine $h = 0$. Cette quantité représente les erreurs de mesures, de localisation et surtout la variabilité résiduel-

le du phénomène qui s'exprime aux échelles inférieures à celles disponibles expérimentalement. Le type de discontinuité est lié au modèle et s'interprète en terme de régularité spatiale du processus étudié. Ainsi dans un modèle sphérique, le comportement à l'origine est linéaire et témoigne d'un processus continu mais non différentiable. La portée a correspond à la distance au-delà de laquelle la corrélation spatiale est nulle et C_1 la valeur maximale atteinte par la covariance, c'est-à-dire $C_1 = C(0) = \gamma(\infty)$.

Effet de support

La formule du variogramme donnée ci-dessus correspond en fait au variogramme ponctuel, autrement dit à une variable mesurée sur un support réduit à un point. En pratique, les données sont en général accessibles à partir d'un volume ou d'une surface, par exemple une carotte de terre ou une placette. Implicitement, on mesure en fait une variable régularisée, c'est-à-dire intégrée (ou moyennée) sur son support. En principe, l'estimation requière l'utilisation d'un variogramme ponctuel déduit du variogramme régularisé (estimé à partir des données) [7] lorsque l'on prédit la variable sur un support de dimensions différentes. Dans notre cas, nous prendrons cet effet en compte lors du krigeage de l'inventaire en plein comme nous l'expliquons dans le paragraphe 3.4.1.

Définition du domaine d'interpolation

L'interpolation par krigeage sur le bloc est effectuée pour un ensemble de sites disposés selon une maille régulière $100 \text{ m} \times 25 \text{ m}$ entre les layons de comptage : ces points représentent le barycentre des placettes interpolées qui possèdent alors une surface identique à celles mesurées. L'estimation est donc ponctuelle, réalisée à partir d'un modèle de variogramme (figure 4) à support considéré comme ponctuel. Cependant, comme les unités d'échantillonnage et les unités prédites sont en fait représentées par un point pour le même support, l'approximation émise ci-dessus n'a aucune conséquence sur la prédiction.

Calcul de la variance globale sur le bloc

La variance du total σ_B^2 prédite pour le bloc ne se déduit pas directement de la somme des variances d'estimation ponctuelles, puisque les placettes ne sont pas indépendantes. Soit n le nombre de sites x_0 d'estimation du bloc. Pour faciliter la lecture, on adopte les notations suivantes : on note pour les n sites d'estimation i , Z_i^* l'estimateur de la variable au site i et σ_{iE}^2 sa variance d'estimation ponctuelle. Les poids affectés à chaque observation x_α dépendent du site i si bien qu'ils sont indexés par α_i . En développant

$$\sigma_B^2 = V\left(\sum_{i=1}^n Z_i^*\right) \sum_i V(Z_i^*) + \sum_i \sum_l \text{cov}(Z_i^*, Z_l^*)$$

on a

$$\sigma_B^2 = \sum_{i=1}^n \sigma_{iE}^2 + \sum_{i=1}^n \sum_{l=1}^n \sum_{\alpha_i} \sum_{\alpha_l} \lambda_{\alpha_i} \lambda_{\alpha_l} C_{\alpha_i \alpha_l}$$

Les covariances $C_{\alpha_i \alpha_l}$ entre les sites de sondage α_i et α_l sont obtenues à partir du modèle de variogramme ajusté (figure 4).

3.3. Comparaison des estimations obtenues pour des sondages stratifiés et post-stratifiés

3.3.1. Explicitation des approximations usuelles

L'évaluation de la ressource à partir d'inventaires tropicaux en transects continus systématiques est usuellement effectuée par l'emploi de statistiques simples mais qui normalement relèvent de plans de sondage soit aléatoire simple ou stratifié. Les estimateurs des totaux et de leurs variances sont alors biaisés. Il nous paraît important d'explicitier les hypothèses presque toujours implicites en pratique, qui permettent, en pratique, l'utilisation de ces statistiques.

i) La variable auxiliaire est utilisée comme une pré-stratification alors qu'en réalité elle apparaît comme une post-stratification.

ii) Les strates sont considérées comme indépendantes alors qu'elles ont des relations spatiales qui laissent supposer une corrélation, même faible.

iii) Le plan de sondage est souvent considéré comme aléatoire simple alors qu'il est de type systématique simple à un degré.

Bien que la stratification soit antérieure à la mise en œuvre de l'inventaire d'aménagement, elle n'a pas été utilisée pour la construction d'un plan de sondage à allocation proportionnelle ou optimale [2] : en effet, les contraintes de terrain en forêt tropicale sont telles qu'il est plus avantageux de progresser de façon continue, une fois le layon ouvert. Cette information auxiliaire doit donc être en toute rigueur utilisée comme une post-stratification.

L'estimateur qui serait a priori bien adapté à ce type de sondage serait donc un estimateur du total pour un sondage systématique simple à un degré où les unités secondaires sont post-stratifiées. Toutefois, l'expression de la variance de cet estimateur n'est pas simple [14] et son estimation resterait également approximative du fait

du nombre de transect trop faible sur le bloc 1 (quatre transects).

Usuellement, l'estimateur employé est celui du total pour un sondage aléatoire stratifié. Il comporte cependant des biais induits par les hypothèses énoncées ci-dessus. Si l'objet de ce travail n'est pas de fournir les expressions théoriques de ces biais, il n'en est pas moins intéressant de comparer les estimations du total et de leur précision obtenues pour les estimateurs du krigeage et de l'estimateur d'un sondage post-stratifié.

On note pour la strate j indiquée par j :

n_j le nombre de placettes, z_{ij} la variable mesurée sur la placette α de surface $s_{\alpha j}$, \bar{s}_j la moyenne des surfaces des placettes, S_j la surface totale de la strate j , $\bar{z}_j = \frac{1}{n_j} \sum_{\alpha=1}^{n_j} z_{\alpha j}$ la

moyenne estimée de Z et $s_j^2 = \frac{1}{n_j - 1} \sum_{\alpha=1}^{n_j} (z_{\alpha j} - \bar{z}_j)^2$ la

variance empirique. Les estimateurs requièrent une estimation du nombre total d'unités élémentaires dans chacune des strates, soit $N_j = \frac{S_j}{s_j}$.

Pour les deux types de sondage, l'estimateur du total est $\hat{Z} = \sum_{j=1}^J N_j \bar{z}_j$ tandis que les estimateurs de la variance du total sont [4] :

- Stratifié (S)

$$\hat{V}(\hat{Z}) = \sum_{j=1}^J N_j \frac{N_j - n_j}{n_j} s_j^2$$

- Poststratifié (P)

$$\hat{V}(\hat{Z}) = \frac{(N-n)}{n} \left(\sum_{j=1}^J \frac{N_j}{N} s_j^2 \right) + \frac{(N-n)}{n^2} \left(\sum_{j=1}^J \left(1 - \frac{N_j}{N} \right) s_j^2 \right)$$

Ces deux estimateurs sont sans biais si N_j et s_j^2 sont connus, or ces termes ne sont pas forcément accessibles dans le contexte de cette étude. En effet, une UE peut très bien contenir une strate non échantillonnée (figure 5) puisque le parcellaire est conçu après l'inventaire. Ainsi, on connaît la surface de la strate 6 mais aucune placette ne la traverse, si bien qu'on ne possède pas d'échantillon « local » de cette strate pour le bloc 1.

Un moyen de quantifier les biais introduits par « l'oubli » de la strate 6 est d'utiliser la moyenne de la surface terrière calculée à partir de l'échantillon de la strate 6 sur l'ensemble de la forêt. Ainsi le total dans la strate 6 est estimé à 16 et la variance est approchée de la façon suivante : on alloue un nombre fictif de placettes secondaires en appliquant à la surface totale de la strate le taux d'échantillonnage moyen obtenu sur le bloc, soit deux à trois placettes : la variance serait alors comprise entre 671 et 1015. Dans cette étude, ce biais est négligeable puisque la surface de la strate 6 était petite au regard de la surface du bloc 1 mais il est évident que celui-ci peut très vite augmenter avec la taille de la surface non-traversée par les layons.

3.4. Méthode de validation

3.4.1. Procédure de krigeage sur l'inventaire en plein

L'inventaire en plein permet de comparer les estimations du total et leur précision obtenues par les techniques de krigeage et de sondage. Les données relevées diffèrent toutefois de celles de l'inventaire d'aménagement puisque d'une part, tous les arbres de diamètres supérieurs à 120 cm sont agrégés dans une classe unique et d'autre part, les dimensions de la placette unitaire de sondage sont de 25 m × 25 m. On a donc ajusté un modèle de variogramme à partir des données de l'inventaire d'aménagement, regroupées dans une seule classe pour les diamètres supérieurs à 120 cm. Le krigeage a ensuite été effectué sur toutes les placettes 25 m × 25 m. La procédure d'interpolation doit tenir compte de la réduction de la taille des placettes : en effet les sites d'estimation représentent en fait une surface (Se) de 0,0625 ha alors que les points de sondages servant à l'estimation du variogramme représentent une surface (S) de 0.25 ha. On utilise alors la technique du krigeage par bloc [7]. Concrètement, chaque surface est discrétisée en un certain nombre de points. On dérégularise le variogramme $\gamma(h)$ estimé sur S pour obtenir le variogramme ponctuel $\gamma_o(h)$ en appliquant les formules suivantes :

$$\gamma_o(h) = \gamma(h) + \dot{\gamma}_0(S, S) \text{ et } a_0 = a - L \text{ où } \dot{\gamma}_0(S, S) = \int_S \int_S \gamma_0(u - u') dud u'$$

et L correspond ici à la différence entre la plus grande longueur des surfaces soit 75 m.

Concrètement, on remplace dans le système de krigeage (4) C_0 respectivement par

$$\bar{C}_0(S, Se) = \int_S \int_{Se} \left(\gamma_0(\infty) - \gamma_0(u_\alpha - u) \right) du_\alpha du.$$

Cette déconvolution est basée sur des règles empiriques et $\bar{\gamma}_0(S, S)$ obtenu à l'aide d'abaques [7].

3.4.2. Mesure de la qualité des estimations par krigeage

La qualité de l'estimation peut être mesurée par les quantités suivantes :

$$B = \frac{1}{n} \sum_{\alpha=1}^n \left(Z(x_\alpha) - Z^*(x_\alpha) \right)$$

et

$$EQP = \frac{1}{n} \sum_{\alpha=1}^n \frac{\left(Z(x_\alpha) - Z^*(x_\alpha) \right)^2}{\sigma^2(x_\alpha)}$$

où $\sigma^2(x_\alpha)$ représente la variance d'estimation au point x_α due au modèle.

B correspond à la moyenne des biais locaux et fournit ainsi une mesure du sur-lissage ou sous-lissage du krigeage tandis que EQP estime l'adéquation du modèle aux données.

4. Résultats

4.1. Analyses exploratoires

Les histogrammes de la surface terrière par placette sont représentés pour les dix strates (*figure 1*).

Le pourcentage de placettes comptant au moins un arbre de l'essence considérée sur l'ensemble des placettes appartenant à une même strate est faible (pour toutes les strates), aussi la classe des valeurs nulles n'est pas représentée afin de pas écraser les autres classes. La distribution de la variable présente une forme leptokurtique classique en forêt tropicale humide. Les *box-plot* (*figure 2*) par strates indiquent des différences dans la distribution de la ressource par strate.

L'analyse de variance à un facteur est significative bien que la disproportion des effectifs entre les strates puisse diminuer la puissance du test. L'analyse a été alors reconduite dans un cadre non-paramétrique : le test de Kruskal-Wallis permet également de rejeter l'hypothèse de l'absence d'un effet strate au seuil de 5 %. Un test de comparaison multiple pour toutes les moyennes de type Kruskal-Wallis a été également effectué au seuil global de 20 % sans qu'aucune différence significative ne soit détectée. Le résultat de ce test n'est guère surprenant puisque la puissance de ce type de test est faible,

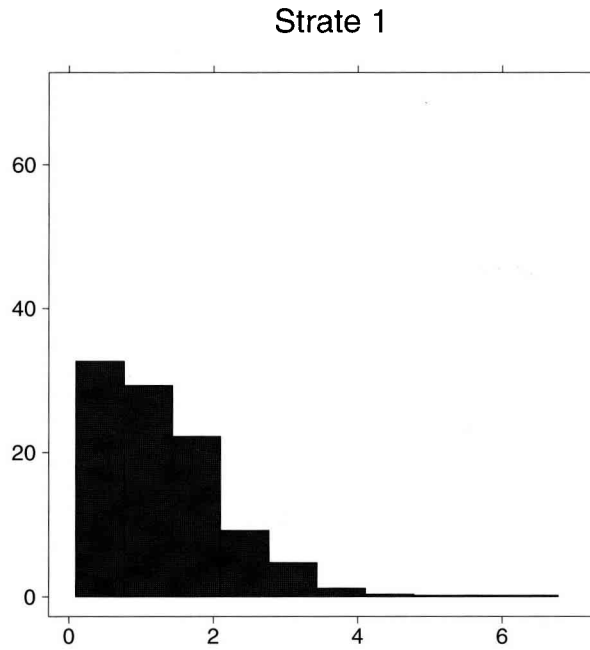


Figure 1. Histogramme de la surface terrière ($m^2 0,25 ha^{-1}$) du Sapelli sur l'ensemble de la forêt.

aucune conclusion hâtive n'est tirée et l'analyse est poursuivie.

Les résultats des *h-scatterplot* sont donnés dans la figure 3. La dispersion des nuages est à la fois assez étalée autour de la bissectrice et assez peu variable avec l'augmentation de l'interdistance h indiquent des corrélations spatiales assez faibles. Par contre, aucun écart à la stationnarité n'est détecté. Quelques points ont des fortes valeurs : ils correspondent à des placettes contenant des arbres de diamètres exceptionnellement importants et ont donc été éliminés de l'analyse.

4.2. Comparaison des estimations obtenues par krigeage et par sondage

Les résultats obtenus par les différentes techniques de krigeage sont comparés en les référant à l'estimation pour les plans de sondage. Les paramètres estimés du variogramme (figure 4) sont $\gamma(0) = 0,3245$, $C_1 = 0,370$ et $a = 410 m$. Z_T et $V(Z_T)$ correspondent respectivement à l'estimation du total de la surface terrière sur la zone d'étude et à l'estimation de la variance du total.

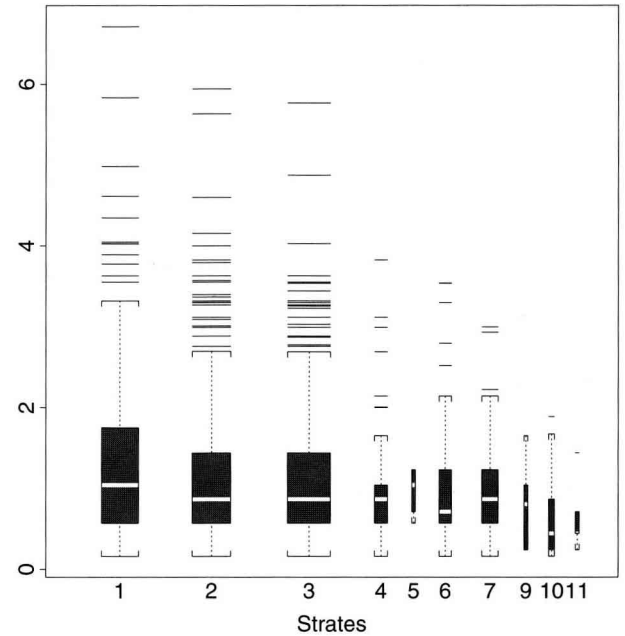


Figure 2. Boxplots des surfaces terrières du Sapelli.

Bloc 1

Le total obtenu pour un plan de sondage aléatoire simple est égal à 2 159 pour une variance de 160 225. La variance obtenue par post-stratification et par pré-stratification sont très proches et restent approximatives à la fois du fait des estimations du nombre de placettes, du taux moyen de sondage et du fait d'une strate non échantillonnée.

Inventaire en plein

À titre indicatif, nous donnons les estimations suivantes. La variance de l'estimateur du total post-stratifié est égale à 62 102. L'estimation aléatoire simple fournit un total de 1 360 avec une variance égale à 58 807. La variance pour un échantillonnage systématique non stratifié est égale à 58 441.

Pour les deux tableaux, les variances totales représentent la somme des variances d'erreurs locales plus les covariances entre les placettes calculées à partir du variogramme. En revanche, si l'on additionne les variances globales obtenues dans les strates, on obtient des totaux légèrement inférieurs à ceux obtenus par le calcul sur le bloc entier : ceci est simplement dû à la non prise en compte des corrélations entre strates.

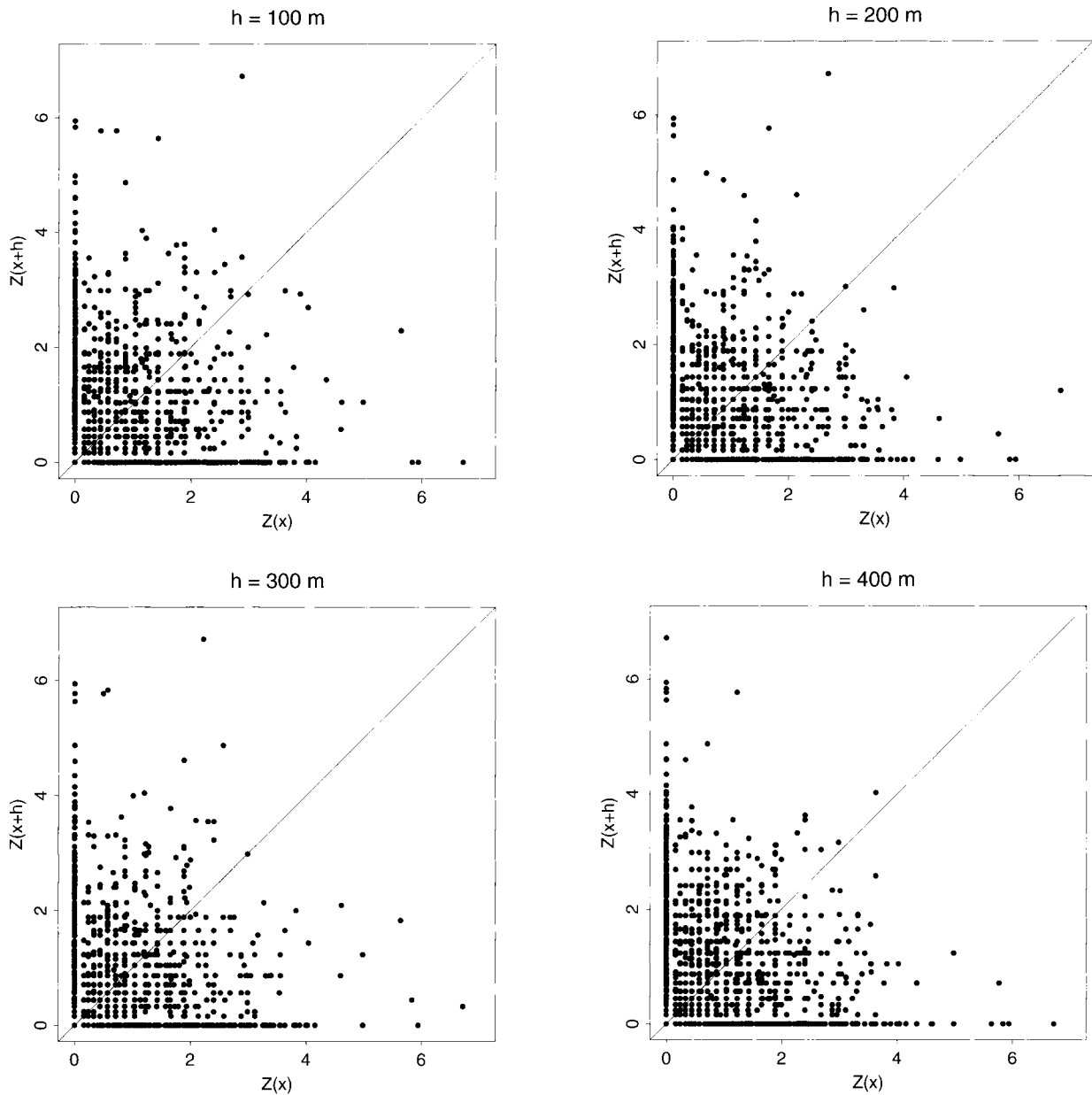


Figure 3. h -scatterplot de la surface terrière (m^2 $0,25 \text{ ha}^{-1}$).

5. Discussion

5.1. Bloc 1

Les variances obtenues par post-stratification et par pré-stratification sont très proches et restent approximatives à la fois du fait des estimations du nombre de

cettes, du taux moyen de sondage et du fait de la strate 6 non échantillonnée. Les totaux obtenus par les trois méthodes sont assez proches sur l'ensemble du bloc : les différences observées sur les estimations pour les plans de sondage sont trop faibles pour préférer une technique à une autre. En revanche, certaines différences sont notables si on considère les totaux obtenus sur chacune

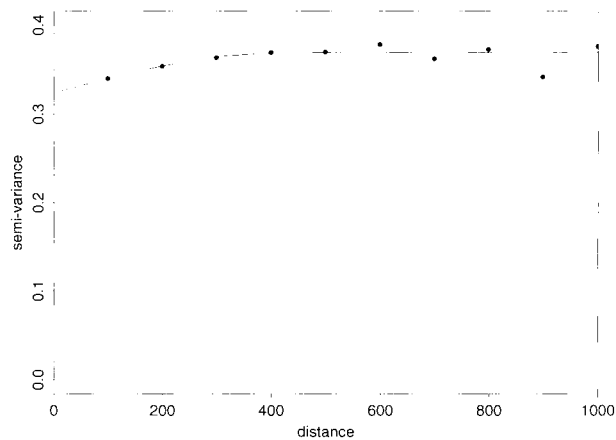


Figure 4. Variogramme de la surface terrière du Sapelli calculé sur l'ensemble de la forêt.

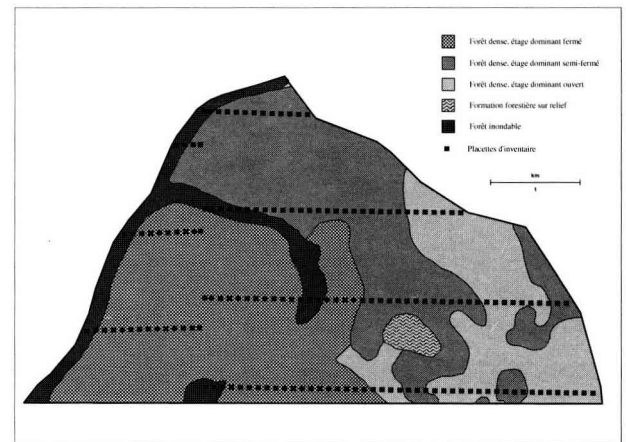


Figure 5. Une unité de prédiction : bloc 1.

des strates pour les techniques de krigeage. Le total dans la strate 6 est particulièrement sous estimé, voire ignoré par la méthode de krigeage ordinaire. L'explication est similaire à celle donnée dans le cas du sondage post-stratifié : le nombre de placettes entrant dans le système de krigeage appartenant à la strate 6 étant très réduit (le voi-

sinage glissant utilisé était supérieur aux dimensions du bloc). Globalement, le krigeage stratifié n'a pas apporté d'améliorations intéressantes, en revanche l'examen des histogrammes des prédictions pour les deux méthodes de krigeage révèle pour quelques strates une assez grande disparité des répartitions (figure 6).

Tableau II. Estimations de la surface terrière sur le bloc 1.

	Krigeage stratifié	Krigeage ordinaire	S	P		\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	$\hat{V}(\hat{Z}_T)$
	\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	$\hat{V}(\hat{Z}_T)$			
Strate 1	1 095	4 396	942	4 395	1 152	2 450		
Strate 2	579	4 362	635	4 358	666	1 698		
Strate 3	315	2 690	324	2 688	291	527		
Strate 6	0,85	143	27	143	0	0		
Strate 7	101	820	180	816	148	190		
Total	2 091	10 581	2 108	10 570	1 657	184 414	180 337	

Tableau III. Estimations de la surface terrière pour l'inventaire en plein.

	Comptage réel	Krigeage stratifié	Krigeage ordinaire S			Surfaces		
	Z_T	\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	\hat{Z}_T	$\hat{V}(\hat{Z}_T)$	S_j
Strate 1	204	160	411	219	409	149	9 414	144
Strate 2	876	985	1 783	900	1 782	732	28 738	796
Strate 3	466	423	830	460	829	356	19 410	479
Strate 7	34	26	28	20	27	52	2 644	23
Strate 10	62	54	114	62	112	50	1 182	48
Total	1 642	1 648	3 254	1 661	3 250	1 340	61 189	1 490

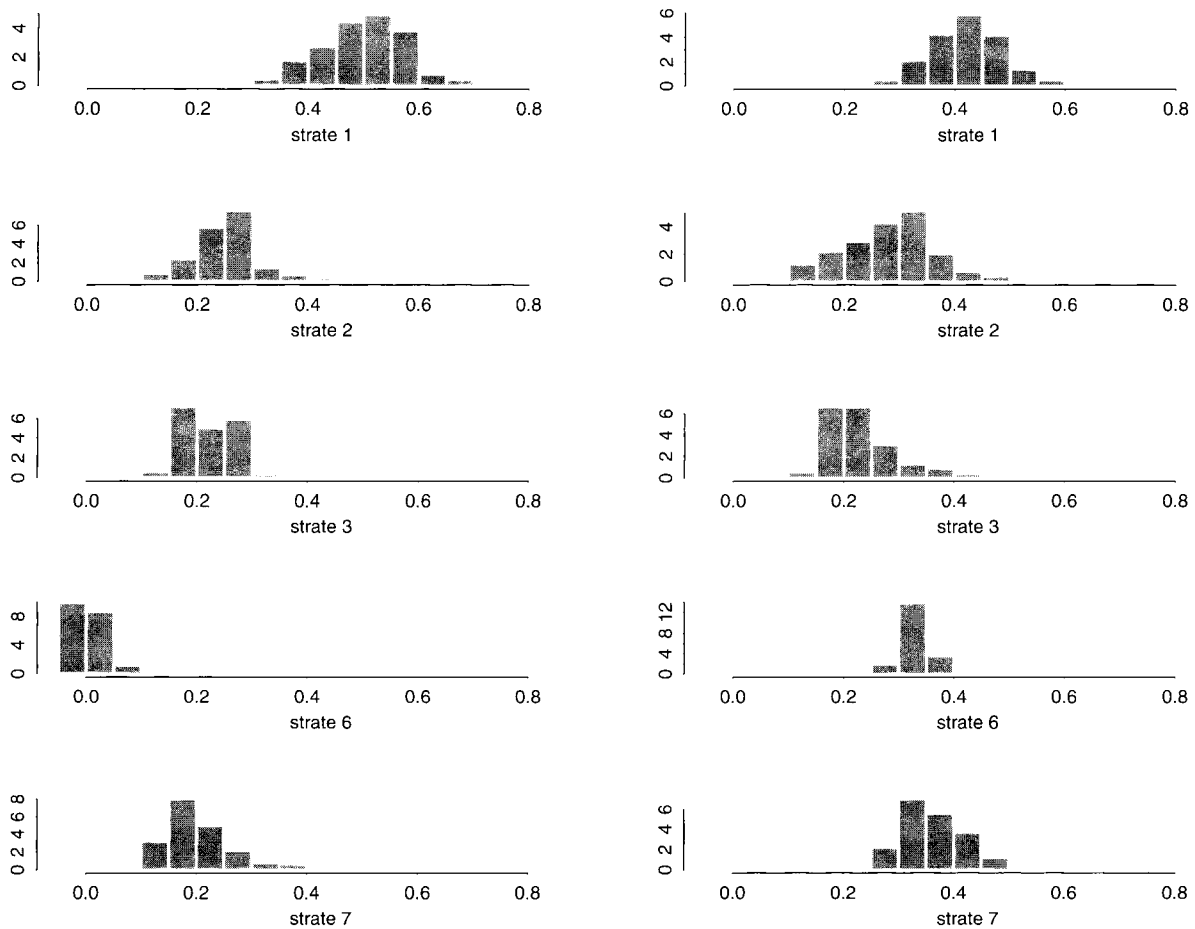


Figure 6. Estimations des surfaces terrières par strate du bloc 1 pour le krigeage ordinaire (à gauche) et le krigeage stratifié (à droite).

Les distributions des prédictions par krigeage stratifié sont plus « uniformes » et symétriques. Ceci est directement lié au système de contraintes sur les poids des observations : globalement les observations appartenant à la même strate que le site d'interpolation sont affectées d'un poids plus fort que sous krigeage ordinaire si bien que l'effet de lissage du krigeage est plus important. Autrement dit, à disposition géométrique égale, les contributions des valeurs faibles des sites de la strates sont plus élevées que sous KO et diminuent donc la disparité des estimations.

Les variances obtenues par krigeage sont plus fortes dans chacune des strates que celles issues d'un plan de sondage : c'est la prise en compte des covariances entre placettes qui augmentent les variances intra-strate.

Cependant, les variances estimées globales pour le bloc, issues du krigeage sont beaucoup plus faibles que les variances estimées issues du sondage. Ceci s'explique toutefois par la sous-estimation du seuil du variogramme, effet classique et inévitable du krigeage, et par la minimisation méthodologique des variances locales. Ainsi [8] trouve des réductions de variances du même ordre dans une étude similaire en forêt tempérée.

5.2. L'inventaire en plein

Globalement, la qualité des résultats obtenus par les deux types de krigeage est identique et satisfaisante, ce qui corrobore les observations précédentes. Le surlissage global est en fait du aux nombreux sites affectés d'une

Tableau IV. Mesure de la qualité des deux krigeages.

	Totalité	Strate 1	Strate 2	Strate 3	Strate 7	Strate 10
KO B	-0,16	-0,16	-0,16	-0,17	-0,18	-0,19
EQP	1,13	1,18	1,14	1,75	0,67	0,75
KS B	-0,16	-0,10	-0,18	-0,15	-0,25	-0,16
EQP	1,13	0,98	1,22	1,10	0,88	0,6

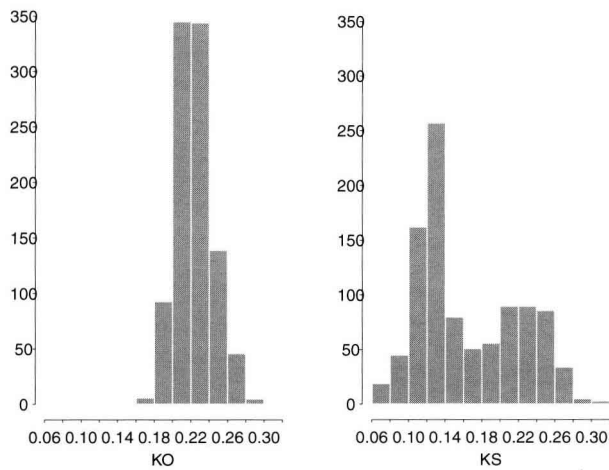


Figure 7. Histogramme des estimations pour la strate 1 sur l'inventaire en plein. KO = krigeage ordinaire/KS = krigeage stratifié.

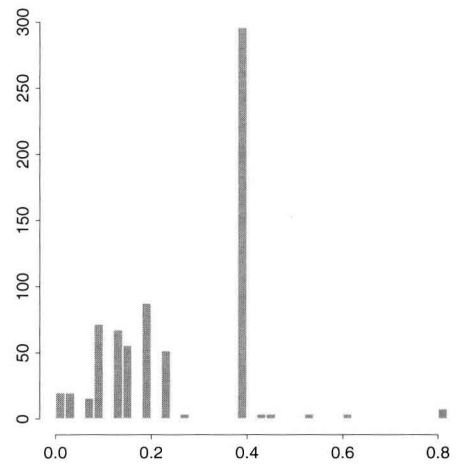


Figure 8. Histogramme des valeurs réelles pour la strate 1 sur l'inventaire en plein.

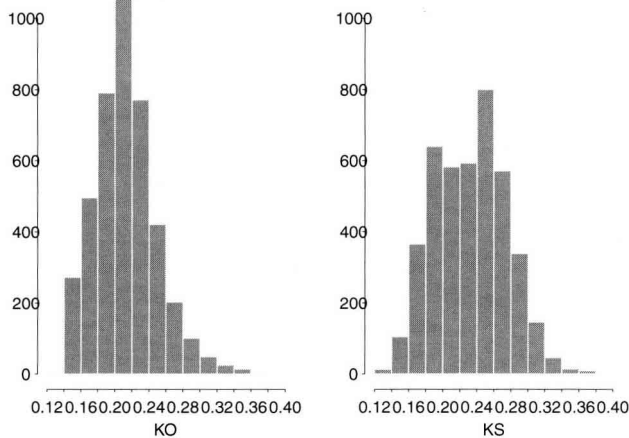


Figure 9. Histogramme des estimations pour la strate 2 sur l'inventaire en plein. KO = krigeage ordinaire/KS = krigeage stratifié.

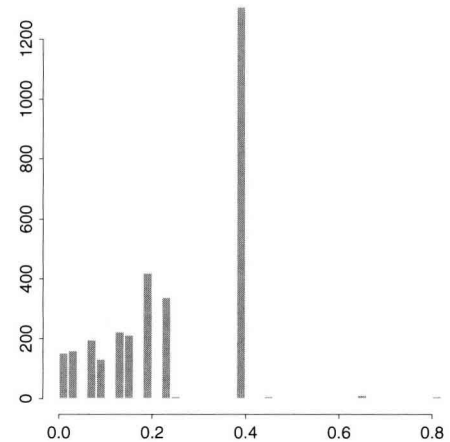


Figure 10. Histogramme des valeurs réelles pour la strate 2 sur l'inventaire en plein.

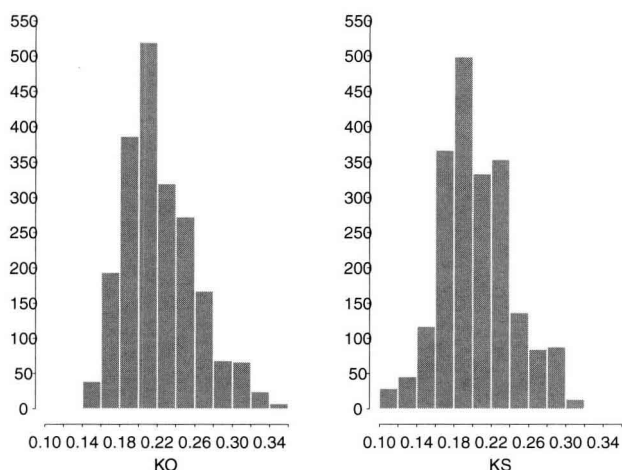


Figure 11. Histogramme des estimations pour la strate 3 sur l’inventaire en plein. KO = krigeage ordinaire/KS = krigeage stratifié.

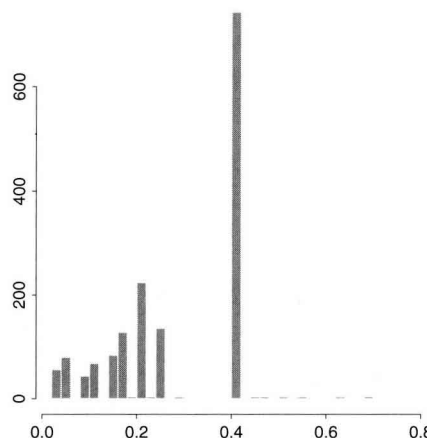


Figure 12. Histogramme des valeurs réelles pour la strate 3 sur l’inventaire en plein.

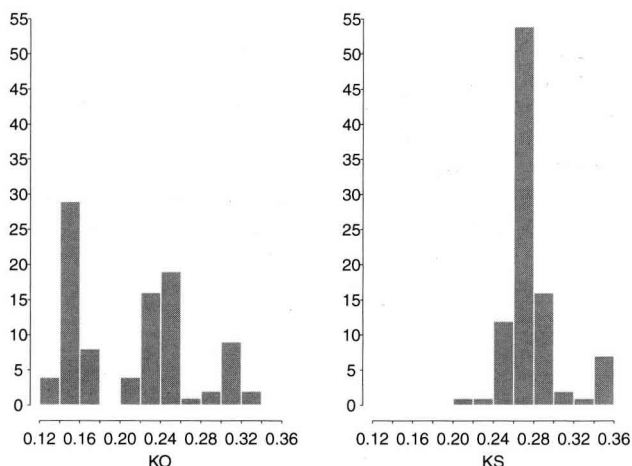


Figure 13. Histogramme des estimations pour la strate 7 sur l’inventaire en plein. KO = krigeage ordinaire/KS = krigeage stratifié.

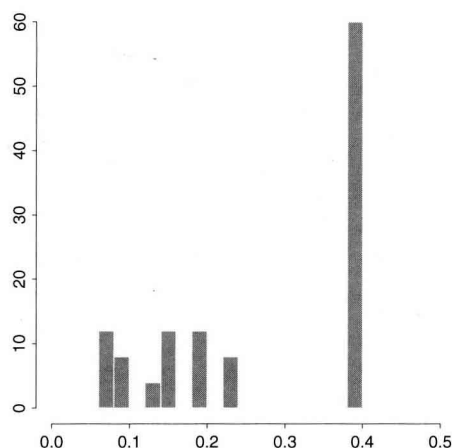


Figure 14. Histogramme des valeurs réelles pour la strate 7 sur l’inventaire en plein.

valeur nulle que le krigeage sur estime en pratique. Le biais moyen par strate reflète essentiellement la proportion de zéros dans chacune avec toutefois un sur lissage du KS plus important dans la strate 7 et moins fort dans la strate 1. Ces différences peuvent être attribuées à une proportion de parcelles sans Sapelli plus importante dans la strate 7 que dans la strate 1. Or, les poids λ_i calculés par KS pour les parcelles à interpoler sont plus élevés, si bien que par rapport au KO, l’effet de la sur estimation

des zéros est forcée dans la strate 7 mais pas dans la strate 1.

La qualité de l’ajustement est en revanche meilleure pour le KS que pour le KO pour les strates 1,3 et 7 : le variogramme étant le même pour les deux types de krigeage, ces différences peuvent s’expliquer par une plus forte contribution de ces strates à la construction du variogramme si bien que la variabilité des estimations par KS est naturellement moins forte.

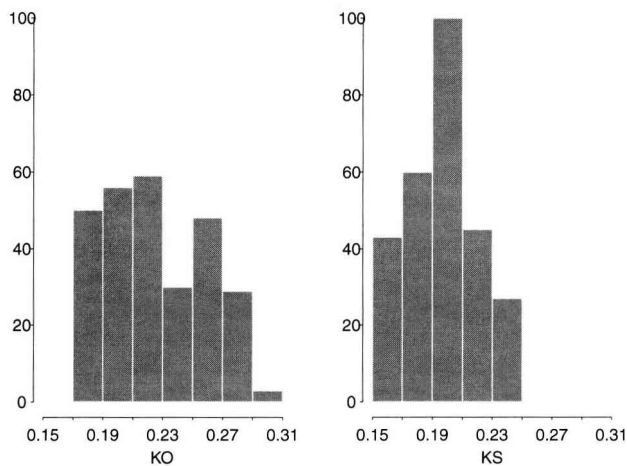


Figure 15. Histogramme des estimations pour la strate 10 sur l'inventaire en plein. KO = krigeage ordinaire/KS = krigeage stratifié.

Une première comparaison intéressante montre que le plan aléatoire simple est celui qui sous-estime le plus le total. Les résultats obtenus pour les prédictions par krigeage sont très légèrement supérieurs aux comptages réels par strates. Ils masquent en réalité un biais local très fort puisque sur toutes les placettes qui ne possédaient pas de Sapelli, la méthode de krigeage stratifié prédit des quantités dont la somme est égale à 1 553 m² (0,25 ha⁻¹). On retrouve évidemment cette caractéristique avec la méthode de sondage où les erreurs auraient une répartition uniforme. On retrouve de nouveau des différences de distribution similaires au bloc 1 des prédictions par strate entre les deux types de krigeage (figures 8 à 16).

Les histogrammes pour la strate 7 montrent une tendance inverse. Ceci peut s'expliquer par le morcellement de la strate 7 en cinq petits îlots : le krigeage stratifié reporte davantage le poids sur les observations les plus proches appartenant à la même strate si bien que les valeurs obtenues sont moins hétérogènes que dans le cas du krigeage ordinaire.

6. Conclusion

Dans le cadre de cette étude, les méthodes de krigeage fournissent des estimations globales moins biaisées que celles fournies par les plans de sondage. Néanmoins, la fiabilité de leur variance reste une source d'interrogation. Il semble aussi que la pertinence de la stratification soit mise en question, du moins pour évaluer la ressource de

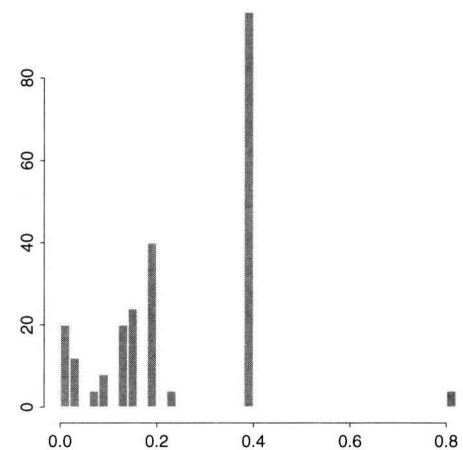


Figure 16. Histogramme des valeurs réelles pour la strate 10 sur l'inventaire en plein.

cette espèce : une faible corrélation avec la surface terrière expliquerait bien le peu d'améliorations constatées sur les prédictions lorsqu'elle utilisée pour les redresser. La détermination des strates est aussi sujette à caution puisque la précision des limites est estimée à 100 mètres sur le terrain. Le pourcentage de placettes mal classées augmente alors assez vite avec la surface de la strate (puisque le nombre de placettes en bordure de strate augmente) et par voie de conséquence la variance du total. Ceci doit expliquer également les faibles différences apportées par les diverses estimations par sondage ou par modèle prenant en compte la stratification. L'intérêt de l'utilisation d'un modèle dépend de l'objectif de l'utilisateur : il est évident que si celui-ci ne s'intéresse qu'à l'estimation du total sur une zone donnée a priori, les méthodes par sondage sont de loin préférables à la fois pour leur simplicité de mise en oeuvre et pour le temps de calcul. En revanche, si le parcellaire n'est pas déterminé a priori, la méthode d'interpolation par krigeage s'avère fructueuse dans les cas suivants :

- si les corrélations entre placettes sont assez fortes pour la variable d'intérêt ;
- si la variable auxiliaire est assez corrélée à la variable d'intérêt, le krigeage stratifié permet une meilleure estimation de la distribution des prédictions, sinon le krigeage ordinaire est suffisant.

Remerciements: Nous remercions G. Cornu pour son aide informatique, J.F Trébuchon et N. Fauvet pour leur travail de cartographie, A. Bar-Hen et D. Allard pour leurs conseils statistiques.

Références

[1] Christman M.C., Comparison of Some Sampling Designs for Spatially Clustered Populations, Springer Verlag, New York, V 122, 1997, pp. 97–101.

[2] Cochran W.G., Sampling Technics, Wiley Series in Probability and Mathematical Statistics-Applied, John Wiley & Sons, New York, 1977, XVI, 428 p.

[3] Cressie N., Statistics for Spatial Data, John Wiley & Sons, Inc. New York, 1993.

[4] De Vries P.G., Sampling Theory for Forest Inventory, Springer-Verlag, 1986.

[5] F.A.O., Manual of forest inventory, with special reference to tropical forests, Technical report, Rome, 1973.

[6] Goovaerts P., Geostatistics for Natural Resources Evaluation, Oxford University Press, 1997.

[7] Journel A.G., Huijbregts C.J., Mining Geostatistics, Academic Press, 1978.

[8] Mandallaz D., Geostatistical methods for double sampling schemes : Application to combined forest inventories,

Technical report, Swiss Federal Institute of Technology, Zurich, 1993.

[9] Navarro-Sanchez I., Monestiez P., Allard D., Courault D., Kriging with categorical external drift: use of thematic maps in spatial prediction and application to local climate interpolation for agriculture, in: Proceedings of the Second European Conference on Geostatistics for Environmental Applications, Valencia, Spain, 28–20 nov. 1998.

[10] Pain-Orcet M., Stratification de la forêt de Sangha-mbaéré, Technical report, Cirad-forêt, 1995.

[11] Petrucci B., Demarquez Y., Plan d'aménagement forestier du pea n163 (Sangha-mbaéré), Manual of forest inventory, Technical report, CIRAD-Forêt, 1997.

[12] Rossi M.E., Assessment of uncertainty using geostatistics, Environ Metrics 3, 71–79, 1992.

[13] Silverman B.W., Density Estimation, Chapman and Hall, New York, 1986.

[14] Valliant R., Poststratification and conditional variance estimation, J. Am. Stat. Assoc. 88 (1993) 89–96.

[15] Wackernagel H., Multivariate geostatistics: an introduction with applications, Springer, Berlin, XIV 1995, 256 p.