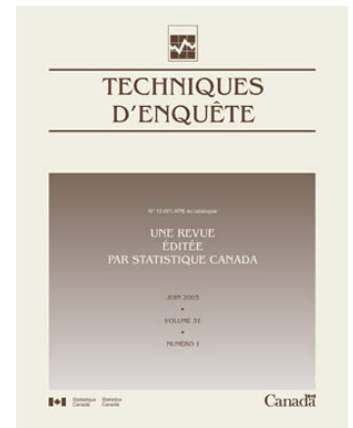


## Techniques d'enquête

# Échantillons maîtres comportant des panels optimisés

par Anton Grafström

Date de diffusion : le 29 juin 2026



Statistique  
Canada

Statistics  
Canada

Canada

---

## Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à [www.statcan.gc.ca](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.ca)

**Téléphone** entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros suivants :

- |                                                                             |                |
|-----------------------------------------------------------------------------|----------------|
| • Service de renseignements statistiques                                    | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur                                                               | 1-514-283-9350 |

## Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site [www.statcan.gc.ca](http://www.statcan.gc.ca) sous « Contactez-nous » > « [Normes de service à la clientèle](#) ».

## Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Publication autorisée par la ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par la ministre de l'Industrie, 2026

L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

*This publication is also available in English.*

---

# Échantillons maîtres comportant des panels optimisés

Anton Grafström<sup>1</sup>

## Résumé

Nous présentons un cadre général pour la construction d'échantillons maîtres qui préservent les propriétés de plan de sondage souhaitables pour tous les panels. La procédure de base consiste à ordonner un échantillon probabiliste initial. La séquence finale devant être robuste à un renouvellement aléatoire uniforme, nous définissons et minimisons une fonction objectif qui agrège le rendement au niveau des panels pour tous les panels circulaires possibles. Un renouvellement aléatoire final est appliqué pour assurer la validité du plan de sondage. Le cadre est souple quant au choix des critères de plan, comme l'équilibre spatial ou marginal, et peut être mis en œuvre efficacement à l'aide de la méthode du recuit simulé pour obtenir des solutions approximatives de grande qualité. Par sa conception, l'approche permet une coordination d'échantillons à la fois positive et négative pour des échantillons spatialement équilibrés, marginalement équilibrés et doublement équilibrés. La polyvalence de la méthode est démontrée par trois applications : la construction d'un échantillon maître comportant des panels spatialement équilibrés, des panels marginalement équilibrés et des panels doublement équilibrés.

**Mots-clés :** Coordination d'échantillons; échantillonnage; échantillonnage équilibré; équilibre spatial; recuit simulé.

## 1. Introduction

Un échantillon maître est un échantillon probabiliste unique de la population qui a été soigneusement sélectionné et qui peut être subdivisé, réutilisé ou coordonné pour répondre à divers objectifs d'enquête au fil du temps (Van Dam-Bates, Gansell et Robertson, 2018; Larsen, Olsen et Stevens, 2008). En tirant des sous-échantillons ou des panels d'un échantillon maître, les spécialistes des enquêtes peuvent réaliser des gains d'efficacité qui seraient difficiles ou impossibles à obtenir si chaque enquête était conçue de façon indépendante.

De nombreuses raisons pratiques motivent l'utilisation d'échantillons maîtres. Une application courante est la construction de plans de sondage avec renouvellement de panel, dans le cadre desquels les vagues d'enquête successives sont fondées sur des sous-ensembles de l'échantillon maître. Cela permet d'équilibrer les objectifs concurrents d'estimation de paramètres transversaux et de mesure de l'évolution au fil du temps. Un autre cas d'utilisation est la coordination d'échantillons générale, dans le cadre de laquelle les panels sont choisis de façon à ce qu'ils ne se chevauchent pas (coordination négative) ou qu'ils se chevauchent (coordination positive), afin de gérer le fardeau de réponse, la rentabilité ou la précision des estimations découlant d'enquêtes répétées. Les échantillons maîtres permettent également une grande flexibilité quant à la taille des échantillons : puisque l'échantillon sous-jacent est fixe, des sous-échantillons de tailles variables peuvent être tirés pour tenir compte de l'évolution des contraintes en matière de ressources ou des objectifs de l'enquête.

Ensemble, ces avantages font des échantillons maîtres une composante très polyvalente de la méthodologie moderne des enquêtes. Parallèlement, les outils statistiques et les outils de calcul pour construire des

---

1. Anton Grafström, Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Suède. Courriel : [anton.grafstrom@slu.se](mailto:anton.grafstrom@slu.se).

échantillons maîtres demeurent assez limités. Diverses méthodes peuvent être utilisées pour sélectionner un échantillon initial unique de grande qualité, comme la méthode du pivot local (Grafström, Lundström et Schelin, 2012), l'échantillonnage de Poisson spatialement corrélé (Grafström, 2012), l'échantillonnage de Poisson localement corrélé (Prentius, 2024), la méthode du cube (Deville et Tillé, 2004) et l'échantillonnage doublement équilibré (Grafström et Tillé, 2013). La mise en œuvre de bon nombre de ces méthodes fondamentales est possible dans le paquet R *BalancedSampling* (Grafström et Prentius, 2024). Des progrès récents ont été réalisés concernant les échantillons maîtres à probabilité égale (Robertson, Price et Reale, 2024), et il y a eu des extensions aux échantillons équilibrés et doublement équilibrés (Robertson, Price, Reale et Davies, 2026). De plus, des méthodes très efficaces ont été élaborées pour construire directement des échantillons spatio-temporels ayant des propriétés d'étalement spatial et temporel (Eustache, Jauslin et Tillé, 2022). Toutefois, il manquait un cadre général permettant de transformer un échantillon probabiliste unique de grande qualité en un échantillon maître dont l'ordonnancement est optimisé selon des critères de plan arbitraires. Le cadre proposé dans la présente étude vise à combler cette lacune.

Nous appliquons la méthode du recuit simulé pour optimiser l'ordonnancement de l'échantillon maître et ainsi améliorer la qualité de tous les panels qui en découlent. Fait important, notre cadre effectue cette optimisation tout en maintenant strictement les probabilités d'inclusion fondées sur le plan de sondage pour les sous-échantillons. Cela contraste avec d'autres applications de la méthode du recuit simulé en échantillonnage, comme l'optimisation de la répartition et de la stratification (O'Luing, Prestwich et Tarim, 2022) ou l'optimisation d'une configuration d'échantillon unique (par exemple Lark, 2016) d'une façon ne constituant pas un échantillonnage probabiliste.

Le reste de l'article est structuré comme suit. À la section 2, nous présentons formellement notre cadre général. La section 3 fournit une description détaillée de l'algorithme d'optimisation et des fonctions objectif modulaires utilisés pour obtenir différentes propriétés de plan. La section 4 présente trois exemples illustrant les possibilités offertes par le cadre. L'article se termine par une discussion à la section 5.

## 2. Création de l'échantillon maître

La procédure proposée pour créer un échantillon maître à haut rendement a plusieurs étapes. On commence par un échantillon initial de grande qualité, on détermine un bon ordonnancement pour cet échantillon et l'on termine par une étape de randomisation qui assure des probabilités valides pour les sous-échantillons.

### 2.1 Population et sélection de l'échantillon initial

Soit  $U = \{1, \dots, N\}$ , une population finie de taille  $N$ . L'objectif est de créer un échantillon maître de taille  $n < N$ .

La première étape consiste à sélectionner un échantillon initial fixe  $S_n \subset U$  de taille  $n$ . Cet échantillon doit être tiré à l'aide d'une méthode assurant une propriété de plan de sondage souhaitée. Par exemple, pour

créer un échantillon maître comportant des panels spatialement équilibrés,  $S_n$  doit être sélectionné à l'aide d'un plan d'échantillonnage spatialement équilibré.

Soit  $\pi_i = \Pr(i \in S_n)$ , les probabilités d'inclusion de premier ordre pour cet échantillon initial, pour toute les  $i \in U$ . Ces probabilités sont choisies par le concepteur de l'enquête, satisfont  $\sum_{i \in U} \pi_i = n$  et peuvent être proportionnelles à une variable auxiliaire. Une fois cet échantillon initial  $S_n$  sélectionné, il devient l'intrant pour l'étape d'ordonnement.

## 2.2 Ordonnement de l'échantillon initial

Même si l'ensemble  $S_n$  peut avoir la propriété souhaitée dans sa globalité, un ordre arbitraire ou aléatoire ne garantit pas que ses sous-séquences partageront également cette propriété. La deuxième étape de notre cadre consiste donc à trouver un ordonnancement de grande qualité des unités de  $S_n$ .

Soit  $S^\circ = (s_1^\circ, s_2^\circ, \dots, s_n^\circ)$ , une permutation des unités de  $S_n$ . L'objectif est de trouver un ordonnancement  $S^\circ$  proche d'être optimal selon certains critères quantifiables, habituellement une forme de robustesse de ses sous-échantillons. On peut y parvenir à l'aide de tout algorithme d'optimisation approprié. À la section 3, nous décrivons en détail une méthode pour trouver un tel ordonnancement à l'aide du recuit simulé. Pour l'instant, nous considérons  $S^\circ$  comme la séquence ordonnée finale résultant de cette recherche.

## 2.3 Renouvellement aléatoire et probabilités d'inclusion finales

La troisième et dernière étape est une étape de randomisation fournissant la structure de probabilités nécessaire pour l'échantillon maître. Nous appliquons un décalage aléatoire circulaire uniforme à la séquence optimisée  $S^\circ$ . Soit  $J$  une position de départ aléatoire tirée uniformément dans  $\{1, \dots, n\}$ . On définit alors la fonction d'indice circulaire

$$\iota(j) = ((j-1) \bmod n) + 1, \quad j \in \mathbb{Z}.$$

L'échantillon maître final, désigné par  $S^*$ , est ensuite défini comme le renouvellement circulaire

$$S^* = (s_{\iota(J)}^\circ, s_{\iota(J+1)}^\circ, \dots, s_{\iota(J+n-1)}^\circ).$$

Après, tout sous-échantillon, comme un panel, est obtenu en prenant un bloc contigu de longueur souhaitée  $k$  dans cette séquence finale  $S^*$ . Cette procédure mène au théorème général suivant concernant les probabilités d'inclusion de tout sous-échantillon de ce type.

**Théorème 1.** (probabilités d'inclusion d'un sous-échantillon). *Soit  $S'_k$  un sous-échantillon obtenu en prenant un bloc contigu de longueur  $k \leq n$  dans la séquence renouvelée de manière aléatoire  $S^*$ . Pour toute unité  $i \in U$ , la probabilité d'inclusion de premier ordre est donnée par :*

$$\pi_{i,k} = \Pr(i \in S'_k) = \frac{k}{n} \pi_i.$$

*Démonstration.* La démonstration repose sur la possibilité d'échange attribuable au renouvellement aléatoire. Soit  $I_i = \mathbf{1}(i \in S_n)$ , l'indicateur montrant que l'unité  $i$  fait partie de l'échantillon initial. Selon le plan de sondage pour l'échantillon initial,  $\Pr(I_i = 1) = \pi_i$ . Le renouvellement aléatoire uniforme permet de s'assurer que, conditionnellement à l'événement  $i \in S_n$ , l'unité  $i$  a une probabilité  $1/n$  de se trouver dans n'importe quelle des  $n$  positions dans la séquence finale  $S^*$ . La probabilité de se trouver dans une « fenêtre » contiguë donnée de taille  $k$  est donc exactement  $k/n$ . Selon la loi de probabilité totale, la probabilité d'inclusion inconditionnelle est le produit de ces deux probabilités :  $\pi_{i,k} = \Pr(i \in S'_k | i \in S_n) \times \Pr(i \in S_n) = (k/n) \pi_i$ .

Ce théorème est puissant parce qu'il est valable pour *tout* ordonnancement  $S^\circ$ . Il permet de bien dissocier le problème du respect des contraintes probabilistes de celui de l'obtention, par exemple, d'une grande qualité spatiale. Cela nous permet d'utiliser une recherche d'optimisation globale pour l'étape d'ordonnement, en étant assurés que le renouvellement aléatoire final rendra les sous-échantillons obtenus valides d'un point de vue probabiliste. La section suivante décrit notre méthode pour trouver un ordonnancement proche de ce qui est optimal.

### 3. Optimisation de l'ordonnement de l'échantillon maître

Une fois l'échantillon probabiliste initial  $S_n$  sélectionné, le seul degré de liberté restant dans la construction de l'échantillon maître est l'*ordre* de ses  $n$  unités. Cet ordonnancement détermine entièrement tous les sous-échantillons possibles pouvant être obtenus après le renouvellement circulaire aléatoire décrit à la section 2. Par conséquent, le problème de construction d'un échantillon maître de grande qualité revient à choisir un ordonnancement de  $S_n$  de sorte que tous les panels en découlant donnent de bons résultats pour un critère de plan choisi.

#### 3.1 Le problème d'ordonnement

Pour une taille de panel donnée  $k \leq n$ , l'ordonnement  $S^\circ$  donne  $n$  panels circulaires. À l'aide de la fonction d'indice circulaire  $\iota(\cdot)$  définie à la section 2, le panel de taille  $k$  commençant à la position  $j$  est

$$S'_{k,j} = (s_{\iota(j)}^\circ, s_{\iota(j+1)}^\circ, \dots, s_{\iota(j+k-1)}^\circ), \quad j = 1, \dots, n.$$

L'échantillon maître final étant obtenu en appliquant un renouvellement circulaire aléatoire uniforme, chacun de ces  $n$  panels a la même probabilité d'être sélectionné. L'ordonnement souhaitable est donc celui pour lequel *tous* les panels circulaires présentent de bonnes propriétés de plan. Cela motive une fonction objectif agrégeant le rendement au niveau des panels pour tous les renouvellements et pour toutes les tailles de panel d'intérêt.

#### 3.2 Fonction objectif

Soit  $Q(S'_{k,j})$  une fonction de qualité non négative qui mesure combien un panel donné  $S'_{k,j}$  satisfait un critère de plan souhaité. De faibles valeurs de  $Q$  correspondent à un meilleur rendement. Le choix de  $Q$

dépend du type de qualité que l'on souhaite avoir pour les panels obtenus. Par exemple, on peut viser l'équilibre spatial, l'équilibre marginal ou une combinaison des deux.

- *Équilibre spatial* : Un échantillon est considéré comme spatialement équilibré si ses unités sont bien étalées dans l'espace des variables auxiliaires. Il existe différentes façons de mesurer l'équilibre spatial, par exemple à l'aide des polytopes de Voronoi (Stevens et Olsen, 2004), d'une mesure fondée sur l'indice I de Moran (Tillé, Dickson, Espa et Giuliani, 2018) et d'une mesure d'équilibre local (Prentius et Grafström, 2024). Dans ce cas-ci, nous utilisons la mesure fondée sur les polytopes de Voronoi, qui quantifie la variation de la masse de probabilité d'inclusion à l'intérieur des polytopes générées par les unités de l'échantillon. Deux approches d'optimisation sont possibles. On peut soit chercher à rendre chaque sous-échantillon aussi bien étalé que possible au sein de la population complète (mesure inconditionnelle), soit traiter le problème comme étant conditionnel à l'échantillon maître réalisé, en visant un étalement au sein de celui-ci (mesure conditionnelle). Dans les deux cas, la forme générale de la mesure est :

$$Q_{\text{étalement}}(S'_{k,j}) = \frac{1}{k} \sum_{i \in S'_{k,j}} (v_i - 1)^2,$$

où  $v_i$  est la masse totale de probabilité dans le polytope de Voronoi  $V_i$  que l'unité  $i$  crée. Pour la mesure inconditionnelle,  $V_i$  contient toutes les unités de population plus proches de l'unité  $i$  (dans l'espace des variables auxiliaires) que de toute autre unité du panel  $S'_{k,j}$ , et  $v_i$  est calculée à l'aide des probabilités d'inclusion inconditionnelles de panel de cette façon :  $v_i = \sum_{u \in V_i} \pi_{u,k}$ . Pour la mesure conditionnelle efficace sur le plan des calculs qui est utilisée dans notre optimisation,  $V_i$  contient toutes les unités de l'échantillon maître  $S_n$  plus proches de  $i$  que de toute autre unité du panel  $S'_{k,j}$ , et  $v_i$  est calculée à partir des probabilités conditionnelles égales de cette façon :  $v_i = \sum_{u \in V_i} (k/n) = (k/n) |V_i|$ . Une valeur plus faible de  $Q_{\text{étalement}}$  indique un degré d'équilibre spatial plus élevé.

- *Équilibre marginal* : Un échantillon est (marginale) équilibré par rapport à un ensemble de variables auxiliaires si les estimations de Horvitz-Thompson des totaux de population concordent avec les totaux de population connus pour ces variables. Il n'est souvent possible d'atteindre qu'un équilibre approximatif, pour lequel la méthode du cube (Deville et Tillé, 2004) fournit une solution. La fonction de qualité  $Q_{\text{équilibre}}(S'_{k,j})$  mesure l'écart par rapport à l'équilibre parfait pour un ensemble de  $P$  variables auxiliaires dont les totaux de population sont connus et non nuls. Nous utilisons l'erreur quadratique moyenne relative :

$$Q_{\text{équilibre}}(S'_{k,j}) = \frac{1}{P} \sum_{p=1}^P \left( \frac{\sum_{i \in S'_{k,j}} \frac{z_{ip}}{\pi_{i,k}} - Z_p}{Z_p} \right)^2,$$

où  $Z_p = \sum_{i \in U} z_{ip}$  est le total de population connu pour la variable  $p$  et  $\pi_{i,k}$  est la probabilité d'inclusion d'un panel selon le théorème 1. Une valeur plus faible indique un meilleur équilibre.

- *Double équilibre* : Il est également possible de sélectionner des échantillons doublement équilibrés qui sont à la fois étalés (équilibrés spatialement) selon un ensemble de variables auxiliaires et équilibrés marginalement de façon approximative selon un ensemble potentiellement différent de variables auxiliaires. Grafström et Tillé (2013) ont proposé la méthode du cube local, qui combine la méthode du pivot local et la méthode du cube afin d'obtenir un double équilibre. Dans les problèmes de double équilibre, la qualité est une combinaison des deux fonctions de qualité individuelles, ce qui nécessite de tenir compte de leurs différentes échelles numériques. La forme générale de la qualité est une somme pondérée :

$$Q_{\text{double}}(S'_{k,j}) = wQ_{\text{équilibre}}(S'_{k,j}) + (1-w)Q_{\text{étalement}}(S'_{k,j}) \quad (3.1)$$

où le poids de préférence  $w \in [0,1]$  permet à un utilisateur d'accorder la priorité à un objectif plutôt qu'à un autre.

Soit  $K$ , l'ensemble des tailles de panel pertinentes pour l'application visée. Pour un ordonnancement  $S^\circ$  donné, nous définissons la fonction objectif globale comme suit :

$$O_d(S^\circ) = \frac{1}{|K|} \sum_{k \in K} \frac{1}{n} \sum_{j=1}^n Q_d(S'_{k,j}), \quad (3.2)$$

où  $d \in \{\text{étalement, équilibre, double}\}$  représente la qualité à optimiser. Pour une taille de panel fixe  $k$ , la moyenne interne représente la qualité attendue d'un panel tiré de l'ordonnancement  $S^\circ$  selon un renouvellement circulaire aléatoire uniforme. La moyenne externe attribue un poids égal à chaque taille de panel cible. Ainsi, on peut interpréter  $O_d(S^\circ)$  comme la qualité attendue au niveau des panels attribuable à l'ordonnancement  $S^\circ$ , dont la moyenne a été calculée sur l'ensemble des tailles de panel prévues et sur tous les renouvellements possibles.

Soit  $\Omega$ , l'ensemble des  $n!$  permutations de l'échantillon initial  $S_n$ . Nous cherchons à trouver la séquence optimale minimisant la fonction objectif (3.2), c'est-à-dire

$$S_{\text{opt}}^\circ \in \arg \min_{S^\circ \in \Omega} O_d(S^\circ).$$

La taille de  $\Omega$  étant  $n!$ , une optimisation exacte est impossible sur le plan des calculs pour des tailles d'échantillon maître réalistes. Par conséquent, le terme « optimisé » fait référence à une minimisation approximative de la fonction objectif (3.2) plutôt qu'à l'obtention d'un optimum global garanti.

**Remarque 1.** (choix du poids pour le double équilibre). *Le choix d'un poids approprié peut s'avérer difficile. Dans les démonstrations présentées dans la présente étude, nous visons un compromis où les deux contribuent également au coût total dans l'état initial. Pour ce faire, nous calculons un poids  $w_{\text{auto}}$  fondé sur des données, en fonction des scores de la séquence aléatoire initiale. Soit  $S^\circ$ , la séquence initiale, alors*

$$w_{\text{auto}} = \frac{O_{\text{étalement}}(S^\circ)}{O_{\text{équilibre}}(S^\circ) + O_{\text{étalement}}(S^\circ)}.$$

*Cette approche sans paramètre fournit une valeur par défaut pour l'optimisation à double équilibre. Nous soulignons toutefois que les spécialistes peuvent fixer  $w$  manuellement dans (3.1) pour affiner le résultat ou cibler délibérément des solutions privilégiant l'équilibre ou l'étalement.*

### 3.3 Méthode du recuit simulé

Pour trouver une séquence d'échantillons maîtres proche de ce qui est optimal pour tous les sous-échantillons prévus, nous utilisons la méthode du recuit simulé. Cette méthode est un puissant algorithme de recherche métaheuristique conçu pour trouver une solution quasi optimale dans un espace de solutions combinatoires vaste et complexe (Kirkpatrick, Gelatt et Vecchi, 1983). Elle est particulièrement bien adaptée aux problèmes d'ordonnancement et permet d'échapper efficacement à des optima locaux médiocres pour trouver une solution finale de grande qualité.

L'algorithme commence par une séquence candidate initiale et propose ensuite de manière itérative des modifications aléatoires structurées. Les mouvements qui réduisent le coût global de la séquence sont toujours acceptés, tandis que ceux qui l'augmentent sont acceptés selon une probabilité qui diminue au fil du temps, ce qui permet d'imiter le processus métallurgique de recuit. Notre mise en œuvre comprend trois composantes principales : l'espace des états et le mécanisme de génération des voisins, la fonction objectif définissant le coût d'une séquence, et le processus du recuit lui-même.

Dans notre système, un « état » correspond à un ordonnancement complet unique des  $n$  unités de l'échantillon initial  $S_n$ . Celui-ci est représenté par un vecteur de permutation  $S^\circ = (s_1^\circ, s_2^\circ, \dots, s_n^\circ)$ . L'espace des états  $\Omega$  est l'ensemble des  $n!$  permutations possibles.

Pour passer d'un ordonnancement actuel  $S_{\text{actuel}}^\circ$  à un ordonnancement voisin, nous utilisons un mécanisme d'échange stochastique qui cible les panels ayant un mauvais rendement tout en conservant la connectivité complète de l'espace des permutations.

À chaque itération, une taille de panel  $k$  est choisie de façon aléatoire selon une loi uniforme dans l'ensemble  $K$ . Avec une probabilité  $q \in (0,1)$ , le panel circulaire de taille  $k$  présentant le plus faible rendement est cerné, et une position  $j_1$  est choisie de façon aléatoire selon une loi uniforme parmi les indices appartenant à ce panel. Avec la probabilité restante  $1 - q$ , la position  $j_1$  est choisie de façon aléatoire selon une loi uniforme dans l'ensemble  $\{1, \dots, n\}$ . Une deuxième position  $j_2$  est ensuite sélectionnée aléatoirement selon une loi uniforme dans  $\{1, \dots, n\} \setminus \{j_1\}$ . Les deux unités situées aux positions  $j_1$  et  $j_2$  sont alors échangées pour produire l'ordonnancement  $S_{\text{candidat}}^\circ$ .

Grâce à ce mouvement ciblé (nous utilisons  $q = 0,9$ ), la plupart des échanges proposés constituent une tentative directe de corriger une faiblesse connue de la solution actuelle. Si l'on attribue une probabilité strictement positive à la sélection aléatoire selon une loi uniforme des deux positions d'échange, chaque transposition de deux positions peut se produire selon une probabilité positive, ce qui garantit que l'espace complet des permutations peut être exploré.

L'algorithme du recuit procède comme suit. À partir d'une séquence initiale aléatoire, il itère pendant un nombre fixe d'étapes. À chaque itération  $t$  :

1. Une nouvelle séquence candidate  $S_{\text{candidat}}^{\circ}$  est générée à partir de la séquence actuelle  $S_{\text{actuel}}^{\circ}$  à l'aide de l'échange ciblé décrit précédemment.
2. La variation de coût,  $\Delta O_d = O_d(S_{\text{candidat}}^{\circ}) - O_d(S_{\text{actuel}}^{\circ})$ , est calculée.
3. La nouvelle séquence est acceptée en tant que séquence actuelle suivante selon le critère de Metropolis :

$$\text{Accepter } S_{\text{candidat}}^{\circ} \text{ si } \Delta O_d < 0 \text{ ou si } \text{Uniform}(0,1) < e^{-\Delta O_d / T_t},$$

où  $T_t$  est la température à l'itération  $t$ .

La température  $T_t$  est progressivement réduite selon un schéma de refroidissement géométrique  $T_{t+1} = \alpha T_t$ , pour lequel le taux de refroidissement  $\alpha$  est une constante très proche de 1 (par exemple 0,999). Lorsque la température est élevée, l'algorithme accepte facilement les solutions moins bonnes, ce qui lui permet d'explorer dans son ensemble l'espace des états. À mesure que la température se refroidit, la probabilité d'accepter des solutions moins bonnes diminue, ce qui force l'algorithme à converger vers un optimum local de grande qualité. Pour améliorer davantage la recherche, nous utilisons un mécanisme de réchauffage : si aucune amélioration de la meilleure solution trouvée n'est observée pendant un grand nombre d'itérations, la température est réinitialisée à une valeur plus élevée afin de redynamiser la recherche à partir de sa position actuelle. L'algorithme arrête après un nombre fixe d'itérations, et la meilleure séquence observée est renvoyée comme ordonnancement optimisé final.

## 4. Démonstration : optimisation selon différents objectifs

Pour démontrer les capacités de notre cadre, nous l'appliquons à trois objectifs de plan : la création d'un échantillon maître optimisé pour l'équilibre spatial, d'un échantillon optimisé pour l'équilibre marginal des variables auxiliaires et d'un échantillon optimisé pour un double équilibre.

### 4.1 Exemple 1 : optimisation pour l'équilibre spatial

Cet exemple reproduit la sélection d'un échantillon temporaire sur cinq ans dans une strate (région 3) de l'Inventaire forestier national suédois (IFN); voir Grafström, Zhao, Nylander et Petersson (2017). Pour permettre l'utilisation de l'information auxiliaire dans le contexte d'une population continue de l'IFN, l'échantillonnage est réalisé en deux étapes. D'abord, un grand échantillon de taille  $N = 100\,000$  est sélectionné au moyen d'une intensité d'échantillonnage uniforme au sein de la strate. Pour les unités d'échantillonnage sélectionnées (chaque unité est un ensemble de parcelles circulaires), les variables auxiliaires sont extraites. Ce grand échantillon sert de base de sondage pour l'échantillon réel sur cinq ans de taille  $n = 360$ , qui est sélectionné au moyen de la méthode du pivot local pour assurer l'équilibre spatial.

L'échantillon sur cinq ans doit ensuite être divisé en cinq échantillons annuels sans chevauchement, chacun d'une taille de  $k = 72$ , et pour chaque année du cycle de cinq ans, les données d'un échantillon sont recueillies. Pour améliorer les estimations annuelles, il est important que chacun des cinq échantillons

annuels soit spatialement équilibré. Bien que, dans le cadre de l'IFN, on utilise habituellement plusieurs variables auxiliaires, nous ne simulons que les coordonnées spatiales dans la présente étude.

Ainsi, nous avons généré une population de taille  $N = 100\,000$  dont les coordonnées spatiales distribuées uniformément sont les variables auxiliaires. Après, l'échantillon maître de taille  $n = 360$  a été sélectionné à l'aide de la méthode du pivot local. L'algorithme du recuit simulé a été exécuté pendant 100 000 itérations, et la température initiale était de  $T_0 = 0,03$ , ce qui correspond à environ 10 % du coût initial. Cela se traduit par une probabilité initiale d'environ 0,37 d'accepter un déplacement augmentant la fonction objectif de 10 %. Le taux de refroidissement a été fixé à  $\alpha = 0,999$  et le réchauffage réinitialisait la température à la moitié de la valeur initiale après 10 000 itérations consécutives sans amélioration.

L'ordonnancement initial était une permutation aléatoire simple de l'échantillon maître. Après l'exécution du recuit simulé, la séquence optimisée a atteint une mesure d'équilibre spatial dans le pire des cas de 0,082 pour des panels de taille  $k = 72$ , évaluée en fonction de l'information auxiliaire de l'ensemble de la population. Étant donné que l'échantillon maître contient 360 panels consécutifs possibles de taille 72, l'optimisation assure que chacun de ces panels est très équilibré sur le plan spatial. Ce résultat valide également l'utilisation de la mesure conditionnelle d'équilibre spatial pour l'optimisation dans de grandes populations et de grands échantillons maîtres. Le rendement en matière d'équilibre spatial des panels selon des ordonnancements optimisés et aléatoires est résumé au tableau 4.1.

Il convient de prendre note que le panel ayant le meilleur rendement parmi tous les panels de l'ordonnancement aléatoire présente une mesure d'équilibre spatial de 0,137, ce qui est moins bon que celle du panel ayant le moins bon rendement de l'ordonnancement optimisé. Ainsi, une position de départ aléatoire selon une loi peut être choisie sans risque dans la séquence optimisée, et les cinq panels annuels consécutifs sans chevauchement de taille  $k = 72$  peuvent être extraits. Selon le tableau 4.1, chacun de ces panels présente nécessairement un degré élevé d'équilibre spatial. Le temps d'exécution total pour cet exemple, y compris l'optimisation et l'évaluation *a posteriori* de l'équilibre spatial par rapport à la population, était inférieur à sept minutes.

**Tableau 4.1**

**Rendement en matière d'équilibre spatial pour l'ensemble des 360 panels consécutifs de taille  $k = 72$  provenant d'un échantillon maître de taille  $n = 360$  (les valeurs plus faibles indiquent un meilleur équilibre spatial)**

| Méthode                   | $\min(Q_{\text{étalement}})$ | $\text{moyenne}(Q_{\text{étalement}})$ | $\max(Q_{\text{étalement}})$ |
|---------------------------|------------------------------|----------------------------------------|------------------------------|
| Originale (aléatoire)     | 0,137                        | 0,271                                  | 0,519                        |
| Optimisée (recuit simulé) | 0,036                        | 0,055                                  | 0,082                        |

## 4.2 Exemple 2 : optimisation pour l'équilibre marginal

Dans certaines enquêtes, l'objectif principal n'est pas l'étalement spatial, mais plutôt de veiller à ce que les sous-échantillons soient bien équilibrés par rapport à un ensemble de variables auxiliaires connues. Notre cadre peut être appliqué directement pour optimiser un échantillon maître en fonction de cet objectif.

Nous avons créé une population de taille  $N = 100\,000$  comportant deux variables auxiliaires,  $z_1 \sim N(10, 2^2)$  et  $z_2 \sim N(100, 25^2)$ . L'objectif était de constituer un échantillon maître de taille  $n = 1\,000$  ayant des probabilités d'inclusion égales. Les probabilités d'inclusion elles-mêmes ont également été incluses en tant que variable auxiliaire pour assurer le respect de la propriété de taille d'échantillon fixe.

D'abord, un échantillon initial de grande qualité a été sélectionné à l'aide de la méthode du cube pour s'assurer qu'il était bien équilibré par rapport aux totaux de population des variables auxiliaires. La qualité de cet échantillon initial était excellente; l'erreur quadratique moyenne relative n'était que de  $4,45 \times 10^{-8}$ .

L'étape critique consiste à ordonner ces  $n = 1\,000$  unités sélectionnées. Pour établir une base de référence, nous avons considéré une permutation aléatoire simple des unités, ce qui assure des probabilités valides, mais ne permet pas d'effectuer une optimisation pour l'équilibre dans les sous-échantillons. Pour trouver une séquence solidement équilibrée parmi les  $1\,000!$  ordonnancements possibles, nous avons utilisé l'algorithme du recuit simulé. La fonction objectif était (3.2), qui vise à minimiser la moyenne du déséquilibre pour un ensemble de tailles critiques  $K = \{100, 200\}$ . Nous avons exécuté l'algorithme du recuit simulé pour  $100\,000$  itérations. La température initiale était de  $T_0 = 0,00003$  (environ 10 % du coût initial). Le taux de refroidissement était de  $\alpha = 0,999$ , et un réchauffage était prévu après  $10\,000$  itérations sans amélioration.

Les résultats, présentés au tableau 4.2, démontrent une amélioration substantielle grâce à l'optimisation de l'ordonnancement. Pour les panels de taille  $k = 200$ , l'erreur quadratique moyenne relative dans le pire des cas, mesurée par  $\max(Q_{\text{équilibre}})$ , est réduite d'environ 99,2 % par rapport à un ordonnancement aléatoire. Pour les panels de taille  $k = 100$ , la réduction correspondante est d'environ 98,7 %. Ces résultats illustrent la capacité du cadre d'optimisation proposé à déterminer un ordonnancement préservant de manière robuste la propriété d'équilibre pour l'ensemble des panels et pour plusieurs tailles de panel. Le temps d'exécution total de cet exemple, y compris l'optimisation et l'évaluation *a posteriori* de l'équilibre, était inférieur à deux minutes.

**Tableau 4.2**

**Rendement en matière d'équilibre marginal pour l'ensemble des panels provenant d'un échantillon maître de taille  $n = 1\,000$ , mesuré selon l'indice d'équilibre  $Q_{\text{équilibre}}$  (les valeurs plus faibles indiquent un meilleur équilibre)**

| Taille de panel ( $k$ ) | Méthode                   | moyenne( $Q_{\text{équilibre}}$ ) | max( $Q_{\text{équilibre}}$ ) |
|-------------------------|---------------------------|-----------------------------------|-------------------------------|
| 100                     | Originale (aléatoire)     | $3,73 \times 10^{-4}$             | $2,39 \times 10^{-3}$         |
| 100                     | Optimisée (recuit simulé) | $6,63 \times 10^{-6}$             | $3,07 \times 10^{-5}$         |
| 200                     | Originale (aléatoire)     | $1,90 \times 10^{-4}$             | $1,10 \times 10^{-3}$         |
| 200                     | Optimisée (recuit simulé) | $2,03 \times 10^{-6}$             | $9,21 \times 10^{-6}$         |

### 4.3 Exemple 3 : optimisation pour un double équilibre

Dans ce cas-ci, nous générons une population spatiale de  $N = 100\,000$  unités en plaçant des points de façon indépendante et uniforme sur le carré-unité. Deux variables auxiliaires ont été générées, soit  $z_1 \sim N(10, 2^2)$  et  $z_2 \sim N(100, 25^2)$ . Nous avons ensuite sélectionné un échantillon maître doublement

équilibré de taille  $n = 500$  à l'aide de la méthode du cube local avec des probabilités d'inclusion égales. Cet échantillon est bien étalé selon les coordonnées dans le carré-unité, équilibré selon les probabilités d'inclusion et approximativement équilibré selon les deux variables auxiliaires. L'équilibrage en fonction des probabilités d'inclusion assure une taille d'échantillon fixe. Après, nous avons utilisé l'algorithme du recuit simulé pour optimiser cet échantillon maître pour des panels de tailles  $K = \{50, 100\}$  et, comme point de référence, nous faisons une comparaison avec la permutation aléatoire simple initiale. Nous avons utilisé 100 000 itérations, une température initiale de  $T_0 = 0,0001$  (10 % du coût initial) et un taux de refroidissement de  $\alpha = 0,999$ , et un réchauffage a été effectué après 10 000 itérations sans amélioration.

Les résultats présentés au tableau 4.3 indiquent une amélioration substantielle pour les deux critères. Pour les panels de taille 100, l'erreur quadratique moyenne relative dans le pire des cas est réduite d'environ 96 %, tandis que l'étalement dans le pire des cas est réduit d'environ 69 %. Pour les panels de taille 50, l'erreur quadratique moyenne relative dans le pire des cas est également réduite d'environ 96 % et l'étalement dans le pire des cas est réduit d'environ 77 %. Le temps d'exécution total pour cet exemple, y compris l'optimisation et l'évaluation *a posteriori* de l'équilibre et de l'équilibre spatial, était de 35 minutes.

**Tableau 4.3**  
**Rendement en matière d'équilibre et d'étalement dans le pire des cas pour l'ensemble des panels (une valeur plus faible est préférable)**

| Taille de panel ( $k$ ) | Méthode                   | $\max(Q_{\text{équilibre}})$ | $\max(Q_{\text{étalement}})$ |
|-------------------------|---------------------------|------------------------------|------------------------------|
| 50                      | Originale (aléatoire)     | $2,97 \times 10^{-3}$        | 0,592                        |
| 50                      | Optimisée (recuit simulé) | $1,15 \times 10^{-4}$        | 0,139                        |
| 100                     | Originale (aléatoire)     | $1,02 \times 10^{-3}$        | 0,392                        |
| 100                     | Optimisée (recuit simulé) | $4,11 \times 10^{-5}$        | 0,122                        |

## 5. Discussion

Nous avons présenté un cadre général permettant de produire des échantillons maîtres. Ce cadre permet la création d'échantillons maîtres tirés sans remise avec probabilité proportionnelle à la taille, dans lesquels des sous-échantillons de toute taille prévue possèdent à la fois des probabilités valides et un degré élevé d'une propriété de plan souhaitée, comme l'équilibre spatial, l'équilibre marginal ou les deux. La possibilité d'adapter la fonction objectif rend la méthode très générale. Par exemple, pour créer un système très efficace de panels se chevauchant, nous incluons à la fois la taille des panels et la taille du chevauchement dans la fonction objectif. Cela assure, par exemple, que les panels et leur chevauchement sont bien équilibrés. Cette approche favorise ainsi à la fois la coordination d'échantillons positive et négative.

La principale limite réside dans le coût de calcul, qui dépend de la taille de l'échantillon maître, de la complexité de la fonction objectif et de l'efficacité de l'algorithme de recherche pour l'ordre optimal. Toutefois, la construction d'un échantillon maître est généralement une opération ponctuelle pour laquelle

le coût de calcul de l'optimisation représente probablement une petite fraction du coût total de l'enquête. Les exemples présentés dans le présent article ont nécessité entre 1 et 35 minutes de calcul. Pour les applications exigeant le plus haut degré de robustesse et de qualité, ce faible investissement initial permet d'obtenir un résultat supérieur et plus fiable.

Le cadre proposé se prête particulièrement bien aux programmes de surveillance de l'environnement par strates, comme l'IFN suédois ou l'Enquête aréolaire sur l'utilisation et l'occupation des sols de l'Union européenne (Ballin, Barcaroli, Palmieri et Zachariadis, 2025). Notre méthode peut être appliquée au sein des strates pour optimiser les plans de renouvellement pour un cycle d'enquête particulier. Toutefois, les données auxiliaires et la stratification évoluant souvent au fil du temps, notre approche doit être considérée davantage comme un outil permettant d'optimiser un horizon de planification fixe, plutôt que comme un mécanisme de gestion d'échantillon d'une durée indéterminée.

Enfin, le cadre ouvre la voie à plusieurs possibilités d'amélioration algorithmique future. L'échange de positions ciblé mis en œuvre dans le présent travail est une stratégie de recherche générale efficace, car on concentre correctement la recherche sur les régions de la séquence qui ont un mauvais rendement. Bien que ce type d'échange soit préférable à un échange aveugle, son caractère aléatoire au sein de la région ciblée signifie qu'il peut tout de même avoir une faible probabilité de proposer une amélioration. Par conséquent, le nombre d'itérations nécessaires pour trouver une séquence réellement bonne peut être élevé.

La prochaine étape importante serait d'élaborer des propositions de mouvements propres à l'objectif et guidées par des diagnostics et par la cause du mauvais score. Dans le cas de l'équilibre spatial, le score élevé d'un panel indique une grande variance de ses masses de polytopes, ce qui reflète à la fois des régions trop denses et des régions trop clairsemées dans l'espace des variables auxiliaires. Un mouvement véritablement intelligent serait donc un échange conçu pour corriger directement cette faiblesse : retirer une unité d'une région dense et la remplacer par une unité extérieure pouvant mieux couvrir une région clairsemée. De même, pour l'équilibre marginal, une approche diagnostique permettrait de déterminer les unités à retirer et le type d'unités à ajouter pour améliorer l'équilibre.

Ces approches heuristiques fondées sur des diagnostics, qui proposent des mouvements ayant une forte probabilité d'être bénéfiques, représentent une classe de stratégies de recherche plus sophistiquée et constituent un domaine de recherche prometteur.

Comme le cadre peut être utilisé pour sélectionner des échantillons maîtres dans lesquels les panels répondent à différents critères de plan de sondage, comme l'équilibre spatial, l'équilibre marginal ou tout autre critère, l'estimation de la variance à partir des panels obtenus n'a pas encore été étudiée. Il est probable que les estimateurs de variance approximatifs couramment utilisés pour des échantillons spatialement équilibrés fonctionneraient également bien pour des panels spatialement équilibrés issus d'un échantillon maître. De façon similaire, pour des panels marginalement équilibrés, un estimateur de variance approximatif pour l'échantillonnage équilibré devrait fonctionner adéquatement. Toutefois, il reste à déterminer de façon

précise, dans des travaux futurs, le rendement d'estimateurs de variance approximatifs pour les différents types de panels.

Pour faciliter l'application du cadre, le code R utilisé pour les démonstrations de la section 4 est accessible sur GitHub (Grafström, 2026). Compte tenu de la modularité de la fonction objectif, qui peut être adaptée à tout critère de plan de sondage quantifiable, ce code vise à servir de modèle adaptable. Cette approche offre aux utilisateurs la souplesse nécessaire pour mettre en œuvre et optimiser des solutions répondant à leurs besoins particuliers.

## Bibliographie

- Ballin, M., Barcaroli, G., Palmieri, A. et Zachariadis, S. (2025). Introduction of a panel component in the LUCAS survey and the use of a composed estimator for statistics production – 2025 edition. *Statistical Working Papers*, Publications Office of the European Union, Luxembourg.
- Deville, J.-C. et Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91(4), 893-912.
- Eustache, E., Jauslin, R. et Tillé, Y. (2022). Spatiotemporal sampling with spatial spreading and rotation of units in time. *Spatial Statistics*, 47, 100613.
- Grafström, A. (2012). Spatially correlated Poisson sampling. *Journal of Statistical Planning and Inference*, 142(1), 139-147.
- Grafström, A. (2026). Master Samples with Optimized Panels [Computer software]. GitHub. Accessible à l'adresse : <https://github.com/antongrafstrom/MasterSamples>.
- Grafström, A., Lundström, N.L. et Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics*, 68(2), 514-520.
- Grafström, A. et Prentius, W. (2024). *BalancedSampling: Balanced and Spatially Balanced Sampling*. R package version 2.1.1.
- Grafström, A. et Tillé, Y. (2013). Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. *Environmetrics*, 24(2), 120-131.
- Grafström, A., Zhao, X., Nylander, M. et Petersson, H. (2017). A new sampling strategy for forest inventories applied to the temporary clusters of the Swedish national forest inventory. *Canadian Journal of Forest Research*, 47(9), 1161-1167.

- Kirkpatrick, S., Gelatt Jr., C.D. et Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.
- Lark, R.M. (2016). Multi-objective optimization of spatial sampling. *Spatial Statistics*, 18, 412-430.
- Larsen, D.P., Olsen, A.R. et Stevens, D.L. (2008). Using a master sample to integrate stream monitoring programs. *Journal of Agricultural, Biological, and Environmental Statistics*, 13(3), 243-254.
- O’Luing, M., Prestwich, S. et Tarim, S.A. (2022). [Un algorithme du recuit simulé pour la stratification et la répartition simultanée de l’échantillon](https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2022001/article/00010-fra.pdf). *Techniques d’enquête*, 48(1), 245-271. Accessible à l’adresse : <https://www150.statcan.gc.ca/n1/fr/pub/12-001-x/2022001/article/00010-fra.pdf>.
- Prentius, W. (2024). Locally correlated Poisson sampling. *Environmetrics*, 35(2), e2832.
- Prentius, W. et Grafström, A. (2024). How to find the best sampling design: A new measure of spatial balance. *Environmetrics*, 35(7), e2878.
- Robertson, B., Price, C. et Reale, M. (2024). Well-spread samples with dynamic sample sizes. *Biometrics*, 80(2), ujae026.
- Robertson, B., Price, C., Reale, M. et Davies, P. (2026). Doubly balanced samples with dynamic sample sizes. *Biometrics*, 82(1), uja011.
- Stevens Jr., D.L. et Olsen, A.R. (2004). Spatially balanced sampling of natural resources. *Journal of the American statistical Association*, 99(465), 262-278.
- Tillé, Y., Dickson, M.M., Espa, G. et Giuliani, D. (2018). Measuring the spatial balance of a sample: A new measure based on Moran’s I index. *Spatial Statistics*, 23, 182-192.
- van Dam-Bates, P., Gansell, O. et Robertson, B. (2018). Using balanced acceptance sampling as a master sample for environmental surveys. *Methods in Ecology and Evolution*, 9(7), 1718-1726.