

Planifier un échantillonnage pour une étude régionale de la végétation

Ph. Daget¹ J. Poissonet¹ B. Toutain²

Mots-clés

Echantillonnage - Végétation - Méthode - Mesure - Burkina Faso.

Résumé

L'échantillonnage de qualité, nécessaire à une bonne étude des relations entre les espèces et les conditions de milieu, est obtenu par une stratification. Dans ce cadre, un échantillonnage est optimal au sens de Neyman (ou N-optimal) lorsque toutes les strates ont le même effectif. Pour l'obtenir, un algorithme ainsi que les relations et les seuils qui permettent d'évaluer la qualité de cet échantillonnage sont proposés. Un exemple de cas réel provenant du Burkina Faso est présenté.

■ INTRODUCTION

L'étude de la végétation et de ses relations avec les conditions de milieu est l'une des étapes préparatoires à des opérations d'aménagement régional et à la mise en place de dispositifs de suivi au sein de l'environnement. Les résultats sont souvent cartographiés car la représentation spatiale de ces connaissances constitue l'outil le plus commode de l'aménagiste. Ils peuvent ensuite être intégrés dans un système d'information géographique (SIG), ce qui permet, par le rapprochement informatique de données de diverses natures, de suggérer des comparaisons, des confrontations et d'envisager des simulations.

■ DE L'INVENTAIRE AU PLAN D'ECHANTILLONNAGE

Un inventaire régional s'appuie sur des relevés agrostologiques, botaniques ou écologiques. Ceux-ci sont constitués, d'une part, de la liste de toutes les plantes végétant à un emplacement déterminé et, d'autre part, de la notation des caractéristiques stationnelles détaillées (1, 6, 8).

Beaucoup d'études d'agrostologie ou d'écologie régionales visent à mettre en évidence des indicateurs biologiques de l'état du milieu ou, tout au moins, de certains de ses descripteurs. A partir de

la confrontation des éléments contenus dans ces relevés et au moyen de méthodes décrites par ailleurs (3, 7,) il est possible de dégager aussi bien une typologie de la végétation, base de la cartographie, que des indicateurs biologiques. Mais l'échantillonnage suppose un nombre suffisant de relevés (de l'ordre de la centaine) pour que les approches statistiques effectuées soient valides. Pour alléger le travail de terrain, qui est onéreux, et pour aider au choix des stations, car les déplacements ne sont pas toujours faciles, il faut planifier l'échantillonnage (4, 5), afin de ne pas faire trop de relevés et surtout de ne faire que ceux qui sont nécessaires et pas plus.

■ ECHANTILLONNAGE STRATIFIÉ

Au moment de préparer les missions de terrain pour collecter les données, l'agrostologue peut être tenté de répartir les emplacements de relevés par un tirage aléatoire ou aux nœuds d'un réseau à mailles carrées pour obtenir une répartition spatiale homogène. De tels choix conduisent à un échantillonnage dans lequel le nombre de relevés de chaque type de milieu est proportionnel à la surface qu'il occupe dans la région. Les milieux les plus fréquents sont suréchantillonnés, tandis que les végétations qui n'occupent que des espaces restreints sont sous-échantillonnées, voire pas du tout inventoriées.

L'agrostologue doit s'inspirer des méthodes des instituts de sondage et utiliser un échantillonnage stratifié (5). Cela consiste à découper l'univers à échantillonner en « sous-univers » appelés strates, relativement homogènes par rapport à un critère jugé *a priori* discriminant. Les échantillons sont ensuite répartis dans ces strates.

1. CIRAD-EMVT/CNRS, Campus international de Baillarguet, BP 5035, 34032 Montpellier Cedex 1, France

2. CIRAD-EMVT, Campus international de Baillarguet, BP 5035, 34032 Montpellier Cedex 1, France

Il est souvent fructueux de retenir deux, parfois trois, critères discriminants. Alors, chaque strate obtenue par le premier critère est à son tour subdivisée sur la base du second. Pour effectuer cette opération, il convient que chacun des paramètres utilisés pour caractériser ces critères soit discret et ait au moins 5 à 7 classes. Ainsi, si la pente est retenue comme paramètre de planification, ce n'est pas sa valeur mesurée au clinomètre (en degré, en grade ou en pourcentage) qui sera utilisée mais des classes de pente.

Par ailleurs, on appelle case létale, ou combinaison létale, une combinaison des deux descripteurs qui, par nature, ne peut pas exister sur le terrain. Ce serait par exemple la combinaison « sommet arrondi-marécage » ou « forêt galerie-cuirasse »...

Tableau optimal

La combinaison de plusieurs strates permet d'établir un tableau d'échantillonnage à double entrée pour deux strates, à triple entrée pour trois... Pour que l'échantillonnage soit le plus efficace et le moins coûteux possible (on dit « pour qu'il soit optimal », ou « optimal au sens de Neyman », ou encore « N-optimal »), il a été démontré que l'allocation des relevés dans les strates doit être uniforme, c'est-à-dire qu'il faut qu'il y ait autant de relevés dans chaque strate, donc, d'une part, sur chaque ligne et, d'autre part, sur chaque colonne, même si les effectifs des cases du tableau sont différents (5, 9).

Algorithme ARISTEA

On obtient aisément un tel tableau, quels que soient le nombre et la position des cases létales en utilisant l'algorithme ARISTEA (2 simplifié) qui donne la succession des manipulations à effectuer étape par étape.

Algorithme de résolution immédiate et systématique des tableaux d'échantillonnage carrés à cases létales (ARISTEA) :

1. Soit n le nombre de classes de chacun des deux descripteurs. Faire un tableau carré de $n.n = n^2$ cases.

2. Soit N le nombre total approximatif de relevés envisagé.

3. Y-a-t-il des cases létales ?

Si oui aller en 5 ; si non aller en 4.

4. L'effectif par case est l'entier le plus proche de $y = N/n^2$. Le tableau est N-optimal. FIN.

5. Où sont-elles placées ?

6. Soit k le nombre maximal de cases létales sur une ligne ou sur une colonne.

7. L'effectif à viser pour chaque ligne ou colonne est $m = N/n$.

8. Donnons provisoirement à chaque case non létale un effectif égal à l'entier le plus proche de $z = m/(n-k)$.

9. Faire les totaux marginaux du tableau T ; soit H le vecteur colonne des sommes faites suivant les lignes et V le vecteur ligne des sommes faites selon les colonnes.

10. Toutes les valeurs de H et toutes celles de V sont-elles égales ?

Si oui aller en 11 ; si non aller en 12.

11. Le tableau T est aussi près que possible d'un résultat N-optimal. FIN.

12. Soit m' le plus petit des totaux de H et de V .

13. Extraire de T le sous-tableau T' contenant les cases de T correspondant aux maximums de H et de V .

14. Les éléments de T' sont-ils tous égaux à 1 ?

Si oui aller à 11, si non aller à 15.

15. Les éléments de T' sont-ils tous nuls ?

Si oui aller à 11, si non aller à 16.

16. Retirer 1 à chaque case de T' ayant plus de 1 relevé, puis retourner en 9.

Exemple

Il s'agit de stratifier un échantillonnage à partir de deux descripteurs de cinq classes pour lesquelles on a noté que les combinaisons {1;1}, {1;5}, {2;3} et {3;3} étaient létales. On construit d'abord un tableau 5 x 5 (question 1), puis on retient (question 2) le nombre de 90 relevés ; les cases létales (question 3) sont repérées (question 5) (figure 1 : a). La première ligne en contient 2 donc $k = 2$ (question 6), par conséquent on place dans chaque case le nombre $6 = (90/5)/3 = 90/15$ (figure 1 : b). Les totaux marginaux sont $H = 18\ 24\ 24\ 30\ 30$ et $V = 24\ 30\ 18\ 30\ 24$. C'est donc aux deux dernières lignes des deuxième et quatrième colonnes qu'il faut retirer 1 (figure 1 : c) ; les totaux deviennent $H = 18\ 24\ 24\ 28\ 28$ et $V = 24\ 28\ 18\ 28\ 24$. L'opération est donc à renouveler dans les mêmes cases puis encore une fois ce qui donne $H = 18\ 24\ 24\ 24\ 24$ et $V = 24\ 24\ 18\ 24\ 24$ ce qui amène à retirer 1 des 4 dernières lignes des colonnes 1, 2, 4 et 5 deux fois de suite. Il vient alors (figure 1 : d) $H = 18\ 16\ 16\ 16\ 16$ et $V = 16\ 16\ 18\ 16\ 16$, ce qui conduit à retirer 1 dans la troisième case de la première ligne deux fois de suite pour aboutir au tableau N-optimal définitif (figure 1 : e) avec 80 relevés.

0 1 1 1 0	0 6 6 6 0	0 6 6 6 0	0 6 6 6 0	0 6 4 6 0
1 1 0 1 1	6 6 0 6 6	6 6 0 6 6	4 4 0 4 4	4 4 0 4 4
1 1 0 1 1	6 6 0 6 6	6 6 0 6 6	4 4 0 4 4	4 4 0 4 4
1 1 1 1 1	6 6 6 6 6	6 5 6 5 6	4 1 6 1 4	4 1 6 1 4
1 1 1 1 1	6 6 6 6 6	6 5 6 5 6	4 1 6 1 4	4 1 6 1 4
a	b	c	d	e

Figure 1 : différentes étapes de la construction d'un tableau d'échantillonnage N-optimal.

■ MESURE DE LA QUALITE D'UN ECHANTILLONNAGE

Modalité de la mesure

On appelle profil d'ensemble d'un descripteur la suite des fréquences absolues des relevés dans les classes de ce descripteur (3) ; le tableau I en donne la forme théorique.

Il est clair que si un descripteur n'a qu'une classe, un relevé au hasard parmi les NR relevés de l'étude n'apporte aucune information particulière vis-à-vis de ce descripteur. Si le descripteur a plusieurs classes, l'étude n'apporte pas d'information sur ce descripteur. En fait, la quantité d'information apportée par un profil d'ensemble est mesurée par son entropie à partir des fréquences relatives des relevés dans les classes :

$$H(L) = -\sum_1^{NR} \frac{R(K)}{NR} \log_2 \frac{R(K)}{NR}$$

Tableau I

Forme théorique du profil d'ensemble

	Descripteur L ayant NK classes						Ensemble des relevés
	Classe 1	Classe 2	...	Classe K	...	Classe NK	
Nombre de relevés	R(1)	R(2)	...	R(K)	...	R(NK)	$NR = \sum_1^{NK} R(K)$

D'après Daget Ph., Godron M., Masson, Paris, France, 1982

Ce qui s'écrit aussi :

$$H(L) = - \log_2 NR - \frac{1}{NR} \sum_1^{NR} R(k) \log_2 R(k)$$

On démontre que H(L) passe par un maximum lorsque toutes les fréquences relatives R(k) sont égales :

$$R(k) = \frac{NR}{NK}$$

On a alors :

$$H(L; \max) = - \log_2 NK$$

L'entropie maximale en binons est égale au logarithme à base deux du nombre de classes.

La qualité de l'échantillonnage se mesure par le rapport entre l'entropie liée à la répartition des relevés suivant le descripteur H(L) et son entropie maximale H(L ; max).

Exemple

Un inventaire effectué il y a quelques années dans une région du Burkina Faso a permis de mettre en évidence que les descripteurs discriminants étaient le type de formation végétale et la situation géomorphologique de la station :

- | | |
|---------------------|-----------------------------|
| 1. Prairie humide | A. Montagne |
| 2. Cordon ripicole | B. Cuirasse |
| 3. Forêt galerie | C. Plaine |
| 4. Forêt | D. Haut de versant |
| 5. Savane arborée | E. Mi-versant |
| 6. Savane arbustive | F. Bas de versant |
| 7. Jachère arborée | G. Bas-fonds et dépressions |
| 8. Culture | H. Talwegs et mares |

Les relevés effectués se répartissent de la manière indiquée dans le tableau II.

Pour un total NR = 115, le profil d'ensemble pour la situation géomorphologique qui rassemble le nombre de relevés pour chacune des huit classes est :

I	II	III	IV	V	VI	VII	VIII
3	4	2	15	48	22	9	12

Tableau II

Echantillonnage effectué dans l'étude au Burkina Faso

	1	2	3	4	5	6	7	8	Σ
A	X	X	X	X	3	0	0	0	3
B	X	X	X	X	2	2	0	0	4
C	X	X	X	X	1	0	1	0	2
D	X	X	X	X	14	1	0	0	15
E	X	X	X	1	44	1	2	0	48
F	0	0	0	1	20	0	1	0	22
G	2	0	0	0	3	0	3	1	9
H	4	4	4	0	2	0	0	0	12
Σ	6	4	4	2	87	4	7	1	115

En divisant toutes ces valeurs par NR = 115, on obtient le profil des fréquences relatives suivant :

0,026 0,035 0,017 0,130 0,417 0,191 0,078 0,104

dont l'entropie est :

$$H = - (0,026 \log_2 0,026 + \dots + 0,104 \log_2 0,104) = 2,401$$

Avec huit situations géomorphologiques (géomor), l'entropie maximale est :

$$H_{\max} = \log_2 8 = 3$$

Par conséquent, la qualité de l'échantillonnage effectué est :

$$Q(\text{géomor}) = \frac{2,401}{3} = 0,800$$

On calculerait de la même manière la qualité de l'échantillonnage des formations végétales (formveg) dans cette étude :

$$Q(\text{formveg}) = \frac{1,439}{3} = 0,480$$

L'analyse de nombreux échantillonnages dans des régions très variées conduit à proposer d'estimer la valeur d'un échantillonnage par référence à l'échelle empirique suivante :

- ≤ 0,50 = très mauvais échantillonnage
- 0,51 à 0,84 = mauvais échantillonnage
- 0,85 à 0,90 = échantillonnage médiocre
- 0,91 à 0,95 = bon échantillonnage
- ≥ 0,96 = très bon échantillonnage

Il apparaît donc que l'échantillonnage des situations géomorphologiques est médiocre tandis que celui des formations végétales est très mauvais, cela étant dû au fait que les trois-quarts des relevés sont dans la même classe (la classe 5 des savanes arborées) qui se trouve, de ce fait, extrêmement suréchantillonnée. Bien entendu, il s'agit de la formation qui couvre la plus grande surface ce qui explique qu'avec un échantillonnage non stratifié elle est la plus représentée. Si elle devait être considérée comme hétérogène, alors il aurait fallu considérer plusieurs types de savane arborée à échantillonner comme tels.

Généralisation

Il est aisé de généraliser aux cas où les deux descripteurs n'ont pas le même nombre de classes, donc aux cas des tableaux d'échantillonnage rectangulaires. Il suffit de remplacer n^2 par $N1.N2$ et $n(n-k)$ par $(N1.N2)-L$, avec L le nombre de cases létales. Le tableau obtenu est très proche du tableau N -optimal cherché qu'il est possible d'atteindre très rapidement par une ou deux inversions.

Ainsi pour le tableau 8×5 avec 4 cases létales en $\{1; 1\}$, $\{1; 2\}$, $\{2; 2\}$, $\{8; 5\}$, on obtient rapidement le tableau de la figure 2 : a dont les totaux marginaux sont $H = 7\ 6\ 6\ 6\ 6\ 6\ 6\ 6$ et $V = 9\ 12\ 10\ 10\ 8$, pour un total général de 49. La qualité de l'échantillonnage a respectivement pour valeur $Q(5) = 0,935$ et $Q(8) = 0,994$; on voit rapidement qu'en intervertissant sur les lignes 3 et 4 les deuxième et cinquième valeurs, on atteint le tableau N -optimal (figure 2 : b) avec $Q(5) = 0,999$, V devenant $9\ 10\ 10\ 10\ 10$ et le total général restant de 49.

0 0 3 3 1	0 0 3 3 1
2 0 1 1 2	2 0 1 1 2
1 2 1 1 1	1 1 1 1 2
1 2 1 1 1	1 1 1 1 2
1 2 1 1 1	1 2 1 1 1
1 2 1 1 1	1 2 1 1 1
1 2 1 1 1	1 2 1 1 1
2 2 1 1 0	2 2 1 1 0
a	b

Figure 2 : tableaux 8×5 à 4 cases létales proches de l'optimum.

Summary

Daget Ph., Poissonet J., Toutain B. Sampling design to study vegetation at the regional level

Stratified sampling provides quality needed in the study of relations between species and environmental conditions. A sampling is then optimal in Neyman sense (or N -optimal) when all strata have the same number of *relevés*. In order to reach it an algorithm as well as the relations and thresholds that help evaluate the quality of the sampling are proposed. A real case example from Burkina Faso is discussed.

Key words: Sampling - Vegetation - Method - Measurement - Burkina Faso.

CONCLUSION

Bien des difficultés se sont manifestées pour l'interprétation de l'échantillonnage ci-dessus ; les calculs qui viennent d'être présentés montrent que ces problèmes viennent d'un échantillonnage très déséquilibré. En fait, il a été réalisé sans planification préliminaire. Ceci explique cela et souligne la nécessité d'une bonne planification préalable. L'algorithme ARISTEA, facile à programmer, peut aider à la préparation de ce préalable, mais il n'est pas suffisant. C'est un exemple de raisonnement qu'il faut adapter aux cas réels (tableaux rectangulaires, tableaux à triple ou quadruple entrées).

BIBLIOGRAPHIE

1. BOUDET G., 1991. Manuel sur les pâturages tropicaux et les cultures fourragères. Paris, France, La Documentation française, 266 p. (Manuels et précis d'élevage n° 4)
2. DAGET Ph., 1989. De la réalisation des plans d'échantillonnage en phytoécologie générale - Quelques algorithmes d'allocation. *Biocénoses (Alger)*, 4 : 96-115.
3. DAGET Ph., GODRON M., 1982. Analyse écologique des espèces dans les communautés. Paris, France, Masson, 172 p.
4. FRONTIER S., 1982. Echantillonnage écologique. Paris, France, Masson, 172 p.
5. GODRON M., 1971. Essais sur une approche probabiliste des végétaux. Thèse doct. Univ. Sci. Tech. Languedoc, Montpellier, France, 247 p.
6. GODRON M. ed., 1968. Code pour le relevé méthodique de la végétation et des milieux. Paris, France, CNRS, 272 p.
7. GUINOCHET M., 1973. Phytosociologie. Paris, France, Masson, 228 p.
8. MUELLERS-DOMBOIS D., ELLENBERG H., 1979. Aims and methods in vegetation science. New York, USA, Wiley, 547 p.
9. NEYMAN N., 1934. On two different aspects of the representative method : the method of stratified sampling and the method of purposive selection. *J. R. Stat. Soc.*, 97: 558-606.
10. TOUTAIN B., DUMAS R., TACHER G., 1978. Zone pastorale d'accueil de Sidéradouougou (Haute Volta). Etude préliminaire. Maisons-Alfort, France, GERDAT-IEMVT, 191 p.

Reçu le 30.1.97, accepté le 4.8.97

Resumen

Daget Ph., Poissonet J., Toutain B. Planificación de un muestreo para un estudio regional de la vegetación

Un muestreo de calidad, necesario para un buen estudio de las relaciones entre las especies y las condiciones del medio, se obtiene gracias a una estratificación. De manera que, desde el punto de vista de Neyman (o N -óptimo), un muestreo es óptimo cuando todos los estratos tienen el mismo efectivo. Con el fin de obtenerlo, un algoritmo así como las relaciones y los umbrales que permiten la evaluación de la calidad de este muestreo son propuestos. Se da el ejemplo de un caso proveniente de Burkina Faso.

Palabras clave: Muestreo - Vegetación - Método - Medición - Burkina Faso.