HAL open science

Variables regionalisees et denombrement d'Insectes : cas unidimensionnel

J. Thioulouse, François Houllier, Jean-Claude Onillon

To cite this version:

J. Thioulouse, François Houllier, Jean-Claude Onillon. Variables regionalisees et denombrement d'Insectes : cas unidimensionnel. Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie, 1985, 301, pp.423-428. hal-00698256

HAL Id: hal-00698256

https://hal.science/hal-00698256v1

Submitted on 20 Jan 2022

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.

ÉCOLOGIE GÉNÉRALE. — *Variables régionalisées et dénombrements d'Insectes : cas unidimensionnel.* Note de **Jean Thioulouse, François Houllier** et **Jean-Claude Onillon,** présentée par Théodore Monod.

La précision d'un échantillon systématique ne peut être calculée dans le cadre de la théorie classique de l'échantillonnage. Nous montrons sur un exemple simple relatif à la biologie des populations (dénombrements d'Insectes) que la théorie transitive des variables régionalisées permet ce calcul. Les résultats sont comparés avec ceux obtenus par simulation d'un sous-échantillonnage dans les données.

GENERAL ECOLOGY. — Regionalized variables and Insects Counting: one—dimensional case.

The accuracy of a systematic sample cannot be calculated within the framework of classical sampling theory. A simple example relating to populations biology (insects counting) is used to show that the transitive theory of regionalized variables allows this calculation. Results are compared with those obtained from sub-sampling simulation within data.

INTRODUCTION. — L'échantillonnage systématique est depuis longtemps considéré comme l'une des meilleures méthodes d'échantillonnage [1]. Il allie, par rapport à l'échantillonnage aléatoire et à l'échantillonnage stratifié, facilité de mise en œuvre sur le terrain et précision des estimateurs.

Dans le cadre de l'étude des structures spatio-temporelles en Biologie des Populations, il constitue même la seule solution envisageable aux problèmes de l'analyse des différentes échelles d'hétérogénéité présentes dans la répartition des individus d'une population naturelle [2].

On lui préfère pourtant très souvent d'autres méthodes, en particulier l'échantillonnage aléatoire en Agronomie, ou l'échantillonnage stratifié, séquentiel, ou avec régression en Écologie ([3], pour ne citer qu'un ouvrage récent). En effet, la théorie classique de l'échantillonnage systématique ne permet le calcul de la précision des estimateurs que si la population étudiée peut être modélisée de manière simple (population aléatoire, stratifiée, à tendance linéaire, ...). Mais ces modèles impliquent des hypothèses fortes sur les structures spatiales présentes (voir même leur inexistence), et restreignent donc considérablement le domaine d'application de la méthode.

Nous montrerons, sur un exemple simple, que la théorie des Variables régionalisées ([4], [5]) fournit tous les outils mathématiques nécessaires au calcul de la précision de la densité moyenne ou de la quantité totale de la variable étudiée. Les applications réalisées jusqu'à présent relèvent essentiellement du domaine minier [5] ou de la sylviculture ([6], [7], [8]), et plus exceptionnellement de la Biologie des Populations ([9], [10]); à notre connaissance aucune ne concerne l'échantillonnage des populations d'Insectes.

MATÉRIEL ET MÉTHODES. — 1. *Matériel expérimental.* — Les données traitées concernent le dénombrement, dans une chapelle de serre de 1 350 m^2, de tous les adultes d'Aleurodes (*Trialeurodes vaporarium*) présents sur la totalité des plants de Tomates répartis sur neuf rangées. Parmi celles-ci, nous en avons choisi une ne présentant pas de plants manquants, et montrant une infestation moyenne (N = 778 insectes pour P = 380 plantes); *tous* les adultes d'Aleurodes présents sont dénombrés.

La figure 1 montre la répartition du nombre d'Insectes dans le rang, et révèle la superposition de deux phénomènes :

— une très forte variabilité des effectifs d'Insectes présents sur deux plantes adjacentes, expression de l'agrégativité des Aleurodes adultes;

— l'infestation globale à l'échelle du rang : les effectifs d'Insectes varient le long du rang, selon la position dans la serre.

0249-6313/85/03010423 $2.00 © Académie des Sciences

Le dénombrement exhaustif des Insectes présents sur 380 pieds permet de calculer la précision d'un sous-échantillon de maille arbitraire. Nous calculerons donc, grâce à la théorie des variables régionalisées, la précision de l'estimation obtenue en échantillonnant selon diverses intensités (1 point sur 2, sur 5, sur 10, etc.) dans les données initiales.

2. *Matériel théorique.* — La théorie des variables régionalisées (V.R.) regroupe deux types de méthodes qui diffèrent par leurs hypothèses sous-jacentes. Seule sera résumée rapidement la partie concernant les représentations transitives; on se reportera à [4] et [5] pour une description complète des éléments théoriques.

Soit $y(x)$ la valeur d'une V.R. au point x; y est définie dans un domaine D, à l'intérieur duquel le champ V représente l'ensemble des points x pour lesquels $y(x)$ est différente de zéro. On se donne un échantillon formé de p points x_i répartis sur D aux p nœuds d'un maillage régulier, supposé ici unidimensionnel, et dont l'origine est aléatoire.

On cherche à estimer la « quantité totale » Q :

$$Q=\int_D y(x)\,dx.$$

Pour un échantillon de maille a donnée, un estimateur de Q est :

$$\hat{Q}_a=a\sum_{i=1}^{p} y(x_i).$$

$\hat{Q}_a$ est un estimateur non biaisé : son espérance lorsque l'origine du maillage varie vaut en effet Q. On veut estimer la variance $\sigma^2(a)$ de $\hat{Q}_a$.

On utilise pour cela le covariogramme transitif $g(h)$. $g(h)$ résume l'information structurale contenue dans y et constitue l'outil de base des représentations transitives :

$$g(h)=\int_D y(x)\,y(x+h)\,dx.$$

La variance théorique de $\hat{Q}_a$ vaut :

$$\sigma^2(a)=a\sum_{k=+\infty}^{-\infty} g(ka)-\int_{-\infty}^{+\infty} g(h)\,dh. \tag{1}$$

La deuxième partie du membre de droite est inaccessible sans la connaissance analytique de $g(h)$. Or, dans la pratique, $g(h)$ n'est pas connu exactement; on ne dispose en effet que d'une estimation $\hat{g}(h)$ de $g(h)$ pour les seules valeurs h multiples de la maille a :

$$\hat{g}(h)=a\sum_{i,\,j} y(x_i)\,y(x_j), \qquad x_i-x_j=h.$$

On est donc amené à modéliser le covariogramme. La qualité du modèle choisi conditionne bien sûr la qualité de l'estimation de $\sigma^2(a)$. Si on choisit de représenter $g(h)$ par son développement limité à l'origine et à l'ordre s. $\hat{g}_s(h)$:

$$\hat{g}_s(h)=\sum_{j=0}^{s} c_j|h|^j$$

Explications de la Planche

Fig. 1. — Répartition du nombre d'Aleurodes adultes sur le rang n° 9. Abscisses : position du plant sur le rang (P=1, 2, . . ., 380). Ordonnées : effectifs d'Insectes pour chaque plant (M).

Fig. 1. — Trialeurodes vaporarium *number along the row of tomato plants. x-axis : position of tomato plant* ($P=1, 2, \ldots, 380$). *y-axis: insects number for each plant* (M).

Fig. 2. — Modèle ajusté au covariogramme calculé sur les données. Abscisses : pas du covariogramme ($h=1$, 2, . . ., 380). Ordonnées : valeurs du covariogramme; traits pleins : covariogramme observé; tirets : modèle linéaire ajusté sur les points n°s 2 à 51.

Fig. 2. — *Model fitting to the observed covariogram. x-axis: covariogram step* ($h=1, 2, \ldots, 380$). *y-axis: covariogram values; unbroken line: observed covariogram. pecked line: linear model fitted with points Nos.* 2 *to* 51.

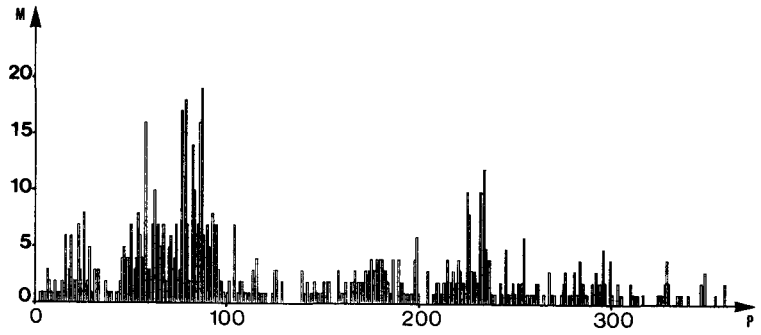

Fig. 1

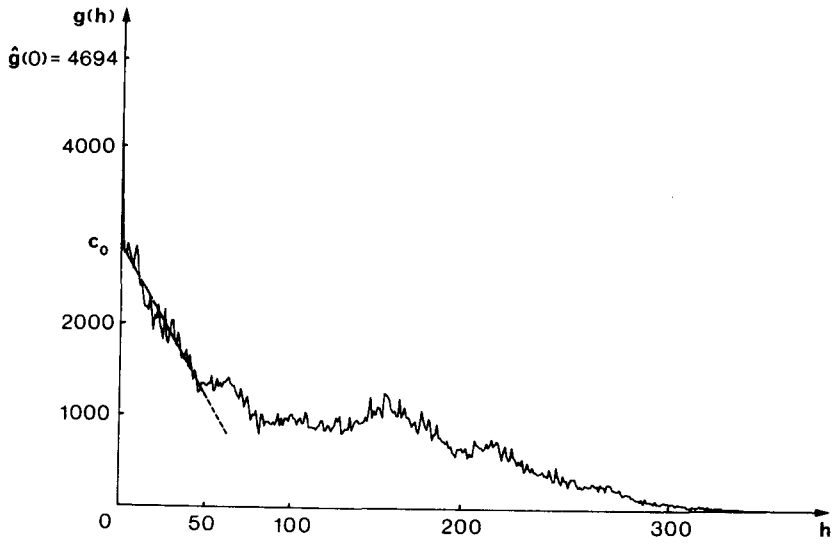

Fig. 2

TABLEAU

Précision des sous-échantillons en fonction de la maille ($a=2$, 5, 10, 20, 38) : comparaison des résultats obtenus par simulation, par la théorie des variables régionalisées et l'échantillonnage aléatoire simple.

Accuracy of sub-samples in terms of sub-sampling ratio ($a=2$, 5, 10, 20, 38) *: comparison of regionalized variables theory, simulation method and simple random sampling.*

Maille de sous-échantillonnage	2	5	10	20	38
Précision (%) :					
V.R.	15	25	35	50	71
Simulations	16	19	38	48	70
Échantillonnage aléatoire	20	32	45	64	88

et si on note C_Δ l'éventuelle discontinuité de g à l'origine :

$$C_\Delta = \hat{g}(0) - \lim_{h \to 0^+} \hat{g}_s(h),$$

on obtient une estimation de $\sigma^2(a)$ par la règle de correspondance suivante, pour a appartenant au domaine de validité du développement limité :

$$\sigma^2(a) = \sum_{j=0}^{s} c_j a^{j+1} \left(-2 \sin\left(j \frac{\pi}{2}\right) \frac{B_{j+1}}{j+1} \right) + a C_\Delta, \tag{2}$$

où les B_j sont les nombres de Bernouilli.

Cette estimation néglige en particulier le « terme fluctuant » qui résulte des irrégularités de $g(h)$ pour h voisin de la dimension du champ V de y; on sait cependant que ce terme fluctuant est en moyenne nul quand la dimension du champ V varie.

Dans le cas d'une variable définie sur un ensemble discret (exemple : plants de Tomates) ces résultats sont valides en remplaçant les intégrales par des sommations.

RÉSULTATS. — Le covariogramme transitif calculé sur les valeurs observées montre une forte discontinuité à l'origine (effet de pépite), due à l'agrégativité des Insectes (*fig.* 2). En considérant les 50 premiers points, à l'exclusion du premier, un ajustement très simple par les moindres carrés conduit au modèle linéaire suivant, acceptable en première approximation :

$$\hat{g}_1(h) = c_0 + c_1 |h|, \tag{3}$$

avec :

$$c_0 = 2\ 881{,}398, \qquad c_1 = -32{,}562$$

et avec une discontinuité à l'origine :

$$C_\Delta = \hat{g}(0) - c_0 = 1\,812{,}062. \tag{4}$$

La règle de correspondance s'écrit alors en utilisant l'équation (2) :

$$\sigma^2(a) = a C_\Delta - \frac{c_1}{6} a^2. \tag{5}$$

Pour une maille donnée, la précision de l'estimateur $\hat{Q}_a$ est estimée par :

$$\Pr(a) = \frac{2\sqrt{\sigma^2(a)}}{Q}.$$

Cette précision vaut 16 % pour une maille égale à 2 (une plante observée sur 2) et atteint 71 % pour une plante échantillonnée sur 38 (tableau). Ces estimations se révèlent proches de celles que l'on peut obtenir par simulation en tirant tous les sous-échantillons possibles pour une maille donnée (tableau). Le gain de précision dû à l'utilisation d'un plan d'échantillonnage systématique par rapport à l'échantillonnage aléatoire est de l'ordre de 20 % pour les mailles considérées (tableau).

Pour expliquer cette différence, remarquons que la décomposition de la formule (5) en deux termes :

$$a C_\Delta \text{ d'une part,}$$

et :

$$\frac{c_1}{6} a^2 \text{ d'autre part,}$$

permet de considérer deux sources dans la variabilité totale des dénombrements, respectivement :

— la *variance intrinsèque*, due à l'agrégativité des Insectes à l'échelle d'une plante (composante aléatoire);

— la variance de sous-échantillonnage au sens strict, ou *variance de régionalisation*, due à la position des points du sous-échantillon par rapport à la répartition générale des Insectes dans le rang.

Le gain observé est dû à la prise en compte de la régionalisation de la variable, négligée dans le cas d'un échantillonnage aléatoire simple.

CONCLUSION. — L'exemple proposé est un cas d'école puisqu'on dispose *a priori* d'un échantillonnage exhaustif de la population étudiée. Le problème est plus compliqué dans le cas pratique général d'un échantillon qui ne représente qu'un fragment de la population. Le choix et l'ajustement d'un modèle du covariogramme doivent alors se fonder à la fois sur l'observation des données et sur les connaissances biologiques de la population étudiée. Ceci souligne l'importance d'échantillons d'investissement plus denses que les échantillons ultérieurs, permettant d'acquérir des informations détaillées sur la structure de la population, puis de choisir et valider un modèle du covariogramme.

La théorie des variables régionalisées, dont seule la partie relative aux méthodes transitives a été abordée, mérite d'être éprouvée sur d'autres exemples de Biologie des Populations. Cette théorie se présente actuellement comme la seule méthode rigoureuse de calcul de la précision d'un échantillon systématique. La connaissance de la précision d'un effectif d'Insectes présents dans un espace donné, et la mise en évidence de structures spatio-temporelles dans la répartition de ces Insectes, constituent deux points fondamentaux de la Biologie des Populations, points qui restent encore peu explorés pour les populations naturelles.

Les auteurs remercient Pascal Monestiez pour ses commentaires sur le manuscrit.

Remise le 3 juin 1985, acceptée le 1er juillet 1985.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] W. G. COCHRAN, *Sampling techniques*, Wiley, New York, 1953.
[2] D. CHESSEL, *Biométrie et Écologie*, J. M. LEGAY et R. TOMASSONE éd., Paris, 1978, p. 45-135.
[3] S. FRONTIER, *Stratégies d'échantillonnage en Écologie*, Masson, Paris, 1982.
[4] G. MATHERON, *Les variables régionalisées et leur estimation : une application de la théorie des fonctions aléatoires aux Sciences de la Nature*, Masson, Paris, 1965.
[5] G. MATHERON, La théorie des variables régionalisées et ses applications, *Cahiers du Centre de Morphologie mathématique de Fontainebleau*, 1970.
[6] J. BACHACOU et N. DECOURT, *Ann. Sc. Forest.*, 33, (4), 1976, p. 177-198.
[7] D. GUIBAL, L'estimation des Okoumés du Gabon. Problèmes méthodologiques, *Cahiers du Centre de Morphologie mathématique de Fontainebleau*, 1973.
[8] G. MARBEAU, Géostatistique forestière, *Thèse d'Ingénieur Docteur*, École des Mines, Paris, 1976.
[9] M. EL BAHI, Représentations transitives des variables régionalisées : Examen des fondements théoriques et essai d'utilisation pour l'analyse spatiotemporelle d'une population d'insectes, *Thèse de 3e cycle*, Université de Lyon-I, 1981.
[10] B. MATHY, D. CHESSEL, H. PLOYE et J. THIOULOUSE, *Biology of the cell*, 48, 1983, p. 74*a*.

J. T. et F. H. : *Laboratoire de Biométrie, U.A.-C.N.R.S. n°* 243,
Université Claude-Bernard, Lyon-I,
43, *boulevard du* 11-*Novembre*-1918, 69622 *Villeurbanne;*
J.-C. O. : *Centre de Recherches agronomiques de Versailles, Station de Zoologie,*
route de Saint-Cyr, 78000 *Versailles.*