

Modélisation Spatio-Temporelle d'une Epidémie par un Processus de Gibbs: Estimation et Tests

J. Chadoeuf

Laboratoire de Biométrie, I.N.R.A., Domaine St Paul, 84140 Avignon, France

D. Nandris, J. P. Geiger, M. Nicole

Laboratoire de Phyto-Pathologie du Centre d'Adiopodoumé, O.R.S.T.O.M.,
BP. V-51 Abidjan, Côte d'Ivoire

et

J. C. Pierrat

Laboratoire de Biométrie, I.N.R.A-C.N.R.F., Champenoux, 54280 Seichamps, France

RÉSUMÉ

Pour modéliser la dynamique d'une épidémie pour laquelle il n'y a pas guérison des plantes malades (présence d'un état absorbant), nous avons utilisé un champ markovien à chaque date sur les individus sains à la date précédente. Nous avons montré que la méthode de codage permet de retrouver les propriétés statistiques classiques (consistance, normalité asymptotique, tests d'hypothèses emboîtées) si le processus est α -mélangeant, c'est à dire si deux individus deviennent quasiment indépendants dès que la distance qui les sépare augmente. Cette condition est toujours assurée dans le cas d'un processus sur \mathbb{Z} et l'est également pour un processus sur \mathbb{Z}^v , $v \geq 2$ si les interactions entre individus voisins sont suffisamment faibles (les conditions de Simon sont vérifiées à chaque date).

L'application présentée porte sur la modélisation de la propagation simultanée de deux champignons du sol (*Phellinus noxius* et *Rigidoporus lignosus*) dans une plantation d'hévéas. Le modèle prend en compte des caractéristiques essentielles de cette propagation (contamination de proche en proche, risque d'attaque lié à l'état du voisinage) et ses paramètres peuvent recouvrir une interprétation biologique. Il a donc été possible de tester différentes hypothèses sur la propagation de ces champignons parasites: influence des arbres trouvés morts à la première ronde, interaction entre les deux champignons, stationnarité de la contamination secondaire. Dans une dernière partie, l'utilisation de ce modèle dans une optique de prévision est illustrée.

1. Introduction

Les problèmes d'épidémie dans le domaine des cultures agricoles sont, de longue date, objets de réflexion en statistique spatiale (Besag, 1974; Ripley, 1981; Jolivet, 1977). Dans ce cadre, nous nous sommes plus particulièrement intéressés à la modélisation d'épidémies où il n'y a pas guérison d'une plante atteinte et où les observations sont disposées régulièrement dans l'espace comme dans le temps (l'observateur visitant régulièrement les parcelles).

Les questions d'ordre agronomique concernent la détermination de la "zone de contamination" d'une plante malade, l'étude de l'évolution dans le temps des caractéristiques de l'épidémie et la prévision du nombre de plantes atteintes.

Key words: Champignons parasites; Epidémiologie; Hevea brasiliensis; Markov; Modèle de Gibbs; Pathologie forestière; Processus spatiaux.

Fonds Documentaire ORSTOM



010015913

1165

Fonds Documentaire ORSTOM

Cote: Bx 15913 Ex: 1

Besag (1977) donne une première réponse à ce type d'épidémie dans le cas d'une seule maladie. Pour cela, il la modélise à la première date par un processus markovien aux plus proches voisins. Aux dates suivantes, il considère que l'état d'une plante à une date donnée ne dépend que de l'état sanitaire de ses voisins à la date précédente. Cette dernière hypothèse reste assez forte car elle suppose qu'à une même date, deux individus voisins sont attaqués indépendamment. Il suggère alors une modélisation globale de ce type de phénomène, basée également sur ce type de modèle (et donc d'interaction à courte portée) à chaque date. Dans une première partie, nous présentons la définition de ce modèle dans le cadre de plusieurs maladies. Les problèmes d'estimations et de tests sont développés dans une deuxième partie. Pour cela, nous nous fondons sur les méthodes de pseudo-vraisemblance et plus spécifiquement sur les résultats de Guyon (1985) qui aborde en particulier les problèmes d'estimation paramétrique dans les modèles markoviens. Nous évoquerons dans une dernière partie les problèmes de prévision et de simulation.

2. Modèle

Chaque plante, repérée par sa position $i \in I \subset \mathbb{Z}^v$, $v = 1$ ou 2 , est suivie au cours de n dates. Par la suite, l'indice i ne sera pas précisé, sauf dans les cas nécessaires. La plantation peut être attaquée par L maladies. Une fois malade, une plante ne peut pas guérir; chaque maladie constitue donc un état absorbant.

Soit k ($0 \leq k \leq n$) une date donnée et n la dernière date d'observation. Notons $X_i(k)$ l'état à la date k de l'individu situé en i ;

$X_i(k) = 0$ si l'individu i est sain à la date k ,

$X_i(k) = (k', l)$ si l'individu i est dans l'état l , $1 \leq l \leq L$, depuis la date $k' \leq k$.

Nous noterons V l'ensemble des plus proches voisins de i (son cardinal est $|V| = 2$ dans le cas d'une seule ligne, $|V| = 4$ dans le cas d'un réseau à maille carrée). A chaque date d'observation k , conditionnellement à l'état lors de la date précédente, le modèle est un modèle de Gibbs (Krickeberg, 1975) aux plus proches voisins sur l'ensemble des plantes saines à la date $k - 1$. Pour cela, on définit en chaque point i un potentiel dépendant de l'état des autres individus dans V à la date k ainsi que de l'état u de i par:

$$W_u = \gamma[u] + \sum_{j \in V} \beta[X_j(k), u]$$

avec la convention $\gamma[0] = 0$ (le potentiel principal d'un individu sain est nul) et $\beta[u, v] = 0$ si u ou $v = 0$ (l'énergie d'interaction entre deux individus voisins est nulle si l'un d'eux est sain).

La probabilité conditionnelle de trouver l'individu i dans l'état l à la date k sachant qu'il était sain à la date $k - 1$ et connaissant l'état des autres individus est:

$$\begin{aligned} P[X_i(k) = (k, l) \mid X_i(k-1) = 0; \{X_j(k)\}, j \neq i] \\ &= P[X_i(k) = (k, l) \mid X_i(k-1) = 0; \{X_j(k)\}, j \in V(i)] \\ &= \frac{\exp(W_{k,l})}{\sum_{l'=0}^L \exp(W_{k,l'})}. \end{aligned}$$

La première égalité exprime que vis-à-vis de i , la connaissance de l'état de son voisinage est équivalente à la connaissance de l'ensemble des individus autres que i . Les paramètres $\gamma(u)$ sont reliés aux probabilités conditionnelles d'être attaqué quand les voisins sont sains. Les paramètres $\beta(u, v)$ décrivent la liaison entre deux voisins respectivement dans les états u et v .

Pour l'ensemble des individus, on définit de même à chaque date un potentiel total, porté uniquement par les individus sains à la date précédente, par:

$$WT(k) = \sum_{i \text{ tels que } X_i(k-1)=0} \gamma[X_i(k)] + \sum_{\substack{i,j \text{ voisins,} \\ \text{tels que } X_i(k-1)=0 \text{ ou } X_j(k-1)=0}} \beta[X_i(k), X_j(k)]$$

alors,

$$P[\{X_i(k)\} = \{x_i(k)\}, i \in I | \{X_i(k-1)\}, i \in I] = \frac{\exp(WT(k))}{\sum_{\substack{i \in I, X_i(k)=0, k_1, \dots, k_L \\ \text{tels que } X_i(k-1)=0}} \exp(WT(k))}$$

et on a pour la probabilité globale entre 1 en n :

$$P[\{X_i(n)\}, i \in I] = \prod_{k=1}^n P[\{X_i(k)\}, i \in I | \{X_i(k-1)\}, i \in I].$$

Cette probabilité reste très lourde à estimer. Dans le cas $n = 1$, celle-ci peut toutefois être estimée si $\nu = 1$ par transformation en une chaîne de Markov (Dacunha-Castelle et Duflo, 1983). Dans le cas $\nu = 2$ et $n = 1$, Younes (1988) a développé des méthodes d'estimation de la vraisemblance fondées sur des méthodes de recuit simulé. Dans notre cas, nous avons cependant préféré la méthode de codage de Besag, celle-ci offrant une réelle souplesse d'utilisation.

3. Estimation et Tests

Les paramètres du modèle ont été estimés par maximum de codage, la fonction utilisée étant le produit des vraisemblances conditionnelles à chaque date:

$$F = \prod_{k \leq n, i \in C} P[X_i(k) | \{X_j(k)\}, j \in V(i)] = \prod_{i \in C} f_i,$$

où C est un codage, c'est à dire un sous-ensemble de points maximal de I tel que deux points de C ne sont pas voisins l'un de l'autre (Besag, 1974).

Guyon (1985) a montré que, sous certaines hypothèses de régularité, les estimations par codage sont consistantes, asymptotiquement normales et que l'on dispose de tests d'hypothèses paramétriques emboîtées. Fondamentalement, le modèle doit vérifier une condition de mélange à décroissance exponentielle: si U et V sont deux parties finies de I , si \mathcal{F}_U est la tribu engendrée par les applications coordonnées sur U , $A \in \mathcal{F}_U$, $B \in \mathcal{F}_V$, on doit avoir:

$$|P(AB) - P(A)P(B)| \leq Mf(|U|, |V|)\alpha^{d(U,V)},$$

où M et α sont des constantes avec $\alpha < 1$, $|U|$ représente le cardinal de U , f est un polynôme et d la distance spatiale entre U et V (Billingsley, 1983). Intuitivement, cette condition exprime le fait que deux événements deviennent rapidement suffisamment indépendants pourvu qu'ils soient observés en des points éloignés. Elle est démontrée dans l'Annexe 1. Dans le cas $\nu = 1$, cette condition est toujours vérifiée. Dans le cas $\nu = 2$, elle l'est dès que les interactions sont suffisamment faibles soit $\sum_{l,l',k' \leq k} \beta[(k, l), (k', l')] < 1$ pour tout k [conditions de Simon (1979)]. Les autres conditions sont des conditions de régularité du processus qui se déduisent immédiatement des résultats de Guyon (1985). D'un point de vue pratique, pour effectuer ces tests on maximise la fonction F qui peut également s'écrire $\prod_{k \leq n} F_k$ où F_k est le produit à la date k des probabilités conditionnelles.

Soient alors $H_q \subset H_p$ deux hypothèses emboîtées et ψ et θ les paramètres associés. Le test de H_q contre H_p se fait en utilisant la statistique $T = 2[\log\{F(\psi)\} - \log\{F(\theta)\}]$ qui suit asymptotiquement un χ^2 à $p - q$ degrés de liberté où p et q sont les dimensions de H_p et H_q . T s'écrit alors $T = \sum_k T_k$ où $T_k = 2[\log\{F_k(\psi)\} - \log\{F_k(\theta)\}]$ est la contribution de la date k à la statistique.

Dans le cas où chacune des deux hypothèses se décompose en un ensemble de sous-hypothèses associées à chacune des dates de façon indépendante par date, alors chaque statistique T_k suit également une loi χ^2 (cf. Annexe 2). C'est par exemple le cas quand on teste l'hypothèse "il n'y a pas d'interaction entre points dans l'état l " soit $\omega = \{\beta[(k, l), (k', l)] = 0, k' \leq k \leq n\}$ pour un état l donné. On décompose en effet cette hypothèse en l'intersection des hypothèses "à la date k , il n'y a pas interaction entre un point dans l'état l à la date k et les points dans l'état l depuis les dates $k' \leq k$ " ce qui se traduit au niveau des paramètres par $\omega^k = \{\beta[(k, l), (k', l)] = 0, k' \leq k\}$. Ces hypothèses interviennent toutes sur des paramètres différents en donc chaque T_k suit un χ^2 à k degrés de liberté (ddl).

4. Prédiction et Simulation

On peut, dans le cas d'une seule date et si $\nu = 1$ (répartition en ligne des individus), étudier explicitement les lois d'un certain nombre de variables synthétiques comme le nombre de plantes atteintes par une maladie. En effet le processus peut être étudié comme une chaîne de Markov à $L + 1$ états en parcourant la ligne dans un sens donné. Sa matrice de transition peut alors être calculée en fonction des paramètres γ et β (Spitzer, 1973). Dans le cas où il n'y a pas interaction, les variables aléatoires décrivant les états des différents individus sont indépendantes et suivent une loi multinomiale. Dans le cas général, les lois de ces variables peuvent être atteintes par simulation.

Si $\nu = 1$, le processus est simulé d'abord à la première date selon la chaîne de Markov précédemment décrite. Aux dates suivantes, la ligne simulée se trouve découpée en différents segments limités par des points attaqués aux dates précédentes. La simulation se fait en tirant successivement l'état des différents points du segment selon la loi conditionnelle à l'état du point précédent à la même date et à l'état du point fermant le segment à la date précédente.

Si $\nu = 2$, la simulation peut être faite à chaque date selon une dynamique de Métropolis (Prum, 1986), utilisant l'échantillonneur de Gibbs: on tire un point au hasard dans I ; s'il est déjà dans un état non nul depuis une date précédente, son état est inchangé, sinon, on calcule les probabilités conditionnelles des différents états possibles de ce point selon son voisinage actuel en on tire son état selon cette loi.

5. Application

5.1 Données et Modèle

La population est formée de 2526 arbres (*Hevea brasiliensis*). Elle est suivie d'un point de vue sanitaire au cours de 7 dates de mesures. Un an sépare les deux premières dates, six mois les suivantes. Ces arbres sont plantés régulièrement en lignes séparées de 8m et tous les 2m sur une ligne. Ils peuvent être attaqués par deux champignons du sol, *Rigidoporus lignosus* (Kl.) Imaz. (représenté par $l = 1$) et *Phellinus noxius* (Corner) G. H. Cunne ($l = 2$). Ces champignons se propagent principalement dans le sol, soit au moyen d'hyphes mycéliens, soit au travers de rhizomorphes et contaminent les arbres de proche en proche (Nandris, Nicole, et Geiger, 1987).

Entre deux dates d'observation, les champignons ne peuvent pas se propager entre deux arbres situés au delà d'une distance limite liée à la vitesse d'avancement des hyphes ou des rhizomorphes dans le sol. Aussi les arbres ou les souches au-delà d'un certain voisinage d'un arbre central n'influent plus sur ce dernier. D'autre part, des analyses en laboratoire sur la vitesse de croissance des hyphes mycéliens ainsi que des analyses préliminaires nous ont permis de montrer qu'à chaque date, la contamination par les deux arbres les plus proches sur la même ligne est prépondérante (Chadœuf et al., 1988). Nous modéliserons donc le processus comme un ensemble de lignes indépendantes (cas $\nu = 1$), le voisinage de chaque arbre étant constitué de deux arbres.

Le modèle est alors paramétré à chaque date k par:

2 paramètres principaux $\gamma(k, 1)$ et $\gamma(k, 2)$ qui caractérisent la contamination primaire par *R. lignosus* et *P. noxius* c'est à dire la probabilité d'attaque d'un arbre entouré d'arbres sains par l'un des deux champignons.

3 paramètres d'interaction $\beta[(k, 1), (k, 1)]$, $\beta[(k, 1), (k, 2)]$, et $\beta[(k, 2), (k, 2)]$ qui traduisent les liaisons entre arbres attaqués à la même date.

$4(k - 1)$ paramètres d'interaction $\beta[(k, 1), (k', 1)]$, $\beta[(k, 1), (k', 2)]$, $\beta[(k, 2), (k', 1)]$, et $\beta[(k, 2), (k', 2)]$ qui traduisent les liaisons avec les arbres attaqués aux dates précédentes. Ces paramètres, ainsi que les paramètres d'interaction à la même date, traduisent plutôt l'extension de la maladie d'arbre en arbre. Ils décrivent plus spécifiquement la notion de contamination secondaire.

A la première date, on relève des arbres morts, soit de l'attaque d'un des deux champignons, soit pour des causes très diverses. On décrit également la structure spatiale de ces arbres par deux paramètres $\gamma(m)$ et $\beta(m, m)$.

A chaque date suivante, deux paramètres d'interaction $\beta[(k, 1), m]$ et $\beta[(k, 2), m]$ décrivent la liaison entre arbres morts en première date et arbres attaqués aux dates ultérieures.

5.2 Tests d'Hypothèses

Le modèle général prenant en compte les deux voisins les plus proches donne un codage de 1134,45 pour 135 ddl et servira de modèle de base pour tester différentes hypothèses.

Les arbres trouvés morts à la première ronde ont une influence dans la propagation de la maladie. Le test $\beta[(k, 1), m] = \beta[(k, 2), m] = 0$ pour $k = 2, \dots, 6$ est rejeté avec $\chi^2 = 45,72$ pour 14 ddl.

L'hypothèse d'interaction entre les deux champignons est également rejetée avec $\chi^2 = 70,3$ pour 49 ddl. Elle correspond à l'hypothèse $\beta[(k, 1), (k', 2)] = 0$. La décomposition du χ^2 par date est intéressante (Tableau 1). Il apparaît en effet que, par date, ce seuil est toujours proche du seuil limite de 5% et n'est pas atteint pour les dernières rondes. Il traduit en fait l'existence d'une interaction de plus en plus faible entre les deux

Tableau 1
Test par dates de l'hypothèse d'interaction entre les deux champignons

Date k	V_k	Seuil à 5%	ddl
1	6.36	3.84	1
2	8.78	7.82	3
3	11.78	11.1	5
4	18.16	14.1	7
5	9.01	16.9	9
6	12.68	19.7	11
7	3.48	22.4	13

champignons, cette décroissance de l'interaction pouvant être due au ralentissement progressif de l'épidémie.

Nous avons également testé la stationnarité temporelle de la contamination secondaire, c'est à dire de l'évolution du pouvoir contaminateur d'un arbre atteint en fonction de sa date d'attaque. Celle-ci s'exprime au niveau des paramètres par $\beta[(k, u), (k', v)] = \beta[(k - k', u), (1, v)]$ pour $u, v = 1, 2$. Cette hypothèse est acceptée avec $\chi^2 = 80$ pour 109 ddl.

5.3 Préviation et Simulation

Dans une optique de prévision de la propagation de l'épidémie, et au vu des résultats précédents, nous avons modélisé l'évolution des différents paramètres en fonction du temps par une tendance linéaire et une composante périodique correspondant à l'existence de deux saisons. La composante périodique des paramètres d'interaction β était non-significative. Pour les paramètres d'interaction avec les arbres morts, les deux composantes étaient non-significatives, d'où le modèle final, pour $u, v = 1, 2$, et $k \geq k'$:

$$\gamma[(k, u)] = a_0(u) + ka_1(u) + (-1)^k a_2(u),$$

$$\beta[(k, u), (k', v)] = b_0(u, v) + (k - k')b_1(u, v),$$

$$\beta[(k, u), m] = b_0(u, m),$$

Ce modèle est accepté contre le modèle général ($\chi^2 = 270,24$ pour 118 ddl). Les estimations sont présentées dans le Tableau 2 avec leurs intervalles de confiance individuels à 95%. Ces derniers ont été construits par maximum de pseudo-vraisemblance. On peut noter qu'ils restent assez importants malgré la taille de l'échantillon mesuré. Les paramètres d'interaction similaires pour chacun des deux champignons sont du même ordre de grandeur. Les différences entre les deux champignons sont principalement observées au niveau des deux paramètres principaux $a_1(0)$ et $a_2(0)$, c'est à dire pour la valeur à l'origine de γ . Les deux paramètres d'interaction $b_1(1, 1)$ et $b_1(2, 2)$ diffèrent à un degré moindre.

Tableau 2
Estimation des paramètres du modèle et intervalles de confiance individuels à 95%

Paramètres	Estimation	Inf	Sup
$a_1(0)$	-4,38	-4,81	-4,00
$a_1(1)$	-0,20	-0,30	-0,10
$a_1(2)$	-0,33	-0,52	-0,14
$a_2(0)$	-5,37	-6,06	-4,78
$a_2(1)$	-0,19	-0,34	-0,03
$a_2(2)$	-0,25	-0,54	0,03
$a_0(m)$	-2,36	-2,59	-2,15
$b_0(m, m)$	0,73	0,40	1,04
$b_0(1, m)$	0,76	0,45	1,06
$b_0(2, m)$	0,83	0,32	1,31
$b_0(1, 1)$	-0,21	-0,37	-0,06
$b_1(1, 1)$	2,86	2,46	3,29
$b_0(2, 2)$	-0,35	-0,68	-0,07
$b_1(2, 2)$	3,88	3,22	4,57
$b_1(1, 2)$	-0,19	-0,66	0,16
$b_0(1, 2)$	1,77	1,09	2,41
$b_1(2, 1)$	-0,12	-0,46	0,15

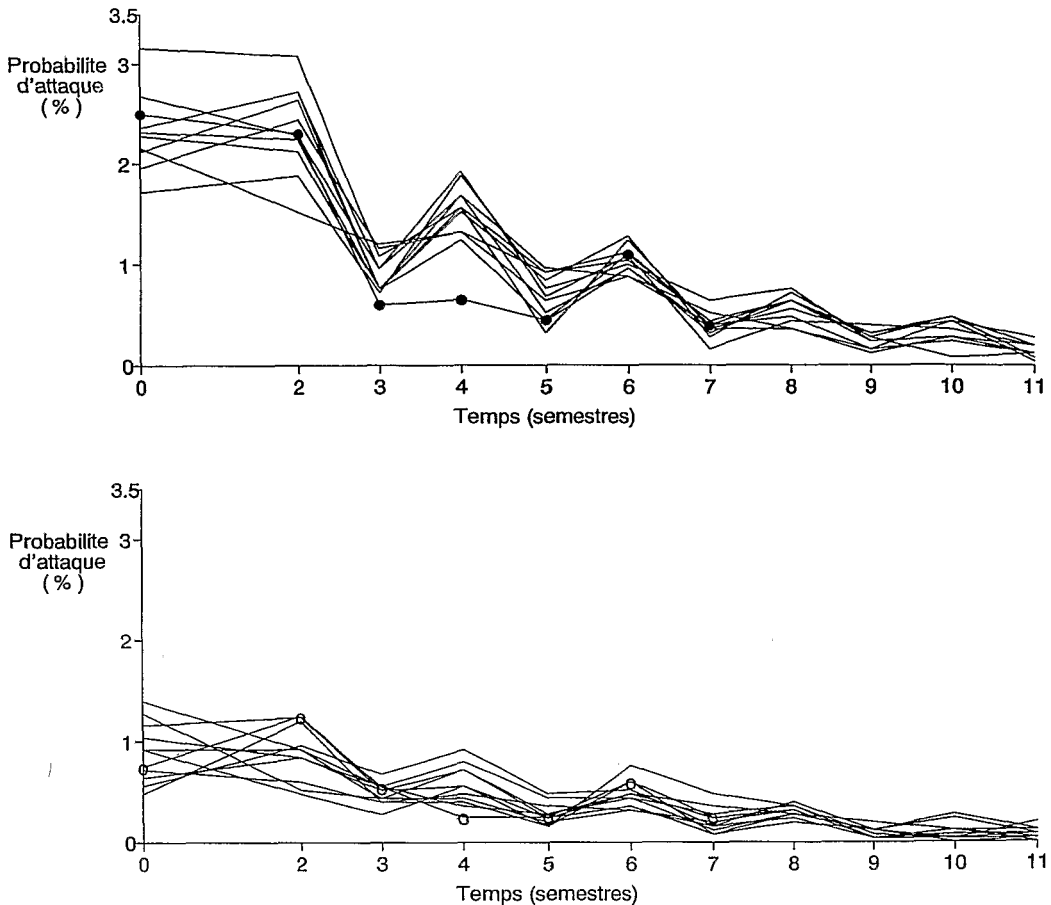


Figure 1. Evolution temporelle de la probabilité d'attaque d'un arbre par *Lignosus* (●: observé) dans le graphique du haut et par *Noxius* (○: observé) dans le graphique du bas. Les dix simulations sont faites chacune sur un ensemble de 2500 arbres.

La première différence peut être due à une forte différence de départ (avant plantation) dans la répartition des deux champignons. La deuxième s'explique par la différence dans les modes de propagation des deux champignons (rhizomorphe mycélien dans le cas de *Rigidoporus Lignosus*, manchon mycélien dans le cas de *Phellinus Noxius*).

La simulation, destinée à prévoir l'évolution de l'épidémie, a été conduite selon la première procédure décrite au paragraphe 4 sur une ligne de 2500 arbres (Figure 1). On peut noter un bon accord global entre données mesurées et simulées, si on excepte la quatrième ronde. L'épidémie est fortement ralentie dès la 9^{ième} ou la 10^{ième} date (soit cinq ans après la première mesure et 9 ans après la mise en place de la plantation). Toutefois, une forte variabilité peut être remarquée autour des valeurs moyennes prédites par le modèle.

6. Conclusions

De tels modèles spatio-temporels permettent de tester une grande variété d'hypothèses biologiques (évolution temporelle, influence de différents facteurs, portée du phénomène, ...) et sont donc bien adaptés à l'étude des épidémies dans le domaine végétal. Les modèles de Gibbs permettent de décrire l'aspect spatial de l'épidémie alors que l'aspect temporel,

décrit par une propriété markovienne, permet de prendre en compte le suivi de l'épidémie dans le temps. D'un point de vue statistique, l'hypothèse principale sur laquelle repose toute l'estimation est une hypothèse d'ergodicité à temps fixé conditionnellement au passé. Dans le cas traité, cette hypothèse a été assurée par la décroissance rapide de la corrélation entre deux points avec leur distance. Dans le cas à deux dimensions, cette hypothèse a été obtenue en supposant que les interactions sont relativement faibles.

Cette dernière hypothèse reste très contraignante. Une approche peut être envisagée en faisant une hypothèse de sous-ergodicité (Geman et Graffigne, 1986). Dans le cas du modèle de Gibbs au plus proche voisin, cette approche permet de retrouver l'ensemble des résultats précédents (Guyon, communication personnelle).

D'un point de vue pratique, l'utilisation de tels modèles se trouve cependant limitée par le nombre de paramètres à mettre en œuvre dès que l'on augmente la taille du voisinage, c'est à dire le nombre d'interactions possibles. D'autre part, le nombre de dates joue également un rôle important dans le nombre de paramètres à estimer, car il croît en n^2 si aucune hypothèse supplémentaire (comme la constance de la contamination secondaire par exemple) n'est faite. Ces modèles s'avèrent donc plus spécialement adaptés au suivi de plantations in situ, où l'on dispose généralement de quelques inspections par an et où l'ensemble relativement important de plantes observées permet un bon ajustement.

SUMMARY

Gibbs models have been introduced in order to describe the spatial separation of epidemiological diseases in plants (Besag, 1974, *Journal of the Royal Statistical Society, Series B* 76, 192–236) when the individuals are located on a regular lattice of point sites. Besag (1977, *Bulletin of the International Statistical Institute* 47, 77–92) extended these models to the spatio-temporal case. In this paper, a more general modeling of such phenomena is proposed. The case under study is that of epidemics without remission of the disease. In the proposed model, the infection probability of an individual depends on the state of health of its neighbors at the same date (i.e., healthy, diseased at this date, or diseased at a preceding date). At each date, the process is then a Gibbs model defined on the individuals who were previously healthy. It takes into account the influence of the individuals who were diseased at a preceding date.

Parameters are estimated using a pseudo-likelihood method. The product, over the dates of observation and over the individuals belonging to a coding set, of the conditional probabilities at each date of the state of health of an individual given the state of health of its neighbors is maximized. This procedure extends the method proposed by Besag (1974) for the case of a single date. It can be shown that the estimators are asymptotically consistent and normally distributed if the interactions are sufficiently weak. This condition is always satisfied for a one-dimensional process. Moreover, with such a method tests of nested parametric hypotheses are available.

The model is applied to the simultaneous propagation of two soil fungi in a rubber tree plantation. It takes into account the principal characteristics of the diseases (propagation from tree to tree, attack risk depending on the state of health of the neighborhood). The parameters of the model have a biological interpretation. It is then possible to test different hypotheses: influence of the trees found dead at the first inspection, interaction between the two fungi, stationarity of the secondary contamination. In the last part of the paper the use of such models for simulation and prediction is pointed out.

BIBLIOGRAPHIE

- Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice system. *Journal of the Royal Statistical Society, Series B* 76, 192–236.
- Besag, J. E. (1977). Some methods of statistical analysis for spatial data. *Bulletin of the International Statistical Institute* 47, 77–92.
- Billingsley, P. (1983). *Convergence of Probability Measures*. New York: Wiley.
- Chadœuf, J., Joannes, H., Nandris, D., and Pierrat, J. C. (1988). Mathematical analysis and modeling of epidemics of rubber tree diseases: Probability of infection of an individual tree. *Forest Science* 34, 831–845.

- Dacunha-Castelle, D. et Duflo, M. (1983). *Probabilités et Statistiques. 2 Problèmes à temps mobile*. Paris: Masson Ed.
- Geman, S. and Graffigne, C. (1986). Markov random field image models and their applications to computer vision. *Proceedings of International Congress of Mathematicians*, A. M. Gleason (ed.). Providence, Rhode Island: American Mathematical Society.
- Guyon, X. (1985). Estimation d'un champ par pseudo-vraisemblance conditionnelle: Étude asymptotique et application au cas markovien. *Actes de la 6ème Rencontre Franco-Belge de Statisticiens*. Edition des Facultés Universitaires de St-Louis, Bruxelles.
- Jolivet, E. (1977). Application de quelques résultats relatifs aux processus ponctuels à un essai de modélisation d'une épidémie de Septoriose du Blé. Thèse de 3e cycle, Université Paris XI n° 2209.
- Karlin, S. (1969). *Initiation aux Processus Aléatoires*. Paris: Dunod.
- Krickeberg, K. (1975). Processus ponctuels en statistique. *Ecole d'été de probabilités de St Flour X*. Springer Lecture Notes in Mathematics 529.
- Nandris, D., Nicole, M., et Geiger, J. P. (1987). Root rot diseases of rubber tree. *Plant Diseases* 71, 298-306.
- Prum, B. (1986). *Processus sur un Réseau et Mesures de Gibbs*. Paris: Masson Ed.
- Ripley, B. D. (1981). *Spatial Statistics*. New York: Wiley.
- Simon, B. (1979). A remark on Dobrushin's uniqueness theorem. *Communications in Mathematical Physics* 68, 183-185.
- Spitzer, F. L. (1973). Introduction aux processus de Markov à paramètres dans \mathbb{Z}^d . *Ecole d'été de probabilités de St Flour III*. Springer Lecture Notes in Mathematics 390.
- Younes, L. (1988). Problèmes d'estimation paramétrique pour des Champs de Gibbs Markoviens. Applications au Traitement d'Images. Thèse Paris XI.

Received October 1989; revised December 1990; accepted June 1991.

ANNEXE 1

La condition d' α -mélange est démontrée en deux étapes. Dans une première étape, nous montrons que le processus est α -mélangeant sur l'ensemble des n dates si à chaque date il l'est conditionnellement à la date précédente. Dans une deuxième étape, nous étudions la condition d' α -mélange à une date donnée conditionnellement à la date précédente.

1. Montrons par récurrence sur le nombre de dates que le processus présenté plus haut vérifie la condition d' α -mélange si cette condition est vérifiée à chaque date conditionnellement à la date précédente (nous présentons ici le cas d'une seule maladie $l = 1$). Notons $\mathcal{F}_k^{U,V}$ la tribu engendrée sur U par les applications coordonnées jusqu'à la ronde $k - 1$, \mathcal{F}_k^F la tribu engendrée par les applications coordonnées à la date k , si F est une réalisation du processus sur U à la date $k - 1$.

L'hypothèse d' α -mélange du processus jusqu'à la date $k - 1$ donne, pour $A \in \mathcal{F}_k^{U,V}$, $B \in \mathcal{F}_k^F$:

$$|P(AB) - P(A)P(B)| \leq M |U| |V| \alpha^{d(U,V)},$$

où $|U|$ représente le cardinal de U .

La condition d' α -mélange à la date k donne, pour F une configuration de U à la date $k - 1$, et pour $A \in \mathcal{F}_k^F$, $B \in \mathcal{F}_k^F$:

$$|P(AB) - P(A)P(B)| \leq M |U| |V| \alpha^{d(U,V)}.$$

Les constantes M et α ne dépendent que des paramètres de la loi de Gibbs à la date k conditionnellement à la date $k - 1$ et sont donc indépendantes de F (Guyon, 1985). Elle donne également, si $A_2 \in \mathcal{F}_k^F$, si $U \subset U_1$ et si $\mathbb{T} \in \mathcal{F}_{\partial U}^F$ où $\partial U = \{u \in L \setminus U, d(u, U) \leq 1\}$:

$$|P(A_2 | \mathbb{T}F) - P(A_2 | F)| \leq M |U| \alpha^{d(U, \partial U)}$$

avec de la même façon M et α indépendantes de F .

De plus, de la propriété markovienne à la ronde k on déduit, si $B_1 \in \mathcal{F}_{k-1}^F$,

$$P(A_2 | B_1) = \sum_{\mathbb{T}} P(A_2 | \mathbb{T}B_1)P(\mathbb{T} | B_1) = \sum_{\mathbb{T}} P(A_2 | \mathbb{T}F)P(\mathbb{T} | B_1)$$

et donc

$$|P(A_2 | B_1) - P(A_2 | F)| \leq M |U| \alpha^{d(U, \partial U)}.$$

Si donc $U \subset U_1, V \subset V_1$, avec $d(U, \partial U_1) \geq d, d(V, \partial V_1) \geq d, d(U_1, V_1) \geq d$, notons A_1, B_1, C_1 , et D_1 configurations respectivement sur U, V, U_1 , et V_1 à la date $k - 1$, C_1 et D_1 compatibles avec A_1 et B_1 . Soit F une configuration à la date $k - 1$ coïncidant avec C_1 et D_1 sur U_1 et V_1 .

Soit $A_2 \in \mathcal{F}_k^b(F)$ et $b_2 \in \mathcal{F}_k^b(F)$, alors,

$$\begin{aligned} P(A_2 B_2 | C_1 D_1) &= P(A_2 B_2 | F) + \varepsilon_1 \\ &= P(A_2 | F)P(B_2 | F) + \varepsilon_2 \\ &= P(A_2 | C_1)P(B_2 | D_1) + \varepsilon_3 \end{aligned}$$

où chaque ε_i est majoré par une inégalité similaire aux précédentes et donc,

$$\begin{aligned} P(A_2 B_2 | A_1 B_1) &= \sum_{C_1, D_1} P(A_2 B_2 | C_1 D_1)P(C_1 D_1 | A_1 B_1) \\ &= \sum_{C_1, D_1} P(A_2 B_2 | C_1 D_1)P(C_1 D_1 | A_1 B_1) + \varepsilon_4 = P(A_2 | A_1)P(B_2 | B_1) + \varepsilon_5. \end{aligned}$$

Et donc, si $A \in \mathcal{F}_k^b, B \in \mathcal{F}_k^b$,

$$\begin{aligned} P(AB) &= P(A_2 B_2 | A_1 B_1)P(A_1 B_1) = P(A_2 | A_1)P(B_2 | B_1)P(A_1)P(B_1) + \varepsilon \\ &= P(A)P(B) + \varepsilon. \end{aligned}$$

2. Prouvons maintenant la propriété d' α -mélange à chaque date:

Si I est un réseau plan à maille carrée, l' α -mélange conditionnellement à la date précédente peut être imposée par les conditions de Simon, soit à chaque ronde k ,

$$\sum_{k' \leq k} |\beta(k', k)| < 1.$$

Si I est porté par une ligne, cette inégalité se déduit du théorème de Frobénius. Regardons le cas où U et V sont réduits à un point. A la première date, soit W la matrice telle que $w_{i,j} = \exp[\alpha(i - 1) + \beta(i - 1, j - 1)]$ avec $\alpha(0) = 0$ et $\beta(i, j) = 0$ si i ou j est nul. Alors

$$P(X_0 = x_0, X_d = x_d | X_{-n} = x_{-n}, X_{d+m} = x_{d+m}) = \frac{(W^n)(x_{-n}, x_0)(W^d)(x_0, x_d)(W^m)(x_d, x_{d+m})}{(W^{d+m+n})(x_{-n}, x_{d+m})}.$$

D'après le théorème de Frobénius, W étant positive, W^n converge vers $\lambda^n l.r'$ où l et r sont les premiers vecteurs propres à gauche et à droite de W et λ la valeur propre associée. De plus, la matrice étant strictement positive, on a (Karlin, 1969):

$$\left| \frac{W^n}{\lambda^n} - l.r' \right| < \left(\frac{\rho}{\lambda} \right)^n \text{ pour } \rho < \lambda$$

et donc

$$\begin{aligned} P(X_0 = x_0, X_d = x_d) &= \lim_{n \rightarrow \infty, m \rightarrow \infty} P(X_0 = x_0, X_d = x_d | X_{-n} = x_{-n}, X_{d+m} = x_m) \\ &= \frac{l(x_0)(W^d)(x_0, x_d)r(x_d)}{\lambda^d} \end{aligned}$$

et donc, quand $d \rightarrow \infty$,

$$|P(X_0, X_d) - P(X_0)P(X_d)| \leq l(x_0)l(x_d)r(x_0)r(x_d)\left(\frac{\rho}{\lambda}\right)^d$$

d'où l' α -mélange à cette date.

Aux dates suivantes, si nous considérons deux points séparés de d , soit il existe un point entre eux déjà attaqué à la date précédente et ces points sont indépendants conditionnellement à la date précédente, soit on construit une inégalité du même type que la précédente. L' α -mélange est donc toujours vérifiée si $\nu = 1$.

ANNEXE 2

Soit Θ l'espace paramétrique correspondant à l'hypothèse nulle, $\theta \in \Theta$. $\theta = (\theta_1, \dots, \theta_n)$ où θ_i est un ensemble de paramètres relatifs à la date i , $\theta_i \in \Theta_i$. Soit $\Theta^{(0)}$ l'espace paramétrique correspondant à la sous-hypothèse. Supposons que l'on a de plus:

$$\{\theta \in \Theta^{(0)}\} \Leftrightarrow \{\forall i, \theta_i \in \Theta_i^{(0)}\}. \quad (1)$$

Le test utilisé est celui du rapport de pseudo-vraisemblance donné par

$$T = 2(\log F(\hat{\theta}^{(0)}) - \log F(\hat{\theta}))$$

où $F(\theta)$ est le produit des probabilités conditionnelles sur le codage et l'ensemble des dates, $\hat{\theta} = \operatorname{argmin}_{\theta \in \Theta} F(\theta)$ et $\hat{\theta}^{(0)} = \operatorname{argmin}_{\theta \in \Theta^{(0)}} F(\theta)$.

D'après les résultats de Guyon, T tend vers un χ^2 de degrés de liberté la différence entre les dimensions des espaces Θ et $\Theta^{(0)}$.

De plus, $F(\theta) = \prod_i F_i(\theta_i)$ où F_i est le produit à la date i des probabilités conditionnelles sur le codage. On a également

$$T_i = 2(\log F_i(\hat{\theta}_i^{(0)}) - \log F_i(\hat{\theta}_i))$$

où $\hat{\theta}_i = \operatorname{argmin}_{\theta_i \in \Theta_i} \log F_i(\theta_i)$ et $\hat{\theta}_i^{(0)} = \operatorname{argmin}_{\theta_i \in \Theta_i^{(0)}} \log F_i(\theta_i)$ tend vers un χ^2 de degrés de liberté la différence des dimensions paramétriques entre Θ_i et $\Theta_i^{(0)}$.

D'après (1), $\hat{\theta}^{(0)}$ le minimum de F sur $\Theta^{(0)}$ vérifie $\hat{\theta}^{(0)} = (\hat{\theta}_1^{(0)} \dots \hat{\theta}_n^{(0)})$ où $\hat{\theta}_i^{(0)}$ est le minimum sur $\Theta_i^{(0)}$ de F_i . De même $\hat{\theta} = (\hat{\theta}_i)$ et donc les T_i sont indépendants. On a donc $T = \sum_i T_i$ est la somme de χ^2 indépendants.

