

# Processus de Cox marqué dirigé par un environnement prédit : application à la répartition spatiale de juvéniles en forêt tropicale humide

Pierrette Chagneau, Frédéric Mortier, Nicolas Picard, Jean-Noël Bacro

## ► To cite this version:

Pierrette Chagneau, Frédéric Mortier, Nicolas Picard, Jean-Noël Bacro. Processus de Cox marqué dirigé par un environnement prédit : application à la répartition spatiale de juvéniles en forêt tropicale humide. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386734>

**HAL Id: inria-00386734**

**<https://hal.inria.fr/inria-00386734>**

Submitted on 22 May 2009

**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.

# PROCESSUS DE COX MARQUÉ DIRIGÉ PAR UN ENVIRONNEMENT PRÉDIT : APPLICATION À LA RÉPARTITION SPATIALE DE JUVÉNILES EN FORÊT TROPICALE HUMIDE

Pierrette Chagneau<sup>1,4</sup> & Frédéric Mortier<sup>2</sup> & Nicolas Picard<sup>3</sup> & Jean-Noël Bacro<sup>4</sup>

<sup>1</sup>*CIRAD, UR Dynamique des forêts naturelles, TA C-37—D,*

*Campus international de Baillarguet, 34 398 Montpellier CEDEX 5, France*

<sup>2</sup>*CIRAD, UR Diversité génétique et amélioration des espèces forestières, TA A-39—C,*

*Campus international de Baillarguet, 34 398 Montpellier CEDEX 5, France*

<sup>3</sup>*CIRAD, UR Dynamique des forêts naturelles, Libreville, Gabon*

<sup>4</sup>*ISM, UMR CNRS 5149, Université de Montpellier 2,*

*CC 51, Place Eugène Bataillon, 34 095 Montpellier CEDEX 5, France*

## Résumé

Un des points faibles des modèles de dynamique forestière spatialement explicites est la modélisation du recrutement. Un inventaire détaillé du peuplement et des conditions environnementales a permis de mettre en évidence les effets de ces deux facteurs sur la densité locale de juvéniles. Mais en pratique, la collecte de telles données est coûteuse et ne peut être réalisée à grande échelle : seule une partie des juvéniles est échantillonnée et l'environnement n'est connu que partiellement. L'objectif est ici de proposer une approche pour prédire la répartition spatiale et le génotype des juvéniles sur la base d'un échantillonnage raisonnable des juvéniles, des adultes et de l'environnement. La position des juvéniles est considérée comme la réalisation d'un processus ponctuel marqué, les marques étant constituées par les génotypes. L'intensité du processus traduit les mécanismes de dispersion à l'origine de l'organisation spatiale et de la diversité génétique des juvéniles. L'intensité dépend de la survie des graines, qui dépend elle-même des conditions environnementales. Il est donc nécessaire de prédire l'environnement sur toute la zone d'étude. L'environnement, représenté par un champ aléatoire multivarié, est prédit grâce à un modèle hiérarchique spatial capable de traiter simultanément des variables de nature différente. Contrairement aux modèles existants où les variables environnementales sont considérées comme connues, le modèle de recrutement proposé prend en compte les erreurs liées à la prédiction de l'environnement. La méthode est appliquée à la prédiction du recrutement des juvéniles en forêt tropicale (Guyane française).

## Abstract

One of the weak points of forest dynamics models is the recruitment. Classically, ecologists make the assumption that recruitment mainly depends on both spatial pattern of mature trees and environment. A detailed inventory of the stand and the environmental conditions enabled them to show the effects of these two factors on the local density of

seedlings. In practice, such information is not available : only a part of seedlings is sampled and the environment is partially observed. The aim of the paper is to propose an approach in order to predict the spatial distribution and the seedlings genotype on the basis of a reasonable sampling of seedling, mature trees and environmental conditions. The spatial pattern of the seedlings is assumed to be a realization of a marked point process. The intensity of the process is not only related to the seed and pollen dispersal but also to the sapling survival. The sapling survival depends on the environment ; so the environment must be predicted on the whole study area. The environment is characterized through spatial variables of different nature and predictions are obtained using a spatial hierarchical model. Unlike the existing models which assume the environmental covariables as exactly known, the recruitment model we propose takes into account the error related to the prediction of the environment. The prediction of seedling recruitment in tropical rainforest in French Guiana illustrates our approach.

*Mots-clés* : Recrutement des juvéniles, Modèle hiérarchique spatial, Processus ponctuel marqué.

## 1 Introduction

Les modèles de dynamique forestière reposent sur trois grands processus biologiques : la croissance, la mortalité et le recrutement. Les processus de régénération des essences forestières étant souvent mal connus, la modélisation du recrutement reste l'un des points faibles de ces modèles. Une étude visant à connaître l'influence des facteurs biotiques et abiotiques sur le processus de régénération a été menée (Flores (2005)). Flores a montré que l'installation des juvéniles dépend des conditions du milieu et de la localisation des adultes reproducteurs. Ce travail a été rendu possible grâce à un inventaire détaillé du peuplement et d'une description fine de l'environnement. En pratique, la collecte de telles données est coûteuse et ne peut pas être réalisée à grande échelle. En général, seule une partie des juvéniles a été échantillonnée et l'environnement n'est connu que partiellement. L'environnement doit donc être prédit sur les zones non échantillonnées. Cette prédiction soulève deux types de difficulté. Toutes les variables environnementales n'étant pas de même nature, les méthodes classiques de prédiction ne peuvent pas s'appliquer. De plus, les variables environnementales obtenues par prédiction sont entachées d'erreur alors qu'elles sont considérées comme connues dans le modèle existant. L'étude effectuée par Flores ne s'intéresse qu'à l'organisation spatiale des juvéniles et n'intègre aucune information génétique. Caractériser la diversité génétique des juvéniles et étudier la relation génotype-environnement pourraient contribuer à une meilleure compréhension des processus de régénération.

Notre étude consiste à prédire la répartition spatiale et le génotype des juvéniles

connaissant la localisation des adultes reproducteurs et l'environnement de manière partielle. D'un point de vue biologique, ce travail a pour but d'améliorer la compréhension des mécanismes de dispersion et de savoir quelles sont les conditions du milieu favorables à l'installation des juvéniles. D'un point de vue mathématique, un des objectifs est de proposer un modèle de recrutement qui tienne compte des variables environnementales. L'environnement est extrapolé sur tout le domaine d'étude grâce à un modèle hiérarchique spatial multivarié. Différentes approches sont envisagées pour mesurer l'impact de l'erreur de prédiction de l'environnement sur l'estimation des paramètres caractérisant le recrutement.

## 2 Modélisation du recrutement

Modéliser le recrutement consiste à caractériser l'organisation spatiale et la diversité génétique des juvéniles connaissant la position et le génotype des adultes reproducteurs ainsi que l'environnement sur tout le domaine d'étude. Il existe plusieurs modèles pour décrire les phénomènes de dispersion dans la littérature, notamment le modèle de Burzyck (2006) basé sur la notion de voisinage et celui de Shimatani (2004) utilisant les processus ponctuels. Ces deux modèles permettent d'estimer les distances de dispersion des graines et du pollen, le succès reproducteur des arbres matures, le flux de graines extérieur, mais seul le modèle de Shimatani permet de déterminer la densité locale de juvéniles. Il offre également l'avantage de mieux décrire les mécanismes biologiques à l'origine du recrutement en modélisant la survie des graines entre leur dispersion et l'installation des plantules (Shimatani *et al.* (2006)). Nous utilisons l'approche basée sur les processus ponctuels pour modéliser le recrutement, mais contrairement à Shimatani, nous intégrons les variables environnementales comme variables explicatives de la survie des graines.

### 2.1 Processus de Cox hétérogène avec prise en compte de variables environnementales

La répartition spatiale des juvéniles est considérée comme la réalisation d'un processus ponctuel marqué, les marques étant constituées par les génotypes.

Soit  $A$  la zone sur laquelle les arbres adultes vivants ont été géoréférencés et géotypés. Soit  $x$  un point de  $\mathbb{R}^2$ . Soit  $G$  le génotype d'un juvénile potentiel situé au point  $x$ . On définit d'abord un processus de Poisson hétérogène dont la fonction d'intensité  $\mu(x, G)$  s'écrit sous la forme :

$$\mu(x, G) = \lambda(x, G)S(x).$$

$\lambda(x, G)$  permet de modéliser la position et le génotype des juvéniles potentiels (graines dispersées) et  $S(x)$  la survie entre la dispersion des graines et l'installation des plantules. L'intensité  $\lambda(x, G)$  se décompose sous la forme d'une somme  $\lambda(x, G) = \lambda_A(x, G) + \lambda_{\bar{A}}(x, G)$ .

Le premier terme correspond aux juvéniles potentiels dont la mère est située dans le domaine  $A$  et le second à ceux dont la mère est en dehors de  $A$  ou déjà morte. Soit  $f$  le noyau de dispersion des graines ;  $f$  est une fonction positive telle que  $\int_0^{+\infty} f(r)2\pi r dr = 1$ . Soit  $x_j$  la position de l'arbre  $j$  et  $U_j$  son succès reproducteur. On note  $d_{x_j}$  la distance euclidienne entre le point  $x$  et l'arbre  $j$ .  $\lambda_A(x, G)$  s'écrit alors :  $\lambda_A(x, G) = \sum_{j: x_j \in A} U_j f(d_{x_j}) \mathbb{P}(G|j)$ .

$\mathbb{P}(G|j)$  est la probabilité qu'un juvénile ayant pour mère  $j$  présente le génotype  $G$ . De la même façon, on obtient l'expression de  $\lambda_{\bar{A}}(x, G)$  :

$$\lambda_{\bar{A}}(x, G) = \sum_{j: x_j \notin A} U_j f(d_{x_j}) \mathbb{P}(G|\text{ext}) = \mathbb{P}(G|\text{ext}) \sum_{j: x_j \notin A} U_j f(d_{x_j}).$$

$\mathbb{P}(G|\text{ext})$  désigne la probabilité qu'un juvénile issu d'un croisement aléatoire soit de génotype  $G$ . Les quantités  $f(d_{x_j})$ , pour  $j : x_j \notin A$ , ne peuvent pas être calculées puisque les positions des adultes à l'extérieur de  $A$  ne sont pas connues. Soit  $m = \sum_{j: x_j \notin A} U_j f(d_{x_j})$ .

Le paramètre  $m$  correspond aux phénomènes de dispersion des graines des mères situées en dehors de  $A$  et des mères potentielles déjà mortes. Finalement,  $\lambda_{\bar{A}}(x, G) = m\mathbb{P}(G|\text{ext})$  et

$$\lambda(x, G) = \sum_{j: x_j \in A} U_j f(d_{x_j}) \mathbb{P}(G|j) + m\mathbb{P}(G|\text{ext}).$$

Pour le calcul de  $\mathbb{P}(G|j)$  et  $\mathbb{P}(G|\text{ext})$ , on fait l'hypothèse que la population considérée se trouve en équilibre d'Hardy-Weinberg. En suivant un raisonnement analogue au calcul de  $\lambda_{\bar{A}}$ , on obtient l'expression de  $\mathbb{P}(G|j)$  :

$$\mathbb{P}(G|j) = \frac{\sum_{i: x_i \in A} g(d_{ji}) \mathbb{P}(G|j, i) + m' \mathbb{P}(G|j, \text{ext})}{\sum_{i: x_i \in A} g(d_{ji}) + m'} \quad \text{avec } m' = \sum_{i: x_i \notin A} g(d_{ji}).$$

$g$  désigne le noyau de dispersion du pollen.  $\mathbb{P}(G|j, i)$  est la probabilité qu'un juvénile présente le génotype  $G$  sachant qu'il a pour mère  $j$  et pour père  $i$  et  $\mathbb{P}(G|j, \text{ext})$  la probabilité qu'il présente le génotype  $G$  sachant qu'il a pour mère  $j$  et un père situé en dehors de la zone  $A$  ou déjà mort.

La survie entre la dispersion des graines et l'installation d'un juvénile dépend de l'environnement. Soit  $E(x)$  un vecteur représentant l'environnement au point  $x$ . Ce vecteur est composé de  $k$  composantes. Soient  $\gamma \in \mathbb{R}^k$  et  $\delta \in \mathbb{R}$ . La survie en  $x$  est modélisée par :

$$S(x) = \frac{e^{\delta + \gamma E(x)}}{1 + e^{\delta + \gamma E(x)}}.$$

L'intensité du processus ponctuel marqué est donc donnée par :

$$\mu(x, G) = \left[ \sum_{j: x_j \in A} U_j f(d_{x_j}) \mathbb{P}(G|j) + m\mathbb{P}(G|\text{ext}) \right] \frac{e^{\delta + \gamma E(x)}}{1 + e^{\delta + \gamma E(x)}}.$$

Si l'environnement est considéré comme connu, le processus ponctuel est un processus de Poisson hétérogène. Si  $E(x)$  est la réalisation d'un champ aléatoire, ce qui est le cas lorsque l'environnement est prédit, le processus est un processus de Cox hétérogène.

## 2.2 Prédiction de l'environnement

En général, les variables environnementales n'ont pas été échantillonnées sur toute la zone d'étude, alors que la modélisation du recrutement décrite ci-dessus, nécessite de connaître leurs valeurs à chaque emplacement potentiel d'un juvénile. L'environnement doit donc être prédit sur toute la zone. Ce dernier est modélisé par un champ aléatoire multivarié, mais les variables qui le composent ne sont pas forcément de même nature (variable gaussienne, variable de Poisson, variable ordinaire, etc). Les méthodes classiques de prédiction (cokrigage, cokrigage disjonctif) ne peuvent pas s'appliquer à ce type de problèmes. Un modèle hiérarchique spatial permettant de prédire simultanément des variables de nature différente est proposé (Chagneau *et al.* (2008)). Le principe de ce modèle est d'associer à chaque variable environnementale, quelle que soit sa nature, une variable latente gaussienne. Autrement dit, la distribution de chaque variable d'environnement  $Y_k(s)$  dépend d'une variable latente  $\beta_k(s)$ . Pour les variables gaussiennes et de Poisson, nous suivons le modèle linéaire généralisé proposé par Diggle *et al.* (1998). Pour les variables ordinaires, nous généralisons le modèle probit ordinal au cas spatial. Conditionnellement à  $\beta_k(s)$  et  $\beta_{k'}(s)$ , les variables  $Y_k(s)$  et  $Y_{k'}(s)$  sont spatialement indépendantes. Les variables latentes  $\beta_k(s)$  sont construites suivant l'approche moyenne mobile de Ver Hoef et Barry (1998) :

$$\beta_k(s) = \int_{\mathbb{R}^2} h_k(x - s) T_k(dx),$$

où  $T_k$  est un mélange de bruits blancs gaussiens et  $h_k$  une fonction moyenne mobile. Les variables  $\beta_k(s)$  sont définies à partir d'un processus latent commun ; elles sont donc dépendantes. La structure de dépendance entre ces variables a l'avantage d'être flexible. Les paramètres du modèle sont estimés grâce à un algorithme Monte Carlo par Chaînes de Markov. Pour de plus détails, voir Chagneau *et al.* (2008).

## 3 Simulations et application

Les variables environnementales interviennent dans la modélisation du recrutement en tant que variables explicatives de la survie. La valeur de ces variables aux points non échantillonnés est prédite. Classiquement, les prédictions sont considérées comme les vraies valeurs de ces variables. Pourtant, celles-ci sont entachées d'erreur et doivent avoir un impact sur l'estimation des paramètres du processus ponctuel modélisant le recrutement. Dans ce papier, plusieurs approches sont envisagées pour prendre en compte et mesurer cette incertitude.

Une première approche consiste à procéder par étapes successives. L'environnement  $\mathbf{Y}$  est prédit aux points non échantillonnés  $\mathbf{s}'$ . Il peut être considéré comme une fonction de l'environnement  $\mathbf{Y}$  aux points échantillonnés  $\mathbf{s}$  et des paramètres estimés  $\hat{\boldsymbol{\theta}}$  relatifs à l'environnement :  $\widehat{\mathbf{Y}}(\mathbf{s}') = f(\mathbf{Y}(\mathbf{s}), \hat{\boldsymbol{\theta}})$ . L'impact de l'erreur commise lors de la prédiction de  $\mathbf{Y}(\mathbf{s}')$  sur les éléments caractéristiques du processus ponctuel est illustré par des simulations.

Une seconde approche consiste à adopter un point de vue hiérarchique. Le modèle proposé pour prédire l'environnement est complété : des niveaux hiérarchiques supplémentaires sont ajoutés pour pouvoir estimer les paramètres du processus ponctuel. Il est alors possible d'estimer simultanément tous les paramètres, ceux relatifs à l'environnement et ceux relatifs aux mécanismes de dispersion. L'avantage de cette méthode est que l'erreur commise sur l'estimation des paramètres associés à l'environnement est directement prise en compte dans la procédure d'estimation des paramètres du processus ponctuel.

Les estimations obtenues par ces deux approches sont comparées à celles obtenues lorsque l'environnement est connu de manière exacte.

L'approche est appliquée à la prédiction du recrutement de l'angélique à partir de données collectées sur le dispositif expérimental de Paracou en Guyane française. Le domaine d'étude est constitué par quatre parcelles situées au sud du dispositif où les adultes et une partie des juvéniles ont été géoréférencés et géotypés. L'environnement de cette zone est caractérisé par plusieurs variables : altitude, teneur du sol en éléments chimiques, drainage du sol, etc.

## Bibliographie

- [1] Flores, O. (2005) *Déterminisme de la régénération chez quinze espèces d'arbres tropicaux en forêt guyanaise : les effets de l'environnement et de la limitation de la dispersion*. Thèse de Doctorat : Biologie des Organismes, Université de Montpellier 2, 298 p.
- [2] Diggle, P., J. Tawn et R. Moyeed (1998) Model-based geostatistics. *J. Roy. Statist. Soc. Ser. C*, 47(3), 299–350.
- [3] Ver Hoef, J. et R. Barry (1998) Constructing and fitting models for cokriging and multivariable spatial prediction. *J. Statist. Plann. Inference*, 69(2), 275–294.
- [4] Chagneau, P., F. Mortier, N. Picard et J.N. Bacro (2008) Hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. *Soumis*.
- [5] Burczyk, J., W. Adams, D. Birkes et I. Chybicki (2006) Using genetic markers to directly estimate gene flow and reproductive success parameters in plants on the basis of naturally regenerated seedlings. *Genetics*, 173, 363–372.
- [6] Shimatani, K. (2004) Spatial molecular ecological model for genotyped adults and offspring. *Ecological modelling*, 174, 401–410.
- [7] Shimatani, K., K. Kitamura, T. Kanazashi et H. Sugita (2006) Genetic inhomogeneous Poisson process describing the roles of an isolated mature tree in forest regeneration. *Population Ecology*, 48, 203–214.