

STATISTIQUES SPATIALES APPLIQUÉES EN FORESTERIE : PRÉDICTION DE VARIABLES ENVIRONNEMENTALES NON GAUSSIENNES PAR APPROCHE HIÉRARCHIQUE BAYÉSIENNE ET MODÉLISATION DE LA RÉGÉNÉRATION À L'AIDE DE PROCESSUS PONCTUELS HÉTÉROGÈNES MARQUÉS

Pierrette Chagneau

Résumé. — Un des points faibles des modèles de dynamique forestière spatialement explicites est la modélisation de la régénération. 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 juvéniles, 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 régénération proposé doit prendre en compte les erreurs liées à la prédiction de l'environnement. La méthode est appliquée à la prédiction de la régénération des juvéniles en forêt tropicale humide.

1. Introduction

Les modèles de dynamique forestière reposent sur trois grands processus biologiques : la croissance, la mortalité et la régénération. Les processus de régénération des essences forestières étant souvent mal connus, la modélisation de la régénération reste l'un des points faibles de ces modèles. Des études visant à connaître l'influence des facteurs biotiques et abiotiques sur le processus de régénération en forêt

tropicale ont montré que l'installation des juvéniles dépend des conditions du milieu, de la localisation des adultes reproducteurs et des phénomènes de compétition. Ces études ont été rendues 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. Toutes les variables environnementales n'étant pas de même nature, les méthodes classiques de prédiction ne s'appliquent pas. De plus, les variables environnementales obtenues par prédiction sont entachées d'erreur alors qu'elles sont considérées comme connues dans les modèles de régénération existants. Ces modèles de régénération ne s'intéressent, en général, qu'à l'organisation spatiale des juvéniles et n'intègre aucune information génétique, ce qui pourrait contribuer à une meilleure compréhension du processus de régénération.

Notre étude consiste à prédire la répartition spatiale et la diversité génétique 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 régénération qui tienne compte des variables environnementales et intègre de l'information génétique. L'environnement est extrapolé sur tout le domaine d'étude grâce à un modèle hiérarchique spatial multivarié.

2. Modélisation de la régénération

Le modèle de régénération proposé est capable de prendre en compte simultanément l'hétérogénéité du milieu et l'information apportée par le génotypage des individus, ce que ne permettent pas les modèles existants. Il s'agit d'une extension d'un modèle développé par Shimatani [4] basé sur les processus ponctuels. Ce modèle permet de déterminer la densité locale de juvéniles. Il offre l'avantage de mieux décrire les mécanismes biologiques à l'origine de la régénération en modélisant la survie des juvéniles [5]. Contrairement à Shimatani, nous intégrons les variables environnementales comme variables explicatives de la survie des juvéniles.

2.1. Processus de Cox marqué avec prise en compte de variables environnementales. — Soit B la zone sur laquelle les arbres adultes ont été géoréférencés et géotypés. On note \bar{B} son complémentaire. Soit A la zone d'échantillonnage exhaustif des juvéniles, $A \subseteq B$. Soit x un point de \mathbb{R}^2 . Soit G le génotype d'un juvénile potentiel situé au point x . Soit x_j la position de l'adulte j . On note $d_{x,y}$ la distance entre les points x et y .

La répartition spatiale des juvéniles est considérée comme la réalisation d'un processus ponctuel de Poisson hétérogène marqué, les marques étant constituées par les génotypes. Ce processus peut être considéré comme un processus de Poisson hétérogène multivarié. La fonction d'intensité $\mu(x, G)$ de chacun des processus de Poisson indépendants qui le composent s'écrit sous la forme : $\mu(x, G) = \lambda(x, G)\mathcal{S}(x)$. La fonction $\lambda(x, G)$ permet de modéliser la position et le génotype des juvéniles potentiels et $\mathcal{S}(x)$ leur survie. Soit g le noyau de dispersion des graines et U_j le succès reproducteur de l'adulte j . $\lambda(x, G)$ se décompose sous la forme d'une somme :

$$\lambda(x, G) = \sum_{j: x_j \in B} U_j g(d_{x; x_j}) \mathbb{P}(G|j) + \sum_{j: x_j \in \bar{B}} U_j g(d_{x; x_j}) \mathbb{P}(G|\text{ext}).$$

$\mathbb{P}(G|j)$ est la probabilité qu'un juvénile ayant pour mère j présente le génotype G et $\mathbb{P}(G|\text{ext})$ celle qu'un juvénile issu d'un croisement aléatoire soit de génotype G . Le premier terme de la somme correspond aux juvéniles potentiels dont la mère est située dans la zone B et le second à ceux dont la mère est en dehors de B . Le second terme ne peut pas être calculé puisque les positions des adultes de \bar{B} ne sont pas connues. Ce terme peut être approché en faisant l'hypothèse que la répartition spatiale des adultes reproducteurs est homogène. $\lambda(x, G)$ s'écrit alors :

$$\lambda(x, G) = \sum_{j: x_j \in B} U_j g(d_{x; x_j}) \mathbb{P}(G|j) + \bar{U} \bar{d} \beta_x \mathbb{P}(G|\text{ext}) \text{ avec } \beta_x = 1 - \int_B g(d_{x; x_j}) dx_j.$$

Le paramètre \bar{U} désigne le succès reproducteur des arbres situés hors de B et \bar{d} la densité d'adultes reproducteurs.

Le calcul de $\mathbb{P}(G|j)$ et $\mathbb{P}(G|\text{ext})$ s'effectue sous l'hypothèse que la population considérée se trouve en équilibre d'Hardy-Weinberg. Soit f le noyau de dispersion du pollen. En suivant un raisonnement analogue au calcul de $\lambda(x, G)$, c'est-à-dire en distinguant les pères issus de B des pères issus de \bar{B} , on obtient l'expression de $\mathbb{P}(G|j)$:

$$\mathbb{P}(G|j) = \frac{\sum_{i: x_i \in A} f(d_{x_j; x_i}) \mathbb{P}(G|j, i) + \eta_j \bar{d} \mathbb{P}(G|j, \text{ext})}{\sum_{i: x_i \in A} f(d_{x_j; x_i}) + \eta_j \bar{d}}.$$

$\mathbb{P}(G|j, i)$ 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 B .

La survie $\mathcal{S}(x)$ des juvéniles dépend de l'environnement. Soit $Y(x)$ un vecteur à k composantes représentant l'environnement au point x . La survie en x est modélisée par : $\mathcal{S}(x) = \frac{e^{\delta + \gamma' Y(x)}}{1 + e^{\delta + \gamma' Y(x)}}$, $\gamma \in \mathbb{R}^k$, $\delta \in \mathbb{R}$.

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

2.2. Prédiction de l'environnement. — Les variables environnementales n'ont été échantillonnées que partiellement, alors que la modélisation de la régénération nécessite de les connaître sur toute la zone A . L'environnement doit donc être prédit. Ce dernier est modélisé par un champ aléatoire multivarié, mais les variables qui le composent ne sont pas forcément de même nature (variables gaussiennes, de Poisson, ordinales, etc). Les méthodes classiques de prédiction multivariée ne s'appliquent pas. Un modèle hiérarchique spatial permettant de prédire simultanément des variables de nature différente est proposé [3]. Ce modèle est une généralisation des modèles linéaires généralisés (MLG) spatiaux [1] au cas multivarié. 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 $S_k(s)$. Cette variable S_k correspond à la composante spatiale apparaissant dans l'expression du MLG spatial associé à la variable Y_k . Conditionnellement à $S_k(s)$ et $S_m(s)$, les variables $Y_k(s)$ et $Y_m(s)$ sont spatialement indépendantes. Les variables latentes $\beta_k(s)$ sont construites suivant une approche dite moyenne mobile [2] : $S_k(s) = \int_{\mathbb{R}^2} f_k(x-s)V_k dx$, où V_k est un mélange de bruits blancs gaussiens et f_k une fonction moyenne mobile. Les variables $S_k(s)$ 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 de MCMC.

3. Application à la prédiction de la régénération de l'angélique

Les deux modèles proposés sont appliqués à la prédiction de la régénération de l'angélique (*Dicorynia guianensis*) à partir de données collectées sur le dispositif de Paracou en Guyane française. Le domaine d'étude, situé au sud du dispositif, couvre environ 40 ha. Différentes mesures ont été réalisées pour caractériser son environnement. L'ensemble des adultes reproducteurs y ont été géoréférencés et géotypés. Une petite zone au centre du domaine d'étude a fait l'objet d'un échantillonnage exhaustif des juvéniles ; ces derniers ont également été géotypés. Le modèle hiérarchique spatial multivarié a été utilisé pour prédire simultanément la pente, l'altitude et le drainage du sol (Figure 1). Les cartes de prédiction obtenues sont cohérentes avec les connaissances que nous avons sur le dispositif. La corrélation entre les variables latentes associées à l'altitude et au drainage est négative : les zones de faible altitude coïncident avec les bas-fonds. Travailler dans un cadre multivarié a permis d'améliorer la qualité des prédictions des variables ordinales. Le modèle de régénération a été testé avec différentes variables environnementales comme variables explicatives de la survie (Tableau 1). Le succès reproducteur est considéré ici comme une fonction du diamètre D : $U_j = e^{bD_j}$ et $\bar{U} = e^{b\bar{D}}$ où \bar{D} désigne le diamètre moyen des adultes reproducteurs. Les noyaux de dispersion du pollen et des graines sont

choisis gaussiens de matrice de variance-covariance respective $\tau_1^2 \mathbf{I}_2$ et $\tau_2^2 \mathbf{I}_2$. La mise

Environnement	b (cm ⁻¹)	τ_1 (m)	τ_2 (m)	\bar{d} (ind.m ⁻²)	δ	γ
Pente	0,069	638,99	152,72	fixée	-4,366	0,740
Altitude	0,063	643,21	147,31	fixée	-4,493	0,988
Drainage	0,080	625,40	164,82	fixée	fixée	$\gamma_1 = -1,772$ $\gamma_2 = 0,432$ $\gamma_3 = -0,861$ $\gamma_4 = -16,202$

TABLE 1. Estimation des paramètres du processus ponctuel modélisant la répartition spatiale des juvéniles. La variable environnementale utilisée dans le terme de survie est indiquée à gauche.

en œuvre du modèle de régénération ne s'est pas révélée entièrement satisfaisante dans le cas de l'angélique. La distance moyenne de dispersion estimée du pollen est comprise entre 783 et 806 m et celle des graines entre 184 et 206 m. Le nombre de juvéniles installés issus d'un adulte de 42 cm varie entre 14 et 28. Les distances de dispersion des graines et du pollen sont surestimées. Les juvéniles s'installent de préférence sur les sols présentant une pente supérieure à 7°. Leur survie sur les sols hydromorphes est quasi nulle.

Références

- [1] P. Diggle, J. Tawn, R. Moyeed, Model-based geostatistics, *J. Roy. Statist. Soc. Ser. C*, 47(3), 299-350, 1998.
- [2] J. Ver Hoef, R. Barry, Constructing and fitting models for cokriging and multivariable spatial prediction, *J. Statist. Plann. Inference*, 69(2), 275-294 1998.
- [3] P. Chagneau, F. Mortier, N. Picard, J.-N. Bacro, Hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields, *A paraître dans Biometrics*.
- [4] K. Shimatani, Spatial molecular ecological model for genotyped adults and offspring, *Ecological modelling* 174, 401-410, 2004.
- [5] K. Shimatani, K. Kitamura, T. Kanazashi, H. Sugita Genetic inhomogeneous Poisson process describing the roles of an isolated mature tree in forest regeneration, *Population Ecology* 48, 203-214, 2006.

Pierrette Chagneau

AgroParisTech, Dépt MMIP, 16 rue Claude Bernard, 75231 Paris Cedex 05.

E-mail : pierrette.chagneau@agroparistech.fr

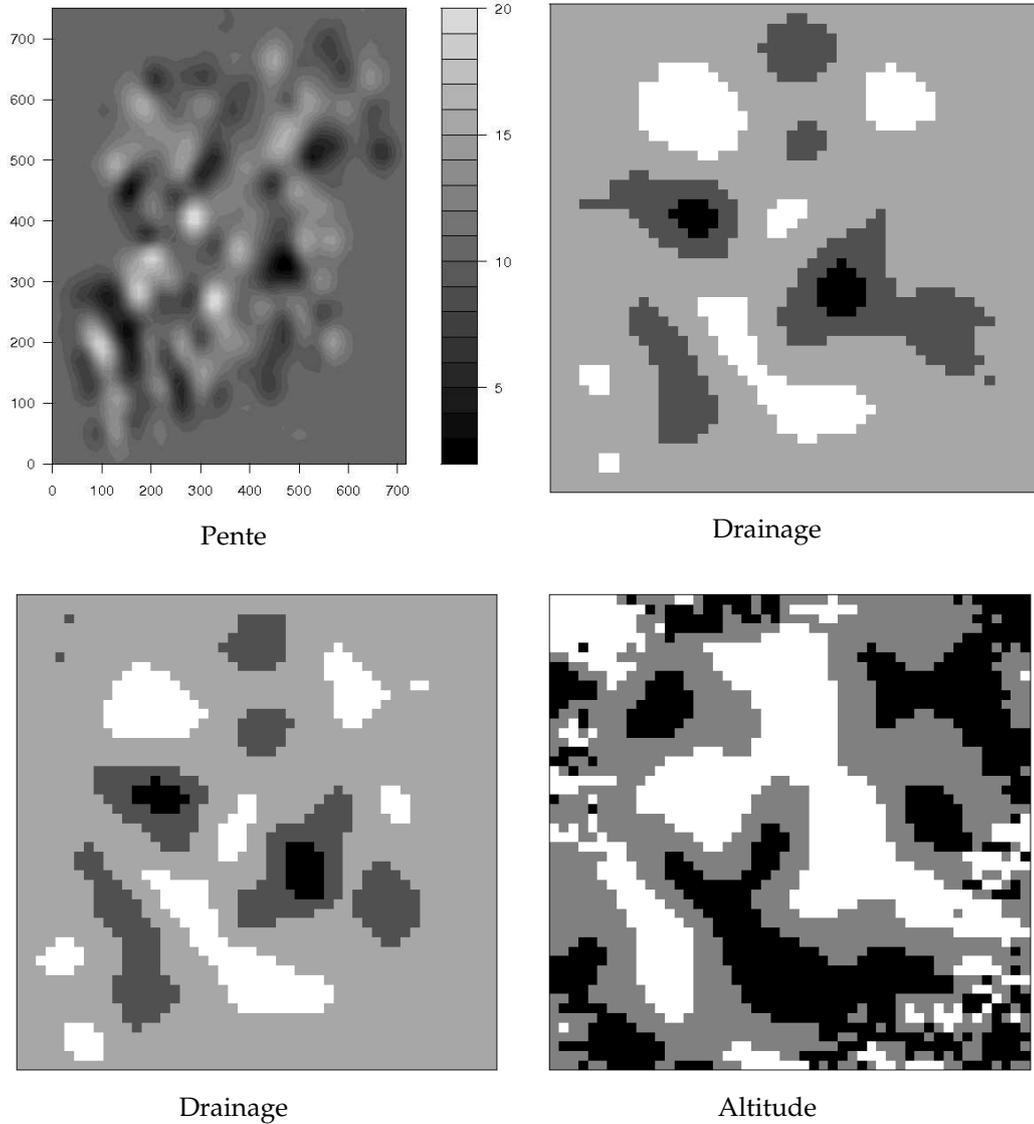


FIGURE 1. Cartes de prédiction obtenues à partir du jeu de données composé de la pente et du drainage (en haut) et du jeu de données composé du drainage et de l'altitude (en bas). La zone d'étude est un rectangle de 717 mètres sur 750. La pente est mesurée en degrés. Les modalités du drainage sont représentées suivant une échelle de gris allant du noir correspondant aux sols bien drainés au blanc correspondant aux sols hydromorphes. Les modalités de l'altitude sont représentées suivant une échelle de gris allant du noir correspondant aux altitudes faibles au blanc correspondant aux altitudes élevées.