

Inférence d’une dynamique d’usage des sols par des modèles semi-markoviens *

Angelo Raherinirina[‡] Dominique Hervé[§] Rivo Rakotozafy[‡]
Fabien Campillo[¶]

1 Introduction

L’étude des dynamiques d’usages des sols peut se faire à grande échelle (province, région) [1] ou à petite échelle (village, parcelle), c’est ce dernier point de vue que nous adoptons ici. À partir des historiques des usages de parcelles située sur la lisière ouest du corridor forestier de Ranomafana–Andringitra (cf. Figure 1), nous proposons des modèles d’usages des parcelles qui seront ensuite utilisés comme outils d’inférence des dynamiques existantes. Contrairement à [1], nous ne tenons compte que de l’aspect temporel et non pas de l’aspect spatial afin de comprendre l’échelle de temps sur la dynamique d’utilisations des parcelles.

Nous faisons appel aux chaînes semi-markoviennes (en temps discret) pour modéliser la dynamique d’occupation des parcelles. Nous disposons d’un jeu de données constitué de l’historique de l’usage de 131 parcelles pendant 50 ans, de 1956 à 2006 (cf. Figure 1).

Toutes les parcelles partent de l’état “forêt naturelle”, nous restent indéfiniment dans cet état (état absorbant), nous écarterons donc également cet état. Il reste alors $P = 112$ parcelles et 5 états possibles : culture annuelle (Ca), jachère (Jc), herbe (He), forêt secondaire (Fs) et culture pérenne (Cp).

2 Description du modèle

Considérons un processus en temps discret $(X_n)_{n \in \mathbb{N}}$ à valeurs dans un espace fini E .

Soit $(T_k)_{k \in \mathbb{N}}$ la suite des instants de changement d’état, $(Y_k)_{k \in \mathbb{N}}$ la suite des états successivement visités par $(X_n)_{n \in \mathbb{N}}$, avec $Y_k = X_{T_k}$, et $(S_k)_{k \in \mathbb{N}}$ la suite des temps de séjour, avec $S_0 = 0$ et $S_k = T_k - T_{k-1}$.

*Ces travaux ont été partiellement soutenus par le “Laboratoire International de Recherche en Informatique et Mathématiques Appliquées” (LIRIMA) et INRIA.

[†]Correspondant, Email : angelo_raherinirina@yahoo.fr

[‡]Université de Fianarantsoa, Madagascar

[§]IRD, MEM (IRD/Université de Fianarantsoa), UMR220 GRED (IRD/Montpellier 3), Montpellier, France.

[¶]Inria, Modemic Team Projet (Inria/Inra), Montpellier, France.

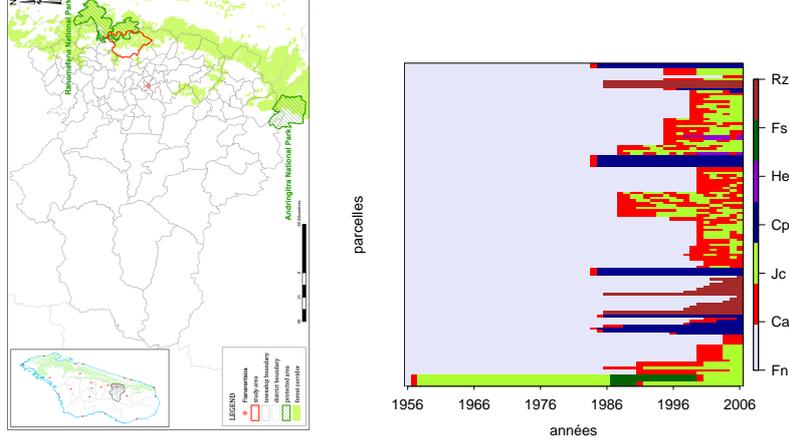


FIGURE 1 – *Jeu de données : Évolution annuelle de l’occupation des 131 parcelles sur la lisière ouest du corridor forestier de Ranomafana–Andringitra. On observe 7 états : Forêt naturelle (Fn), Culture annuelle (Ca), Jachère (Jc), Culture pérenne (Cp), Herbe (He), Rizière (Rz) et Forêt secondaire (Fs).*

On suppose que le couple $(Y_k, T_k)_{k \in \mathbb{N}}$ est une chaîne de renouvellement markovienne, c’est-à-dire :

$$\begin{aligned} \mathbb{P}(Y_k = i_k, T_k - T_{k-1} = s_k | Y_{0:k-1} = i_{0:k-1}, T_{0:k-1} = t_{0:k-1}) \\ = \mathbb{P}(X_k = i_k, T_k - T_{k-1} = s_k | Y_{k-1} = i_{k-1}). \end{aligned} \quad (1)$$

pour tout $k \geq 1$, $i_0, \dots, i_k \in E$, $0 = t_0 < t_1 < \dots < t_{k-1} < t_k \in \mathbb{N}$ et $S_k \geq 1$. Le processus $(X_n)_{n \in \mathbb{N}}$ définit une chaîne semi-markovienne. De même le processus $(Y_k)_{k \in \mathbb{N}}$ est une chaîne de Markov homogène, appelée chaîne incluse.

Notons $\mathcal{M}_E(\mathbb{N})$ l’ensemble des applications de \mathbb{N} dans \mathcal{M}_E où \mathcal{M}_E est l’ensemble des matrices carrées dans $E \times E$. On définit le noyau semi-markovien $q \in \mathcal{M}_E(\mathbb{N})$ par $q_{ij}(n) = \mathbb{P}(Y_k = j, T_k - T_{k-1} = n | Y_{k-1} = i)$.

Le noyau semi-markovien vérifie

$$q_{ij}(n) = Q_{ij} f_{ij}(n), \quad \forall n > 0, \text{ pour } i \neq j, \quad (2)$$

où, $Q_{ij} = \mathbb{P}(Y_{k+1} = j | Y_k = i)$, $i, j \in E$, $k \in \mathbb{N}$ la matrice de transition de la chaîne incluse et $f_{ij}(n) = \mathbb{P}(S_{k+1} = n | Y_k = i, Y_{k+1} = j)$ la loi conditionnelle des temps de séjour. On en déduit que la loi du processus semi-markoviens $(X_n)_{n \in \mathbb{N}}$ est entièrement caractérisée Q et f .

Nous supposons que la dynamiques des parcelles sont indépendantes et qu’elles correspondent toutes à une chaîne semi-markovienne $(X_n)_{n \in \mathbb{N}}$. Soit $(Y_k)_{k \in \mathbb{N}}$ la chaîne de Markov incluse de matrice de transition Q et f la loi conditionnelle des temps de séjour.

3 Inférence des modèles

Dans les conditions citées ci-dessus, la fonction de vraisemblance associé au modèle s’écrit :

$$L(Q, f) = \prod_{p=1}^P \prod_{k=1}^{M^{(p)}} \prod_{i,j \in E} [Q_{ij}]^{n_{ij}^{(p)}} [f_{ij}(k)]^{n_{ij}^{(p)}(k)} \quad (3)$$

où $M^{(p)}$ le nombre des changements d'état observés pour la parcelle p et $n_{ij}^{(p)}$ le nombre de transition entre deux états i et j .

3.1 Modèles absorbants déduits de la méthode empirique et la méthode de maximum de vraisemblance

L'estimateur empirique et l'estimateur du maximum de vraisemblance de la matrice de transition Q coïncident. En effet ces deux estimateurs se basent sur le nombre des transitions observées entre les états. Dans ce cas, la chaîne incluse $(Y_k)_{k \in \mathbb{N}}$ correspond à une chaîne de Markov absorbante.

Il y a une légère différence entre les deux méthodes au niveau de l'estimation de la loi conditionnelle des temps de séjour, mais les résultats sont toutefois cohérents. Ainsi Les deux modèles semi-markoviens déduits de l'estimateur empirique et de l'estimateur du maximum de vraisemblance sont quasiment identiques. En temps long, les processus sont absorbés par l'état "culture pérenne" (C_p) (cf. Figure 2).

3.2 Modèle régulier déduit de l'estimateur bayésien

Nous avons repris la technique utilisée dans [3] pour calculer l'estimateur bayésien de la matrice de transition Q , à l'aide d'une méthode de Monte Carlo par chaîne de Markov (MCMC). Nous avons utilisé cette même technique pour calculer les estimateurs des paramètres des lois des temps de séjour. Dans les deux cas, la loi a priori utilisée est la loi non informative de Jeffreys.

Avec les estimateurs bayésiens de Q et f , la chaîne de Markov incluse est irréductible et les temps de séjour sur les états ont tous une moyenne finie. On obtient donc un modèle semi-markovien régulier admettant une loi limite [2] (cf. Figure 2).

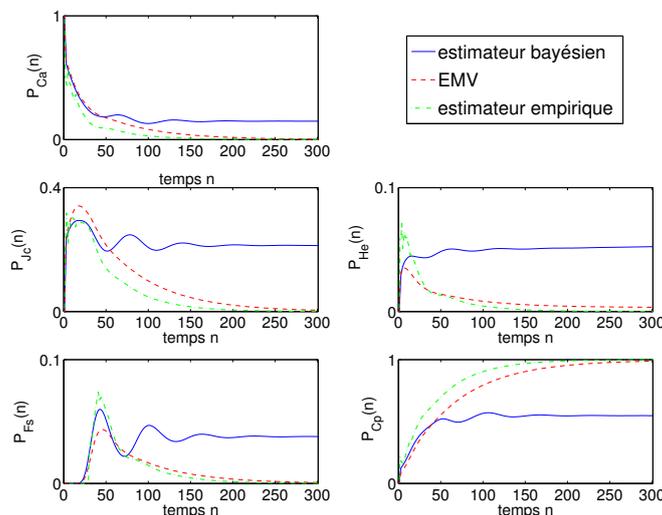


FIGURE 2 – Comparaison des lois de transition semi-markoviennes estimées selon les trois modèles déduits de l'estimateur empirique, de l'estimateur du maximum de vraisemblance et de l'estimateur bayésien.

4 Discussion

Les trois méthodes d'estimation utilisées conduisent à trois modèles semi-markoviens différents. Les modèles déduits de l'estimateur empirique et de l'estimateur du maximum de vraisemblance sont absorbants. Celui obtenu par la méthode bayésienne est régulier et admet une loi limite qui correspond aux proportions des occupations des parcelles à l'équilibre. À court terme, ces trois modèles donnent des résultats semblables mais les différences sont nettes à partir de la 50^{ième} année d'observation, voir la Figure 2.

Si on considère la proportion d'usages des parcelles hormis l'état absorbant (C_p), celle-ci converge vers une proportion stable entre "culture annuelle" et "jachère" alors que les états "forêt secondaire" et "herbe" disparaissent. Pour le modèle régulier, les proportions d'usages des parcelles se stabilisent pour tous les états, voir la Figure 3.

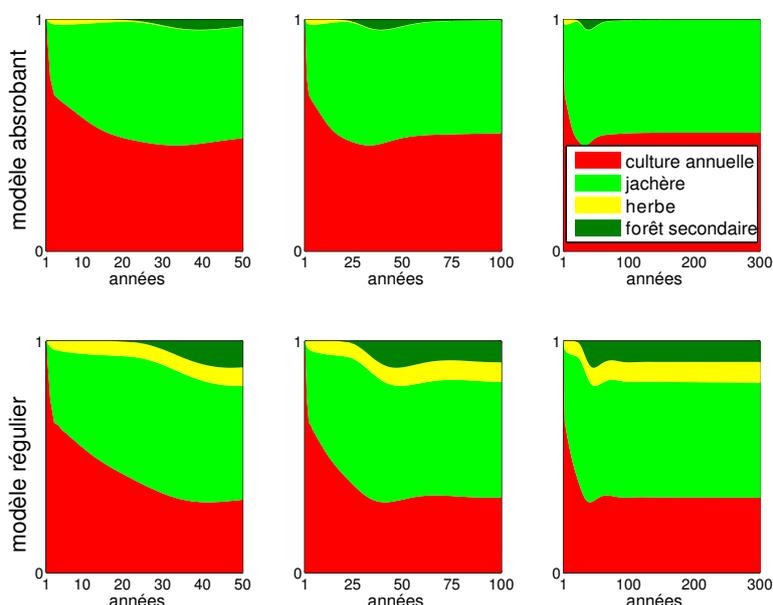


FIGURE 3 – Proportions d'usages des parcelles sans l'état (C_p) pour les modèles déduits de l'EMV et de l'estimateur bayésien.

De plus, nous remarquons que, pour les deux modèles, les proportions des parcelles en "jachère" sont presque identiques. Les sommes des proportions des parcelles sur les états "culture annuelle", "herbe" et "forêt secondaire" sont les mêmes, voir la Figure 4. Dans tous les cas, la proportion des parcelles en "jachère" atteint son équilibre vers la 50^{ième} années d'observations.

5 Conclusion

Nous avons développé une démarche de modélisation de dynamique d'usage de parcelles agricoles fondée sur les chaînes semi-markoviennes. Dans ce cadre il faut identifier la matrice de transition de la chaîne incluse ainsi que la loi des temps de séjour conditionnées par l'état suivant. L'estimateur empirique ne correspond plus exactement à l'estimateur du maximum de

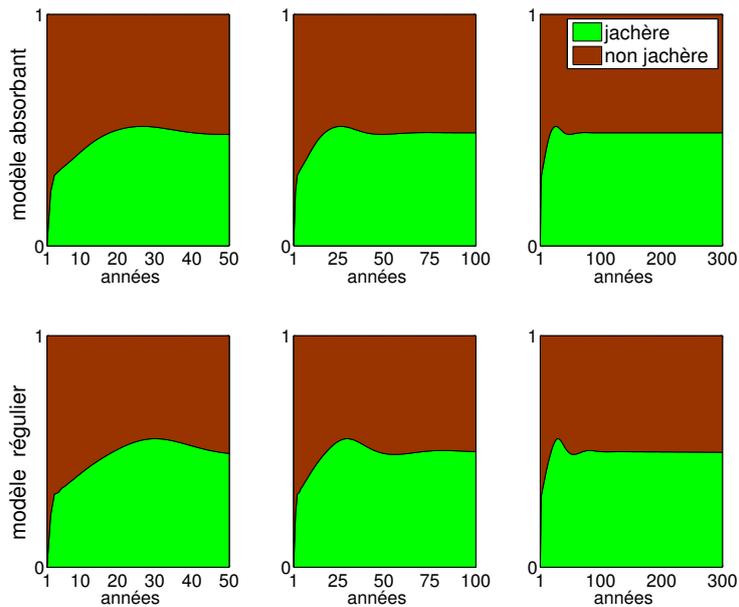


FIGURE 4 – Proportions des parcelles en état “jachère” et “non jachère” pour le modèle absorbant déduit de l’EMV et le modèle régulier déduit de l’estimateur bayésien.

vraisemblance mais les deux modèles ainsi identifiés sont très proches. L’approche bayésienne nous a permis d’intégrer des transitions non-présentes dans les données mais jugées pertinentes par les spécialistes dans les échelles de temps considérées. Ce second cas comprend notamment un état de forêt secondaire et notre approche nous a permis de quantifier les temps de retour à cet état et l’équilibre qui s’instaure entre “culture/jachère” d’une part et “forêt secondaire” d’autre part. L’équilibre concernant la forêt secondaire est atteint en environ 50 ans, celui entre culture et jachère en environ 25 ans. Les détails de cette étude sont publiées dans [4].

Références

- [1] D. K. Agarwal, J. A. Silander Jr., A. E. Gelfand, R. E. Deward, and J. G. Mickelson Jr. Tropical deforestation in Madagascar : analysis using hierarchical, spatially explicit, Bayesian regression models. *Ecological Modelling*, 185 :105–131, 2005.
- [2] V. S. Barbu, J. Bulla, and A. Maruotti. Estimation of the stationary distribution of a semi-markov chain. *Journal of Reliability and Statistical Studies*, 5 :15–26, 2012.
- [3] F. Campillo, Hervé D., A. Raheiririna, and R. Rakotozafy. Markov analysis of land use dynamics : A case study in madagascar,. In *XI Colloque Africain sur la Recherche en Informatique en Mathématiques Appliquées (CARI 2012)*, 2012.
- [4] A. Raheiririna. *Modélisation markovienne des dynamiques d’usage des sols – Cas de parcelles situées sur le bord du corridor forestier Ranomafana-Andringitra*. Thèse, Université de Fianarantsoa, à paraître.