

MODÉLISATION DE LA DYNAMIQUE VERTICALE DU CARBONE DANS LE SOL

Rana JREICH ¹ & Chritine HATTÉ ² & Éric PARENT ³

¹ CEA/LSCE, UMR 8212 GeoTrAc, F-91198, Gif-Sur-Yvette rana.jreich@lsce.ipsl.fr

² CEA/LSCE, UMR 8212 GeoTrAc, F-91198, Gif-Sur-Yvette christine.hatte@lsce.ipsl.fr

³ AgroParisTech/INRA, UMR 518 MIA, F-75005, Paris Eric.parent@agroparistech.fr

Résumé. Le groupe d'experts intergouvernemental sur l'évolution du climat, a mis en évidence une grande incertitude sur les estimations futures du carbone face au changement climatique, en particulier pour le stockage en profondeur dans les sols. La compréhension de la dynamique verticale du carbone est donc un défi majeur.

Nos données de 122 profils de carbone sont issues de nombreux articles de la littérature. Pour chaque site (profil), on disposons d'observations de $\Delta^{14}\text{C}$ à différentes profondeurs, un indicateur de la potentialité de séquestration du C par le sol. On bénéficie aussi d'autres informations relatives au site telles que : la température, le type de sol, le type d'écosystème, l'indice d'aridité...

Dans un premier temps, on modélise les profils de concentration de $\Delta^{14}\text{C}$ par un modèle hiérarchique à effet aléatoire. Pour l'inférence, on a utilisé les algorithmes MCMC afin d'évaluer les incertitudes intra et inter sites.

La deuxième partie de notre travail consiste à développer un modèle de sélection de variables. Un échantillonneur de Gibbs sur des variables indicatrices permet d'identifier la probabilité a posteriori de chacune des variables potentiellement explicatives. Cette technique de sélection nous permet de recommander de nouvelles stratégies pour augmenter la séquestration du carbone dans le sol et mitiger le réchauffement global.

Mots-clés. Modèle hiérarchique à effet aléatoire, recherche stochastique pour la sélection de variable, MCMC, carbone du sol.

Abstract. Understanding the vertical dynamics of soil carbon is a major challenge, especially since the Intergovernmental Panel on Climate Change has highlighted a large uncertainty about future carbon estimates with the climate change. Data are collected from 122 sites through literature articles. For each site (profile), we have observations of $\Delta^{14}\text{C}$ at different depths [2]. We also benefit from other information for each site such as : temperature, soil type, ecosystem type, index of aridity and others ...

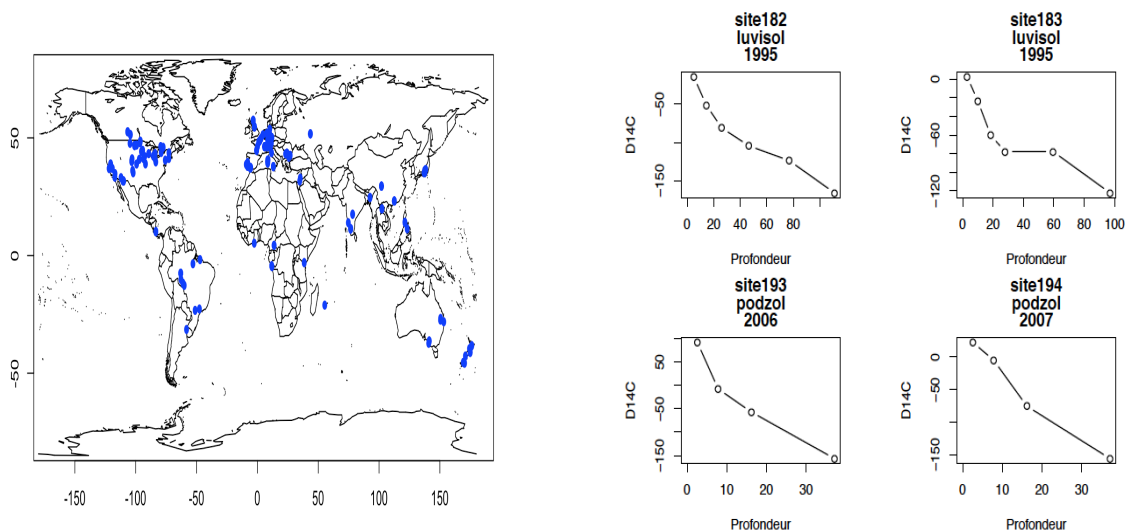
As a first step, the profiles are modeled by a hierarchical linear model with random effects. For the estimation part, MCMC algorithms are used to take into account the intra and inter-site uncertainties. The second part of our work consists in developing a variable selection model in order to identify via Gibbs Sampling the probabilistic influence of each of the potentially explanatory variables on the carbon dynamics. This technique allows

us to suggest new strategies decisions in order to increase soil carbon sequestration and reduce the global warming.

Keywords. Hierarchical Model with Random Effects, Stochastic Search Variable Selection, MCMC, soil carbon.

1 Introduction

Le carbone dans le sol constitue le plus grand réservoir terrestre en interaction avec l'atmosphère. La modélisation mécaniste de la dynamique de carbone dans le sol se heurte à la méconnaissance des façons de hiérarchiser les processus de la protection de la matière organique dans le sol et aux incertitudes sur la répartition du carbone dans les compartiments labile, lent et passif. Notre objectif est ici de proposer un modèle statistique qui nous permette de prédire la quantité de carbone dans le sol en essayant d'intégrer tous les facteurs climatiques et environnementaux potentiellement explicatifs dont on dispose. Un autre point d'intérêt est de donner un jugement probabiliste sur la contribution de chaque facteur à la dynamique du carbone. Par exemple, identifier si le passage des terres agricoles à des forêts nous permettrait de stocker plus de carbone dans le sol. Les profils de carbone de 122 sites distribués sur le globe terrestre représentent la variation de $\Delta^{14}\text{C}$ en fonction de la profondeur. On cherche à modéliser la variation intra et inter site par un modèle linéaire hiérarchique à effet aléatoire. Voici pour exemples la répartition de nos 122 sites sur un planisphère et les profils de certains sites typiques :



2 Modèle

Le modèle non linéaire multinormal à variance homogène d'occurrence locale des mesures, s'écrit de la façon hiérarchique suivante [4] :

Pour un site $s \in [1 : S]$, et pour une mesure $m_s \in [1 : n_s]$, on modélise l'évolution de $y = \Delta^{14}\text{C}$ du sol en fonction de la profondeur $x(m_s)$ par :

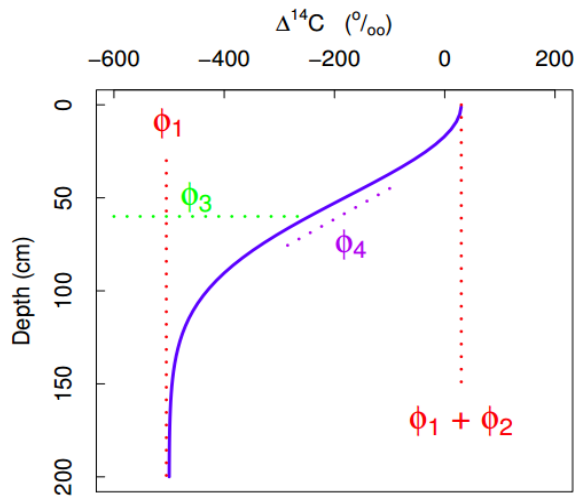
$$y(m_s) = \varphi_1(s) + \varphi_2(s) \exp - \left(\frac{x(m_s)}{\varphi_3(s)} \right)^{\varphi_4(s)} + \sigma^2 \times \varepsilon(m_s) \quad (1)$$

$$\varepsilon(m_s) \sim N(0, 1)$$

- ϕ_1 : $\Delta^{14}\text{C}$ en grande profondeur.
- $\phi_1 + \phi_2$: $\Delta^{14}\text{C}$ en surface.
- ϕ_3 : lien avec la bosse à mi-hauteur.
- ϕ_4 : décroissance plus ou moins forte.

On voit que ce modèle possède la même structure pour chaque site, mais celui-ci est non linéaire car la réponse y est une fonction non linéaire des coefficients $\varphi_1, \varphi_2, \varphi_3, \varphi_4$. Dans nos premiers essais, le bruit de mesure ε est supposé normal homogène. Pour des raisons de positivités, on considère les variables caractéristiques transformées suivantes :

$$\begin{aligned} \theta_1(s) &= \varphi_1(s) \\ \theta_2(s) &= \varphi_2(s) \\ \theta_3(s) &= \log \varphi_3(s) \\ \theta_4(s) &= \log \varphi_4(s) \end{aligned}$$



Les variables latentes $\theta_1(s), \theta_2(s), \theta_3(s)$ et $\theta_4(s)$ sont reliées par un modèle linéaire multivarié aux variables potentiellement explicatives X_1, X_2, \dots, X_p . Ce modèle peut s'écrire sous la forme matricielle suivante :

$$\Theta = X\beta + E$$

Autrement dit, en détaillant pour chaque site $s \in [1, S]$, pour chaque caractéristique $j \in [1, 4]$, on suppose une réponse multivariée normale :

$$\theta(s) = \beta^T F(s) + E(s)$$

- $\theta(s) \in \mathbb{R}^4$: vecteur colonne correspondant aux variables caractéristiques du site s .
- $\beta \in M_{(p,4)}(\mathbb{R})$.
- $F(s) \in \mathbb{R}^p$: vecteur colonne des variables explicatives relatives au site s .
- $E(s) \in \mathbb{R}^4$: effet aléatoire désignant la variabilité inter-sites de matrice 4×4 de variance-covariance Ω , $E(s) \sim N(0, \Omega)$.

3 Estimation

Les paramètres du modèle sont estimés selon l'approche bayésienne, grâce aux méthodes Markov chain Monte Carlo [1]. L'inférence de ce modèle hiérarchique à effets aléatoires est implémentée sous Jags en supposant des priors vagues pour les paramètres.

4 Sélection des variables

Dans un deuxième temps, on rajoute une couche supplémentaire de variables indicatrices au modèle afin de calculer la probabilité de sélection a posteriori de chacun des prédicteurs. On a recours à la technique de sélection de variables via l'échantillonneur de Gibbs proposé par George et McCulloch [3].

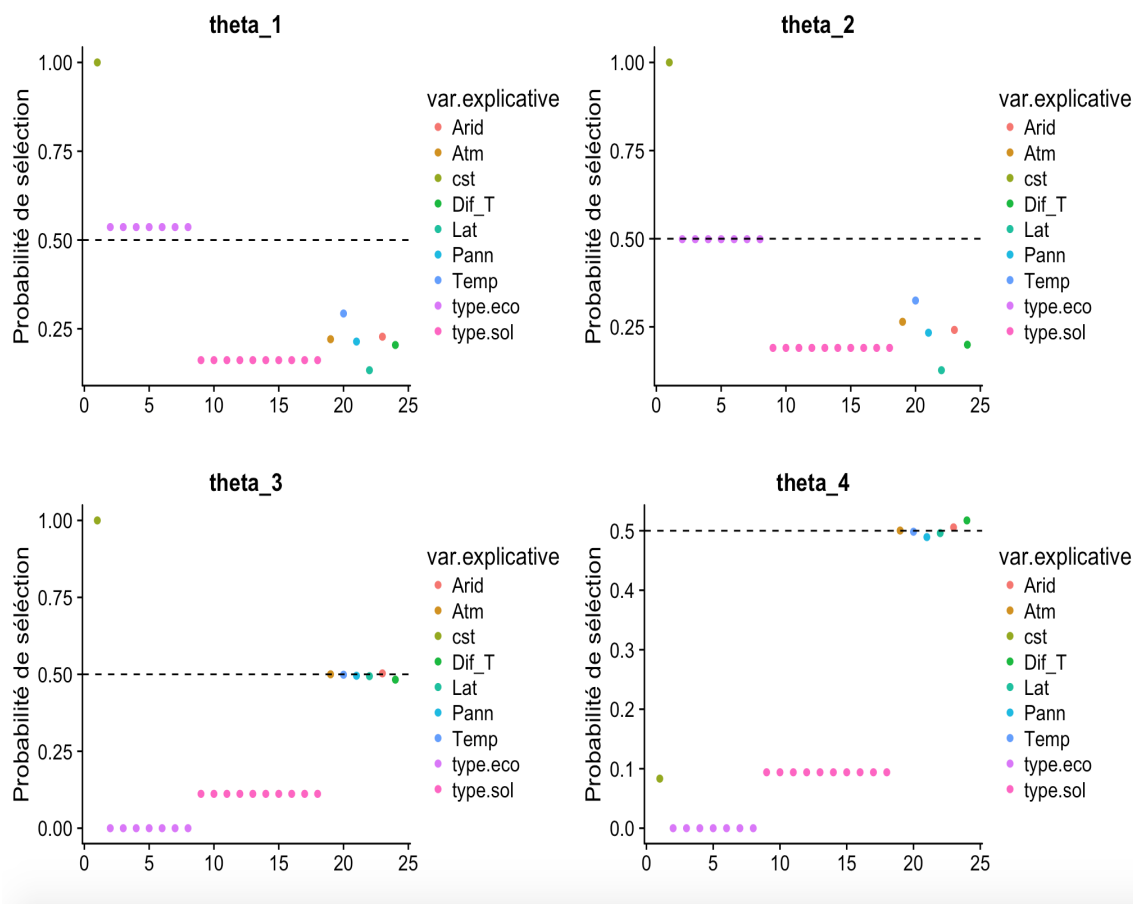
Cette technique revient à imaginer un modèle de mélange Gaussien a priori pour chacun des termes de la matrice $\beta \in M_{(p,4)}(\mathbb{R})$, relatifs aux coefficients de la variable numéro $i = 1, \dots, p$ potentiellement explicative du paramètre $j = 1, 2, 3, 4$:

$$\beta(i, j) / \lambda(i, j) \sim (1 - \lambda(i, j))N(0, \tau_i) + \lambda(i, j)N(0, c_i^2 \tau_i^2)$$

$$\lambda(i, j) = \begin{cases} 1 & \text{si la variable X(i,j) est sélectionnée} \\ 0 & \text{sinon} \end{cases}$$

$$1 - P(\lambda(i, j) = 0) = \pi_i$$

τ_i doit être très petit de telle sorte que si la variable binaire $\lambda(i, j)$ est nulle, on a tendance à tirer une valeur de $\beta(i, j)$ proche de 0, autrement dit, la variable X(i,j) n'aura pas d'influence sur notre réponse. La figure ci-après illustre les probabilités de sélection ainsi obtenues pour les $p = 9$ variables explicatives sur chacune des 4 variables latentes.



Références

- [1] Gerhard Arminger and Bengt O. Muthén. A Bayesian Approach To Nonlinear Latent Variable Models Using The Gibbs Sampler And The Metropolis Hastings Algorithm. *Psychometrika-Vol.63, No.3, 271-300*, Sep., 1998.
- [2] Jérôme Balesdent and Bernard Guillet. Les datations par le carbone 14 des matières organiques des sols. *Association française pour l'étude du sol - www.afes.fr*, 2010.
- [3] Edward I. George and Robert E. McCulloch. Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association, Vol.88, No.423, 881-889*, Sep., 1993.
- [4] Jordane Mathieu, Christine Hatté, Jerome Balesdent, and Eric Parent. Deep soil carbon dynamics are driven more by soil type than by climate : A worldwide meta-analysis of radiocarbon profiles. *Global Change Biology*, June 2015.