



HAL
open science

Modélisation spatiale multi-sources du carbone organique dans le sol

Mounia Zaouche, Liliane Bel, Jessica Tressou, Emmanuelle Vaudour

► To cite this version:

Mounia Zaouche, Liliane Bel, Jessica Tressou, Emmanuelle Vaudour. Modélisation spatiale multi-sources du carbone organique dans le sol. 50èmes Journées de Statistique, May 2018, Paris Saclay, France. hal-01814121

HAL Id: hal-01814121

<https://hal.science/hal-01814121>

Submitted on 5 Jun 2020

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.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

MODÉLISATION SPATIALE MULTI-SOURCES DU CARBONE ORGANIQUE DANS LE SOL

Mounia Zaouche ¹ & Liliane Bel ¹ & Jessica Tressou ¹ & Emmanuelle Vaudour ²

¹ *UMR MIA-Paris, AgroParistech, INRA, Université Paris-Saclay, 75005, Paris, France. E-mail: mounia.zaouche@agroparistech.fr, liliane.bel@agroparistech.fr,*

jessica.tressou@agroparistech.fr

² *UMR ECOSYS, Agroparistech, INRA, Université Paris-Saclay, 78850, Thiverval-Grignon, France. E-mail: emmanuelle.vaudour@agroparistech.fr*

Résumé. La réduction des intrants chimiques d'usage agricole et la valorisation des matières organiques d'origine urbaine en substitution nécessitent une connaissance fine des propriétés des sols agricoles. Le carbone étant un bon indicateur de la fertilité des sols, il s'avère nécessaire de disposer d'une cartographie précise des teneurs. Dans cette étude, on cherche à produire une cartographie des teneurs en carbone dans la plaine de Versailles et le plateau des Alluets, région agricole de 220 km². La nouvelle approche d'inférence Bayésienne Integrated Nested Laplace Approximation et Stochastic Partial Differential Equation (INLA-SPDE) nous permet de mettre en cohérence les différentes sources d'information disponibles (prélèvements au sol et image satellite) et d'obtenir en peu de temps des estimations a posteriori des paramètres et du champ de carbone considéré comme un champ latent. Nous avons évalué et comparé via une procédure bootstrap les performances plusieurs modèles utilisant comme covariable l'altitude issue d'un modèle numérique de terrain incluant ou non les données des images satellites. L'intégration de l'image améliore légèrement la qualité de prédiction en terme de RMSE (Root Mean Square Error) dont la médiane passe de 3.17 g.kg⁻¹ à 3.15 g.kg⁻¹. Mais surtout la carte de prédiction du carbone issue de la modélisation jointe présente de façon beaucoup plus réaliste la structure spatiale du carbone.

Mots-clés. Statistique spatiale, INLA, SPDE, carbone, sol, modélisation jointe.

Abstract. In order to reduce chemical fertilizer in agricultural use and to value the urban organic matter in substitution, a precise knowledge of soil properties is mandatory. Carbon is a good indicator of soil fertility and having at our disposal a precise mapping of its content is useful. In this study we aim at spatially estimate the soil carbon content (SOC) in the Versailles plain and the Alluets plateau, a 220 km² agricultural area. The novel Bayesian inference approach called Integrated Nested Laplace Approximation with Stochastic Partial Differential Equation (INLA-SPDE) allows us to ensure consistency between the various available sources of information (soil samples and satellite image) and to produce in a short time a posteriori estimations of the parameters and the SOC field, considered as a latent field. Several models were evaluated and compared using

the elevation covariate stemming from a Digital Elevation Model (DEM), including or not the data from the satellite image. Adding the image slightly improves the prediction quality in terms of RMSE (Root Mean Square Error RMSE) since the median goes from 3.17 g.kg^{-1} to 3.15 g.kg^{-1} . Overall the carbon prediction map from the joint model represents more realistically the spatial structure of the carbon field.

Keywords. Spatial statistics, INLA, SPDE, SOC, joint modelling.

1 Introduction

Les préoccupations actuelles pour la valorisation des matières organiques d'origine urbaine susceptibles de pallier à la disparition des sources traditionnelles de matières organiques résiduelles amènent à étudier de manière fine les propriétés des sols. Le carbone étant un bon indicateur de la fertilité des sols, car contenu par des matières organiques, une application raisonnée des pratiques passe par une bonne connaissance de la cartographie des teneurs en carbone des sols. Le but de cette étude est de réaliser une cartographie des teneurs en matière organique à l'échelle d'une petite région agricole (220 km^2), la plaine de Versailles et le plateau des Alluets en Ile de France. Les données à disposition sont de nature et d'échelle très différentes : analyse chimique de prélèvements effectués dans les parcelles au cours de plusieurs années, mesures de réflectance réalisées au niveau de la parcelle et diverses images satellites obtenues sur la région. Ces diverses mesures ne sont que partiellement synchrones. Une première étude a permis de produire des cartes de teneur en carbone, par des techniques géostatistiques à partir des seuls prélèvements au sol (Zaouche et al., 2017). Ces mesures étant très coûteuses, elles ne peuvent être qu'en nombre limité (quelques dizaines de relevés par an) et ne suffisent pas à apporter une idée précise des variations locales à cette échelle. Les images satellites sont obtenues à des coûts qui diminuent sensiblement et couvrent toute la région. Diverses études (Vaudour, 2013) ont montré qu'une transformation simple des réflectances (rapport entre le flux lumineux réfléchi et le flux lumineux incident) fournies par les pixels de ces images permettent d'évaluer les teneurs en carbone des sites en sol nu au moment de la prise de vue. L'approche envisagée dans ce travail est de mettre en cohérence deux sources d'information (prélèvements au sol et image satellite SPOT4) dans un cadre bayésien hiérarchique. La teneur en carbone est considérée comme un processus gaussien latent, dont la moyenne dépend d'une covariable de terrain. Etant donné la grande dimension des données que nous manipulons (plus de 100000 points), l'inférence a été réalisée en utilisant l'approche INLA-SPDE (Rue et al. (2009), Blangiardo et al. (2013)), une nouvelle méthode représentant une véritable alternative aux méthodes classiques MCMC car elle permet d'obtenir en peu de temps des estimations a posteriori précises. Pour évaluer l'apport de l'image, nous avons comparé deux modèles via une procédure bootstrap, l'un incluant uniquement les prélèvements, l'autre intégrant conjointement les deux sources d'information.

2 Données

Un total de 253 prélèvements au sol répartis dans la zone d'étude ont été collectés entre 0 et 8 cm de profondeur de 2010 à 2013 chaque année à la campagne de printemps. Par ailleurs une image SPOT4 (composée de 4 bandes spectrales) d'une résolution de 20m a été acquise en avril 2014 sur la région d'étude. Les sols nus de la Plaine de Versailles représentent environ 10% de la région d'étude soit 105,941 pixels . Les quatre bandes spectrales de l'image étant très corrélées seule la première bande a été conservée dans le modèle. La figure 1 représente pour les sites de prélèvements situés dans un pixel de l'image (149 sur 253 prélèvements), la valeur de la réflectance donnée par la première bande spectrale au pixel en fonction de la teneur en carbone dans le sol du prélèvement au site. On constate que la relation entre le carbone et la réflectance n'est clairement pas linéaire et sera modélisée par un lien exponentiel.

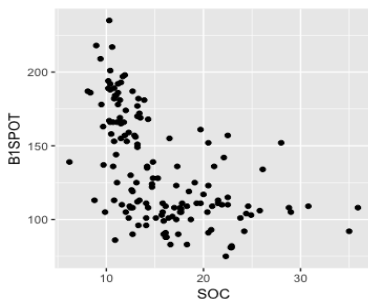


Figure 1: Relation entre la réflectance de la première bande et la teneur en carbone prélevée.

3 Modèles et Inférence

Plusieurs modèles ont été évalués et comparés. On en présente ici deux: un premier basé sur les seuls prélèvements et un second intégrant les données fournies par la première bande spectrale de l'image. Les deux modèles utilisent l'altitude comme covariable.

3.1 Modèle prélèvements seuls

Ce modèle est une simple régression spatiale sur l'altitude.

$$y_i^p \sim Normal(\beta_0 + \beta_1 E_i + \xi_i, \epsilon_i^p)$$

y_i^p est la mesure du prélèvement au site i , β_0, β_1 sont des paramètres, E_i est la valeur de l'altitude au site i , ξ est le champ spatial supposé gaussien et de fonction de covariance Matern de paramètres d'échelle κ_ξ , de régularité ν et de variance σ_ξ^2 , et ϵ^p est un bruit blanc gaussien de variance $\sigma_{\epsilon^p}^2$. Ce modèle sera noté *Elev*¹.

3.2 Modèle intégrant prélèvements et image

On considère les données des pixels de $20\text{m} \times 20\text{m}$ comme des données ponctuelles obtenues au centre du pixel. Etant donnée la nature de la relation entre les données de réflectance et les mesures de carbone (Figure 1) on suppose que les données de réflectances suivent une loi exponentielle. Les deux jeux de données sont reliés à un même effet spatial ξ et on ajoute un effet spatial supplémentaire γ à la modélisation des reflectances. Le modèle joint est donc défini par :

$$y_i^p \sim \text{Normal}(\beta_0 + \beta_1 E_i + \xi_i, \epsilon_i^p)$$

$$y_j^s \sim \text{Exponential}(\lambda_j)$$

avec $1/\lambda_j = \exp(\beta_2 + \beta_3 E_j + \beta_4 \xi_j + \gamma_j)$.

où y_i^s est la valeur de la réflectance de la première bande spectrale de SPOT4 au site i , β_3 et β_4 sont des paramètres, et γ est un champ aléatoire gaussien de covariance Matern de paramètre d'échelle ρ_γ , de régularité ν et de variance σ_γ^2 . Ce modèle sera noté *SPOTb1 – Elev*¹².

3.3 Inférence

L'algorithme INLA proposé par Rue et al. (2009) permet l'inférence des modèles hiérarchiques à processus gaussiens latents. Ces modèles se composent de trois couches : hyperparamètres θ , le vecteur des effets latents x , et modèle de vraisemblance. Si $\xi = \{\xi(s), s \in D \subseteq \mathbb{R}^2\}$ désigne un champ aléatoire de fonction de covariance de Matérn de paramètres d'échelle et de lissage κ et ν , l'objectif de l'approche SPDE est de trouver un champ gaussien markovien (GMRF) (Rue et Held, 2005) $\tilde{\xi}$ avec un voisinage local et une matrice de précision creuse, qui approxime au mieux le champ ξ . Une solution approchée de la SPDE s'obtient en divisant le domaine en un ensemble de triangles qui ne s'intersectent pas. La Figure 2 représente la triangulation de la région d'étude.

Le champ ξ est alors approché par $\tilde{\xi}(s) = \sum_{g=1}^G \tilde{A}_{sg} \tilde{\xi}_g$ où G est le nombre de sommets de la triangulation et \tilde{A}_s est le vecteur des coefficients dans la base (ξ_g) au site s . Par défaut dans R-INLA le paramètre ν est égal à 1, dans le cas du modèle *Elev*¹ on a $x = (\tilde{\xi}, \beta_0, \beta_1)$ et $\theta = (\kappa_\xi, \sigma_\xi^2, \sigma_{e_p}^2)$, dans le cas du modèle *SPOTb1 – Elev*¹² $x = (\tilde{\xi}, \tilde{\gamma}, \beta_0, \beta_1, \beta_2, \beta_3)$ et $\theta = (\beta_4, \kappa_\xi, \sigma_\xi^2, \sigma_{e_p}^2, \kappa_\gamma, \sigma_\gamma^2)$.

4 Résultats et cartes de prédiction

Pour évaluer la qualité des modèles on calcule la RMSE à partir de 183 points de calibration et 70 points de validation, obtenus par une procédure de rééchantillonnage itérée 40

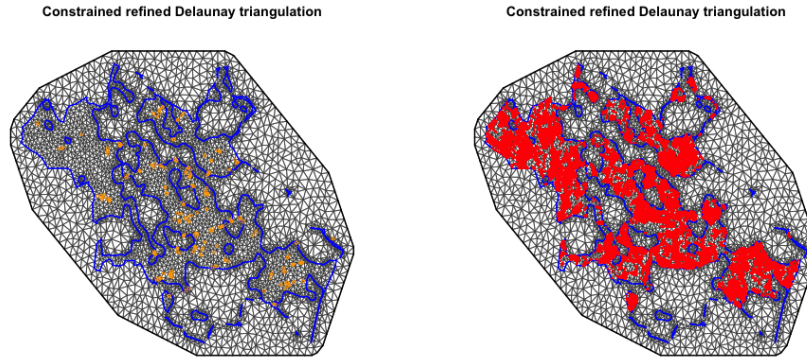


Figure 2: Triangulation de la région d'intérêt. En orange à droite les 253 prélèvements au sol. En rouge à gauche les centres de pixels en sols nus de l'image SPOT4.

fois. La (Table 1) montre que les résultats du modèle $SPOTb1 - Elev^{12}$ sont légèrement meilleurs.

Table 1: Statistiques RMSE en $g.kg^{-1}$ du Bootstrap

	mean	median	min	max
$Elev^1$	3.23	3.17	2.78	4.00
$SPOTb1 - Elev^{12}$	3.22	3.15	2.53	4.98

On constate par ailleurs que plus on dégrade la résolution de l'image SPOT, plus la RMSE augmente (Table 2).

Resolution	mean	median	min	max
400x400	3.29	3.23	2.83	3.99
100x100	3.24	3.18	2.72	4.05
40x40	3.19	3.16	2.65	4.03

Table 2: Statistiques RMSE en $g.kg^{-1}$ du Bootstrap pour $SPOT - Elev^{12}$ à des résolutions plus grossières.

Sur un ensemble de validation de 26 points collectés à la campagne de prélèvements de 2016, et un ensemble de calibration des 253 points issus des campagnes précédentes, l'écart est plus important, la RMSE passant de 3.98 ($Elev^1$) à 3.76 ($SPOTb1 - Elev^{12}$). La figure 3 représente les cartes de prédiction de la teneur en carbone associées aux deux modèles. La carte de prédiction du carbone issue de la modélisation jointe (à droite) présente des variations locales qui semblent d'après les spécialistes refléter de façon plus réaliste la structure spatiale du carbone.

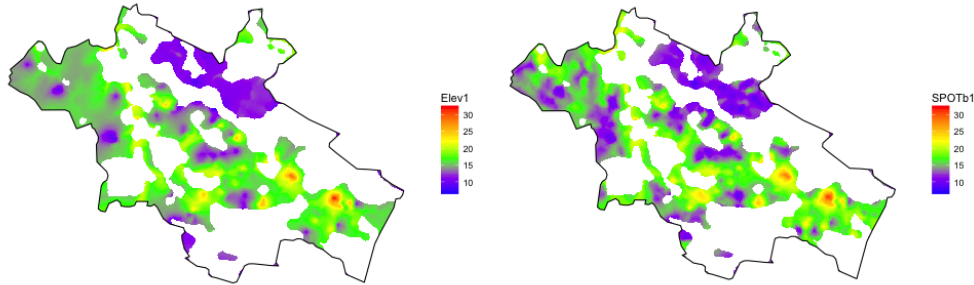


Figure 3: Cartes de prédiction du carbone. À gauche $Elev^1$, à droite $SPOTb1 - Elev^{12}$.

5 Perspectives

Le choix de la structure hiérarchique du modèle permet de combiner des observations issues de sources différentes, l'inférence étant facilitée par la méthode SPDE-INLA. D'autres sources d'information sont disponibles : des images satellites hyperspectrales, des mesures au champ prises par des drones, ... La structure du modèle doit permettre de les intégrer ce qui devrait soit affiner les prédictions, soit permettre de se passer d'une partie des prélèvements au sol. L'inférence devient dans ce cas encore plus complexe, et des procédures d'approximation supplémentaires sont à envisager.

Bibliographie

- [1] Blangiardo, M., Cameletti, M., Baio, G., Rue, H. (2013), *Spatial and spatio-temporal models with R-INLA*, Spatial and spatio-temporal epidemiology, 4, 33-49.
- [2] Rue, H., Held, L. (2005), *Gaussian Markov random fields: theory and applications*, CRC press.
- [3] Rue, H., Martino, S., Chopin, N. (2009), *Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations*, Journal of the royal statistical society: Series b (statistical methodology), 71(2), 319-392.
- [4] Vaudour, E., Bel, L., Gilliot, Coquet, Y. (2013), *Potential of SPOT multispectral satellite images for mapping topsoil organic content over a peri-urban croplands*, SSAJ, 77, 2122–2139.
- [5] Zaouche, M., Bel, L., Vaudour, E. (2017), *Geostatistical mapping of topsoil organic carbon and uncertainty assessment in Western Paris croplands (France)*, Geoderma Regional, 10, 126-137.