

APPROCHE HIÉRARCHIQUE POUR ESTIMER LA TENEUR EN CARBONE DANS LE SOL À L'ÉCHELLE D'UNE PETITE RÉGION AGRICOLE

Mounia Zaouche ¹ & Liliane Bel ² & Jessica Tressou ³ & Emmanuelle Vaudour ⁴

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

² UMR MIA-Paris, Agroparistech, INRA, Université Paris-Saclay, 75005, Paris, France. E-mail: liliane.bel@agroparistech.fr

³ UMR MIA-Paris, Agroparistech, INRA, Université Paris-Saclay, 75005, Paris, France. E-mail: jessica.tressou@agroparistech.fr

⁴ 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, nécessite 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. L'approche envisagée est le cadre bayésien hiérarchique qui doit permettre de mettre en cohérence les différentes sources d'information disponibles (prélèvements au sol et images satellites) et de produire des distributions a posteriori des paramètres et du champ de carbone, considéré comme un champ latent. Le modèle proposé suppose que toutes les variables sont gaussiennes, ce qui permet d'utiliser l'algorithme d'inférence de Gibbs, il permet d'intégrer les différences d'échelle entre les mesures et des covariables issues d'un modèle numérique de terrain. Des simulations dans un cadre simplifié donnent de bons résultats, en particulier la convergence de la plupart des paramètres et la bonne identification de la structure spatiale du champ latent.

Mots-clés. Statistique spatiale, modèle hiérarchique, changement d'échelle

Abstract. In order to reduce chemical fertilizer in agricultural use and to value the urban organic matter in substitution, the knowledge of soil properties is necessary. Carbon is a good indicator of soil fertility and disposing of 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 approach we consider is the hierarchical Bayesian framework to ensure consistency between the different sources of information (laboratory soil measurements and satellite images) and produce posterior distributions of the parameters and the SOC field, considered as a latent field.

In the model we investigated we assume all distributions are Gaussian, so that the Gibbs inference algorithm can be used, it enables integrating the different scales induced by the data and covariates related to a numerical terrain model. Simulations performed in a simplified framework give encouraging results, especially for most of the parameter convergence and the good identification of the spatial structure for the latent field.

Keywords. Spatial statistics, hierarchical modelling, scale change

1 Introduction

Les préoccupations actuelles pour la réduction des intrants chimiques d'usage agricole et 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 la disposition d'une bonne 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 (Hamiache, 2012). Ces mesures étant très coûteuses, elles ne peuvent être qu'en nombre limité (plusieurs 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 comptes numériques (CN) fournis 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 les diverses sources d'information dans un cadre bayésien hiérarchique. La teneur en carbone est considérée comme une variable latente, dont la moyenne dépend de covariables de terrain, une relation affine est envisagée entre le carbone et les CN de l'image satellite, tandis que les prélèvements au sol sont considérés comme une mesure bruitée de la teneur en carbone.

2 Modèle

On note $CO(s)$ la teneur en carbone au site s . s est un point d'une grille régulière de pas δ_x et δ_y qui recouvre la plaine de Versailles. On dispose plusieurs séries de mesures :

- n_1 mesures de type 1, notées Y_i^1 , acquises aux sites $s_{1,1}, s_{1,2}, \dots, s_{1,n_1}$ qui représentent la valeur du pixel centré en $s_{1,i}$ d'une image satellite acquise en mars 2013,
- n_2 mesures de type 2, notées Y_j^2 , acquises aux sites $s_{2,1}, s_{2,2}, \dots, s_{2,n_2}$ qui représentent les prélèvements au sol de CO effectués de 2010 à 2013.
- des covariables X issues d'un modèle numérique de terrain (altitude, pente, ...), disponibles en tout point de la grille.

On suppose que les valeurs (CN compte numérique) obtenues aux pixels de l'image satellite sont une fonction affine bruitée des teneurs en CO des quatre plus proches voisins du centre du pixel, que les prélèvements donnent une mesure bruitée des teneurs en CO et que l'espérance des teneurs en CO est une combinaison linéaire des covariables de terrain. On suppose également que les teneurs en CO sur la grille sont spatialement corrélées, avec une fonction de covariance exponentielle. En supposant la normalité des différentes variables on obtient le modèle suivant :

$$CO \sim \mathcal{N}(X\beta, \Sigma) \quad (1)$$

$$Y^1(s_1)|CO, \sigma_1 \sim \mathcal{N}(a + b\overline{CO(s_{1v})}, \sigma_1^2) \quad (2)$$

$$Y^2(s_2)|CO, \sigma_2 \sim \mathcal{N}(CO(s_2), \sigma_2^2) \quad (3)$$

$\overline{CO(s_{1v,i})}$ est la moyenne des teneurs en CO des quatre plus proches voisins de $s_{1,i}$, on suppose que la corrélation entre deux sites est donnée par un modèle exponentiel et la matrice Σ s'écrit $\Sigma = \sigma_0^2 \Sigma_0$, avec

$$\Sigma_0(s, s') = \exp(-|s - s'|/\alpha).$$

L'inférence de ce modèle peut se réaliser dans un cadre bayésien avec l'algorithme de Gibbs à partir de lois a priori classiques pour obtenir les bonnes conjuguaisons.

3 Simulations

Pour évaluer les performances de l'algorithme, on réalise une étude sur simulation dans un cadre simplifié. La grille est de taille 1252×885 ($N = 1108020$ localisations s), soit la taille réelle de l'image satellite, le champ CO simulé est de moyenne $\mu_0 = 0$, de fonction de covariance exponentielle de portée $\alpha = 10$, et les valeurs choisies pour les autres paramètres sont $a = 10$, $b = 5$, $\sigma_1 = 10$, $\sigma_2 = 2$, et $\sigma_0 = 5$.

$n_1 = 400$ et $n_2 = 300$ valeurs de Y^1 et Y^2 sont simulées suivant ces paramètres en des points de grille. Dans un premier temps on suppose que les valeurs de Y^2 sont obtenues par une transformation affine du CO au même site.

Le paramètre de portée réputé difficile à estimer dans un cadre bayésien est fixé à la valeur estimée sur le variogramme des Y^2 .

L'algorithme d'estimation a été lancé sur plusieurs chaînes de façon à permettre un contrôle de la convergence en loi (Robert, 2006), et avec plusieurs valeurs d'hyper-paramètres afin d'étudier la sensibilité aux priors.

Les résultats sont assez satisfaisants comme le montrent les figures ?? et ?? qui présentent les distributions a posteriori pour l'initialisation la plus éloignée des valeurs réelles. La convergence des paramètres vers leur vraie valeur est atteinte sauf pour σ_0 et le champ latent moyen présente les mêmes caractéristiques spatiales que celui qui a servi à initier la simulation.

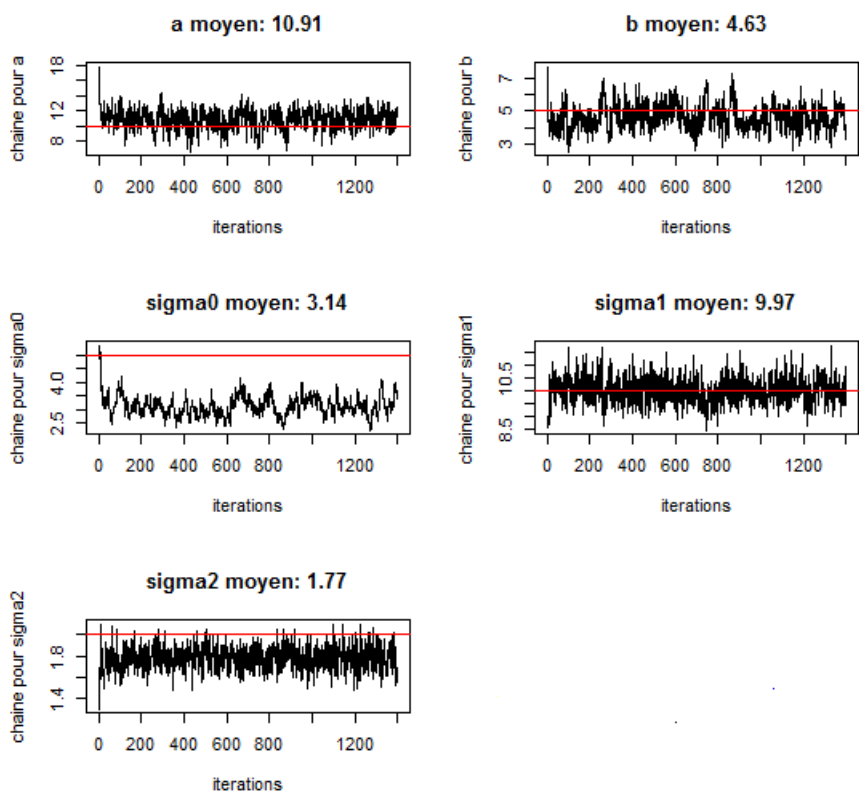


Figure 1: Distribution à posteriori des paramètres pour la chaîne 5. En rouge la vraie valeur du paramètre.

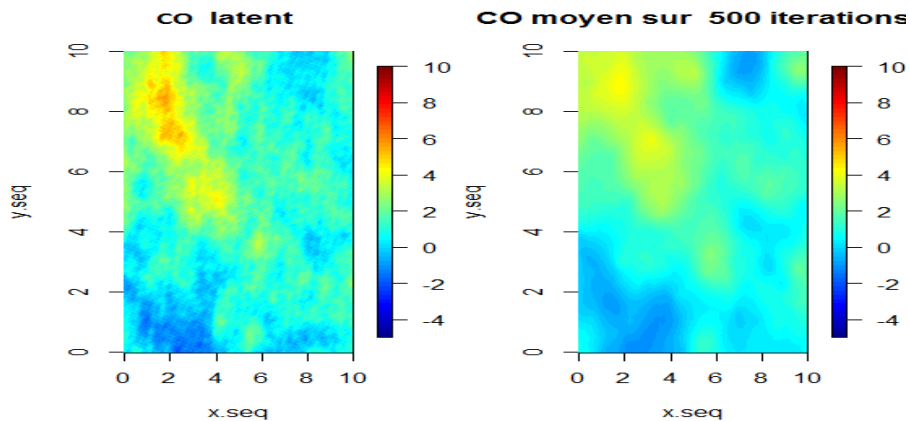


Figure 2: Comparaison du champ latent initial et du champ moyen des CO sur les 500 dernières itérations

4 Perspectives

Les premiers essais sur simulations donnent des résultats satisfaisants mais ne reflètent que partiellement la situation réelle des données à traiter. L'image satellite, bien que réduite aux seules localisations en sols nus le jour de la prise de vue contient près de 10000 pixels à convertir. Cela amène à inverser des matrices de taille 10000×10000 plusieurs fois par itérations, ce qui porte le temps d'exécution à un seuil rédhitoire.

L'intégration dans le modèle des covariables et la prise en compte de la représentativité des pixels de l'image satellite n'ont pas été effectuées dans les simulations, et introduiront certainement une complexité supplémentaire qu'il faudra gérer probablement au prix de quelques approximations.

Bibliographie

- [1] Hamiache, J., Bel, L., Gilliot, J.M., Vaudour, E. (2012), *Spatial stochastic modeling of topsoil organic carbon content over a cultivated peri-urban region, using soil properties, soil types and a digital elevation model*, Minasny, Malone, and McBratney, Eds., CRC press, Balkema, 2012, pp. 161-166.
- [2] Robert, C. (2006), *Le choix bayésien*, Springer Science & Business Media.
- [3] Vaudour, E., Bel, L., Gilliot, J.M., Coquet, Y. (2013), *Potential of SPOT multispectral satellite images for mapping topsoil organic content over a peri-urban croplands*, SSAJ, 77, 2122-2139.