
Cartographie des ETMS dans les sols de France combinant le quantile regression forest et le krigeage avec dérive externe

Nicolas Saby^{*1}, Chantal De Fouquet², Xavier Freulon³, Thomas Loiseau², and Manon Caubet⁴

¹INFOSOL – Institut national de la recherche agronomique (INRA) : US1106 – France

²Centre de Géosciences, Equipe Géostatistique – MINES ParisTech - École nationale supérieure des mines de Paris – 35 rue Saint-Honoré 77305 Fontainebleau cedex, France

³MINES ParisTech - École nationale supérieure des mines de Paris – Ministère de l’Economie, des Finances et de l’Industrie – 60, boulevard Saint-Michel 75006 Paris, France

⁴INRAE, INFOSOL – INRAE – Av Pomme de Pin 45000 Orléans, France

Résumé

La cartographie numérique des sols propose de produire une estimation spatialisée de classes ou de propriétés du sol à partir de modèles statistiques utilisant les relations quantitatives entre des variables d’intérêt, issues d’observations sur les sols, et des variables environnementales spatialement exhaustives. Ces variables, considérées comme les facteurs du processus de répartitions du phénomène observé, correspondent à des cartes géologiques, des données climatiques, des informations sur le relief, etc. Dans ce cadre, nous avons entrepris de cartographier des teneurs des sols en éléments traces métalliques et métalloïdes sur l’ensemble du territoire métropolitain à partir des observations collectées dans le cadre du Réseau de Mesures de la Qualité des Sols (RMQS). Ce réseau comporte 2200 sites de mesures répartis selon une grille systématique de 16km de côté.

Deux démarches sont ainsi comparées: un modèle de *machine learning* seul utilisant le *quantile regression forest* (QRF) et la géostatistique à travers le krigeage avec dérive externe (KED). Pour ce dernier, les prédictions issues du QRF ont été utilisées comme dérive externe. Afin de tenir compte de la forte asymétrie des distributions, une transformation Quantile Quantile a été effectuée, impliquant l’utilisation de l’espérance conditionnelle pour la transformation inverse. Nous avons fondé notre comparaison sur la base des résultats d’une validation croisée par groupe (10 groupes).

Les résultats démontrent une forte variation de la qualité des prédictions selon l’élément considéré, ainsi qu’une amélioration significative de la qualité des cartes produites en utilisant le KED. Ce résultat est cohérent avec les structures spatiales attendues des pollutions correspondant à une origine géogénique des contaminants mais également à des gradients liés aux activités humaines n’ayant pas forcément de lien déterministe simple avec les covariables utilisées (distance aux routes).

L’absence de prise en compte efficace des corrélations spatiales dans les algorithmes de *machine learning* justifie le couplage au KED.

*Intervenant