

Analyse de la diversité des sols et des micro-organismes telluriques à l'échelle d'un paysage

Approche par cartographie numérique

C. Swiderski⁽¹⁾, N.P.A. Saby⁽¹⁾, C. Ratié⁽¹⁾, C. Jolivet⁽¹⁾, D. Arrouays⁽¹⁾, S. Dequiedt⁽³⁾, et P-O. Redon⁽²⁾

1) INRA, US 1106 InfoSol, F-45075 Orléans, France

2) Andra, (Agence nationale pour la gestion des déchets radioactifs), direction Recherche et Développement, Centre de Meuse Haute Marne, Route départementale 960, F-54280 Bure, France

3) INRA, UMR 1347 Agroécologie – plateforme GenoSol, 17 rue de Sully, BP 86510, F-1065 Dijon Cedex, France

*: Auteur correspondant: Nicolas.Saby@orleans.inra.fr

RÉSUMÉ

L'Observatoire pérenne de l'environnement (OPE) de l'ANDRA a mis en œuvre depuis 2007 un réseau d'inventaires et d'observations à long terme des différents milieux de l'environnement sur un territoire de 240 km² dans les départements de la Meuse et de la Haute-Marne. Un réseau de suivi et d'observation de la qualité des sols a ainsi été mis en place selon un maillage systématique en suivant un protocole d'échantillonnage et d'analyse de type « RMQS » conduit à l'échelle du paysage sur une grille de 1,5 km x 1,5 km. Des prélèvements d'échantillons composites ont été effectués lors de 4 campagnes successives entre 2009 et 2012 sur un total de 127 sites, dont 57 disposaient en plus d'une fosse pédologique. Les propriétés physico-chimiques (texture, carbone, pH, calcaire, azote, CEC, phosphore assimilable, cations échangeables et éléments majeurs totaux) et microbiologiques (structure des communautés microbiennes) ont été analysées dans le but d'établir un état de référence des sols de la zone d'observation. Ces observations ont servi de base à une analyse statistique afin de comprendre la distribution spatiale de la diversité pédologique et des micro-organismes telluriques à l'échelle d'un paysage, et les facteurs qui la pilotent. Pour cela, nous avons eu recours aux techniques de cartographie numérique. Nous avons ainsi couplé des analyses en composantes principales (ACP) sous contraintes spatiales avec des outils de régression et de géostatistique (krigeage universel). Les ACP ont été réalisées sur trois matrices différentes regroupant respectivement les données des principales propriétés pédologiques, les communautés bactériennes et les communautés de champignons toutes deux caractérisées par les empreintes moléculaires issues des analyses de type ARISA. Nous avons ensuite tenté d'expliquer les trois premiers axes de chacune des ACP en les mettant en relation avec des covariables environnementales dont la couverture spatiale est exhaustive sur la zone d'étude. Deux algorithmes ont été testés: les arbres de régressions boostés implémentés dans la librairie « gbm » et dans la librairie Cubist du logiciel de statistique R. Les covariables environnementales retenues sont: (1) les dérivées morphométriques issues du MNA à la résolution de 25 mètres (15 variables), (2) l'occupation du sol, (3) le fond géologique (carte géologique à 1/50000) et (4) les unités cartographiques du sol issues de la carte pédologique à 1/50000. Enfin, les résidus des modèles de régression ont été interpolés par des techniques géostatiques. L'ensemble de la procédure a été validée par validation croisée.

Malgré une relative homogénéité, les propriétés physico-chimiques des sols se distribuent selon la nature géologique et pédologique de la zone ainsi que selon la géomorphologie du paysage (MNA et ses dérivées) et l'occupation du sol. La structure génétique des communautés bactériennes et fongiques présente des structures spatiales moyennement marquées à l'échelle du paysage. Les communautés bactériennes s'organisent selon la morphologie du relief (rugosité, orientation des pentes), les unités cartographiques des sols et les structures hydrologiques (distance au plus proche court d'eau). La distribution des communautés fongiques est impactée par la topographie, et les structures hydrographiques. La performance des 6 modèles construits mesurée par le R^2 varie de 0,13 à 0,94. Nos résultats confirment qu'il est possible d'identifier et de hiérarchiser les filtres environnementaux qui pilotent les diversités physico-chimiques et biologiques du sol à l'échelle d'un paysage.

Mots clés

Krigeage universel, géostatistique, réseau de suivi des sols, propriétés physico-chimiques, micro-organismes telluriques, ACP spatiale, analyse multi-variée.

SUMMARY

ANALYSIS OF SOIL AND MICRO-ORGANISMS DIVERSITY AT LANDSCAPE SCALE: approach by digital mapping

In 2007, the Perennial Observatory of the Environment (OPE) was established by ANDRA to support further research on long term evolution of the environment. As a part of its overall project, a soil monitoring and observation network was seted within a 240 km² area between Meuse and Haute-Marne. This network is based on installation and observation monitoring sites installed at regular spatial intervals, in accordance with a 1,5 km-square systematic sampling grid and following the RMQS (National Soil Quality Monitoring Network) program's sampling and analysis strategy. Composite samples have been collected on 127 sites between 2009 and 2012 and 57 soil pits have been opened. In order to define the initial state of environmental quality of the local area, we study the topsoil physical and chemical characteristics (particle-size distribution, organic carbon, pH, nitrogen, cation exchange capacity, available phosphorus, exchangeable cations and total major elements) and also some microbiological activity (fungal and bacterial structure communities analyzed through their molecular imprint by ARISA test).

We used a spatially constrained multivariate analysis combined with regression tools and geostatistics (Universal Kriging) to study the spatial distribution and correlation on topsoil characteristics at landscape scale. Digital Soil Mapping is performed on multispati-PCA's three first axis from pedological properties, fungal and bacterial communities matrix. We chose several morphological indexes extracted from digital elevation model (DEM, 25 meters squares), land use, geological and pedological maps as covariates because its cover the entire study area and we run with GBM and Cubist methods using R software. Morphometric data extracted from numeric field model (25 meters), land use, geology (map 1/50000) and map units of the soil map (1/50000) are retained as covariates. Residuals are treated with geostatistics and the entire process is approved by "leave-one-out" cross validation.

Even if the area is quite homogeneous, physical and chemical properties distribution depends on geology, pedology and morphology. Microbial communities' distribution is less pronounced within landscape but we can assess some relations with hydrological structures, reliefs (slope orientation, surface roughness) and soil map units regarding bacterial communities. The spatial distribution of fungal communities is affected by topography and hydrological structures. Correlation informs us about models performance, it ranges from 0.13 to 0.94. Thanks to this methodology, we showed that finding and prioritizing environmental pilots of physical, chemical and biological soil properties are available on landscape scale.

Key-words

Universal Kriging, geostatistics, soil monitoring network, physico-chimical properties, soil micro-organisms, spatially constrained PCA, multivariate analysis.

RESUMEN

ANÁLISIS DE LA DIVERSIDAD DE LOS SUELOS Y DE LOS MICROORGANISMOS TELÚRICOS A ESCALA DE UN PAISAJE:

Anfoque por cartografía numérica

El observatorio perene del medio ambiente (OPE) de ANDRA implementó desde 2007 una red de inventarios y de observaciones a largo plazo de los diferentes medios del medio ambiente sobre un territorio de 240 km² en los departamentos de Mosa y de Alta Marne (Francia). Se desarrolló una red de seguimiento y de observación de la calidad de los suelos según una malla sistemática siguiendo un protocolo de muestreo y de análisis de tipo "RMQS" conducido sobre una reja de 1,5 km por 1,5 km. Se realizó una recogida de muestras compuestas durante 4 campañas sucesivas entre 2009 y 2012 sobre un total de 127 sitios, de los cuales 57 tenían además

una fosa pedológica. Se analizaron las propiedades físico-químicas (textura, carbono, pH, calcáreo, nitrógeno, CIC, fosforo asimilable, cationes intercambiables y elementos mayores totales) y microbiológicos (estructura de las comunidades microbianas) con la finalidad de establecer un estado de referencia de los suelos de la zona de observación. Estas observaciones sirvieron de base a un análisis estadístico para entender la distribución espacial de la diversidad pedológica y de los microorganismos telúricos a escala de un paisaje, y los factores que la pilotan. Para eso, utilizamos las técnicas de cartografía numérica. Así acoplamos análisis de componentes principales (ACP) bajo limitaciones espaciales con herramientas de regresión y de geoestadística (kriega universal). Los ACP se realizaron sobre tres matrices diferentes que agrupan respectivamente los datos de las principales propiedades pedológicas, las comunidades bacterianas y las comunidades de hongos las dos caracterizadas por las huellas moleculares resultantes de los análisis de tipo ARISA. Luego, intentamos explicitar los tres primeros ejes de cada ACP poniéndolas en relación con covariables ambientales cuyas la cobertura espacial es exhaustiva sobre la zona de estudio. Se testaron dos algoritmos: los arboles de regresión reforzados implementados en la librería "gbm" y en la librería "Cubist" del programa de estadística R. Las covariables retenidas son: (1) las derivadas morfométricas resultantes de MNA con resolución de 25 metros (15 variables), (2) la ocupación del suelo, (3) el fondo geológico (mapa geológica al 1/50 000) y (4) las unidades cartográficas de suelo resultando del mapa pedológico al 1/50 000. En fin, se interpolaron los residuos de los modelos de regresión por técnicas geoestadísticas. Se validó el conjunto del procedimiento por validación cruzada. A pesar de una relativa homogeneidad, las propiedades físico-químicas de los suelos se distribuyen según la naturaleza geológica y pedológica de la zona así que según la geomorfología del paisaje (MNA y sus derivadas) y la ocupación del suelo. La estructura genética de las comunidades bacterianas y fúngicas presenta estructuras espaciales medianamente marcadas a escala del paisaje. Las comunidades bacterianas se organizan según la morfología del relieve (rugosidad, orientación de las pendientes), las unidades cartográficas de suelos y las estructuras hidrológicas (distancia al curso de agua más cerca). La distribución de las comunidades fúngicas está impactada por la topografía, y las estructuras hidrográficas. El rendimiento de los 6 modelos construidos medido por el R² varía de 0,13 a 0,94. Nos resultados confirman que es posible identificar y jerarquizar los filtros ambientales que pilotan las diversidades físico-químicas y biológicas del suelo a escala de un paisaje..

Palabras clave

Kriega universal, geoestadística, red de seguimiento de los suelos, propiedades físico-químicas, microorganismos telúricos, ACP espacial, análisis multivariante.

Le sol est une composante importante de l'environnement car il constitue le support direct de nombreuses activités humaines et de la plupart des écosystèmes continentaux. Il est à l'interface du sous-sol, de la biosphère, de l'atmosphère et de l'hydrosphère. Ses nombreux services et fonctions écosystémiques subissent de nombreuses pressions et se trouvent menacés principalement par les activités humaines. Les caractéristiques et les propriétés des sols évoluent sous l'effet de processus naturels (pédogénèse) et anthropiques (activités agricoles, industrielles, artificialisation...). Connaître l'état actuel du sol et se doter des moyens pour suivre ses évolutions, les comprendre et les prédire nécessite la mise en place d'un réseau d'observations et de prélèvements, d'une stratégie d'échantillonnage adaptée et d'identifier les paramètres et les indicateurs qui le caractérisent.

A la limite des départements de la Meuse et de la Haute-Marne, l'Observatoire pérenne de l'environnement (OPE) de l'Andra a mis en œuvre depuis 2007 un réseau d'inventaires et d'observations à long terme des différents milieux de l'environnement sur un territoire de 240 km². Le réseau de surveillance de la qualité des sols établi à partir d'un maillage systématique de 1,5 km sur 1,5 km se base sur le protocole du Réseau de Mesure

de la Qualité des Sols (Jolivet et al., 2006). Les analyses portent à la fois sur les propriétés physico-chimiques et sur le compartiment biologique des sols.

Comprendre et quantifier la distribution spatiale des propriétés du sol est possible par le recours aux techniques de cartographie ou de modélisation (Grunwald, 2008). Cette dernière repose sur la mise en œuvre d'équations qui simulent et prédisent des processus tandis que la cartographie se concentre sur la production de cartes. L'essor important qu'ont connu les techniques numériques (télédétection des sols, puissance de calculs des ordinateurs, gestion des données géographiques, méthodes statistiques et mathématiques) a permis le développement de la cartographie numérique des sols. Cette démarche repose sur l'usage de méthodes statistiques ou mathématiques pour dériver un modèle compréhensif des relations du sol (type ou propriété) avec le paysage en utilisant le lien entre des données d'observations relativement rares avec des données auxiliaires à forte résolution spatiale. Ces données auxiliaires sont également appelées les variables SCORPAN. Elles ont été d'abord formalisées par Jenny (1941), puis ensuite par McBratney et al. (2003). Elles comprennent les données sur le sol (S), le climat (C), la biolo-

gie (O), la topographie (R), le matériau parental (P), l'âge (A) et les relations spatiales (N).

L'objectif de cette étude est d'appliquer la démarche de cartographie numérique pour appréhender la distribution spatiale des propriétés édaphiques et biologiques de la couche de surface (0-30 cm) des sols à l'échelle de l'OPE. Nous nous sommes appuyés sur un modèle statistique couplant une analyse multivariée intégrant des contraintes spatiales à des procédures d'apprentissage automatique supervisé et de la géostatistique.

L'analyse multi variée a en effet l'avantage de réduire la dimension des tableaux de données étudiées en quelques facteurs qui peuvent ensuite être interprétés.

MATÉRIEL ET MÉTHODE

Zone d'Etude

La zone d'étude s'étend sur un territoire de 240 km² situé à la frontière entre le département de la Meuse et de la Haute-Marne. Ce territoire est caractérisé par un paysage essentiellement agricole (67.2 % de la surface de la zone est couvert en SAU dont 76.5 % en terres arables, 22.1 % en prairies et 1.4 % en zones agricoles hétérogènes) entrecoupé par des zones de forêts de feuillus et de conifères (31.8 %). Les zones urbaines occupent

quant à elles seulement 1 % de la zone. La zone d'étude se caractérise principalement par des sols bruns calciques et des rendzines brunifiées. On note la présence de quelques sols bruns alluvio-colluviaux en fond de vallon et de sols podzolitiques présents par taches sur les plateaux forestiers. Ces sols se sont développés sur un socle géologique calcaire daté du Tithonien et du Kimméridgien Terminal. Au nord-ouest de la zone, on distingue un étage argilo-sableux du Crétacé. Enfin, les fonds de vallons reposent sur un socle d'alluvions récentes.

Le réseau de surveillance de la qualité des sols sur la zone est constitué d'un réseau de 127 points établi à partir d'un maillage systématique de 1,5 km sur 1,5 km (*figure 1*). Le suivi du site prévoit l'échantillonnage de l'ensemble des points du réseau durant une campagne de 5 années consécutives et un retour sur chaque site d'une fréquence de 10 ans.

Les sites du dispositif ne sont pas tous prélevés avec le même protocole. On distingue ainsi deux ensembles. Premièrement, sur chaque site, une surface d'échantillonnage est définie par un carré de 20 m x 20 m divisé en 100 unités d'échantillonnage (placettes). Lors de la première campagne, 25 prélèvements sont effectués, aléatoirement, au sein des 25 unités d'échantillonnage portant le numéro 1, répondant au plan d'échantillonnage aléatoire non aligné. Les prélèvements se font sur les premiers horizons de sol 0-30 cm et 30-50 cm. Deuxièmement, sur certains sites (choisis de manière experte par le pédologue pour décrire au mieux la diversité pédologique de la zone et au nombre de 57),

Figure 1 - Carte du Réseau de la qualité des sols.

Figure 1 - Map of the soil monitoring network

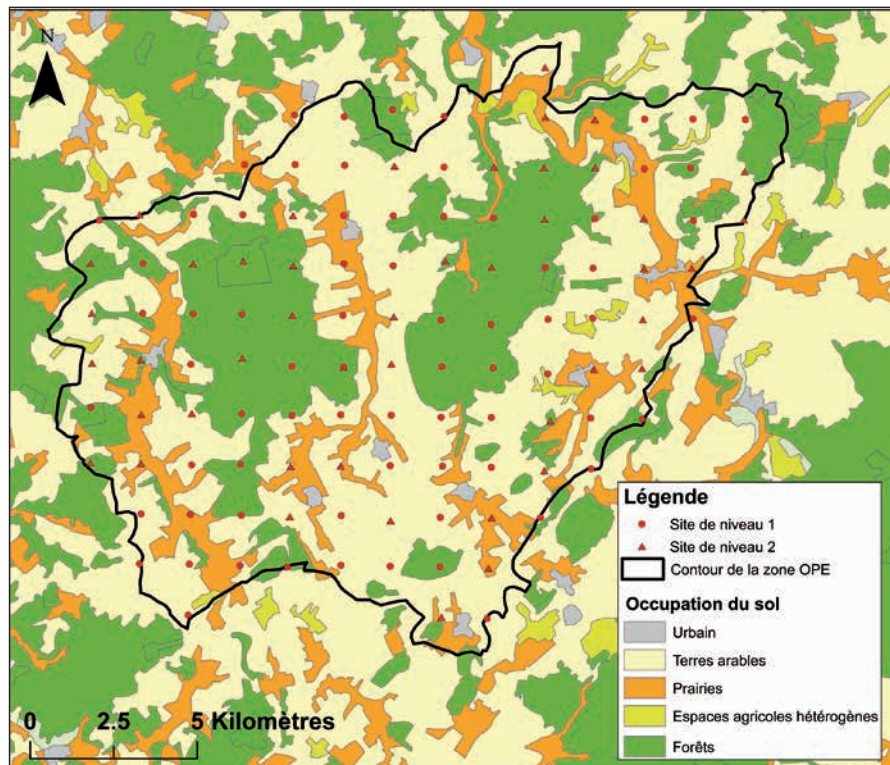
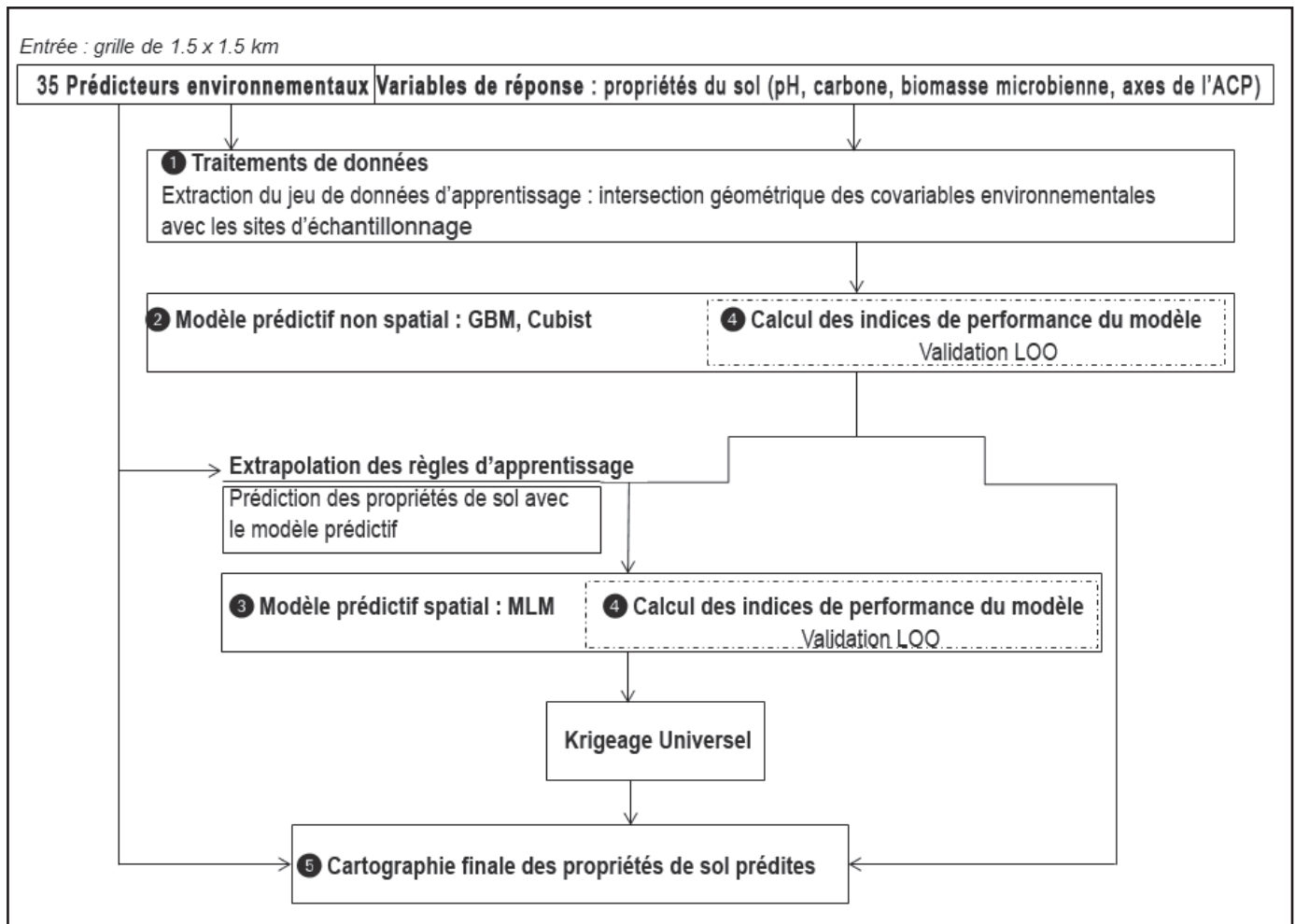


Figure 2 - Etapes de cartographie des variables pédologiques.
Figure 2 - Soil variables mapping steps.



nommés sites de niveau 2, une fosse pédologique est réalisée pour obtenir une description pédologique complète.

Les analyses physico-chimiques classiques telles que granulométrie, pH, teneur en carbone, en azote et en nutriments sont complétées par l'analyse des éléments traces et des micro-polluants organiques (16 hydrocarbures aromatiques polycycliques et 20 polychlorobiphényles). La séparation des communautés bactériennes, d'une part, et des communautés de champignons, d'autre part, est réalisée par génotypage ARISA (Automated method of Ribosomal Intergenic Spacer Analysis). Cette méthode permet l'estimation rapide de la structure de chacune des communautés microbiennes. C'est une technique de typage moléculaire utilisant la taille de l'intergène ADN ribosomal 16S et 23S comme marqueur des modifications de composition des communautés de bactéries et de champignons. Les analyses physico-chimiques sont réalisées par le

laboratoire d'analyse des sols de l'INRA à Arras tandis que les analyses biologiques sont réalisées par l'UMR Agroécologie et la plateforme GenoSol de l'INRA de Dijon.

Dans ce travail, nous nous intéressons aux 7522 résultats d'analyses caractérisant le premier horizon de sols.

Les ACP (Analyses en Composantes Principales) classiques et spatiales sont menées sur une matrice des propriétés pédologiques du sol comprenant la granulométrie, les éléments totaux et échangeables majeurs, la teneur en carbone organique, en azote, phosphore, la capacité d'échange cationique, les cations échangeables (Ca, Na, Mg, K) et le pH.

Algorithme de cartographie

La méthode utilisée pour cartographier des variables pédologiques et biologiques repose sur les étapes suivantes (figure 2):

Tableau 1 - Liste des covariables entrées dans les modèles de prédiction.**Table 1** - List of covariates used for modelling.

Covariables	Remarques
Pente	Issues du MNT 25 mètres
Orientation des pentes	
Courbures (globales, verticales et horizontales)	
Rugosité (écart-type, étendue)	
Topography Position Index (TPI)	
Classification Landform	
Indice de Beven	
Multiresolution Valley Bottom Flatness (MRVBF)	
Multiresolution Ridge Top Flatness (MRRTF)	
Distance au plus proche réseau (DPPR)	
Hauteur au plus proche réseau (HPPR)	
Indice de Développement et de Persistance des Réseaux (IDPR)	
Carte pédologique	1/50000
Carte géologique	1/50000
Occupation du sol	Issue du Registre Parcellaire Graphique, Corinne Land Cover, contours de forêt de l'ONF

1 Préparation des jeux de données

Extraction du jeu de données d'apprentissage correspondant aux sites appartenant à la grille.

2 Création d'un modèle prédictif non spatial

Le modèle est basé sur tous les sites préparés lors de l'étape 1 et les covariables environnementales spatialement exhaustives (listées dans le *tableau 1*). Pour chaque variable étudiée, nous avons construit deux modèles basés sur des algorithmes d'arbre de régression: Cubist v2.04 (Kuhn et al., 2013) et GBM v2.1 (Ridgeway et al., 2013). L'algorithme de Cubist est basé sur des règles permettant l'accumulation de modèles prédictifs basée sur des partitions du jeu de données. GBM est un algorithme qui réalise une multitude d'arbres élémentaires binaires de classification. À l'issue de la régression, l'arbre final correspond à une combinaison linéaire de ces différents arbres. Ces deux algorithmes impliquent la mise en œuvre d'une étape de réglage des paramètres. Nous avons utilisé pour cela le package caret (Khun et al., 2008, 2014).

3 Création d'un modèle spatial en combinant le modèle de l'étape 2 et la géostatistique

Le modèle issu de 2 permet d'obtenir une prédiction en chaque site du jeu de calibration. Ces prédictions ont été introduites comme covariables dans un modèle linéaire mixte (MLM) (Stein, 1999). Un MLM sépare la variabilité spatiale observée en deux parties. La première est une variation déterministe issue de la relation linéaire de la propriété avec une ou plusieurs covariable(s) identique(s) pour chaque individu. La deuxième est une variation dite aléatoire correspondant à la réalisation d'une variable aléatoire suivant une loi normale multivariée de moyenne nulle et dont la covariance peut s'exprimer à l'aide d'une fonction dépendant de la distance entre les observations. Dans notre étude, nous avons retenu le modèle Matérn (Diggle et al., 2007 ; Paulo et al., 2001) qui est connu pour sa grande capacité à modéliser une large gamme de processus spatiaux.

4 Performance des modèles

La performance des modèles non spatiaux et spatiaux a été évaluée à partir d'une validation croisée LOO (Leave-One-Out). Cette dernière consiste à apprendre sur (n-1) observations puis à valider le modèle sur la *nième* observation et l'on répète cette

opération n fois. Plusieurs indices de qualité ont été calculés à partir de la valeur observée et la valeur prédite à la position k pour les n échantillons:

1) L'erreur moyenne:

$$MSE = \frac{1}{n} \sum_{k=0}^n (\hat{Y}_k - Y_k)^2$$

2) La racine carrée de l'erreur moyenne:

$$RMSE = \sqrt{\frac{1}{n} \sum_{k=0}^n (\hat{Y}_k - Y_k)^2}$$

3) Le coefficient de détermination qui mesure le degré d'accord entre les mesures et les prédictions:

$$R^2 = \frac{||\hat{Y} - \bar{Y}||^2}{||Y - \bar{Y}||^2} = \frac{SCE}{SCT}$$

avec $SCT = SCE + SCR$ (où SCE et SCR désignent respectivement la somme des carrés expliquée par le modèle et la somme des carrés des résidus).

4) La moyenne et la médiane des carrés des erreurs standardisées (CES), avec

$$CES_k = \frac{(Y_k - \hat{Y}_k)^2}{\text{var}[\hat{Y}_k]}$$

avec $\text{var}[\hat{Y}_k]$ la variance de krigeage.

Ces deux derniers indicateurs renseignent sur la capacité du modèle à reproduire le processus spatial. Dans ce cas, la moyenne doit être égale à 1 et la médiane à 0.455 (Saby *et al.*, 2011). Dans le cas contraire, le modèle spatial n'est pas validé et les prédictions se limitent aux relations non spatiales (étape 2).

5 Production des cartes: A l'issue de l'étape 4, le meilleur modèle est appliqué au raster des variables environnementales. Dans le cas des modèles spatiaux, une étape de krigeage avec dérive externe est ajoutée. Dans ce cas, il est possible d'adjoindre à la carte des prédictions, la carte de la variance du modèle (Webster et Oliver, 2007).

RÉSULTATS

Propriétés physico-chimiques

Afin de rendre compte de l'état des sols de la zone OPE, les statistiques paramétriques et non paramétriques sont calculées sur l'ensemble des sites d'échantillonnage. Elles permettent d'ap-

préhender de manière synthétique le comportement de la zone pour une propriété mesurée: la valeur centrale de la propriété (médiane, moyenne), sa dispersion (écart-type) et son étendue (quartiles, déciles). Certaines statistiques sont notées « LQ » ce qui signifie « Limite de Quantification », à savoir que la statistique prend la valeur sous la limite du seuil de détection (tableau 6).

Du point de vue de la texture, la zone est essentiellement argilo-limoneuse avec des moyennes de taux d'argile à plus de 40 % et des limons fins à 30 %. Le taux de sable est en moyenne inférieur à 10 %. Le taux de calcaire total est faible (moins de 5 % en moyenne) mais sa dispersion relativement importante (écart-type à 8.6 %). Les teneurs en carbone sont en moyenne de 33 % relativement peu dispersées autour de cette valeur. Afin de parfaire la description de l'état initial du sol, les statistiques descriptives de l'OPE sont confrontées aux statistiques descriptives nationales métropolitaines (soit 2176 sites) et locales (les analyses RMQS de la Meuse et de la Haute-Marne, soit 47 sites). Les concentrations en éléments échangeables et totaux sont dans la moyenne nationale, à l'exception du calcium échangeable dont les concentrations sont supérieures à celles rencontrées sur l'ensemble de la France métropolitaine. Néanmoins, elles restent cohérentes avec les teneurs rencontrées dans la région (25.65 cmol+/kg en moyenne). De même la CEC (Capacité d'Echange Cationique) et le pH affichent des valeurs légèrement au-dessus des valeurs nationales (respectivement 25.6 cmol+/kg et 7.1). Aucune caractéristique des sols de l'OPE ne témoigne d'une anomalie par rapport aux résultats des statistiques nationales et régionales. Certaines propriétés sont marquées par un comportement local, c'est-à-dire que les statistiques OPE et régionales vont différer conjointement avec les statistiques nationales (pH plus basique, concentration en calcium échangeable plus élevée...). Le tableau 3 donne la comparaison des résultats des ACP classiques et spatiales en termes de variance expliquée. Pour l'ACP « classique », la variance est égale à la valeur propre. Pour l'ACP sous contrainte spatiale, la variance correspond à la variance spatiale. Ce tableau nous permet de comparer la perte d'inertie (variance spatiale versus variance totale) et le gain en information spatiale (indice de Moran de l'ACP spatiale versus celui de l'ACP « classique »). L'ACP conduite sous contrainte spatiale permet un gain d'information surpassant largement la perte d'inertie pour les deux premiers axes des paramètres physico-chimiques.

La figure 3 illustre les quatre premiers axes de l'ACP sous contrainte spatiale. L'axe 1 est principalement influencé par la teneur en argile et en éléments totaux d'une part (corrélé négativement) et les sables fins d'autre part (corrélé positivement). Cet axe reflète donc une richesse minérale globale du premier horizon de sol. L'axe 2 oppose la matière organique et un fort C/N au pH basique et au phosphore assimilable élevé (tableau 7).

Cet axe témoigne de l'effet anthropique de l'occupation du sol avec, d'un côté, les sols acides, riches en matière organique et au C/N élevé qui sont plutôt typiques des sols forestiers, et

Tableau 2 - Indices de performance de l'ACP sous contrainte spatiale face à l'ACP classique (paramètres physico-chimiques).**Table 2** - Compared performance index of spatially constrained PCA and classical PCA (physico-chemical parameters).

Élément	D1	Q1	moyenne	écart-type	médiane	Q3	D9
argile	305,00	378,00	434,10	104,48	440,00	500,00	557,20
Limons fins	237,00	276,00	307,53	55,91	311,00	341,00	377,80
Limons grossiers	111,60	132,00	159,69	41,03	154,00	186,00	214,80
Sables fins	17,00	24,00	64,82	81,36	35,00	58,00	182,20
Sables grossiers	3,00	6,00	33,85	34,87	21,00	51,00	82,80
Calcaire tot	LQ	LQ	46,83	86,31	3,20	53,80	182,80
Carbone	17,68	21,50	33,10	15,01	29,30	41,50	55,20
CEC	10,54	19,20	25,60	10,32	26,80	31,50	38,70
C/N	9,71	10,10	11,35	1,82	10,50	12,90	14,10
MO	30,66	37,20	57,26	25,96	50,60	71,90	95,52
pH	5,46	6,61	7,14	1,00	7,51	7,95	8,06
Al ech	0,03	0,04	0,27	0,65	0,06	0,13	0,78
Al tot	4,57	5,55	6,16	1,14	6,22	6,97	7,47
Ca ech	8,65	18,20	25,65	11,81	27,00	32,00	40,72
Ca_tot	0,27	0,53	2,52	3,51	0,92	2,98	7,93
Fe ech	0,01	0,01	0,01	0,01	0,01	0,02	0,02
Fe tot	3,19	3,58	4,40	1,93	3,97	4,43	5,55
K ech	0,26	0,33	0,49	0,23	0,46	0,56	0,77
K tot	1,06	1,37	1,57	0,36	1,58	1,82	1,96
Mg ech	0,40	0,52	0,73	0,32	0,66	0,90	1,14
Mg tot	0,33	0,43	0,48	0,11	0,51	0,56	0,59
Mn ech	LQ	LQ	0,06	0,09	0,01	0,08	0,17
Mn tot	591,20	766,00	963,69	318,84	947,00	1170,0	1346,0
Na ech	0,02	0,03	0,05	0,05	0,04	0,06	0,07
Na tot	0,15	0,18	0,25	0,09	0,24	0,31	0,37
N tot	1,47	1,89	2,95	1,30	2,77	3,70	5,19
P ass	0,01	0,02	0,06	0,05	0,05	0,07	0,11
P tot	0,18	0,22	0,28	0,10	0,27	0,31	0,41

de l'autre des sols moins acides (chaulage) et riches en P assimilable qui sont propres aux sols de grandes cultures. L'axe 3 classe les sols selon la saturation ou la désaturation de leur CEC une fois les effets des deux premiers axes enlevés. D'un côté, nous avons un groupe de sols sableux très acides et très désaturés contenant probablement peu de minéraux altérables (pH très faibles, fortes teneurs en aluminium échangeable et fortes teneurs en sables) et un groupe de sols calcaires mais relati-

vement moins pourvus en calcium échangeable que les autres sols calcaires (donc contenant probablement une fraction calcaire résistante qui ne passe pas en solution). De l'autre côté de l'axe 3, on trouve des sols sur-saturés argilo-limoneux développés sur matériau calcaire tendre (CEC élevées, forts taux de calcium et magnésium échangeable, teneur en argile élevée).

Pour chaque axe des ACP, les deux modèles (GBM et Cubist) ont été testés et les indicateurs de performance

Tableau 3 - Indices de performance de l'ACP sous contrainte spatiale face à l'ACP classique (paramètres physico-chimiques).

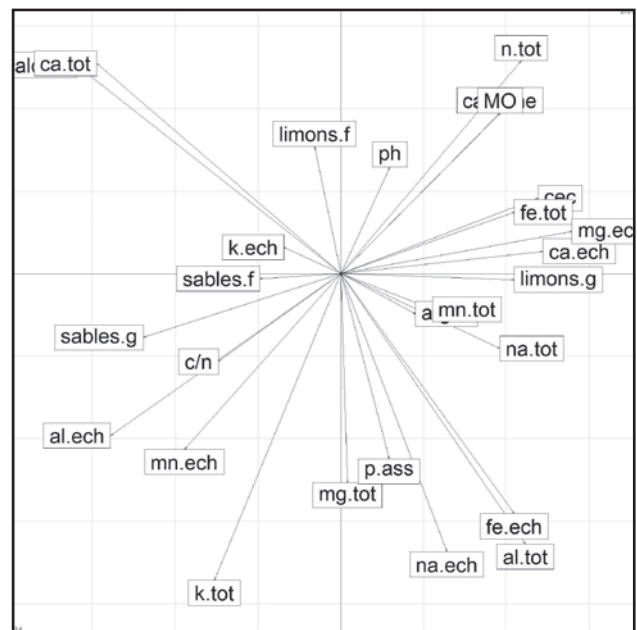
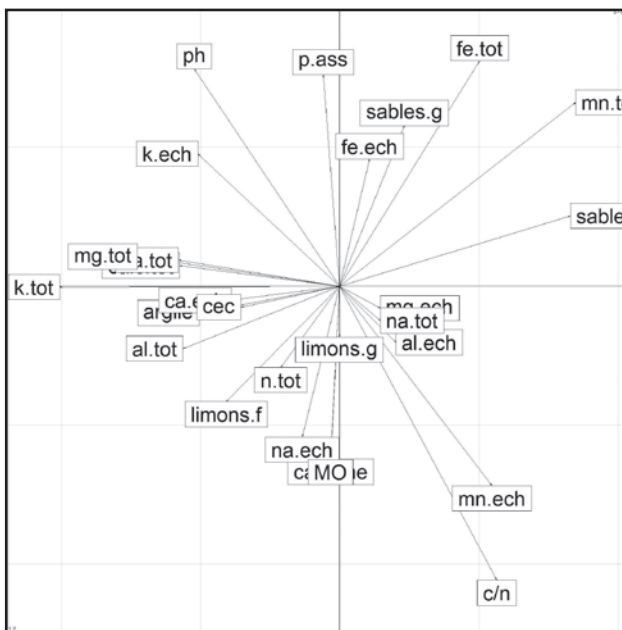
Var, la variance expliquée par l'axe de l'ACP spatiale (sp) ou classique (cl); Moran, l'indice de Moran traduisant le degré de corrélation spatiale; % cum, le pourcentage cumulé d'explication de la variance par les axes des ACP; Gain spatial, le gain entre le l'indice de Moran de l'axe de l'ACP spatiale et celui de l'ACP classique; Perte inertie, la différence (en pourcent) de la part d'explication de la variance de l'ACP spatiale et de l'ACP classique.

Table 3 - Compared performance index of spatially constrained PCA and classical PCA (physico-chemical parameters).

Axes	ACP spatiale				ACP classique			Gain spatial	Perte inertie
	Valeur propre	Var sp	Moran sp	% cum	Var cl	Moran cl	% cum		
Pédo axe 1	2,97	7,27	0,41	39,15	9,65	0,20	9,65	51,2 %	24,7 %
Pédo axe 2	0,57	2,79	0,20	54,18	4,10	0,09	13,74	55 %	31,9 %
Pédol axe 3	0,52	3,46	0,15	72,82	3,48	0,24	17,23	-60 %	0,6 %

Figure 3 - Graphique des variables composant les 4 premiers axes de l'ACP spatiale des critères pédologiques. Le graphique 'a' représente les axes 1 (abscisse) et 2 (ordonnées), le 'b' représente les axes 3 (abscisse) et 4 (ordonnées).

Figure 3 - Chart of variables composing the first 4 axis of soil criteria's PCA. (a) Axis 1 and 2, (b) axis 3 and 4.



des modèles ont permis la sélection du meilleur modèle pour la prédiction de chaque paramètre. Pour les deux premiers axes, la valeur des moyennes des carrés des erreurs standardisées et des médianes des carrés des erreurs standardisées renseignent sur l'existence d'un processus spatial et la capacité du modèle à le reproduire. Ainsi, pour ces deux axes, le krigeage universel améliore la prédiction (tableau 4).

Les coefficients de détermination varient entre 0.28 pour le troisième axe de l'ACP et 0.94 pour l'axe 1 sur les critères pédo-

logiques (après krigeage). Le krigeage universel améliore systématiquement le R^2 des prédictions de manière considérable. Le modèle GBM est plus souvent privilégié au modèle Cubist. Compte-tenu des gammes de variation des prédictions, les RMSE sont assez faibles et témoignent donc d'un bon succès des modèles.

À l'issue de la modélisation, les variables utilisées sont ordonnées selon leur importance qui est fournie par la fonction `varImp()` de la librairie `Caret` (tableau 4). Le critère d'importance varie selon le modèle, pour GBM l'importance de la variable

Tableau 4 - Modèles et indicateurs de performance pour chaque paramètre physico-chimique prédit. CES signifie carrés des erreurs standardisées.

Table 4 - Models and their performance index for each predicted physico-chemical variables. CES means squared standardised prediction errors.

Paramètres	Modélisation				Krigage		
	Modèle	R ²	Gamme (min/ max prédit)	RMSE SD	R ²	Moyenne des CES	Médiane des CES
Pédologie axe1	Cubist	0,48	-5,02 / 8,71	2,22	0,9411	0,955	0,149
Pédologie axe2	GBM	0,577	-3,41 / 3,44	1,09	0,9405	1,048	0,2787
Pédologie axe3	GBM	0,288	-4,8 / 3,69	0,391			

est exprimée en fonction de l'utilisation de la variable dans le modèle. Pour résumer, cette quantité correspond au nombre d'arbres où la variable a été retenue. Pour le modèle Cubist, l'importance de la variable correspond au pourcentage de fois où la variable est utilisée comme condition et/ou comme variable d'un modèle linéaire.

L'axe 1 de l'ACP, correspondant à la richesse minérale, dépend principalement de la distance et de la hauteur par rapport au réseau hydrographique (les alluvions et les colluvions sont plus riches en éléments minéraux), de la nature pédologique et de l'occupation du sol (action anthropique de la fertilisation minérale). L'axe 2 est hautement impacté par l'occupation du sol. Comme nous l'avons vu, cet axe traduit l'action de l'activité agricole et forestière sur les propriétés du sol. Enfin, la distribution spatiale des propriétés traduites par l'axe 3 de l'ACP (saturation/désaturation relative de la CEC est orientée par la courbure verticale des pentes (affleurement des calcaires tendres), le matériau parental (qui rend compte notamment de la présence de calcaire dur, peu soluble sur l'axe négatif de l'ACP), puis, l'orientation des pentes, certains indices de topographie (TPI, BEVEN) et de rugosité des pentes.

Les sols sableux se distribuent sur les plateaux au nord-ouest de la zone (valeurs positives sur l'axe de l'ACP) et les sols riches en éléments totaux et argileux se localisent principalement au sud et à l'est de la zone sur les sols de coteaux (figure 4a). Les incertitudes maximum se situent en dehors de la zone quadrillée par le réseau de mesure, au nord et au sud de la zone, mais également sur le plateau au nord de la Saulx (non montré).

Les sols forestiers riches en C/N et en matière organique, acides, apparaissent en clair sur la carte tandis que les sols agricoles chaulés (pH basiques, à fort taux de phosphore assimilable) sont foncés (figure 4b).

La carte qui spatialise les valeurs du troisième axe de l'ACP présente les sols peu saturés sableux du plateau du Crétacé au nord-ouest et les sols calcaires du Kimméridgien au sud et

à l'extrême est. Les sols de coteaux riches en éléments échangeables sont en jaune (figure 4c).

Enfin, la figure 5 synthétise la répartition de l'ensemble des trois axes de l'ACP sur les propriétés physico-chimiques par composition colorée. Chaque axe est représenté par une des trois couleurs primaires en synthèse additive (rouge, vert, bleu). Ainsi, pour chaque pixel, la saturation sur les 3 couleurs primaires dépend de sa valeur sur les trois axes de l'ACP.

Propriétés biologiques

En ce qui concerne l'ACP sur la matrice des empreintes moléculaires issues des analyses de type ARISA, le pourcentage de la variance expliquée est plus faible que pour l'ACP sur les caractéristiques physico-chimiques. La variabilité au sein des communautés des microorganismes telluriques est en effet expliquée également par des dissemblances spatiales. L'interprétation des axes n'a pas de sens dans ce cas car les variables représentent un nombre relatif d'individus. L'interprétation des axes se limite à une ressemblance génétique des individus partageant les mêmes scores. Le gain fourni par l'ACP spatiale est de l'ordre de 100 % alors que la perte d'inertie se situe entre 50 % et 60 %.

Le premier axe de l'ACP sur les communautés bactériennes et les deux premiers axes des communautés de champignons présentent des indicateurs témoignant de l'existence d'un processus spatial dans la distribution des paramètres. Le krigage permet donc d'améliorer les modélisations en prenant en compte la dimension spatiale.

Les coefficients de détermination varient entre 0.08 pour le troisième axe de l'ACP concernant les communautés de champignons et 0.88 pour l'axe 1 (communautés de champignons) après krigage. Le modèle GBM est plus souvent privilégié au modèle Cubist (5/6).

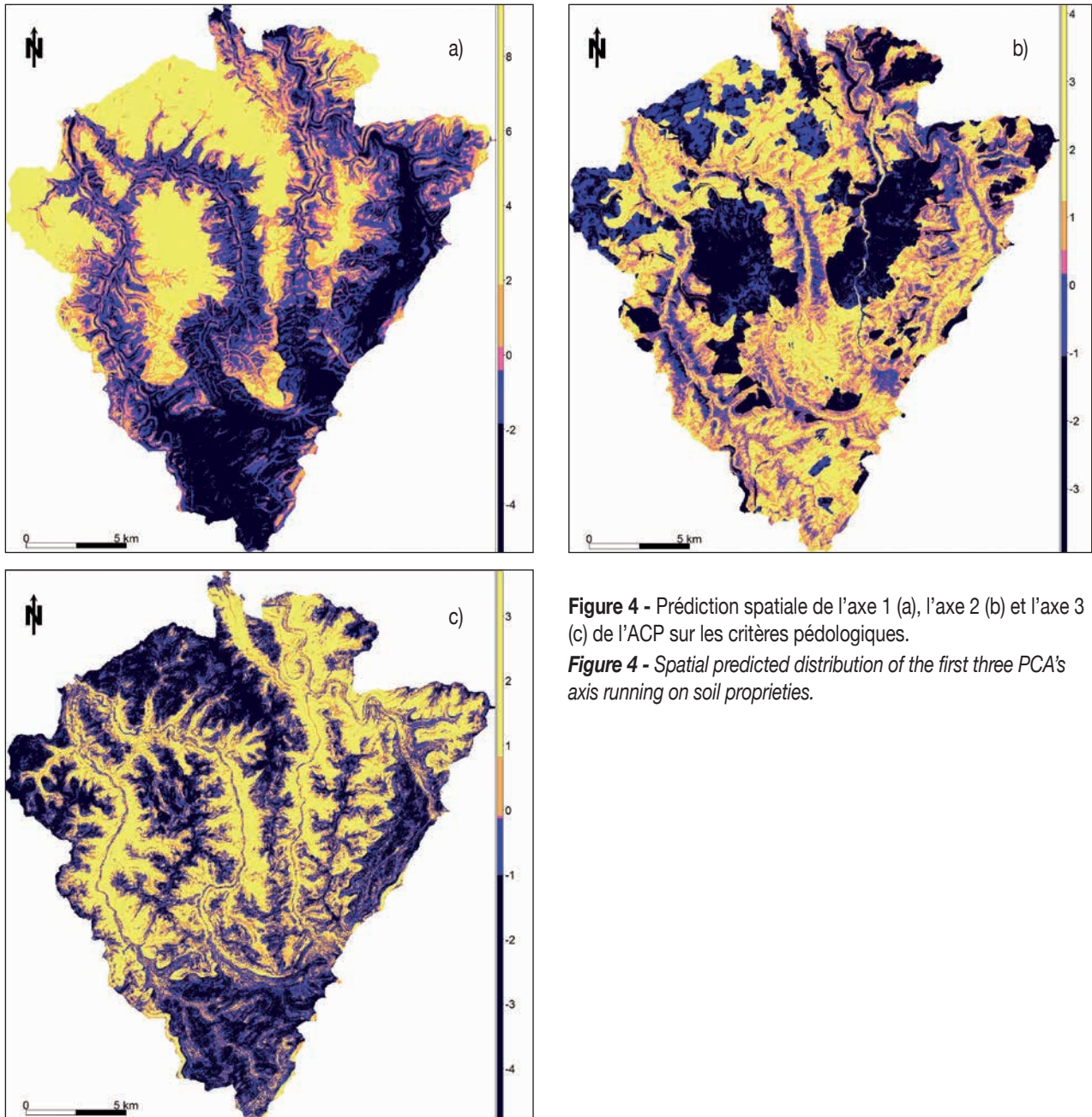


Figure 4 - Prédiction spatiale de l'axe 1 (a), l'axe 2 (b) et l'axe 3 (c) de l'ACP sur les critères pédologiques.

Figure 4 - Spatial predicted distribution of the first three PCA's axis running on soil properties.

Tableau 5 - Principales variables explicatives de la distribution spatiale des paramètres physico-chimiques.

Table 5 - Main explanatory variables of physico-chemical's parameters spatial distribution.

Axe1	Axe2	Axe3
HPPR (100)	Occupation du sol (100)	Courbure verticale (100)
DPPR (90)	MNT (9,3)	Géologie (99)
Occupation du sol (70)	Rugosité	Orientation pente (90)

Figure 5 - Représentation cartographique des trois axes des ACP sur les propriétés physico-chimiques.
Figure 5 - Mapping of the first three PCA's axis running on physical and chemical properties.

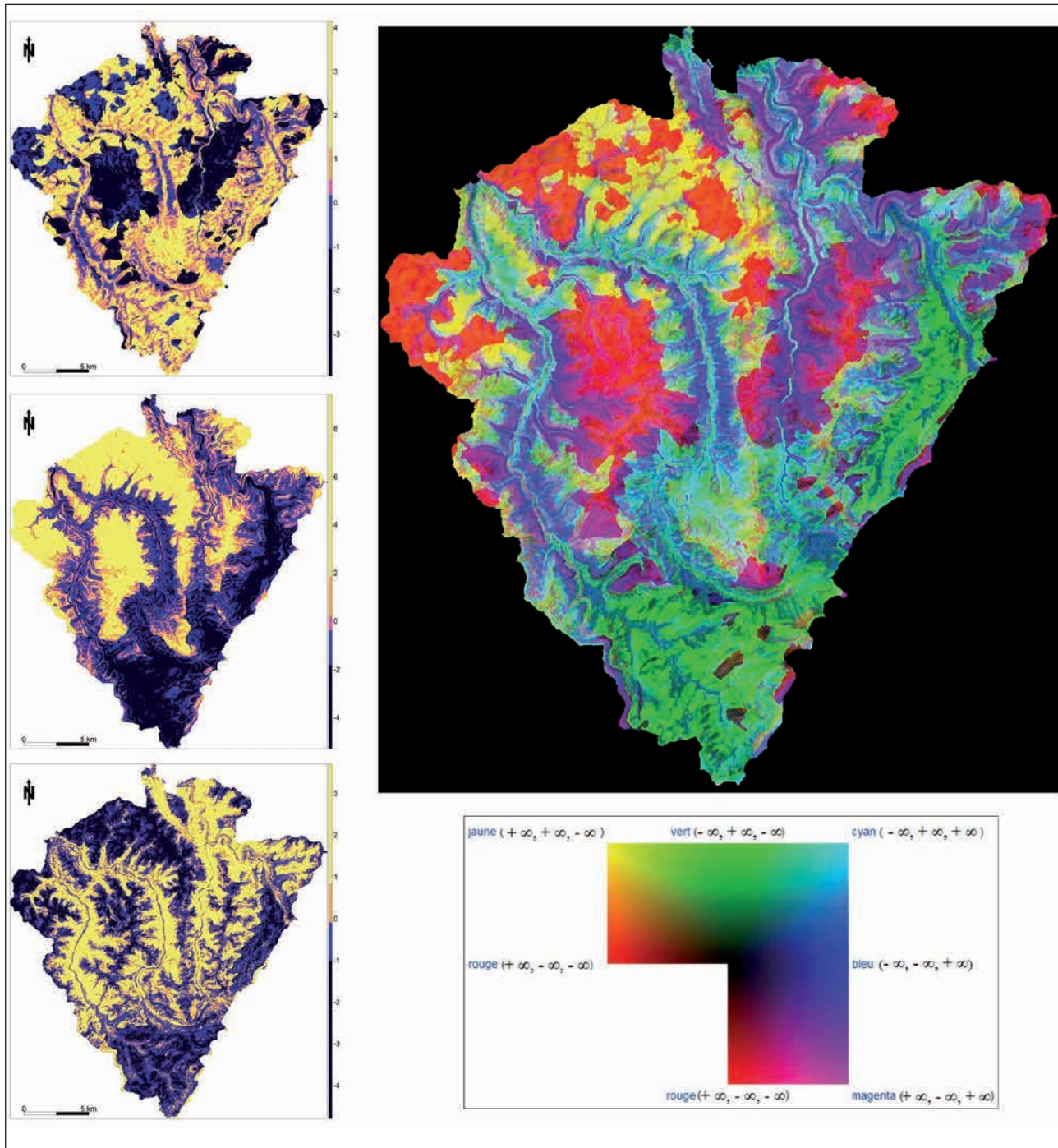


Tableau 6 - Indices de performance de l'ACP sous contrainte spatiale face à l'ACP classique (paramètres biologiques).

Var, la variance expliquée par l'axe de l'ACP spatiale (sp) ou classique (cl); Moran, l'indice de Moran traduisant le degré de corrélation spatiale; % cum, le pourcentage cumulé d'explication de la variance par les axes des ACP; Gain spatial, le gain entre l'indice de Moran de l'axe de l'ACP spatiale et celui de l'ACP classique; Perte inertie, la différence (en pourcent) de la part d'explication de la variance de l'ACP spatiale et de l'ACP classique.

Table 6 - Compared performance index of spatially constrained PCA and classical PCA (biological parameters).

Axes	ACP spatiale				ACP classique			Gain spatial	Perte inertie
	Val. propre	Var sp	Moran sp	% cum	Var cl	Moran cl	% cum		
Bactéries axe 1	2,92	9,91	0,29	14,24	27,25	-0,02	27,25	106,9 %	63,6 %
Bactéries axe 2	1,90	6,38	0,30	23,41	15,43	-0,02	42,68	106,7 %	58,7 %
Bactéries axe 3	1,89	4,70	0,40	30,16	14,98	0,01	57,66	97,5 %	68,6 %
Champignons axe 1	3,34	5,91	0,57	9,32	14,15	-0,02	14,15	103,5 %	58,2 %
Champignons axe 2	2,51	5,34	0,47	17,73	11,36	-0,01	25,52	102,1 %	52,9 %
Champignons axe 3	2,34	5,27	0,44	26,03	11,19	-0,02	36,71	104,5 %	52,9 %

Tableau 7 - Modèles et indicateurs de performance pour chaque paramètre physico-chimique prédit. CES signifie carrés des erreurs standardisées.**Table 7** - Models and their performance index for each predicted physico-chemical variables. CES means squared standardised prediction errors.

Paramètres	Modélisation				Krigage		
	Modèle	R ²	Gamme (min/max prédit)	RMSE SD	R ²	Moyenne des CES	Médiane des CES
Bactéries axe1	GBM	0,175	-3,65 / 7,19	2,9	0,524	1,107	0,402
Bactéries axe2	GBM	0,165	-2,73 / 3,35	0,79			
Bactéries axe3	Cubist	0,138	-0,79 / 2,33	0,799			
Champignons axe1	GBM	0,228	-5,44 / 5,38	0,601	0,875	1,138	0,326
Champignons axe2	GBM	0,298	-4,04 / 4,89	0,451	0,819	1,044	0,484
Champignons axe3	GBM	0,08	-2,39 / 3,44	0,481			

Tableau 8 - Principales variables explicatives de la distribution spatiale des paramètres biologiques.**Table 8** - Main explanatory variables of biological parameters spatial distribution.

Bactéries			Champignons		
Axe1	Axe2	Axe3	Axe1	Axe2	Axe3
Rugosité (100)	MNT (100)	Rugosité (100)	MNT (100)	MNT (100)	MNT (100)
Orientation pente (83)	HPPR (56)	Orientation pente (86)	DPPR (60)	HPPR (26)	TPI (99)
Rugosité (73)	Rugosité (36)		HPPR (60)	Beven (26)	

Les variables explicatives de la distribution des axes des ACP menées sur les paramètres biologiques sont également fournies par les modèles.

Les distributions des communautés bactériennes et des communautés de champignons sont essentiellement dirigées par des facteurs morphologiques (MNT, rugosité, orientation des pentes, plus courte hauteur ou distance avec le plus proche cours d'eau...) (tableau 8).

Les cartes spatialisant les axes des ACP des communautés de bactéries et de champignons sont directement représentées sous forme de vignettes à gauche de la composition colorée (figure 6). Les compositions colorées des communautés bactériennes et des champignons, peu tranchées et résultant d'un mélange des couleurs primaires additives, représentent différentes niches écologiques et donc différents équilibres où la proportion des différentes populations de la communauté varie. En aucun cas les zones ne discriminent un habitat pur abritant une seule communauté.

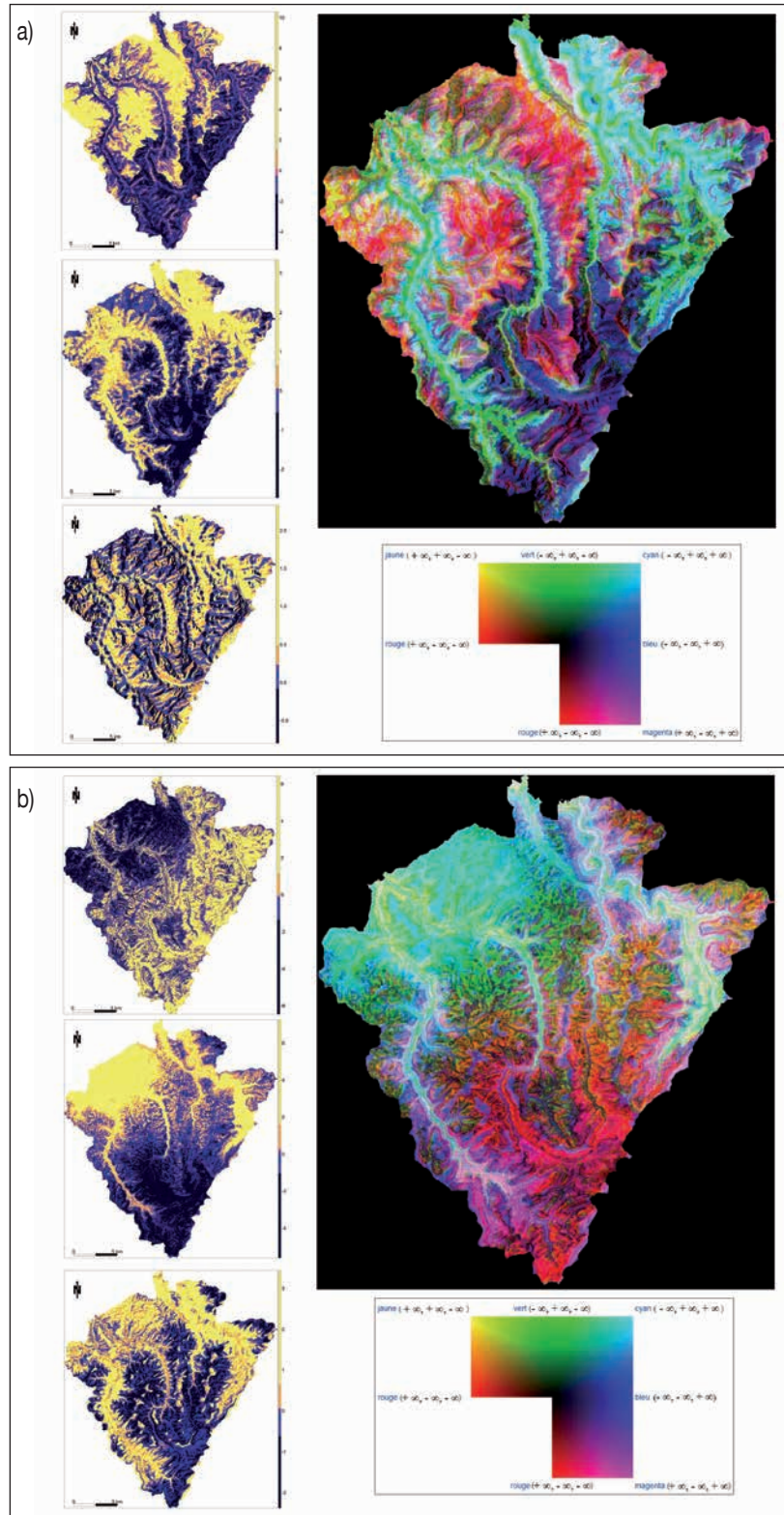
DISCUSSION

La procédure: modélisation et prédiction

La combinaison de méthodes statistiques de régression et de krigeage pour l'obtention d'un modèle compréhensif des propriétés du sol pour la spatialisation des propriétés de sols est une procédure explicite et répétable puisque les prédicteurs et leurs contributions respectives sont clairement identifiés. Des indicateurs de performances peuvent facilement être calculés, ce qui nous permet d'avoir un jugement qualitatif à l'issue de la production cartographique. Néanmoins, l'absence d'un jeu de données externe pour la validation du modèle peut faire défaut. Ce dernier a été simulé par validation croisée dite LOOCV (Leave-One-Out Cross-Validation). Mais comme l'ont montré Lacoste et al., 2011, les taux d'erreurs du modèle diffèrent selon le jeu de validation (taux d'erreur légèrement plus faible avec une validation interne plutôt qu'une validation externe). Ce taux d'erreur s'expliquerait

Figure 6 - Représentation cartographique des trois axes des ACP sur les communautés bactériennes (a) et fongiques (b).

Figure 6 - Mapping of the first three PCA's axis running on bacterial (a) and fungal (b) communities.



par une autocorrélation spatiale des pixels utilisés à la fois pour ajuster le modèle et pour la validation du point.

Les modèles Cubist et GBM ont l'avantage de permettre de combiner des variables qualitatives et quantitatives. Ce sont des outils puissants qui combinent différents modèles linéaires et différentes règles d'apprentissage avec un temps de calcul relativement faible.

Aussi, comme l'ont montré Odeh *et al.*, (1995) et plus tard, de Souza *et al.*, (2014) l'utilisation conjuguée de régression multilinéaire et du krigeage universel offre plus de performances.

Modélisation et prédiction à l'échelle du paysage

La modélisation et la prédiction des propriétés des sols à l'aide de procédures d'apprentissage automatique supervisé et de géostatistiques a déjà fait preuve d'efficacité à l'échelle du territoire national, tant pour la prédiction de caractéristiques physico-chimiques (Arrouays *et al.*, 2011) que biologiques (Dequiedt *et al.*, 2009). Cette méthode a également été testée à des échelles plus fines, par exemple Lacoste *et al.*, (2011) ont prédit le matériau parental à l'échelle de la région de la Bretagne et le carbone organique sur un petit territoire hétérogène de 10 km² au nord-est de l'Ille-et-Vilaine (Lacoste *et al.*, 2014). Dans cette étude, nous avons adapté la méthode à l'échelle d'un paysage de 240 km² plutôt homogène. L'homogénéité des variables SCORPAN sur la zone ne nous permet pas de prédire de manière fiable les propriétés physico-chimiques du sol une à une. Néanmoins, à l'aide d'une ACP spatialement contrainte, nous pouvons définir des typologies de sol que nous modélisons et spatialisons par la suite avec une bonne vraisemblance (r^2 compris entre 30 % et 95 %).

Malgré ces bons résultats, certains paramètres restent difficiles à modéliser. En effet, la méthode et l'échelle de travail ne s'adaptent pas forcément à l'échelle des processus que nous cherchons à expliquer. D'une part, pour deux modèles (axe 1 pédologie et axe 1 champignon), les valeurs des indicateurs de qualité du krigeage (RCES) sont trop éloignées des valeurs attendues, ce qui nécessite donc d'améliorer la modélisation. D'autre part, les processus microbiens se raisonnent d'une part à des échelles très fines: la diversité des communautés microbiennes dépend de la diversité des habitats, qui plus est, l'échelle de la niche microbienne est l'échelle de la motte de terre (Constans *et al.*, 2015). Ensuite, les micro-organismes sont sensibles aux actions anthropiques, aussi les différents modes de conduite des parcelles peuvent intervenir dans la diversité microbienne. Enfin, il convient de considérer le poids des paramètres physiques et chimiques du sol (texture, pH, C...) dont la répartition spatiale est fortement liée à la pédogénèse observable à l'échelle du paysage ou de régions écologiques entières. Ainsi, la répartition de ces communautés

microbiennes résulte de plusieurs processus ayant lieu à différentes échelles imbriquées les unes dans les autres (paysages > parcelles > motte de terre > micro-agrégats). Cette difficulté d'interprétation liée à l'échelle d'observation se matérialise très bien au vu des résultats et de la cartographie des communautés microbiennes présentées précédemment (certains $r^2 < 20$ %). Des travaux antérieurs (Ranjard *et al.*, 2013 ou Dequiedt *et al.*, 2009) ont montré l'influence supérieure des paramètres locaux (caractéristiques du milieu environnant) sur des paramètres environnementaux globaux (climat...). De fait, nous pouvons imaginer que les variables prédictives choisies par les modèles ne sont pas forcément adaptées de par l'échelle de leur définition, ou encore, c'est ce que nous allons voir dans le paragraphe suivant, leur nature même.

L'importance des variables explicatives

D'une part, comme le mentionne Grundwald, (2009), les modèles de prédiction se concentrent sur des facteurs externes: les facteurs du modèle SCORPAN. Ils ne prennent pas en compte les facteurs internes tels que la minéralogie, la structure du sol, la compaction, etc. Or, comme l'ont déjà montré Dequiedt *et al.*, 2009, la diversité génétique microbienne est influencée par le pH du sol en premier lieu mais aussi par la texture (Dequiedt *et al.*, 2009). Ces covariables ne sont pas disponibles selon une grille fine comme peuvent l'être les données sur l'altitude. Elles ne peuvent donc être intégrées à nos prédictions spatiales. La seule variable pédologique utilisée dans les modèles de prédiction spatiale est l'unité cartographique de la carte pédologique. Cependant, cette covariable n'apporte pas beaucoup à la qualité des modèles. En effet, au sein d'une même unité cartographique, les propriétés physico-chimiques des sols peuvent être variables. Il semble ainsi difficile d'améliorer à l'heure actuelle la finesse des cartographies sur la distribution des communautés de bactéries et de champignons.

D'autre part, certaines covariables explicatives représentent des effets induits car elles traduisent l'effet d'une autre covariable: la distribution de l'axe 2 de l'ACP sur les critères pédologiques est dictée par l'occupation du sol en premier lieu puis par l'altitude en second, or les forêts sont principalement situées sur les plateaux, il s'agit là donc d'une redite. De la même manière, les trois premiers axes de l'ACP effectuée sur les communautés de champignons sont structurés par le MNT, en réalité on peut penser qu'il s'agit de la texture du sol: plateaux sableux *versus* coteaux et fond de vallon argileux.

Retour sur l'ACP spatiale et l'interprétation des axes

Comme nous l'avons déjà exprimé, la combinaison d'une ACP sous contrainte spatiale et de techniques de régression et de géostatistique est une stratégie qui semble adaptée à l'échelle à laquelle nous travaillons. Certes, ce ne sont plus des valeurs de propriétés de sol que nous spatialisons mais les scores des axes de l'ACP. Pourtant, les modèles de prédiction ajustés sur ces données se révèlent performants au vu des indicateurs de qualité. Cette approche permet ainsi d'avoir une vision cartographique synthétique de la distribution multivariée des propriétés du sol.

Les axes de l'ACP s'interprètent comme de grands ensembles de caractéristiques des sols. Concernant les variations de compositions des communautés microbiennes, l'ACP nous permet de rapprocher les échantillons par leur empreinte génétique (l'électrophorèse renvoie un code-barre propre à chaque échantillon). L'analyse ARISA ne nous renseigne cependant pas sur la nature fonctionnelle ou taxonomique des communautés mais simplement sur la présence/absence et l'intensité relative des groupes microbiens caractérisés par la longueur de leur espace intergénique. Le test ARISA est un test rapide, peu fastidieux et peu coûteux dont la robustesse et la sensibilité ne sont plus à démontrer; il fournit des profils permettant des comparaisons entre les échantillons traduisant des modifications de composition. Suite à l'ACP et à la production des cartes de distributions des communautés, il nous est possible de conclure uniquement sur la répartition globale des communautés sans connaître la nature de leur fonction. Néanmoins, on observe des ruptures superposables avec les grands types d'occupation du sol et la nature pédologique des sols (plateaux sableux pour les communautés fongiques, ou encore les fonds de vallons pour les communautés bactériennes).

Concernant les caractéristiques physico-chimiques, l'interprétation des deux premiers axes est relativement facile. A l'échelle du paysage de l'OPE, la richesse minérale se structure selon la proximité avec les cours d'eau (alluvions), rugosité de la pente (colluvions), la nature du sol (unités cartographiques de la carte pédologique) et l'occupation du sol. L'axe 2 met en surbrillance ce dernier facteur d'influence puisque l'on observe que les distributions du pH, du C/N et du phosphore assimilable se structurent pour l'essentiel en fonction de la classe d'occupation du sol. L'axe 3 représente la part résiduelle de ce qui n'a pas encore été expliqué par les deux premiers axes. Il traduit un niveau relatif de saturation ou de désaturation de la CEC: il discrimine d'un côté des sols sur matériau calcaire tendre qui présentent une CEC sur-saturée en éléments échangeables. De l'autre côté, le groupe de sols n'est pas homogène et caractérise deux situations très différentes: des sols sableux très acides et très désaturés, et des sols calcaires ayant un

taux de saturation moins élevé que les autres sols calcaires de la région. Le calcaire dans ces sols est très probablement constitué de particules dures et résistantes ne passant pas (ou très faiblement) en solution. En d'autres termes, ce pôle rassemble des sols qui ont des teneurs en calcium échangeable plus faibles que leurs « voisins » sur les deux premiers axes.

CONCLUSION

En utilisant des échantillons de sols du réseau de suivi de la qualité des sols de maille de 1,5 km de la zone de référence de l'OPE, nous avons montré qu'il est possible de caractériser la distribution des propriétés physico-chimiques et biologiques de la couche superficielle du sol à l'échelle du paysage.

La réalisation d'une ACP sous contrainte spatiale sur les propriétés physico-chimiques des sols a permis d'identifier et de décrire différentes typologies de sols à l'échelle du paysage de l'OPE. La combinaison de procédures d'apprentissage automatique supervisé sur des covariables SCORPAN avec la géostatistique (procédure adaptée de travaux basés sur des échelles nationales ou régionales) a permis de définir des modèles prédictifs pour la cartographie à une précision de 25 mètres. Les modèles statistiques nous permettent d'identifier et de hiérarchiser des facteurs explicatifs. La distribution spatiale de la richesse minérale s'explique par la proximité (verticale et horizontale) avec le réseau hydrographique, la nature du sol et du substrat. L'impact de l'activité agricole se lit parfaitement dans la distribution des propriétés physico-chimiques du sol. En ce qui concerne la biologie du sol, les conclusions sont surtout qualitatives et non quantitatives du fait de la nature des analyses menées sur le territoire de l'OPE. Cette méthode nous permet toutefois d'identifier et de cartographier des zones aux équilibres microbiens différents à moindre coût.

REMERCIEMENTS

Cette étude a été financée par l'ANDRA dans le cadre de l'accord spécifique relatif à la mise en place et à la valorisation de l'étude de l'état de référence de la qualité des sols dans le secteur de référence de l'OPE. Nous remercions tous les techniciens qui ont permis le suivi, le prélèvement, la préparation et l'analyse des échantillons du sol sur la zone de référence. Nous remercions également Jean-Paul Party (Sol-Conseil) pour sa participation à la collecte et la description pédologique de la zone.

BIBLIOGRAPHIE

- Arrouays D., Saby N.P.A., Thioulouse J., Jolivet C., Boulonne L., Ratié C., 2011- Large trend in French topsoil characteristics are revealed by spatially constrained multivariate analysis. *Geoderma* 161. pp. 107-114.
- Brus D., De Gruijter J., Bierkens M., Knotters M., 2011. Sampling for natural resource monitoring. Springer. 2011. 332 p.
- Burrough, P.A., 1986 - Principles of Geographical Information Systems for Land Resources Assessment. Clarendon Press. Oxford.
- Constancias, F., Saby, N., Terrat, S., Dequiedt, S., Horrigue, W., Nowak, V., Ranjard, L. 2015- Contrasting spatial patterns and ecological attributes of soil bacterial and archaeal taxa across a landscape. *MicrobiologyOpen*. pp. 518-531.
- Dequiedt S. Thioulouse J., Jolivet C., Saby N.P.A., Maron P.A., Martin M., Chemidlin Prévost-Bouré N., Toutain B., Arrouays D., Lemanceau P., Ranjard L., 2009 - Biogeographical patterns of soil bacterial communities, *Environmental microbiology reports*. 1 (4). pp. 251-255.
- Dequiedt S., Lelièvre M., Jolivet C., Saby N.P.A., Martin M., Thioulouse J., Maron P.A., Mougél C., Chemidlin Prévost-Bouré N., Arrouays D., Lemanceau P., Ranjard L., 2009. ECOMIC-RMQS: biogéographie microbienne à l'échelle de la France. Etat d'avancement et premiers résultats. *Etude et Gestion des Sols*, 16, pp. 219-231.
- De Souza E., Hengl T., Fernandes E.I., 2014 - Comparing spatial prediction methods for soil property mapping in Brazil: A case study for the Rio Doce Basin. *GlobalSoilMap*. pp. 267-271.
- Fierer N., Morse J.L., Berthrong S.T., Bernhardt E.S., Jackson R.B., Environmental controls on the landscape-scale biogeography of stream bacterial communities. 2007. *Ecology*. 88.2162-2173. Grunwald, 2009 - Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, pp. 195-207.
- Diggle, P.J., Ribeiro Jr, P.J. 2007 - Model Based Geostatistics Springer, New York.
- Huybens N., Mainil J., Marlier D., 2009 - Les techniques de biologie moléculaire d'analyse des populations bactériennes complexes, *Ann. Méd. Vét.* 153. pp. 112-128.
- Jenny H., Factors of Soil Formation. 1941 - McGraw-Hill. New York. 279 p.
- Jolivet C., Boulonne L., Ratié C., Edition 2006 - Manuel du Réseau de Mesures de la Qualité des Sols (RMQS). Unité InfoSol. INRA Orléans. France. 190 p.
- Khun M., 2008 - Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28, 5. 26 p.
- Kuhn M., Weston S., Keefer C., Coulter C., 2013. C code for Cubist by Ross Quinlan - Cubist: Rule- and Instance-Based Regression Modeling. R package version 0.0.13. <http://CRAN.R-project.org/package=Cubist>
- Kuhn M. Contributions from Jed Wing, Steve Weston, Andre Williams, Chris Keefer, Allan Engelhardt, Tony Cooper, Zachary Mayer and the R Core Team. 2014. caret: Classification and Regression Training. R package version 6. pp. 0-22. <http://CRAN.R-project.org/package=caret>
- Lacoste M., Lemerrier B., Walter C., 2011 - Regional mapping of soil parent material by machine learning based on point data, *Geomorphology*. 133, pp. 90-99.
- Nocker A., et al., 2007. genotypic microbial community profiling: a critical technical review. *Microb. Ecol.* 54. 276-289.
- Odeh, I.O.A., McBratney A.B., Chittleborough D.J., 1995 - Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma* 67, pp. 3-4, pp. 215-226.
- Paulo J. Ribeiro Jr & Peter J. Diggle geoR: a package for geostatistical analysis R-NEWS, 1(2):15-18. June, 2001
- Ranjard L., Lejon D.P.H., Mougél C., Schehrer L., Merdinoglu D., Chaussod R., 2003 - Sampling strategy in molecular microbial ecology: influence of soil sample size on DNA fingerprinting analysis of fungal and bacterial communities, *Environmental Microbiology*, 5 (11), pp. 1111-1120.
- Ranjard L., Dequiedt S., Chemidlin Prévost-Bouré N., Thioulouse J., Saby N.P.A., Lelièvre M., Morin F.E.R., Bispo A., Jolivet C., Arrouays D., Lemanceau P., 2013 - Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity, *Nature Communications*. 4, 1434. 10.
- Ridgeway G. with contributions from others. 2013 - gbm: Generalized Boosted Regression Models. R package version 2.1. <http://CRAN.R-project.org/package=gbm>
- Saby N.P.A., Thioulouse J., Jolivet C., Ratié C., Boulonne L., Bispo A., Arrouays D., 2009 - Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data. *Science of the Total Environment*. 407, pp. 5644-5652.
- Saby, N.P.A., Marchant B.P., Lark R.M., Jolivet C., Arrouays D., 2011 - Robust geostatistical prediction of trace elements across France. *Geoderma*, 162 (3-4), pp. 303-311.
- Stein, M. Interpolation of Spatial Data. 1999 - New York: Springer. 73 p.
- Van der Gucht K., Cottelie K., Muylaert K., Vloemans N., Cousin S., Declerck S., Jeppesen E., Conde-Porcuna J. M, Schwenk K., Zwart G., Degans H., Vyverman W., De Meester L., 2007 - The power of species sorting: Local factors drive bacterial community composition over a wide range of spatial scales. Proceedings of the National Academy of Sciences of the USA. 104. 20404-20409. Webster et Oliver, 2007, Geostatistics for environmental scientists. 2nd edition. John Wiley & Sons Ltd, Chichester UK.

