



HAL
open science

Essais de représentation cartographique de l'incertitude pour les utilisateurs de cartographies des sols par modélisation statistique

Thomas Loiseau, Anne C Richer-De-Forges, Pierre Roudier, Christophe Ducommun, Songchao Chen, Philippe Lagacherie, Dominique. Arrouays

► To cite this version:

Thomas Loiseau, Anne C Richer-De-Forges, Pierre Roudier, Christophe Ducommun, Songchao Chen, et al.. Essais de représentation cartographique de l'incertitude pour les utilisateurs de cartographies des sols par modélisation statistique. *Étude et Gestion des Sols*, 2020, 27 (1), pp.257-275. hal-03141430

HAL Id: hal-03141430

<https://hal.inrae.fr/hal-03141430v1>

Submitted on 15 Feb 2021

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 - NonCommercial - NoDerivatives 4.0
International License

Essais de représentation cartographique de l'incertitude pour les utilisateurs de cartographies des sols par modélisation statistique

T. Loiseau⁽¹⁾, A. C. Richer-de-Forges⁽¹⁾, P. Roudier⁽²⁾, C. Ducommun⁽³⁾, S. Chen⁽¹⁾, P. Lagacherie⁽⁴⁾ et D. Arrouays⁽¹⁾

1) INRAE, InfoSol, 45075, Orléans, France.

2) Manaaki Whenua - Landcare Research, Palmerston North, Nouvelle-Zélande.

3) Institut Agro, Agrocampus-Ouest, Angers, France.

4) LISAH, University Montpellier, INRAE, IRD, Institut Agro, Montpellier 34060, France.

* : Auteur correspondant : dominique.arrouays@inrae.fr

RÉSUMÉ

La cartographie des sols par modélisation statistique (CSMS), est passée aujourd'hui au stade de la production opérationnelle. L'un de ses avantages majeurs est la possibilité de cartographier, au-delà de simples valeurs moyennes ou modales, des incertitudes, souvent matérialisées par des intervalles de confiance des valeurs prédites. Pour autant, l'incertitude reste une notion qui est souvent difficile à transmettre aux utilisateurs pour différentes raisons. Ces derniers sont souvent troublés par le lien implicite, bien qu'incorrect en CSMS, entre la résolution spatiale et la précision sémantique de l'information. Nous proposons ici quelques modes de visualisation de l'incertitude possibles afin de mieux transmettre cette notion aux utilisateurs. En discussion, nous insistons sur le fait que l'objectif final de la CSMS devrait être de produire des cartes de probabilités de distribution de paramètres et que la communication sur l'importance de l'incertitude serait probablement efficace si elle était réalisée, non seulement sur les paramètres pédologiques, mais aussi sur les sorties des modèles ou des simulations les utilisant.

Mots-clés

Cartographie des Sols par Modélisation Statistique, cartographie numérique des sols, incertitude, communication, visualisation, utilisateur.

SUMMARY

SEARCHING FOR WAYS TO VISUALIZE UNCERTAINTY FOR DIGITAL SOIL MAPPING END-USERS

Digital soil mapping (DSM) has now moved from the academic and research domains to operational use. DSM methods bring one major advantage compared to conventional soil maps: in

Comment citer cet article :

Loiseau T., Richer-de-Forges A.C., Roudier P., Ducommun C., Chen S., Lagacherie P et Arrouays D., 2020 - Essais de représentation cartographique de l'incertitude pour les utilisateurs de cartographies des sols par modélisation statistique, *Etude et Gestion des Sols*, 27, 257-275

Comment télécharger cet article :

<https://www.afes.fr/publications/revue-etude-et-gestion-des-sols/volume-27/>

Comment consulter/télécharger

tous les articles de la revue EGS :

<https://www.afes.fr/publications/revue-etude-et-gestion-des-sols/>

top of providing quantitative prediction of soil attributes, it also provides estimates of the uncertainty associated to these predictions, often expressed as prediction intervals. End users and decision makers, however, frequently feel uneasy with uncertainty. End users are often confused by the implicit, yet incorrect, link between spatial resolution and uncertainty. Here we propose some simple ways to visualize the uncertainty in order to better communicate to end users about it. In the discussion we stress that the final goal of DSM should be to provide the complete probability distribution functions of the parameters of interest. Moreover, we feel that communication about uncertainty would be much more efficient if it was based, not only on the primary soil attributes, but on the consequences of their uncertainties when they are used as inputs for final products delivered by end-users or modelers.

Key-words

Digital soil mapping, uncertainty, communication, visualization, end-user.

RESUMEN**ENSAYOS DE REPRESENTACIÓN CARTOGRÁFICA DE LA INCERTIDUMBRE PARA LOS USUARIOS DE CARTOGRAFÍAS DE SUELOS POR MODELIZACIÓN ESTADÍSTICA**

La cartografía de suelos por modelización estadística (CSMS), pasó ahora al estadio de la producción operacional. Uno de sus ventajas mayores es la posibilidad de cartografiar, más allá de simples valores promedios o modales, incertidumbres, generalmente materializadas por intervalos de confianza de los valores predichos. No obstante, la incertidumbre queda una noción que es generalmente difícil a transmitir a los usuarios por diferentes razones. A menudo, los utilizadores están preocupados por el vínculo implícito, aunque incorrecto en CSMS, entre la resolución espacial y la precisión semántica de la información. Proponemos aquí algunos modos posibles de visualización de la incertidumbre a fin de transmitir mejor esta noción a los utilizadores. En discusión, insistimos sobre el hecho que el objetivo final de la CSMS debería ser producir mapas de probabilidades de distribución de parámetros y que la comunicación sobre la importancia de la incertidumbre sería probablemente eficaz si se realizaba, no solamente sobre los parámetros pedológicos, pero también sobre las salidas de los modelos o de las simulaciones que los utilizan.

Palabras clave

Cartografía de suelos por modelización estadística, cartografía numérica de suelos, comunicación, visualización, usuarios.

1. INTRODUCTION

La cartographie des sols par modélisation statistique (CSMS), souvent aussi appelée cartographie numérique des sols (CNS, Lagacherie *et al.*, 2013), et en anglais « digital soil mapping » (DSM, McBratney *et al.*, 2003), est passée progressivement du stade de la recherche académique au stade de la production opérationnelle (Minasny et McBratney, 2016 ; Arrouays *et al.*, 2017a ; Hengl *et al.*, 2017). Face à l'impossibilité financière de réaliser des cartes à moyenne échelle sur l'ensemble du territoire Français, elle a été proposée comme une méthode à mettre en œuvre, déclinée selon différents scénarios (Voltz *et al.*, 2018, 2020), et semble également la principale voie méthodologique recommandée pour produire des prédictions harmonisées des caractéristiques des sols à l'échelle mondiale (Arrouays *et al.*, 2014a ; Hengl *et al.*, 2017). Ses avantages ont été décrits dans de nombreuses publications (e.g., McBratney *et al.*, 2003 ; Lagacherie *et al.*, 2013 ; Minassy et McBratney 2016 ; Arrouays *et al.*, 2017b, 2020) et de nombreux ouvrages (e.g., Lagacherie *et al.*, 2006 ; Hartemink *et al.*, 2008 ; Boettinger *et al.*, 2010 ; Minasny *et al.*, 2012 ; Arrouays *et al.*, 2014b ; Arrouays *et al.*, 2018). Ils peuvent se résumer en quelques points essentiels :

- i) Elle permet une prédiction quantitative de propriétés des sols.
- ii) Elle est moins coûteuse et plus rapide à produire qu'une carte conventionnelle.
- iii) Elle est reproductible, peut être facilement mise à jour, voire automatisée.
- iv) Elle offre la possibilité de relier et de quantifier explicitement l'importance des facteurs de contrôle sur la distribution des sols et/ou leurs propriétés.
- v) Elle permet de produire des estimations des incertitudes associées aux prédictions.

Ce dernier point est d'une importance capitale pour les utilisateurs finaux des cartes produites :

- i) Il permet de juger de la fiabilité d'une carte avant de l'utiliser comme outil d'aide à la décision.
- ii) Il permet de mettre en évidence, voire de quantifier, la propagation des erreurs dans les résultats lorsqu'une donnée sur les sols est introduite comme donnée d'entrée dans un modèle.
- iii) Il permet de mettre en évidence les zones les plus imprécises où l'acquisition de données complémentaires devrait être prioritaire.

Pour autant, l'incertitude, reste une notion qui est souvent difficile à transmettre aux utilisateurs pour différentes raisons. Les utilisateurs sont souvent troublés par le lien implicite, bien qu'incorrect en CSMS, entre la résolution spatiale et la précision sémantique de l'information. Certains pensent qu'une information délivrée à une résolution spatiale fine est nécessairement plus juste et plus précise qu'une information

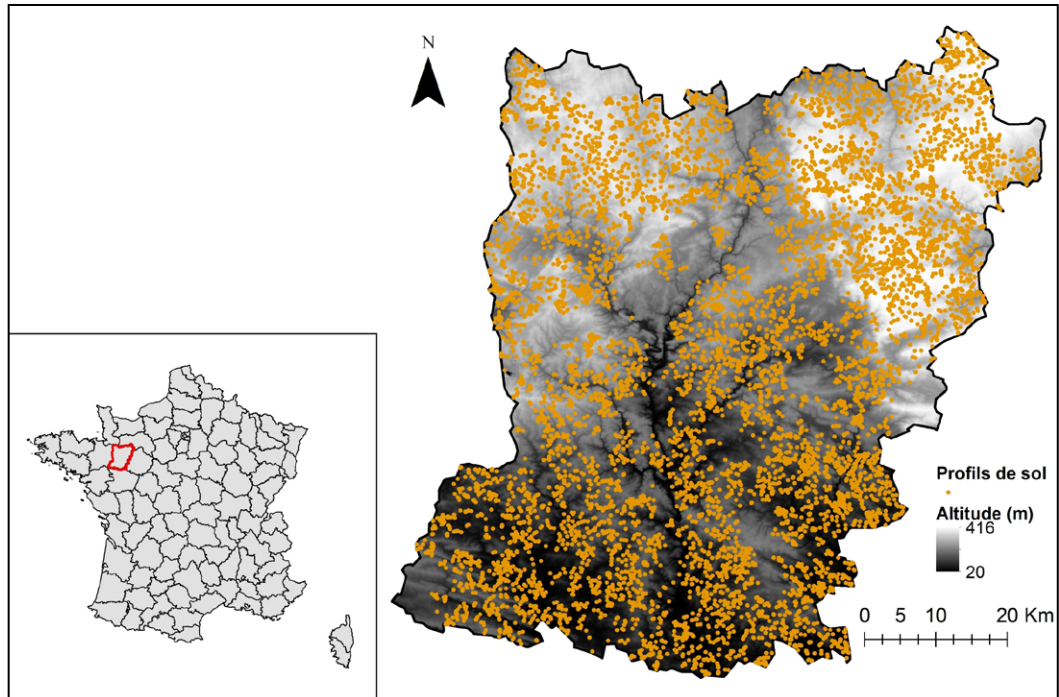
délivrée à une résolution plus grossière. D'autres, conscients de cet état de fait, craignent qu'une carte délivrée à une résolution très fine soit considérée comme la plus juste possible - représentant, en quelque sorte, la « vérité » - et que l'information sur l'incertitude soit simplement ignorée (Richer-de-Forges *et al.*, 2019 ; Arrouays *et al.*, 2020).

Bien souvent aussi, les utilisateurs sont mal à l'aise avec la notion d'incertitude et ne savent, ni comment l'utiliser, ni comment la communiquer (Voltz *et al.*, 2018 ; 2020). Il existe ainsi un besoin certain de mieux expliquer ce que constitue l'incertitude et de la rendre plus facile à comprendre, en particulier visuellement. Il est vrai que la multiplication des indicateurs de performance des prédictions (intervalles de confiance, coefficients de détermination, erreurs moyennes, erreurs quadratiques moyennes, coefficients de variation, pour ne citer que les plus simples) ne facilite pas la compréhension par un néophyte en la matière. En cartographie traditionnelle, des efforts ont été conduits depuis longtemps pour définir les différentes conceptions de la précision d'une carte des sols (Legros, 1973 ; 1978a, 1978b), sans qu'il soit véritablement possible de juger si ces efforts ont réellement permis de faire passer ces notions aux utilisateurs finaux. La notion de « pureté » d'une carte traditionnelle, ou encore celle de la définition des proportions surfaciques de différents types de sols au sein d'une unité complexe (cas le plus fréquent dans les Référentiel Régionaux Pédologiques (RRP) à 1/250 000 réalisés en France) reste parfois mal interprétée ou mal utilisée. Le cas le plus fréquent d'utilisation erronée est de prendre uniquement en compte les caractéristiques modales de l'unité typologique de sols dominante pour appliquer un modèle et de considérer que le résultat de ce modèle est le résultat dominant sur l'unité cartographique complexe.

Il semble qu'avec l'apparition de la cartographie numérique des sols des situations en partie similaires puissent se produire ; autrement dit, que l'utilisateur ne prenne en compte que la valeur moyenne ou modale d'une ou plusieurs propriétés prédites sur un « pixel » comme donnée d'entrée d'un modèle. Devant ce risque, il est évident qu'il faut mieux communiquer sur cette incertitude. L'un des moyens possibles est de tenter de la rendre plus visuelle tout en expliquant au mieux ce que cette visualisation signifie. Il s'agit ici de tenter d'améliorer la sémiologie des cartes d'incertitudes en fonction du message que l'on cherche à faire passer à l'utilisateur. Nous prenons ici pour exemple les CSMS de la granulométrie (teneurs en sable, limon et argile) des horizons de surface du département de la Mayenne. Nous avons choisi ce département car c'est celui qui dispose de la plus forte densité de mesures ponctuelles et nous avons illustré cet exemple par la granulométrie de surface (0-30 cm) car c'est une information très fréquemment demandée par les utilisateurs (applications à l'érosion, aux contraintes vis-à-vis de certaines cultures, aux jours disponibles, au réservoir en eau utile...).

Figure 1 :
Représentation des données pédologiques ponctuelles sur le département de la Mayenne.

Figure 1: Maps of point data on the Mayenne department.



2. MATÉRIEL ET MÉTHODES

2.1 Site d'étude

Le département de la Mayenne présente la densité d'échantillonnage sur les données pédologiques la plus importante en France métropolitaine, environ 8 100 points soit une densité de 1,6 point par km² (figure 1). La densité de points n'est pas homogène, car elle résulte d'une compilation de nombreuses études locales et non d'une étude systématique, ce qui est fréquemment le cas pour les RRP (Laroche *et al.*, 2014). De plus, ce département possède une grande diversité texturale de surface, héritée de matériaux lithologiques variés (argilites, limons des plateaux, grès sableux, schistes calcaireux), ou issue d'une évolution pédologique poussée (horizons éluviaux albiqes issus de Luvisols érodés). Par conséquent, la diversité des sols et de leurs teneurs (figure 2) en chaque fraction granulométrique s'en retrouvent grandement impactées et nous permettent une étude des variations de prédiction en CSMS.

2.2 Prédiction ponctuelle de la granulométrie de surface

Afin d'obtenir une prédiction spatiale tenant compte de la complémentarité des fractions granulométriques dans les sols (Lark et Bishop, 2007), ces données ont été log-transformées en suivant la méthodologie des données compositionnelles ou

couramment appelée « transformation alr » (Aitchison, 1982). Cette transformation a été effectuée comme suit :

$$Argile_{atr} = \ln\left(\frac{Argile}{Sable}\right)$$

$$Limon_{atr} = \ln\left(\frac{Limon}{Sable}\right)$$

Dans une seconde étape et pour exprimer les résultats de prédiction finaux en g.kg⁻¹, les valeurs ont été transformées en suivant la méthodologie inverse :

$$Argile = \frac{e^{Argile_{atr}}}{(1 + e^{Argile_{atr}} + e^{Limon_{atr}})} \times 1000$$

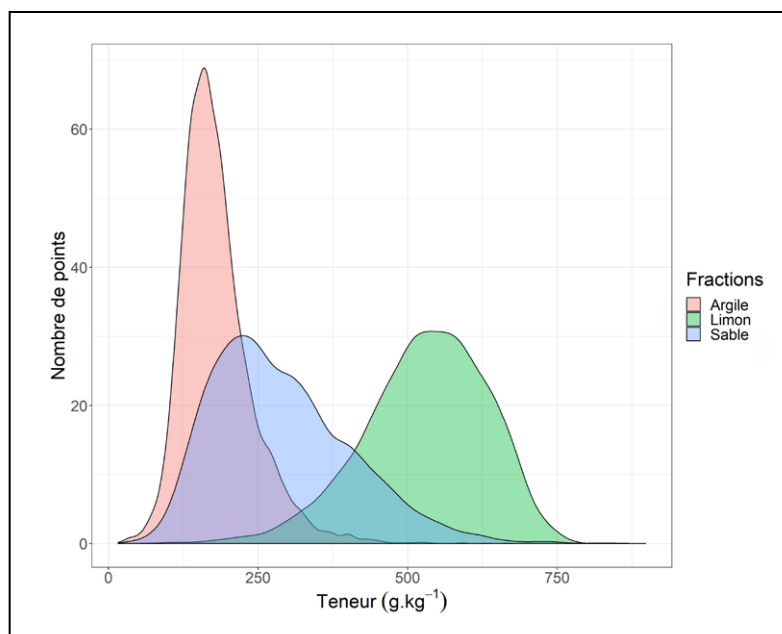
$$Limon = \frac{e^{Limon_{atr}}}{(1 + e^{Argile_{atr}} + e^{Limon_{atr}})} \times 1000$$

$$Sable = \frac{1}{(1 + e^{Argile_{atr}} + e^{Limon_{atr}})} \times 1000$$

Ces fractions granulométriques ont par la suite été modélisées en utilisant des données environnementales, sélectionnées au préalable pour leur capacité à prédire la granulométrie (Loiseau *et al.*, 2019 ; 2020) et les redistributions de surface (érosion, colluvionnement...) (tableau 1), et injectées dans un modèle non linéaire de type forêt d'arbres aléatoire : « Quantile Regression Forest » (QRF, Meinshausen, 2006, en français : régression quantile sur des forêts d'arbres aléatoires) pour une spatialisation des prédictions à 90m de résolution (Arrouays *et al.*, 2018).

Figure 2 : Distribution de la teneur de chaque fraction granulométrique sur les 8 100 profils de sol (horizon de surface).

Figure 2: Distributions of the values of each particle-size fraction for the 8,100 soil profiles (topsoil).



En CSMS, la séparation du jeu ponctuel global, inféré des données environnementales, en un jeu de calibration et en un jeu de validation, est fréquemment utilisée. Le jeu de calibration est injecté dans le modèle pour l'ajuster en fonction de la distribution des données observées. La particularité des modèles du type forêt d'arbres aléatoire est la construction des arbres de décisions, au fur et à mesure des simulations, par l'apprentissage par échantillonnage aléatoire (Breiman, L. 2001). Le modèle, une fois calibré après n arbres réalisés, peut être appliqué pour fournir une prédiction des fractions granulométriques pour chaque pixel présentant des données environnementales aussi appelées covariables.

La particularité du modèle QRF est de permettre l'accès à la totalité de la distribution de prédiction pour les n arbres simulés pour la calibration, dans notre cas 1000, pour l'estimation en chaque pixel. C'est une méthode jugée robuste en comparaison de certains modèles ne reposant que sur l'estimation d'une valeur moyenne de prédiction (Vaysse et Lagacherie, 2017). En effet, les prédictions peuvent alors être déterminées par la méthode de la médiane (0,5) et de différents quantiles dont certains peuvent être choisis, plus ou moins arbitrairement, pour évaluer des intervalles de confiance des prédictions (IC).

La performance des prédictions est ensuite évaluée par la méthode de validation croisée « 16-fold » nous permettant d'estimer des indicateurs statistiques communément utilisés en CSMS. Cette méthode consiste à diviser aléatoirement le jeu de données en 16 groupes d'effectifs égaux, puis à n'utiliser que 15 groupes pour prédire, et le groupe restant pour valider. L'opération est donc conduite 16 fois. Pour obtenir

des indicateurs statistiques fiables, le partage en 16 groupes aléatoires est réalisé 100 fois, et les indicateurs de performance correspondent à la moyenne des valeurs observées sur ces ensembles de prédiction/validation. Ces indicateurs visent à comparer les prédictions avec les valeurs observées sur le terrain en estimant leur adéquation (R^2), leur différence statistique (erreur moyenne), leur dispersion (racine carrée de l'erreur quadratique moyenne ou RMSE) et leur distribution (concordance, pourcentage de points dans l'intervalle de confiance à 90 % ou PICP90 %).

2.3 Calcul et représentations des incertitudes

L'utilisation des valeurs extrêmes pour représenter les incertitudes est communément admise en cartographie traditionnelle, comme par exemple, en France, dans le cadre du volet Référentiel Régional Pédologique du programme Inventaire Gestion et Conservation des Sols (Laroche *et al.*, 2014), ou aux Etats-Unis, dans le cas du programme STATSGO2 (Helmick *et al.*, 2014). En CSMS, ces valeurs bornes inférieures et supérieures sont généralement estimées par des valeurs de quantiles à 5 et 95 % (Heuvelink, 2014 ; Arrouays *et al.*, 2014 ; Helmick *et al.*, 2014 ; Odgers *et al.*, 2014 ; Libohova *et al.*, 2016). L'estimation des bornes inférieures et supérieures d'une propriété à un pixel donné nous permet de connaître l'étendue de la variation de la prédiction du modèle à un endroit précis et donc d'estimer sa précision. Dans cette optique, chaque spatialisation s'accompagne d'une prédiction de l'intervalle de confiance (IC) à 90 %, par la spatialisation des quantiles 5 % et 95 %.

Tableau 1 : Covariables sélectionnées pour calibrer le modèle QRF puis prédire la distribution spatiale des fractions granulométriques.**Table 1**: Selected co-variates for the prediction model.

Co-variables	Résolution ou échelle	Données	Référence
Climat			
Type de climat (8 Classes)	250 m	Qualitative	Joly <i>et al.</i> (2010)
Végétation			
Type de forêt (10 classes)	Départementale	Qualitative	Inventaire Forestier National (2006)
Occupation des sols 2016 (17 classes)	10 m	Qualitative	CESBIO (2016)
NDVI : 3 premières composantes principales*	500 m	Quantitative	Loiseau <i>et al.</i> (2019)
Données sols			
Teneur en argile LUCAS (%)	500 m	Quantitative	Ballabio <i>et al.</i> (2016)
Teneur en limon LUCAS (%)	500 m	Quantitative	Ballabio <i>et al.</i> (2016)
Teneur en sable LUCAS (%)	500 m	Quantitative	Ballabio <i>et al.</i> (2016)
Taux de développement et de persistance du réseau fluvial (IDPR)	1/50 000	Quantitative	BRGM (2014)
Topographie			
Altitude (BDAlti, donnée primaire)	25 m	Quantitative	IGN (2014)
Compound topographic index (CTI)	25 m	Quantitative	Calculé d'après BDAlti 25m
Multi-Resolution Valley Bottom Flatness (MRVBF)	25 m	Quantitative	Calculé d'après BDAlti 25m
Multi-Resolution Ridge Top Flatness (MRRTF)	25 m	Quantitative	Calculé d'après BDAlti 25m
Rugosité	25 m	Quantitative	Calculé d'après BDAlti 25m
Courbure totale	25 m	Quantitative	Calculé d'après BDAlti 25m
Exposition	25 m	Quantitative	Calculé d'après BDAlti 25m
Position dans la pente	25 m	Quantitative	Calculé d'après BDAlti 25m
Pente, Cos(Pente)	25 m	Quantitative	Calculé d'après BDAlti 25m
Matériau parental			
Thorium (Gamma)	250 m	Quantitative	Bonijoly <i>et al.</i> (1999)
Uranium (Gamma)	250 m	Quantitative	Bonijoly <i>et al.</i> (1999)
Potassium (Gamma)	250 m	Quantitative	Bonijoly <i>et al.</i> (1999)
Th/K ratio (Gamma)	250 m	Quantitative	Bonijoly <i>et al.</i> (1999)
Th/U ratio (Gamma)	250 m	Quantitative	Bonijoly <i>et al.</i> (1999)
Lithologie simplifiée (14 classes)	1/50 000	Qualitative	Loiseau <i>et al.</i> , 2020, d'après Vernhet (2010)

* 3 premières composantes principales issues de l'analyse en composantes principales des données NDVI mensuelles (2003 et 2016) des missions MODIS et Copernicus.

Notre travail vise à proposer des solutions pertinentes de communication de l'incertitude auprès d'utilisateurs n'étant pas familiers avec les techniques utilisées en cartographie numérique. Quatre types de représentation ont été testés :

- (A) : Une représentation de la prédiction médiane représentée conjointement avec deux quantiles extrêmes spatialisés. Cette

méthode, désormais relativement classique (Mulder *et al.*, 2016a, 2016b ; Grundy *et al.*, 2015 ; Helmick *et al.*, 2014), a pour objectif de comparer de manière cartographique les zones les plus variables au sein d'une étendue géographique et de fournir à l'utilisateur 3 scénarios possibles (valeur prédite la plus faible, valeur médiane (ou moyenne), et valeur la plus élevée).

- (B) : Une technique de nuance de la symbologie de valeur de la variable prédite en fonction du niveau d'incertitude de l'estimation (Correl *et al.*, 2018). Le rendu consiste en une carte unique de la propriété cible spatialisée, mais utilisant une rampe de couleur multivariée, l'une pour l'attribut, l'autre pour l'incertitude (<https://research.tableau.com/sites/default/files/2018-UncertaintyPalettes-CHI.pdf>), dotée d'un gradient plus ou moins terne de la valeur prédite associé à son intervalle de confiance (R package "multiscales" (Wilke, 2020)).
- (C) : Une représentation de la donnée médiane accolée à une cartographie de l'incertitude relative (ex : IC 90 % / prédiction). De la même manière que la première méthode, cette représentation vise à estimer spatialement les zones présentant le plus de variations, mais de manière plus simple.
- (D) : Une représentation de la prédiction médiane avec une carte d'incertitude globale (ex : IC 90 %), comparée à une carte de référence. Un exemple peut être une utilisation de la carte du Référentiel Régional Pédologique obtenue en faisant une moyenne pondérée par la surface des valeurs modales des strates de surface des différentes unités typologiques constituant une unité cartographique à 1/250 000. Cette représentation a pour but de comparer la structuration spatiale de la prédiction et de l'incertitude avec la référence issue de l'expertise des auteurs des cartes traditionnelles.

Ces essais de représentations cartographiques visent à faciliter la compréhension de l'incertitude grâce à sa spatialisation, mais aussi à connaître plus en détail les zones les plus imprécises justifiant l'acquisition de profils pédologiques complémentaires ou de cartes de type « secteurs de référence » (Favrot, 1989, Voltz *et al.*, 1997 ; Voltz *et al.*, 2020).

3. RÉSULTATS

3.1 Performances statistiques

Afin d'évaluer les performances statistiques des modèles d'estimation, des indicateurs issus de validations croisées ont été calculés et moyennés. Ils sont présentés dans le tableau 2.

Nous pouvons constater que les performances statistiques des modèles peuvent être classées du plus performant au moins performant en suivant cet ordre : Argile, Limon et Sable. En effet, l'erreur moyenne est légèrement négative pour les estimations de la teneur en argile, ce qui dénote une légère sous-estimation, et elle augmente pour les estimations des teneurs en sable (sous-estimée) et en limon (surestimée). Le RMSE suit également cette tendance et atteste d'une dispersion plus importante des valeurs estimées des fractions limon et sable par rapport aux valeurs de terrain. Néanmoins, il est également à noter que d'autres indices statistiques suggèrent le contraire. En effet, nous constatons que les valeurs du R^2 et de la concordance montrent une meilleure performance en ce qui concerne la restitution de la distribution d'origine des fractions sable et limon, en comparaison avec l'argile. Seul le PICP reste stable et proche de la valeur théorique idéale de 90 %, ce qui laisse penser que l'incertitude liée au modèle est relativement fiable pour les estimations de l'argile et du limon. L'absence d'évaluation du PICP pour le sable est due au fait qu'aucun modèle QRF n'a été calibré pour son estimation. En effet, les valeurs obtenues pour cette fraction granulométrique ont été obtenues par transformation inverse (*voir section 2.2*) à partir des prédictions des deux autres fractions.

Même si l'importance des co-variables n'est pas l'objet de l'article, Loiseau *et al.*, (2020) ont récemment détaillé cette importance et montré que les variables les plus importantes étaient celles issues de la spectrométrie gamma aéroportée, suivies de la lithologie et du relief. Les variables issues de LUCAS originellement au pas de 500 mètres sortent dans le premier tiers des variables explicatives mais ne sont pas dominantes, ce qui montre assez logiquement que la densité de points de LUCAS et sa résolution ne capturent pas toute la variabilité observée avec une forte densité de points.

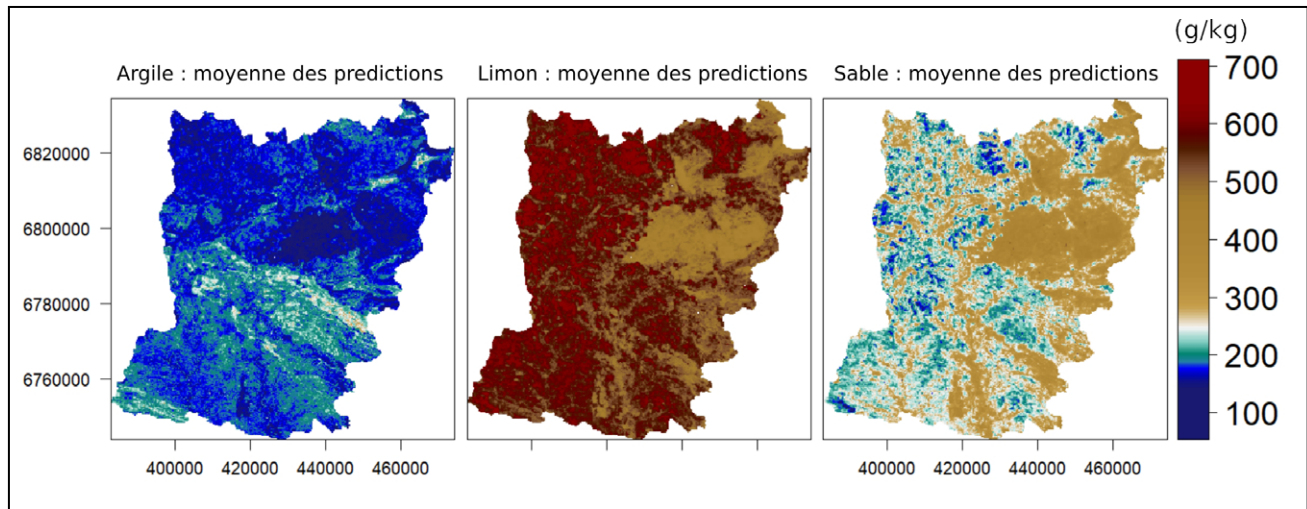
Tableau 2 : Performances moyennes des estimations des fractions granulométriques sur le département de la Mayenne.

Table 2: Average performance of the predictions of particle-size fractions on the Mayenne department.

Indicateur	Argile	Limon	Sable
R^2	0,26	0,32	0,40
RMSE (g.kg ⁻¹)	52,04	85,98	89,84
Erreur moyenne (g.kg ⁻¹)	-2,17	11,07	-8,89
Concordance (%)	38,73	48,54	55,47
PICP90% (%)	92,52	92,16	/

Figure 3 : Prédictions des médianes moyennées des fractions granulométriques.

Figure 3: Predictions of the average medians of particle-size fractions.



3.2. Représentations cartographiques

Les valeurs des prédictions médianes observées *figure 3* donnent un aperçu de la distribution des fractions granulométriques sur le département.

En effet, nous pouvons observer pour l'argile une zone centrale axiale orientée Nord-Ouest/Sud-Est au centre du département, ainsi qu'une zone périphérique au sud-ouest, présentant des valeurs fortes pour la prédiction médiane, mais pouvant atteindre des valeurs minimales proches de 150 g.kg⁻¹ et des valeurs maximales supérieures à 300 g.kg⁻¹. Ces zones sont surtout caractérisées par l'action combinée d'une lithologie riche en argiles ou ses précurseurs (argilites et roches micaschisteuses), avec une morphologie globalement concave jouant le rôle de réceptacle (et de néoformation) pour les particules fines provenant des zones en amont.

Nous pouvons également constater la présence d'une limite assez nette Est-Ouest concernant les fractions limoneuses, pouvant atteindre des valeurs minimales d'estimations proches de 400-450 g.kg⁻¹ et des valeurs maximales supérieures à 600 g.kg⁻¹. La zone de plus faible concentration est située au nord-est du département avec des valeurs minimales pouvant atteindre 100-200 g.kg⁻¹ et des valeurs maximales proches de 500 g.kg⁻¹. Cette distribution spatiale trouve là encore son origine à la fois dans la morphologie et la lithologie du territoire. De manière générale, les zones planes favorisent les dépôts limoneux et l'exportation des argiles par lessivage, tandis que les zones granitiques et pentues sont à l'origine d'une tendance plus sableuse.

Enfin, la fraction sableuse présente également une zone de forte concentration au nord-ouest et selon certains axes orientés préférentiellement nord/sud, situés au sud, avec des valeurs d'estimations minimales pouvant atteindre les 300 g.kg⁻¹ et des

valeurs maximales supérieures à 700 g.kg⁻¹, pointant le contexte à la fois granitique septentrional, particulièrement pentu au nord-ouest, et les axes des plaines alluviales se dirigeant vers le sud du département (Oudon, Mayenne, Vaigé...).

La représentation de l'incertitude sous la forme d'une cartographie vise à s'assurer, en plus des résultats statistiques, de la bonne estimation spatiale d'une propriété de sol. Les *figures 4 à 7* représentent des essais cartographiques, issus des mêmes prédictions spatiales.

La représentation des intervalles de confiances à 90 %, ainsi que de la valeur médiane, dans la *figure 3*, nous informent d'une disparité spatiale attendue pour les fractions granulométriques, mais également de leur variation d'estimation au sein du modèle.

Cette représentation nous montre également que les structures spatiales des intervalles de confiance suivent en grande partie la structuration spatiale de l'estimation médiane. Cette structuration spatiale étant par construction dérivée des covariables (ou du moins des relations entre les caractéristiques à prédire et les covariables, il semble donc que l'espace des covariables puisse servir à structurer l'estimation des incertitudes, hypothèse déjà formulée par Malone *et al.* (2011). Ce résultat semble logique, puisque les incertitudes sont celles des prédictions, elles-mêmes définies par le modèle faisant appel aux covariables. Dans le même ordre d'idée, en se fondant sur des résultats de cartographie conventionnelle impliquant l'expertise du pédologue, Murciano *et al.* (2015) ont montré que l'étendue des variables prédites (valeur maximale-valeur minimale) pouvait dépendre à la fois de la valeur modale prédite et de certaines co-variables environnementales.

Ces constatations sont tout autant visibles dans la *figure 5*. Dans cette représentation, les zones les plus ternes (proches du blanc) correspondent aux zones où les estimations des fractions

Figure 4 : Prédictions des médianes moyennées des fractions granulométriques et de leur intervalle de confiance à 90 %.
Figure 4: Predictions of the average particle-size fractions and of their 90% confidence interval.

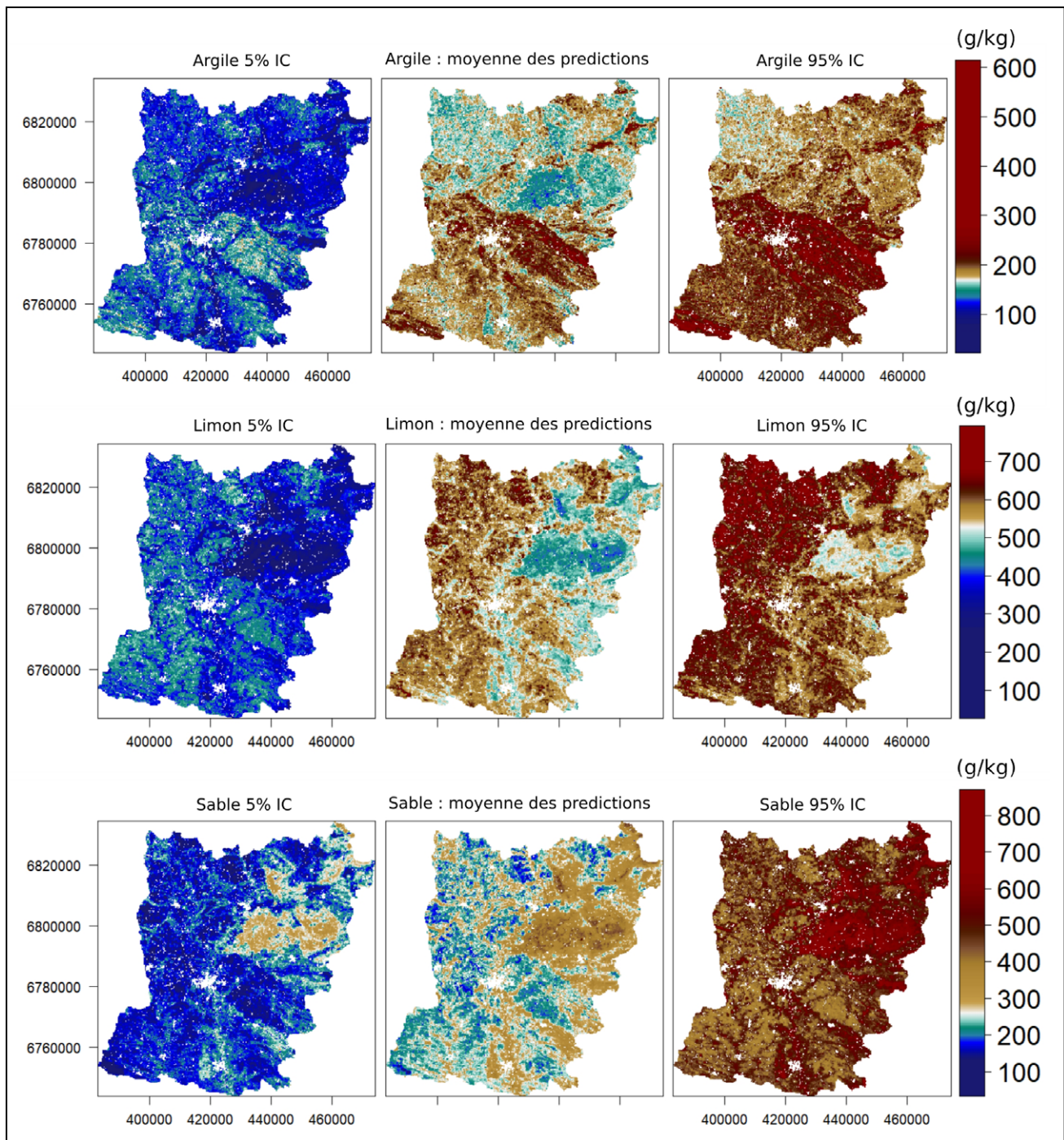


Figure 5 : Représentation des prédictions des fractions granulométriques et de leurs intervalles de confiance combinés sur une même carte, en utilisant une rampe de couleur multivariée, l'une pour l'attribut, l'autre pour l'incertitude (https://research.tableau.com/sites/default/files/2018-UncertaintyPalettes-CHI.pdf), dotée d'un gradient plus ou moins terne de la valeur prédite associée à son intervalle de confiance de 90 %.

Figure 5: Maps of different particle-sizes using the “value suppressing uncertainty visualization” (https://research.tableau.com/sites/default/files/2018-UncertaintyPalettes-CHI.pdf) using the 90% confidence interval.

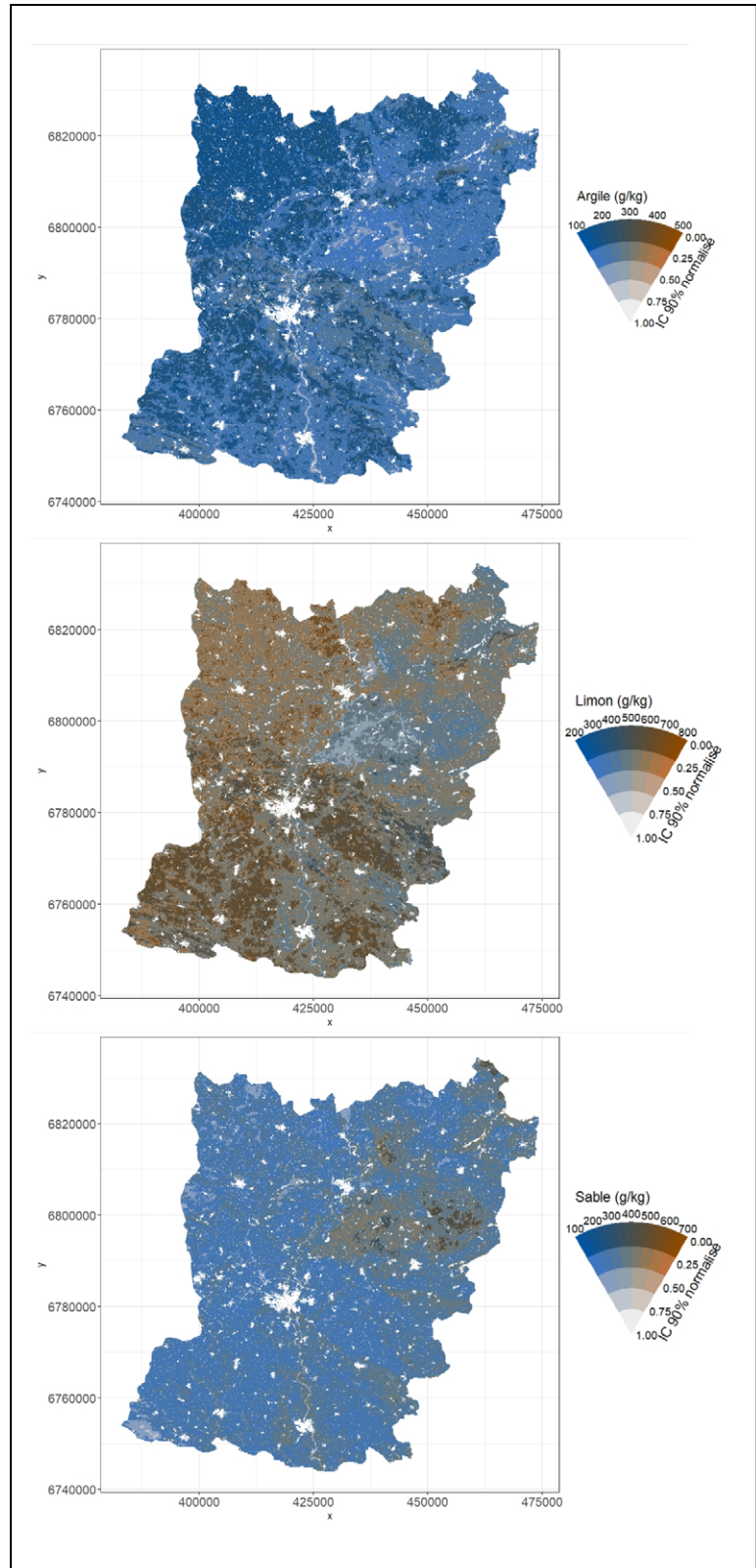
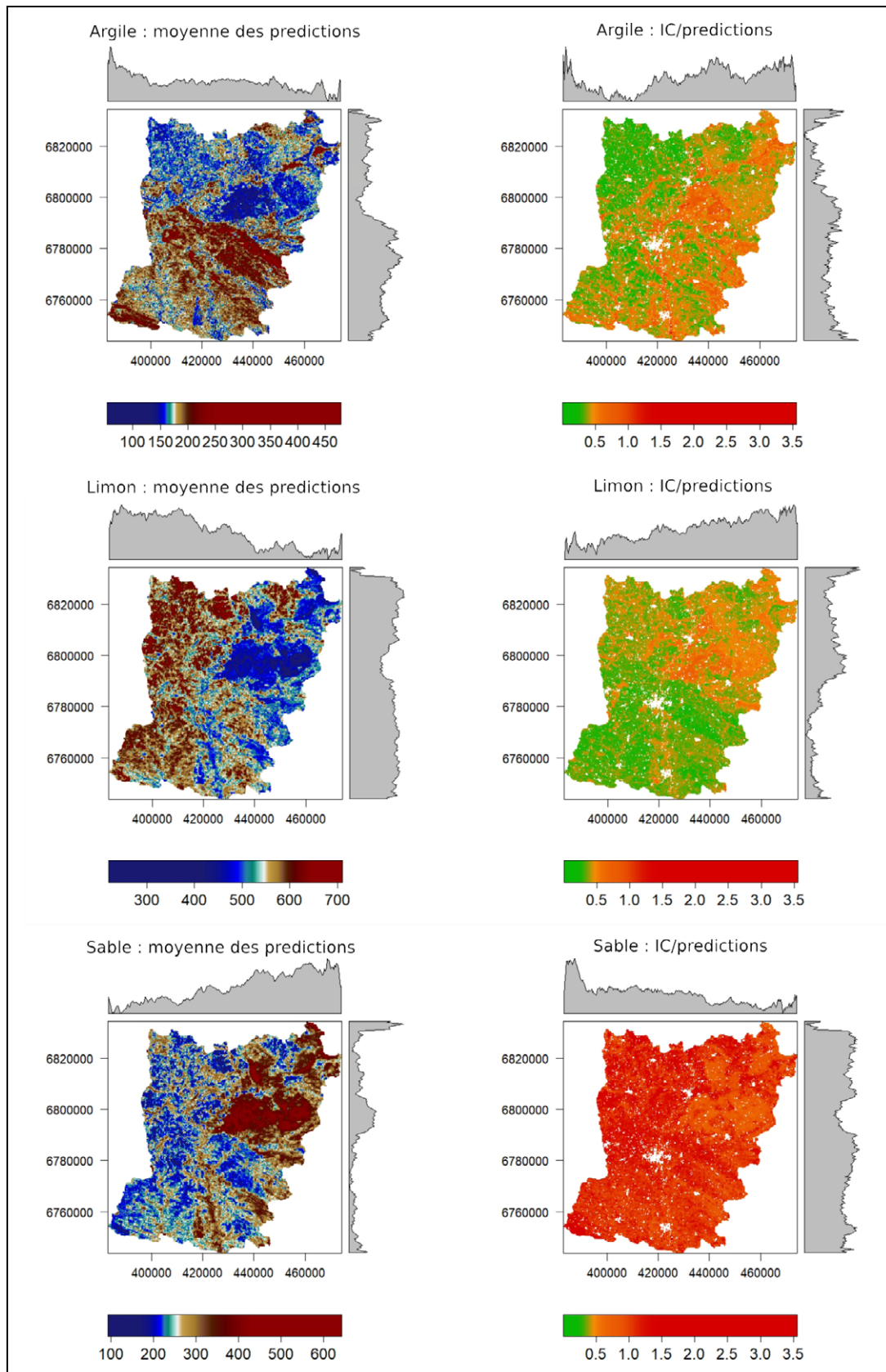


Figure 6 :
Représentation de la variabilité spatiale des prédictions des fractions granulométriques.

Figure 6:
Representation of the spatial variability of the prediction of particle-size fractions.



granulométriques sont les plus variables en fonction des valeurs d'estimations de l'intervalle de confiance choisi (90 %).

Nous pouvons constater que la structuration spatiale reste inchangée, mais que des zones plus marquées se détachent du reste des cartes. En effet, pour certaines fractions, des zones plus pâles présentent des variations d'estimation importantes et correspondent à des zones de forte variabilité observée dans la *figure 4*. De plus, les zones alluviales semblent présenter une incertitude plus importante pour l'argile et les limons. L'avantage de ce type de représentation est qu'elle permet à la fois une visualisation graphique des grandes tendances de la granulométrie mais qu'elle y superpose une information concernant l'incertitude.

La mise en relation de la prédiction médiane et de son intervalle de confiance apporte donc une nouvelle perspective au rendu cartographique. En regroupant ces informations, dans la représentation en *figure 6*, leur comparaison est également possible.

Cette méthode cartographique regroupe à la fois les constatations précédemment observées de la *figure 1*, mais également une relativisation des valeurs de l'intervalle de confiance en fonction de la valeur médiane. En effet, sur cette représentation, plus une zone tend vers le rouge, plus cette zone présente une forte variabilité dans son estimation, c'est-à-dire une distribution plus étalée d'estimations possibles par rapport à la valeur médiane pour un pixel donné. Inversement, plus une zone tend vers le vert, plus les prédictions réalisées sur cette zone sont proches de la valeur médiane d'estimation. Nous pouvons donc juger de la robustesse de la valeur prédite pour chaque pixel de la zone d'étude.

En effet, en reprenant l'exemple de la zone centrale pour la fraction argile, nous pouvons constater que l'intégralité de la zone ne présente pas de fortes variabilités d'estimations. Des zones avec un ratio inférieur à 0,4, donc ayant des valeurs de l'IC deux fois moins importantes que leur valeur médiane, sont voisines de zones où l'IC présente des valeurs égales ou supérieures à la valeur médiane.

A contrario, la zone présentant des teneurs plus faibles, visible à la fois sur l'argile et le limon, au Nord-Est, semble quant à elle présenter une incertitude importante, par rapport à sa valeur médiane, pour l'ensemble de son étendue spatiale.

Il est à noter qu'en utilisant la même méthode pour la fraction sable, c'est la grande majorité du département qui se retrouve avec un ratio supérieur à 1. Néanmoins, nous pouvons y retrouver certaines structures moins incertaines, particulièrement au Nord-Est.

La dernière représentation, en *figure 7*, nous permet de mettre en relation cette structuration spatiale avec le RRP à l'échelle 1/250 000, traité ici en faisant une moyenne des valeurs modales des strates de surface, pondérée par la surface occupée par chaque type de sol, pour chaque unité cartographique de sol. Cette comparaison reste donc à prendre avec précaution mais

permet de contrôler globalement si les structures observées sur les cartes de prédiction présentent des similarités avec celles de la carte au 1/250 000.

En observant les cartes thématiques pour chaque fraction, nous pouvons constater que les grandes structures sont plus ou moins visibles dans les prédictions de CSMS. Nous retrouvons les zones aux teneurs plus élevées en argile au centre et au Sud-Est du département, ainsi que le gradient limoneux Est-Ouest. La correspondance est beaucoup moins évidente pour le sable.

4. DISCUSSION

La représentation de l'incertitude pour les utilisateurs de la CSMS reste un défi tant pour la vulgarisation au plus grand nombre des notions d'incertitudes que pour son utilisation comme outil d'aide à la décision. Dans ce contexte, le travail réalisé visait à suggérer différentes représentations permettant à la fois de visualiser et de comprendre la distribution des prédictions des fractions granulométriques, ainsi que l'incertitude de leurs estimations.

Ces quatre représentations remplissent leur rôle en tant que résultats cartographiques et ont permis l'extraction d'informations nécessaires à la compréhension de la cartographie des propriétés des sols par l'utilisation d'un intervalle de confiance fixe. Cependant, cet intervalle de confiance utilisé ici (90 %), communément utilisé dans le monde scientifique en CSMS, n'a pas forcément son utilité pour les décisionnaires. Nous sommes donc en droit de nous demander si un intervalle de confiance plus faible (60, 65 ou 70 %) ne pourrait pas suffire aux acteurs locaux dans leurs prises de décisions (*figures 8 et 9*). Plus prosaïquement, une information du type « la valeur qui m'intéresse a deux chances sur trois d'être dans un intervalle donné » n'est-elle pas suffisante pour bon nombre d'applications ?

On constate sur la *figure 8* que les structures spatiales se maintiennent mieux entre les cartes extrêmes lorsque l'on passe à un intervalle de confiance de 65 % (soit approximativement 2 chances sur 3) par rapport à ce qui était visible sur la *figure 4*. Les cartes de la *figure 9* sont également plus contrastées que celles de la *figure 5* car l'atténuation des couleurs y est moins forte.

Plus généralement, le choix d'un intervalle de confiance très large conduit à des cartes de valeurs minimales et maximales qui sont extrêmement dissemblables et peuvent amener l'utilisateur à douter fortement de l'utilité des prédictions de valeurs moyennes ou modales. Lorsque l'on sait, par exemple, que la pureté d'une carte traditionnelle, même détaillée, est très rarement supérieure à 80 % (Marsman et de Gruijter, 1986), et qu'à défaut de produits de type CSMS, on utilise fréquemment les valeurs des propriétés des sols décrites dans les bases de

Figure 7 : Comparaison des prédictions granulométriques avec les cartes thématiques issues du RRP 1/250 000. A gauche ; moyenne pondérée des valeurs modales issues du RRP pour les teneurs en argile, limon et sable. A droite : valeurs prédites par CSMS et intervalles de confiance (IC) à 90 % (quantile 95 - quantile 05).

Figure 7: Comparison of the prediction of particle-size distributions with values calculated from the 1:250,000 map. Left side: mean of the modal values predicted for clay, silt and sand contents. Right side: small up-side maps, predicted values by DSM; small down-side maps: 90% confidence intervals.

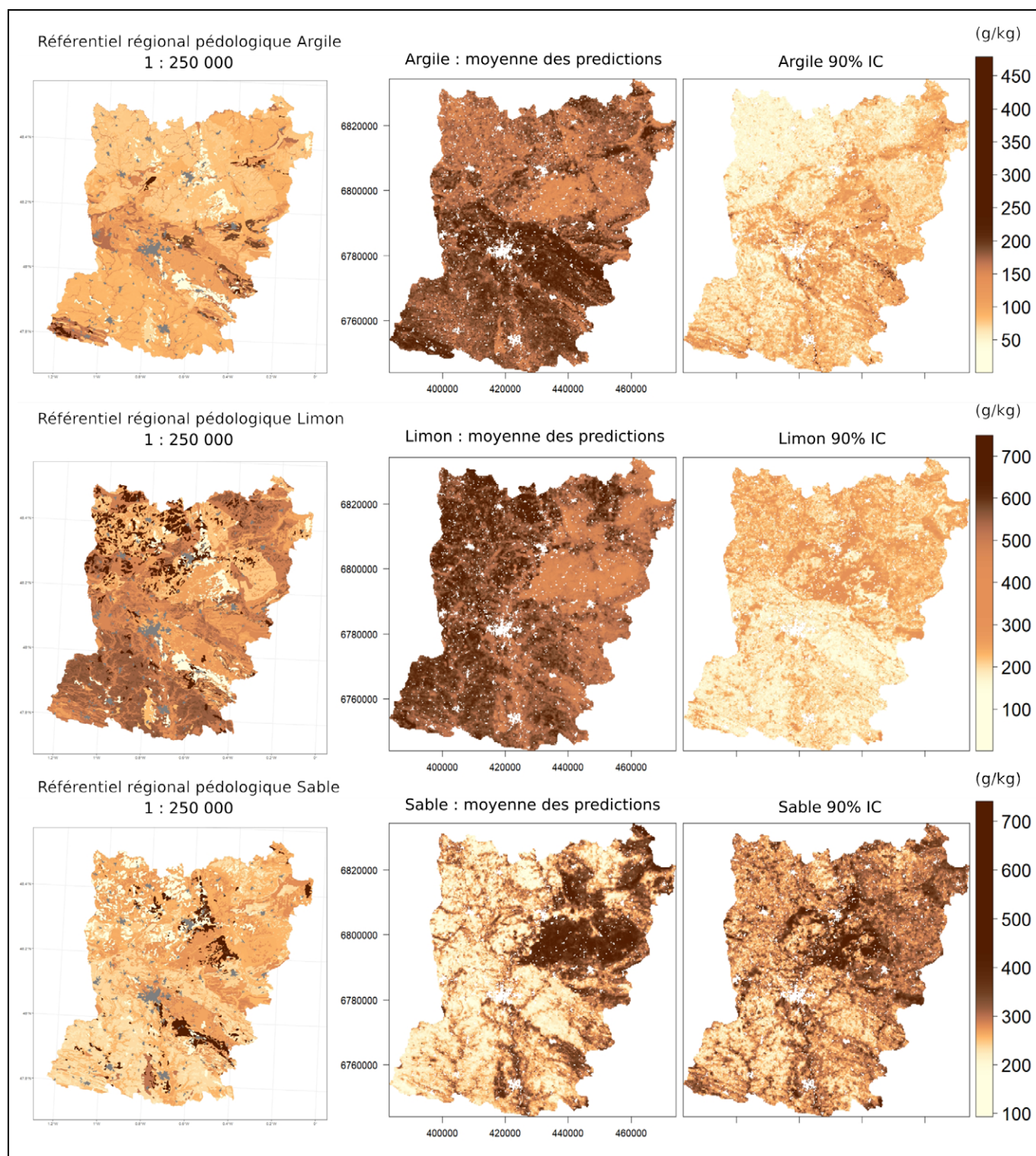


Figure 8 : Prédications des fractions granulométriques et de leur intervalle de confiance à 65 %.

Figure 8: Predictions of the particle-size distributions and their 65% confidence interval.

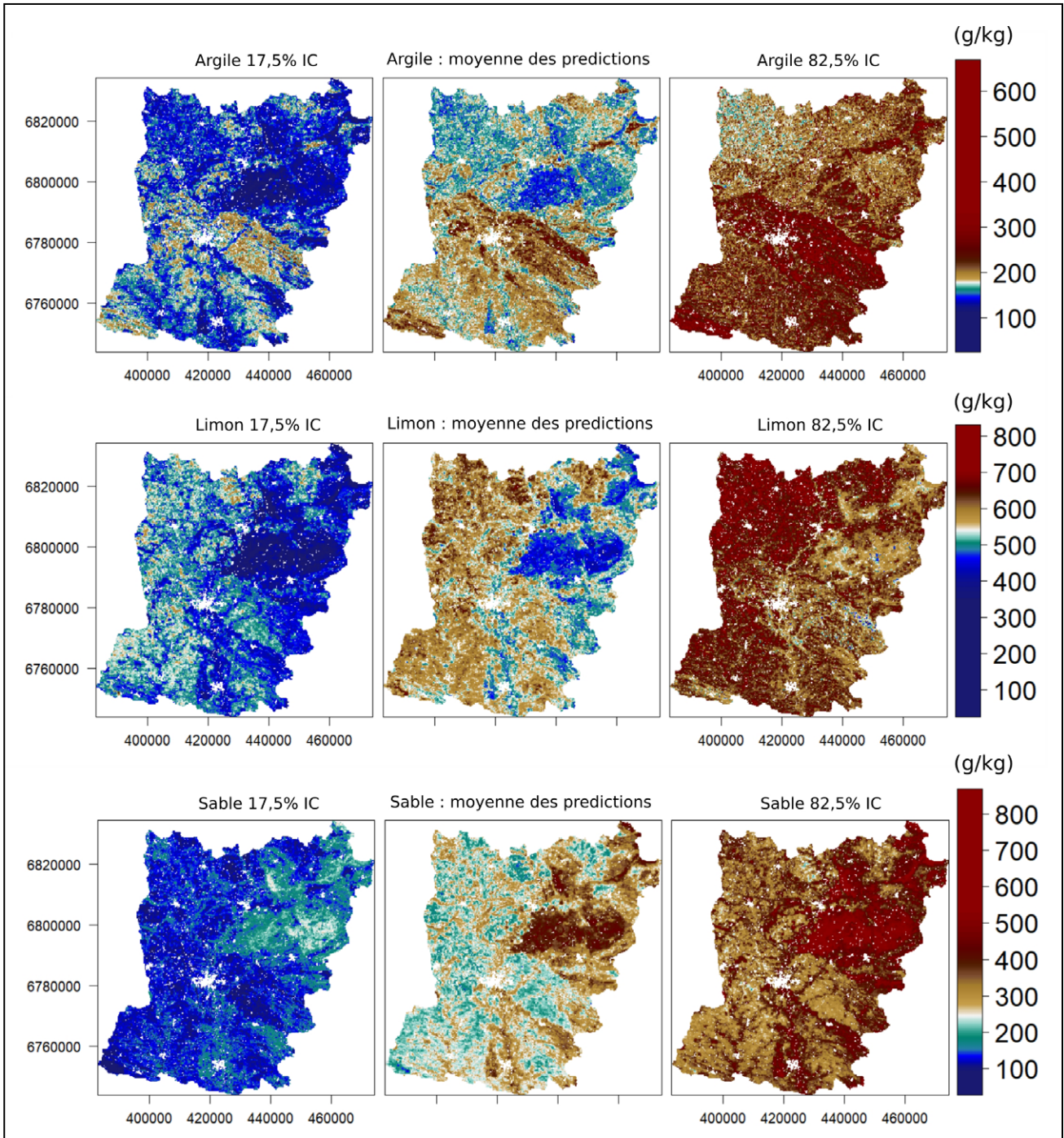


Figure 9 : Représentation des fractions granulométriques par des cartes uniques de la propriété cible spatialisée, en utilisant une rampe de couleur multivariée, l'une pour l'attribut, l'autre pour l'incertitude (<https://research.tableau.com/sites/default/files/2018-UncertaintyPalettes-CHI.pdf>), dotée d'un gradient plus ou moins terne de la valeur prédite associée à son intervalle de confiance de 65 %.

Figure 9: Maps of different particle-sizes using the "value suppressing uncertainty visualization" (<https://research.tableau.com/sites/default/files/2018-UncertaintyPalettes-CHI.pdf>) using the 65% confidence interval.

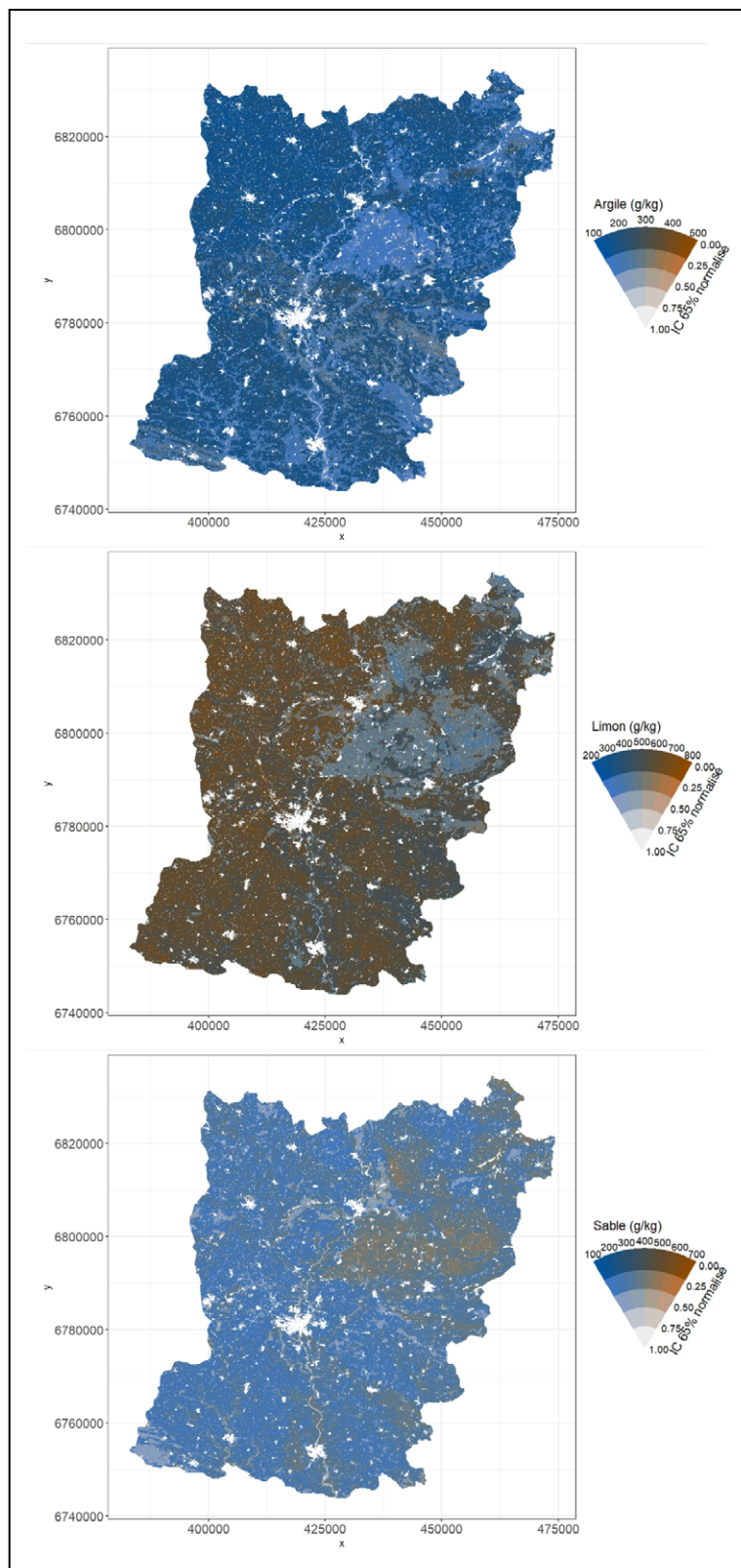
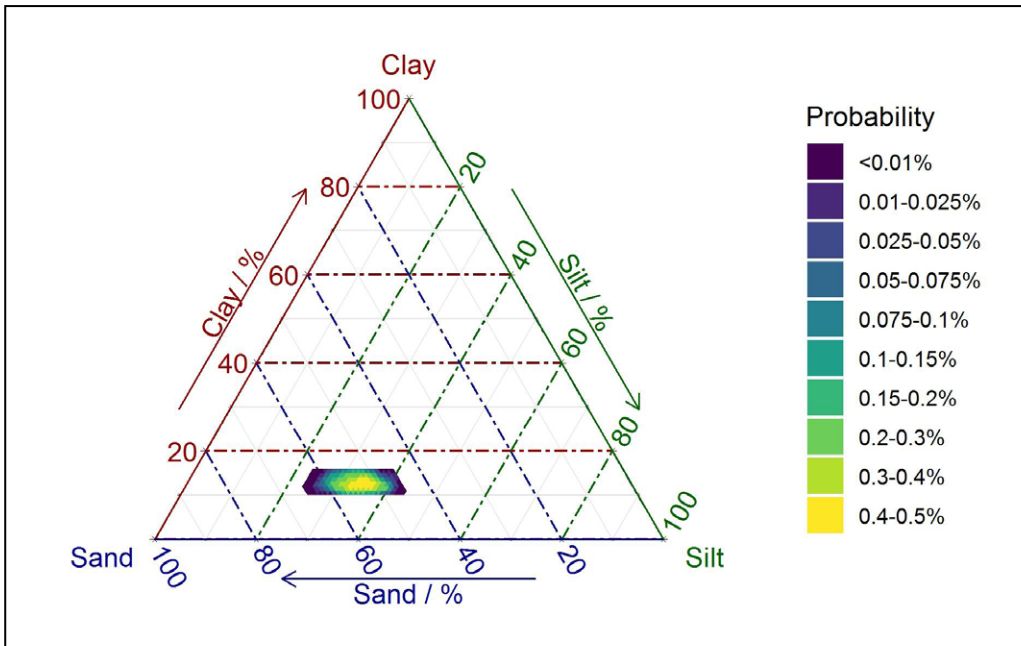


Figure 10 : Exemple de distribution de probabilité de la granulométrie d'un horizon à partir des valeurs extrêmes et modales de ses strates et sous l'hypothèse de distributions de densités triangulaires de chaque classe granulométrique (argiles, limons, sables).

Figure 10: Example of a probability distribution function of a layer using extreme and modal values defined by the soil surveyor and making the hypothesis of a triangular distribution function for each particle size fraction (clay, silt, sand).



données issues de la cartographie traditionnelle, on peut se poser à juste titre la question de l'intérêt de prédire des valeurs extrêmes. Cette question rejoint en partie les inquiétudes vis-à-vis de la cartographie numérique, domaine parfois incompris par les non-initiés (Arrouays *et al.*, 2020). Cependant, l'avantage de ce domaine est sa malléabilité dans ses modes de représentation et dans la flexibilité des méthodes qui peuvent être utilisées pour représenter l'incertitude en fonction des besoins, et ce d'une façon explicite.

En effet, une incertitude forte peut exprimer différentes informations, comme des zones lacunaires en informations pédologiques d'apprentissage induisant une augmentation de la variabilité des prédictions dans les zones à faible densité de points, soit un manque de pertinence locale de l'information issue des co-variables injectées dans le modèle. Enfin, cette incertitude peut refléter une réalité pédologique, en particulier si la variabilité naturelle intra-pixel est forte. Dans ce dernier cas, l'incertitude traduit simplement une réalité de terrain qui nécessiterait une densité de points plus importante et des co-variables beaucoup plus fines, pour obtenir des pixels plus homogènes et de surface élémentaire beaucoup plus réduite.

Pour le modélisateur, la visualisation au sens strict de l'incertitude a sans doute moins d'importance que pour les autres utilisateurs. Connaître une valeur moyenne ou modale et un intervalle de confiance lui sera sans doute peu utile, à moins qu'il ne puisse faire une hypothèse sur la fonction de

distribution de probabilité de la variable d'intérêt. Dans certains cas, des hypothèses simples, comme des distributions log-normales, voire triangulaires, peuvent permettre une première approximation. Notons, au passage, que les informations contenues dans les « strates » des bases de données issues de cartes traditionnelles pourraient être traitées de cette façon, c'est-à-dire sous forme de distributions triangulaires ou autres, à la condition que les valeurs considérées comme minimales et maximales dans les strates puissent être supposées comme représentant, soit de véritables minimum et maximum, soit comme des « proxys » d'un intervalle de confiance donné tel que cela a été réalisé aux USA (Helmick *et al.*, 2014) ou en France (Cazemier *et al.*, 2001). L'exemple présenté en Figure 10 montre comment on peut, à l'aide de distributions triangulaires et de valeurs minimales, maximales et modales de fractions granulométriques, passer de l'appartenance à une classe de valeurs à une distribution probabiliste en 3 dimensions. L'intérêt pour le modélisateur est de pouvoir disposer d'une distribution statistique et donc de pouvoir juger de la sensibilité des sorties de ses modèles aux valeurs de cette distribution.

Un indice IC/valeur médiane prédite fournit une représentation intéressante. Pour une meilleure utilisation et appropriation, on pourrait imaginer de discréditer cet indice pour faire des classes, et avoir une interprétation qualitative pour chacune de ces classes, du type : incertitude très faible, ...

incertitude très élevée, donnant une indication sur les limitations des possibilités d'utilisation de la donnée.

Nos exemples se fondent sur des prédictions à l'échelle départementale, qui représente une étendue très grande pour un grand nombre d'applications. Il serait intéressant de voir ce que ces types de représentations donneraient au niveau d'entités spatiales plus restreintes (petits bassins-versants, par exemple) dans lesquelles la variabilité à la fois des valeurs prédites mais aussi des IC est beaucoup plus faible.

Il reste à tester les modes de représentation de l'incertitude auprès des utilisateurs potentiels. Ceci était prévu dans le cadre de ce travail mais ne pourra être fait qu'ultérieurement. Des réunions de retour d'expérience permettront de les faire s'exprimer sur leurs attentes, sur leurs besoins, et sur l'utilité de ces modes de représentation pour leurs usages thématiques. Elles permettront également de mieux savoir si nos propositions de représentations sont compréhensibles par ces utilisateurs.

Pour les modélisateurs, Le « Graal » ultime serait de disposer de cette fonction de distribution, ce qui est préconisé dans les tiers les plus élevés du programme *GlobalSoilMap* (Arrouays *et al.*, 2014a ; Heuvelink, 2014). Ceci est théoriquement possible en utilisant des approches de type QRF (Vaysse et Lagacherie, 2017) mais reste en l'état actuel très coûteux en terme de temps de calcul. La perfection serait même de disposer des distributions conjointes de toutes les propriétés (Heuvelink, 2014), de façon à ne pas créer de sols « chimériques » présentant des combinaisons de distributions irréalistes, voire impossibles.

CONCLUSION

Notre objectif était de tenter de rendre l'incertitude des prédictions plus visuelle tout en expliquant au mieux ce que cette visualisation signifie. Il s'agissait d'essayer d'améliorer la sémiologie des cartes d'incertitudes en fonction du message que l'on cherche à faire passer à l'utilisateur.

Pour l'utilisateur qui ne cherche à obtenir qu'un « degré de confiance » dans les cartes qui lui sont proposées, les solutions que nous venons de présenter sont différentes variantes possibles. En pratique, ce devrait être à l'utilisateur lui-même de définir le degré de confiance dont il a besoin pour l'aider dans sa décision.

Il est difficile de juger si un mode de représentation de l'incertitude est plus pertinent que les autres. En pratique, cela dépendra probablement de l'utilisation qui voudra être faite de cette incertitude.

Plus généralement, plutôt que de vouloir communiquer sur l'incertitude des prédictions elles-mêmes, une solution devrait être de communiquer quelles sont les conséquences de cette incertitude sur les sorties des modèles ou sur les décisions prises en prenant en compte les variables concernant le sol dans des outils d'aide à la décision de types divers (gestion de l'eau,

lutte contre l'érosion, planification de l'usage des sols, évaluation des services écosystémiques rendus par les sols, etc.).

Par ailleurs, les incertitudes calculées par les modèles de CSMS concernent des prédictions de propriétés sur un site ponctuel, ce qui constitue un niveau de détail rarement utile. Il faudrait pouvoir propager ces incertitudes au niveau des unités spatiales qui ont un réel intérêt pour un utilisateur, par exemple une parcelle, un bassin-versant, une section cadastrale ou une commune (Vaysse *et al.*, 2017). En plus de mieux correspondre au besoin d'un utilisateur, ceci aurait l'avantage d'atténuer l'incertitude en supprimant l'effet des variabilités locales de sol, toujours difficiles à prédire par les modèles.

REMERCIEMENTS

Ce travail a bénéficié d'un financement de la part du Ministère en charge de l'agriculture (convention MAA 2102 5164 92). Cette recherche a été conduite avec l'appui du « Centre d'Expertise Scientifique Cartographie numérique des Sols » soutenu par le programme TOSCA du CNES ». Dominique Arrouays est coordinateur, Pierre Roudier est membre, et Anne C. Richer-de-Forges et Philippe Lagacherie sont collaborateurs du Consortium de Recherche GLADSOILMAP financé par le « LE STUDIUM Loire Valley Institute for Advanced Studies ». Songchao Chen a été financé par le « China Scholarship Council » pour une thèse de 3 ans à INRAE et Agrocampus Ouest (grant agreement no. 201606320211). Les auteurs remercient Blandine Lemerrier et Joël Moulin pour leur relecture critique très constructive d'une première version de cet article.

BIBLIOGRAPHIE

- Aitchison J., 1982 - The statistical analysis of compositional data. *J. R. Stat. Soc. Ser. B Methodol.* 44 (2), pp. 139-177.
- Arrouays D., Grundy M.G., Hartemink A.E., Hempel J. W., Heuvelink G.B.M., Hong S.Y., Lagacherie P., Lelyk G., McBratney A.B., McKenzie, N.J., Mendonça-Santos M.D., Minasny B., Montanarella L., Odeh, I.O.A., Sanchez P.A., Thompson J.A., Zhang G.-L., 2014a - *GlobalSoilMap: towards a fine-resolution global grid of soil properties*. *Advances in Agronomy*, 125, pp. 93-134.
- Arrouays D., Lagacherie P., Hartemink A., 2017a - Digital soil mapping across the globe. *Geoderma Regional*, 9, pp. 1-4.
- Arrouays D., Leenaars J., Richer-de-Forges A.C., Adhikari K., Ballabio C., Greve M., Grundy M., Guerrero E., Hempel J., Hengl T., Heuvelink G., Batjes N., Carvalho E., Hartemink A., Hewitt A., Hong S.-Y., Krasilnikov P., Lagacherie P., Lelyk G., Libohova Z., Lilly A., McBratney A., McKenzie N., Vasques G., Mulder V.L., Minasny B., Montanarella L., Odeh I., Padarian J., Poggio L., Roudier P., Saby N., Savin I., Searle R., Stolbovov V., Thompson J., Smith S., Sulaeman Y., Vintila R., Viscarra Rossel R., Wilson P., Zhang G.-L., Swerts M., van Oorts K., Karklins A., Feng L., Ibbelles, Navarro A.R., Levin A., Laktionova T., Dell'Acqua M., Suvannang N., Ruam W., Prasad J., Patil N., Husnjak S., Pásztor L., Okx J., Hallet S., Keay C., Farewell T., Lilja H., Juilleret J., Marx S., Takata Y., Kayusuki Y., Mansuy N., Panagos P., van Liedekerke M., Skalsky R., Sobocka J.,

- Kobza J., Eftekhari K., Kazem Alavipanah S., Moussadek R., Badraoui M., da Silva M., Paterson G., da Conceição Gonçalves M., Theocharopoulos S., Yemefack M., Tedou S., Vrscaj B., Grob U., Kozak J., Boruvka L., Dobos E., Taboada M., Moretti L., Rodriguez D. 2017b - Soil legacy data rescue *via* GlobalSoilMap and other international and national initiatives. *GeoRes J*, 14, pp. 1-19.
- Arrouays D., McBratney A.B., Bouma J., Libohova Z., Richer-de-Forges A.C., Morgan C., Roudier P., Poggio L., Mulder V.L., 2020 - Impressions of digital soil maps: the good, the not so good, and making them ever better. *Geoderma Regional*. 20, e00255.
- Arrouays D., McKenzie N.J., Hempel J., Richer de Forges A.C., McBratney A.B. (eds), 2014b - *GlobalSoilMap. Basis of the global spatial soil information system*. CRC Press, Taylor&Francis, 478 p.
- Arrouays D., Savin I.Y., Leenaars J.G.B., McBratney A.B., (eds), 2018 - *GlobalSoilMap. Digital soil mapping from country to globe*. Taylor&Francis CRC Press, London. 174 p.
- Boettinger J.L., Howell D.W., Moore A.C., Hartemink A.E., Kienast-Brown S., 2010 - *Digital Soil mapping. Bridging Research, Environmental Application and Operation. Progress in Soil Science*. Springer.
- Cazemier D.R., Lagacherie P., Martin-Clouaire R., 2001 - A possibility theory approach for estimating available water capacity from imprecise information contained in soil databases. *Geoderma* 103.
- Correl M., Moritz D., Heer J., 2018 - Value-Suppressing Uncertainty Palettes. *CHI 2018*, April 21-26, 2018, Montreal, QC, Canada. DOI: <https://doi.org/10.1145/3173574.3174216>
- Favrot J.C., 1989 - Une stratégie d'inventaire cartographique à grande échelle : la méthode des secteurs de référence. *Science du sol*, 27 (4): 351-36.
- Grundy, M.J., Viscarra Rossel, R.A., Searle, R.D., Wilson, P.L., Chen, C., Gregory, L.J., 2015 - Soil and landscape grid of Australia. *Soil Res.* 53, 835-844.
- Hartemink A.E., McBratney A.B., Mendonça-Santos MdL., 2008 - *Digital Soil Mapping with Limited Data*. Springer.
- Helmick J.L., Nauman T.W., Thompson J.A., 2014 - Developing and assessing prediction intervals for soil property maps derived from legacy databases. 2014. *In*: Arrouays D., McKenzie N.J., Hempel J., Richer-de-Forges A.C., McBratney A.B. (eds), 2014 - *GlobalSoilMap. Basis of the global soil information system*. Taylor&Francis, CRC Press, London, pp. 359-366.
- Hengl T., de Jesus J.M., Heuvelink G.B.M., Gonzalez M.R., Kilibarda M., Blagotic A., Shangguan W., Wright M.N., Geng X.Y., Bauer-Marschallinger B., Guevara M.A., Vargas R., MacMillan R.A., Batjes N.H., Leenaars J.G.B., Ribeiro E., Wheeler I., Mantel S., Kempen B., 2017 - SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12, 2, e0169748. DOI: 10.1371/journal.pone.0169748.
- Heuvelink G.B.M. 2014. Uncertainty quantification of GlobalSoilMap products. *In*: Arrouays D., McKenzie N.J., Hempel J., Richer-de-Forges A.C., McBratney A.B. (eds). *GlobalSoilMap. Basis of the global soil information system*. Taylor&Francis, CRC Press, London, pp. 335-340.
- Lagacherie P., Arrouays D., Walter C., 2013 - Cartographie numérique des sols: principes, mise en œuvre, potentialités. *Etude et Gestion des Sols*, 20(1), 83-98.
- Lagacherie P., McBratney A.B., Voltz M., 2006 - Digital soil mapping. An introductory perspective. *Developments in Soil Science*, N°31. Elsevier.
- Lark R.M., Bishop T.F.A., 2007 - Cokriging particle size fractions of the soil. *Eur. J. Soil Sci.* 58 (3), 763-774. <https://doi.org/10.1111/j.1365-2389.2006.00866.x>
- Laroche B, A.C. Richer de Forges, S. Leménager, D. Arrouays, N. Schnebelen, M. Eimberck, B. Toutain, S. Lehmann, E. Tientcheu, F. Héliès, J-P. Chenu, S. Parot, S. Desbordes, G. Giroit, M. Voltz, M. Bardy., 2014 - Le programme Inventaire Gestion et Conservation des Sols. *Volet Référentiel Régional Pédologique. Etude et Gestion des Sols*, 21, pp. 125-140.
- Legros J.-P., 1973 - Précision des cartes pédologiques, la notion de finesse de caractérisation. *Sciences du Sol*, 2, pp. 115-118.
- Legros J.-P., 1978a - Recherche et contrôle numérique de la précision en cartographie numérique : 1. Précision dans la délimitation des sols. *Annales Agronomiques*, 29, pp. 499-519.
- Legros J.-P., 1978b - Recherche et contrôle numérique de la précision en cartographie numérique : 2. Précision dans la caractérisation des unités de sols. *Annales Agronomiques*, 29, pp. 583-601.
- Libohova Z., Odgers N.P., Ashtekar J.M., Owens P.R., Thomson J.A., Hempel J., 2016 - Some challenges on quantifying soil property predictions uncertainty for the GlobalSoilMap using legacy data. G. Zhang et al. (eds), *Digital Soil Mapping Across Paradigms Scales and Boundaries*, Springer Environmental Science and Engineering, DOI 10.1007/978-981-10-0415-5_11.
- Loiseau T., Chen S., Mulder V.L., Román Dobarco M., Richer-de-Forges A.C., Lehmann S., Bourenanne H., Saby N.P.A., Martin M.P., Vaudour E., Gomez C., Lagacherie P., Arrouays D., 2019 - Satellite data integration for soil clay content modelling, at a national scale. *International Journal of Applied Earth Observation and Geoinformation*. 82, 101905.
- Loiseau T., Richer-de-Forges A., Martelet G., Bialkowiak A., Nehlig P., Arrouays, D., 2020 - Could airborne gamma spectrometric data replace lithological maps as co-variables for digital soil mapping of topsoil particle-size distribution? A case study in Western France. *Geoderma Regional*. 22, e00295.
- Malone B.P., McBratney A.B., Minasny B., 2011 - Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. *Geoderma*, 160, pp. 614-626.
- Marsman B.A., de Gruijter J.J. 1986. Quality of soil maps. A comparison of survey methods in a sandy area. *Soil Survey Papers*, N° 15. Netherlands Soil Survey Institute, Wageningen. 103 pp.
- McBratney A.B., Mendonca Santos M.dL., Minasny B., 2003 - On digital soil mapping. *Geoderma*, 117, pp. 3-52.
- Meinshausen N. 2006. "Quantile Regression Forests", *Journal of Machine Learning Research* 7, pp. 983-99.
- Minasny B., Malone B.P., McBratney A.B. (Eds), 2012 - *Digital Soil Assessments and Beyond: Proceedings of the 5th Global Workshop on Digital Soil Mapping 2012*, Sydney, Australia. CRC Press, Taylor&Francis.
- Minasny B., McBratney A.B., 2016 - Digital soil mapping: A brief history and some lessons. *Geoderma*, 264, Part B, pp. 301-311.
- Mulder V.L., Lacoste M., Richer-de-Forges A.C., Arrouays D. 2016. *GlobalSoilMap France: high-resolution spatial modelling the soils of France up to two meter depth*. *Science of the Total Environment* 573, pp. 1352-1369.
- Murciano V., Paroissien J.-B., Saby N.P.A., Richer-de-Forges A.C., Martin M.P., Emilion R., Arrouays D., 2015 - Estimation des incertitudes liées à la prédiction ponctuelle de variables pédologiques à partir de bases de données géographiques sur les sols. Exemple de l'utilisation des strates issues du programme français Inventaire, Gestion et Conservation des Sols (IGCS). *Etude et Gestion des Sols*, 22, pp. 9-18.
- Odgers N.P., McBratney A.B., Minasny B., 2014 - Digital soil property mapping and uncertainty estimation using soil class probability rasters. *In*: Arrouays D., McKenzie N.J., Hempel J., Richer-de-Forges A.C., McBratney A.B. (eds). *GlobalSoilMap. Basis of the global soil information system*. Taylor&Francis, CRC Press, London, pp. 341-346.
- Richer-de-Forges A.C., Arrouays D., Bardy M., Bispo A., Lagacherie P., Laroche B., Lemerrier B., Sauter J., Voltz M., 2019 - Mapping of Soils and Land-Related Environmental attributes in France: analysis of end-users'needs. *Sustainability*, 11, 2940; doi:10.3390/su1102940.
- Vaysse K., Heuvelink G.B.M., Lagacherie P., 2017 - Spatial aggregation of soil property predictions in support of local land management. *Soil Use Manag.* 33, pp. 1-12.
- Vaysse K., Lagacherie P., 2017 - Using quantile regression forest to estimate uncertainty of digital soil mapping products. *Geoderma*, 291, 55-64. <https://doi.org/10.1016/j.geoderma.2016.12.017>.

- Voltz M., Arrouays D., Bispo A., Lagacherie P., Laroche B., Lemerrier B., Richerde-Forges A.C., Sauter J., Schnebelen N. 2018. La cartographie des sols en France : Etat des lieux et perspectives, INRA, Paris, France, 112 p.
- Voltz M., Arrouays D., Bispo A., Lagacherie P., Laroche B., Lemerrier B., Richerde-Forges A.C., Sauter J., Schnebelen N., Soumis - Possible futures of soil mapping in France. Geoderma Regional.
- Voltz M. Lagacherie P., Louchart X., 1997 - Predicting soil properties over a region using sample information from a mapped reference area. European J. Soil Sci., 48(1), pp. 19-30.
- Wilke C., 2020 - Multiscales R Package : <https://github.com/clauswilke/multiscales>

