Approches géostatistiques pour l'extraction de l'information pertinente dans des données géo-électriques en vue de la cartographie de propriétés des sols

Thomas Loiseau

▶ **To cite this version:**

Thomas Loiseau. Approches géostatistiques pour l'extraction de l'information pertinente dans des données géo-électriques en vue de la cartographie de propriétés des sols. [Stage] France. Université François Rabelais (Tours), FRA. 2015, 35 p. hal-02801221

HAL Id: hal-02801221
https://hal.inrae.fr/hal-02801221

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Master Sciences de l'Eau et de l'Environnement

Hydrosystèmes et Bassins Versants
Diagnostic et Risque Environnemental

Rapport de stage pour l'obtention de la 1ère année de Master

Approches géostatistiques pour l'extraction de l'information pertinente dans des données géo-électriques en vue de la cartographie de propriétés des sols

LOISEAU Thomas

Mai – Août 2015

Maître de stage : Hocine BOURENNANE et Marie-Noël MISTOU
INRA Orléans
Unité de Recherche de Science du sol

INRA SCIENCE & IMPACT

Géosciences Environnement

UR SOLS Unité de Recherche de Science du Sol d'Orléans

Remerciements

Je tiens tout d'abord à remercier Catherine Hénault, Directrice de l'Unité de Recherche Science du sol de l'INRA d'Orléans pour m'avoir permis de réaliser mon stage au sein de cette unité.

Je remercie ensuite Marie-Noël Mistou, responsable du projet CAREX, qui a rendu possible la création de ce stage.

Je remercie particulièrement Hocine Bourennane, pour ses conseils et pour le temps qu'il a consacré pour mon encadrement et aux questions que j'ai pu lui poser. Je le remercie également pour les connaissances en géostatistique, que j'ai pu acquérir à son contact.

Je tiens également à remercier toutes les personnes qui ont contribué de près ou de loin à mon intégration et au bon déroulement de ce stage : Maud Séger, Ghislain Girot et Isabelle Cousin pour certaines questions sur CAREX auxquelles ils ont su répondre, Lionel Cottenot qui a bien voulu partager son bureau avec moi pendant 3 mois, Sébastien, Florent, Charlène, Antoine, Nicolas, Cyrile, Laëtitia, Hakima, Anne-Fleur, Catherine, Eugénie, Jordan,…

Et enfin, un grand Merci à l'ensemble de l'équipe UR Sols et InfoSols pour la bonne ambiance de travail, entrecoupée de pauses, de pots et de sorties, dans laquelle j'ai pu effectuer mon stage de manière agréable à Orléans et au sein de l'INRA particulièrement.

Sommaire

Remerciements .....

Sommaire 1

Résumé / Abstract 3

Présentation de l'organisme d'accueil 4

1. Introduction 5

1.1. Contexte et problématiques 5

1.2. Objectifs du stage 5

2. Site d'étude et campagnes réalisées 6

2.1. Présentation du site d'étude 6

2.2. Mesures de la résistivité électrique des sols 7

2.3. Campagnes pédologiques 8

3. Méthodes statistiques mises en œuvre pour le traitement des données 9

3.1. Approche développée pour la modélisation et la spatialisation de l'épaisseur du sol 9

3.2. Indices statistiques pour la validation 11

3.3. Qualité prédictive du modèle en fonction du jeu de données de prédiction 12

3.4. Taille de la maille de spatialisation de la variable explicative 12

4. Résultats et discussions 13

4.1. Détermination du modèle prédictif par régression linéaire simple 14

4.2. Spatialisation de l'épaisseur de sol par régression linéaire et régression-krigeage 15

4.2.1. Spatialisation par régression linéaire simple 15

4.2.2. Spatialisation par régression-krigeage 18

4.3. Evolution de la qualité prédictive 22

4.3.1. Sensibilité du modèle à la taille du jeu de donnée de prédiction 22

4.3.2. Sensibilité du modèle de spatialisation à la résolution de la variable explicative 26

5. Conclusion 29

Références bibliographiques 30

Liste des figures 31

Liste des tableaux 32

Liste des annexes 32

Annexes 33

Résumé

Dans le cadre du projet CAREX de l'INRA (CARactérisation environnementales des essais conduits dans les unités EXpérimentales), le travail présenté dans ce document a pour principal objectif de cartographier l'épaisseur du sol (Ep) sur un site expérimental de l'INRA. Deux méthodes de spatialisation, la régression linéaire simple et la régression-krigeage, de Ep à partir de mesures de résistivités électriques du sol ont été comparées. Une prospection électrique exhaustive par ARP a été effectuée en automne 2011 sur l'ensemble du site expérimental de l'INRA d' Epoisses. Plusieurs campagnes de sondage pédologique ont été effectuées de 2013 à 2015, pour obtenir une information sur Ep en 80 points. Les résultats ont montré que la résistivité de la voie 3 (R3) était la plus corrélée à Ep. Par ailleurs, la spatialisation en tenant compte d'une composante spatiale de R3 et des résidus entre cette dernière et Ep (régression-krigeage) produit les résultats les plus proches des valeurs mesurées du jeu de donné de validation (ME = -2,87 cm contre -7,32 cm sans résidus). Les résultats ont aussi montré que la méthode de spatialisation par régression-krigeage n'est pas sensible à : (1) la taille de l'échantillon de prédiction, (2) la résolution spatiale de la variable auxiliaire. Ces deux aspects sont d'une importance capitale en termes de coût de spatialisation des propriétés des sols au-delà du cas traité dans ce travail.

Mots clefs : Epaisseurs des sols ; Résistivité des sols ; Régression-krigeage ; Régression linéaire simple ; Qualité prédictive.

Abstract

In the framework of CAREX project, the work presented in this document aims at map soil thickness (Ep) on one INRA's experimental site. Two spatial methods, simple linear regression and regression-kriging, for Ep from soil electrical resistivity measurements were compared. An exhaustive electrical resistivity prospecting was carried out in autumn 2011 throughout the experimental site of INRA Epoisses. In addition, several soil survey campaigns were conducted between 2013 and 2015 over the study site to obtain information on Ep in 80 locations. Results showed that the third way of resistivity (R3) was the most correlated with Ep. Moreover, accounting for a spatial component of R3 and residuals between this component and Ep (regression-kriging) allow obtaining obtain closest results to measured values of validation data set (ME = -2, 87 cm *versus* -7, 32 cm). Finally, results also showed that regression-kriging is insensitive to: (1) the size of prediction sample, (2) the spatial resolution of auxiliary variable. These two aspects are of major matter in terms of spatializing cost of soil properties beyond the case treated in this work

Keywords: Soil thickness; Soil resistivity; Regression-kriging; Simple linear regression; Sensitivity analysis.

Présentation de l'organisme d'accueil

Présentation de l'INRA

L'institut National de la Recherche Agronomique (INRA), est un établissement public à caractère scientifique et technologique placé sous la tutelle conjointe des ministères de l'Agriculture et de l'Enseignement supérieur et de la Recherche. Créé au lendemain de la seconde guerre mondiale, en 1946, il avait pour mission de mettre la science et la technologie au service du développement et de l'agriculture en améliorant les techniques de productions (végétales et animales) afin de répondre aux besoins alimentaires de la population. De nos jours, l'INRA, oriente ses travaux sur trois domaines fortement liés : l'alimentation, l'agriculture et l'environnement avec pour objectif de garder une agriculture compétitive respectueuse de l'environnement, des ressources naturelles et des besoins de l'homme en faisant évoluer les méthodes agricoles. Afin de relever ces différents défis, l'INRA se regroupe en plusieurs centres régionaux dans l'ensemble de la France.

Le centre de recherche INRA Val de Loire et l'Unité de Recherche de Science du sol

Le centre INRA Val de Loire est réparti sur trois sites, Orléans, Tours et Bourges, et mène des recherches autour de quatre pôles :

- Dynamique des sols et gestion de l'environnement
- Biologie intégrative des arbres et organismes associés
- Biologie intégrative animale et gestion durable des productions animales
- Santé animal et santé publique

L'Unité de Recherche de Sciences du sol (l'UR SOLS), située sur le site d'Orléans dans la commune d'Ardon, est l'unité dans laquelle j'ai effectué mon stage. Leur thématique de recherche se situe dans le premier pôle « Dynamique des sols et gestion de l'environnement ». Cette Unité caractérise les qualités physiques des sols en étudiant leurs propriétés hydriques et biogéochimiques, tout en tenant compte de la diversité à grande et petite échelle de la nature des sols. L'UR SOLS approfondit également les connaissances sur l'érosion hydrique des sols, la réserve en eau et les mécanismes de production et de libération de gaz à effet de serre par les sols.

Afin de réaliser ses objectifs, elle s'appuie sur des observations *in situ*, ainsi que sur des expérimentations et des modélisations, grâce à une équipe composée de chercheurs, d'ingénieurs et de techniciens qui ont pour vocation à contribuer à une meilleur préservation des sols, à une meilleur gestion de la ressource en eau et à la protection de la qualité de l'air.

1. Introduction

1.1. Contexte et problématiques

Le sol correspond à la partie superficielle de la croûte terrestre. Il est formé par la dégradation de la roche mère en profondeur et par la dégradation de la matière organique en surface. Il est en contact direct avec l'ensemble des milieux présent à la surface de la Terre (Biosphère, Hydrosphère, Atmosphère, Lithosphère,…). Il représente de ce fait une surface d'échange entre ces différents milieux.

Il ne correspond pas à un milieu stable et possède une variabilité spatiale et temporelle de ses caractéristiques (épaisseur, granulométrie, type d'humus, perméabilité, …) dépendante de contraintes physiques, chimiques et biologiques qui lui sont imposées (activités biologiques, nature de la roche mère, nature de la végétation, climats, hydrologie, …). L'utilisation du sol à des fins culturales ne peut donc pas être indépendante de la caractérisation de l'ensemble de ces variables.

L'INRA réalise des expérimentations agronomiques sur plusieurs sites d'unités expérimentales (UE) dans le but de faire évoluer les méthodes agricoles. Cependant certaines caractéristiques de sol, sur ces sites d'études, ne sont pas forcément connues. Afin de proposer une méthodologie de caractérisation environnementale applicable à toutes ces unités expérimentales, un projet nommé CAREX a vu le jour en 2010. Ce projet s'articule autour de la quantification des stress hydriques et azotés, ainsi que de leur suivie, pour ajuster des programmes d'interventions culturales. Cependant, dans un premier temps, une connaissance approfondie de la nature des sols de ces sites est nécessaire.

C'est dans cette dernière partie que s'inscrit mon stage au sein de l'INRA, et plus particulièrement sur la caractérisation des épaisseurs des sols sur le site de l'UE d'Epoisses, en Bourgogne.

1.2. Objectifs du stage

Dans un premier temps, l'objectif de ce stage était d'établir un modèle de prédiction de l'épaisseur du sol (Ep) en fonction de la résistivité électrique et de spatialiser Ep sur tout le domaine d'étude. L'épaisseur du sol a été mesurée ponctuellement en quelques endroits sur le site d'étude, en revanche la résistivité électrique a été mesurée de façon quasi-exhaustive. Dans un deuxième temps, il s'agissait de valider le modèle établi à partir d'un jeu de données indépendant d'épaisseur de sol. La dernière étape était d'étudier l'évolution de la qualité prédictive du modèle en fonction : (1) de la taille des échantillons de mesure de la variable d'intérêt ; (2) de la résolution utilisée lors de l'organisation des données géophysiques sur une grille régulière pour la spatialisation de l'épaisseur du sol.

La méthodologie développée pour répondre à ces objectifs s'inspire en partie de celle développée par Ndjiba-Mitombo (2013). Elle sera expliquée plus en détails dans la suite de ce document.

2. Site d'étude et campagnes réalisées

2.1. Présentation du site d'étude

Le domaine Expérimentale de l'INRA d'Epoisses se situe à 10 km au Sud-Est de Dijon à la bordure occidentale de la vallée de l'Ouche sur la commune de Bretenière (*Figure 1).* Il regroupe un ensemble de 70 parcelles agricoles expérimentales, en système de grande culture, répartie sur 120 ha. Les expérimentations qui y sont menées sont en lien avec l'amélioration des plantes en agroécologie et portent sur la culture d'espèce très variées en conditions réelles (maïs, blé, orge, betterave, ...).

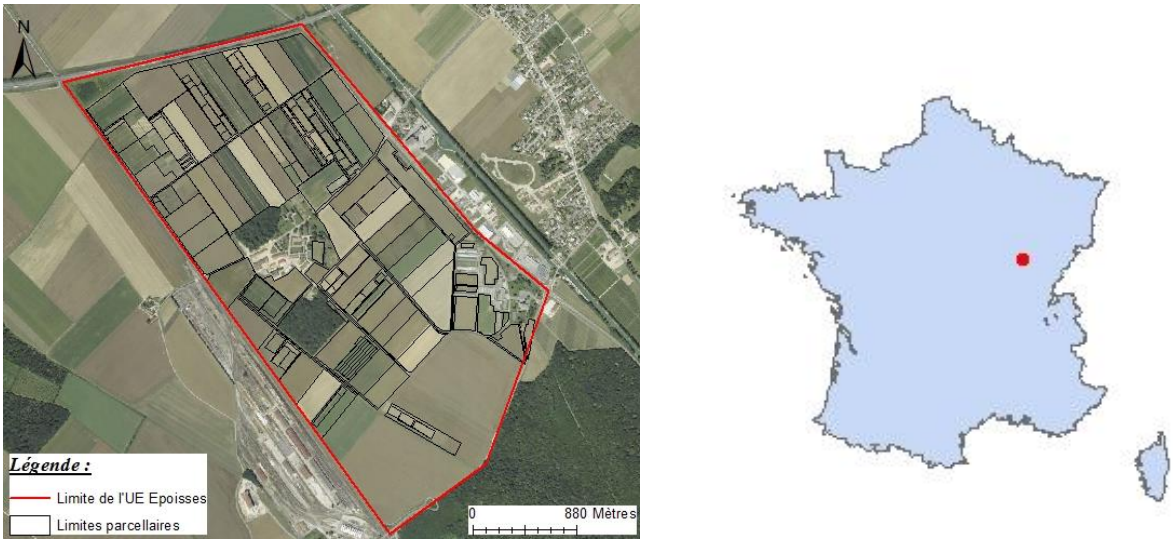

Figure 1 : Localisation de l'UE d'Epoisses, Source : Vue aérienne, Géoportail

Le site se situe dans une zone d'effondrement dans le calcaire du Jurassique, remplie par des dépôts du quaternaire. Des dépôts alluvionnaires y sont présents et proviennent de l'écoulement de l'Ouche et de la Tille s'écoulant plus à l'Est. Le climat de la région correspond au type semi-continental avec une pluviométrie annuelle comprise entre 700 et 800 mm.

La carte pédologique de Dijon, au 1/100000^{e} (Chrétien, 1976), indique que la zone d'étude présente deux types de sols : des sols de type rendzines brunifiées, ou Calcisols humiques (Référentiel pédologique, 2008), et des sols bruns calciques épais (argilo-limoneux, localement humifères sur cailloutis calcaires).

La présence d'une nappe a également pu être détectée dans les horizons comprenant des cailloutis. Il semblerait que sont écoulement soit de direction Nord-Ouest/Sud-Est.

2.2. Mesures de la résistivité électrique des sols

Une campagne de terrain a été effectuée au cours de l'automne 2011 (du 15 au 25 novembre), dans le but de mesurer la résistivité électrique des sols présents sur l'UE d'Epoisses. L'appareil utilisé fut un ARP (Automatic Resistivity Profiler), qui a pris des mesures de résistivités sur des tracés rectilignes, espacés de 6 à 12 m, tous les 10 cm. Ces mesures ont été géoréférencées au GPS différentiel.

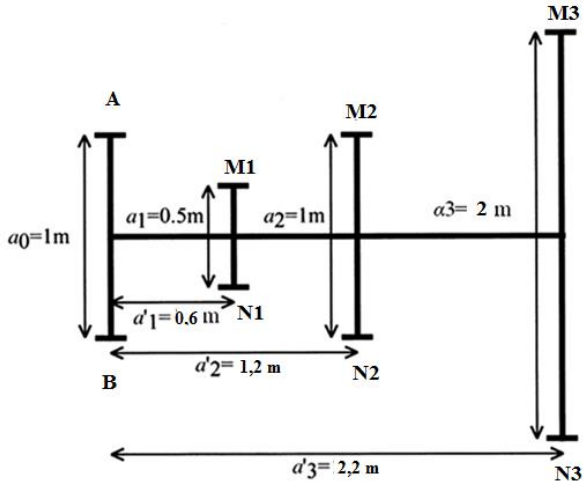

Figure 2 : Représentation du système ARP, modifié à partir de 3D Inversion of Automated Resistivity Profiling (ARP)

Ce système de mesure se base sur un montage de quatre dipôles (*Figure 2)*, montés sur des roues dentelées, éloignés les uns des autres, et tractés par un véhicule. Un dipôle AB, situé à 1m de l'arrière du véhicule, diffuse un courant électrique dans le sol. Un ensemble de 3 autres dipôles récepteurs M1N1, M2N2 et M3N3, placés respectivement à 0,5m, 1m et 2m du premier dipôle, mesurent la différence de potentiel électrique. Ces différences de potentielles, entre le courant envoyé dans le sol et reçu par les autres dipôles, ont permis de déterminer la résistivité électrique du sol selon trois profondeurs.

Les mesures de l'ARP sont intégratives et dépendent de la distance entre les dipôles récepteurs et le dipôle qui injecte le courant dans le sol. Les mesures à grande profondeur intègrent les résultats de surfaces. La campagne de mesure de 2011, a donc permis de déterminer trois valeurs de résistivités pour chaque point de mesure (*Tableau 1)* : une résistivité de surface R1 (0-0,5m), une résistivité semi-profonde R2 (0-1m) et une résistivité profonde R3 (0-2m). Ce jeu de données sera nommé par la suite comme le jeu de données N1.

Tableau 1 : Statistiques descriptives des données de résistivités N1.

Résistivités (ohm.m)	**Effectif**	**Minimum**	**Maximum**	**Moyenne**	**Variance**	**Ecart-type**
R1	1609676	8,85	624,80	31,76	133,93	11,57
R2	1609676	1,26	214,40	32,55	108,30	10,41
R3	1609676	5,38	1025,60	42,86	311,17	17,64

Cette résistivité, en moyenne, augmente avec la profondeur. La résistivité profonde R3 est plus importante que la résistivité semi-profonde R2 et de surface R1. Cependant les valeurs de résistivité de R2 n'atteignent qu'un maximum de 214 Ω.m. Cette différence est très certainement due à la nappe de surface présente dans ces horizons.

2.3. Campagnes pédologiques

En se basant sur les connaissances préalables sur les sols de la région (Chrétien, 1976) la résistivité de la voie 2 a été choisie pour positionner des points de sondages pédologiques. Ainsi, 80 sondages pédologiques (*Figure 3*) ont été effectués afin de déterminer les variations de l'épaisseur des sols sur l'UE d'Epoisses. Ils ont été prélevés sur plusieurs campagnes de terrain réparties sur trois ans (2013 – 2015).

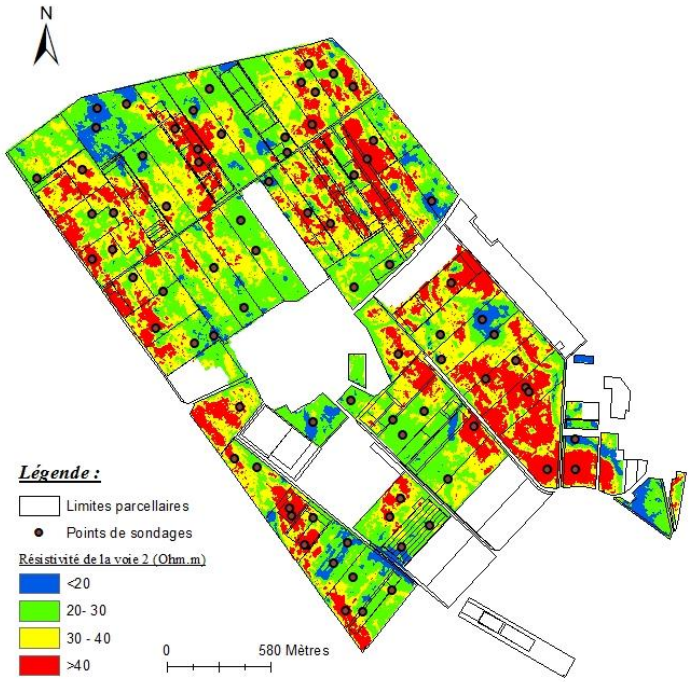

Figure 3 : Carte de la résistivité électrique issue de la voie 2 (R2) et localisation des sondages pédologiques
Source : Couche «résistivité de la voie 2 », Maud Seger INRA Orléans

Ces sondages ont permis d'observer une grande variabilité (de 23 à 224 cm) des épaisseurs des sols (Ep) du site d'étude. Ces données, nommées N2 dans la suite de ce rapport, sont présentées en *Annexes 1*.

3. Méthodes statistiques mises en œuvre pour le traitement des données

3.1. Approche développée pour la modélisation et la spatialisation de l'épaisseur du sol

L'approche développée permet de spatialiser l'épaisseur du sol sur le domaine d'étude selon deux scénarios dont certaines étapes sont communes. Les deux scénarios sont qualifiés respectivement de spatialisation par régression linéaire simple et spatialisation par régression-krigeage.

Le scénario « spatialisation par régression linéaire simple » comporte les étapes présentées ci-dessous.

1. Transformation des données : préalablement à toute analyse, les valeurs des variables (R1, R2, R3 et Ep) sont transformées dans l'espace Gaussien. Cette transformation permet de limiter l'influence des valeurs extrêmes sur la création des différents modèles (régression, variogramme) au cours de la modélisation et la spatialisation de l'épaisseur du sol.
2. Les épaisseurs des sols mesurées au niveau des 80 points de sondages (N2), sont informées des valeurs de résistivités R1, R2 et R3 correspondantes à partir de (N1).
3. Examen des corrélations entre Ep et les trois séries de valeurs de résistivité (R1, R2 et R3) : identification de la variable explicative (R1, R2 ou R3) la plus pertinente pour la modélisation et la spatialisation de Ep.
4. Nous sélectionnons aléatoirement 45 points de sondages, parmi N2, pour former un jeu de donnée de prédiction N4. Les 35 points restant nous permettent la création d'un jeu de données de validation N5. Ces 35 points ne participent pas à la création des modèles. Ils permettent de tester les modèles créés à partir de N4. Les deux jeux de données N4 et N5 sont présentés en *Annexes 2* et *3*
5. Régression linéaire simple en utilisant N4 : Ep *versus* la série de valeur de résistivité la plus corrélée à Ep.
6. Validation du modèle de l'étape 5 sur la base du jeu de donnée N5 après une transformation inverse permettant de revenir dans l'unité initiale de mesure des épaisseurs.
7. Spatialisation de la variable explicative identifiée dans l'étape 4. Cette spatialisation est réalisée par krigeage ordinaire (Wackernagel, 1995). Cette méthode d'estimation nécessite l'estimation et la modélisation du variogramme de la variable à spatialiser.
8. Spatialisation de l'épaisseur du sol en appliquant le modèle de l'étape 5 au résultat de l'étape 7.
9. Validation en projetant les mesures ponctuelles de N5 sur la carte issue de l'étape 8 : épaisseurs estimées sur la carte *versus* épaisseurs mesurées de N5 après avoir réalisé une transformation inverse permettant de revenir dans l'unité initiale de mesure des épaisseurs.

Dans le scénario « spatialisation par régression-krigeage » on se focalise sur la variable explicative R3 transformée dans l'espace Gaussien. Le choix de cette variable s'appuie sur le fait qu'elle intègre les valeurs de résistivité profond, semi-profonde et de surface. On suppose donc qu'elle est représentative de la résistivité du sol sur toute son épaisseur.

1. Le variogramme de R3 est modélisé par une structure emboîtée (généralement 3 variogrammes théoriques représentant le bruit, les hautes fréquences et les basses fréquences dans les données). R3 est ensuite spatialisé sur une grille régulière ainsi que ses composantes spatiales identifiées grâce au variogramme.
2. Les épaisseurs des sols mesurées au niveau des 80 points sondages (N2), sont informées des valeurs de résistivités issues de l'étape 1 et qui sont notées R3*, S0-R3*, S1-R3* et S2-R3* dans la suite du document.
3. Examen des corrélations entre Ep et les quatre séries de valeurs de résistivité de l'étape ci-dessus : identification de la variable explicative (R3*, S0-R3*, S1-R3* ou S2-R3*) la plus pertinente pour la modélisation et la spatialisation de Ep.
4. Les résidus entre Ep et la variable explicative identifié dans 3 sont calculés puis spatialisés par krigeage ordinaire. Cela suppose que ces résidus soient structurés dans l'espace géographique.
5. Krigeage de l'épaisseur du sol à partir de N4 avec prise des résultats des étapes 3 et 4 dans le système d'équations du krigeage.
6. Validation de la carte de l'étape précédente à partir de N5 après avoir réalisé une transformation inverse permettant de revenir dans l'unité initiale de mesure des épaisseurs.

Notions statistiques

Estimation par régression-krigeage

La méthode régression-krigeage pour la spatialisation (e.g. Bourennane, 1997 ; Watt and Palmer, 2012 ; Sun et al., 2012) consiste à utiliser une variable régionalisé, connue sur l'ensemble de la zone, dans le but de guider et dessiner les grandes lignes de la variable d'intérêt, qui est connue précisément mais en peu de point. Par ailleurs, l'estimation de la variable d'intérêt est optimisée en tenant compte des résidus qui existent entre la variable d'intérêt et la variable régionalisée.

Le variogramme expérimental

Un variogramme expérimental se défini par la moyenne des dissemblances entre deux valeurs d'une même variable régionalisée. Il calcule la moitié du carré de la différence entre deux valeurs de cette variable régionalisée, en tenant compte de la distance qui les sépares $\boldsymbol{h}$ et de leur orientation l'une par rapport à l'autre. La moyenne de ces dissemblances, pour un pas de distance défini, forme une courbe $\boldsymbol{\gamma^*(h)}$ appelée variogramme expérimental ***(1)***.

$$\boldsymbol{\gamma*(h) = \frac{1}{2nh}\sum_{\propto=1}^{nh}[z(x+h) - z(x)]^2} \quad \boldsymbol{(1)}$$

Le variogramme permet de savoir si une variable est structurée dans l'espace géographique ou bien si elle possède une distribution aléatoire.

Décomposition du variogramme expérimental

Le comportement d'une variable régionalisée peut être différent en fonction de l'éloignement spatial des points de mesure. La dissemblance des valeurs augmente en moyenne en fonction de l'éloignement spatial des points de mesure et atteint un palier de variation aux grandes distances. Ces différences de comportement peuvent être modélisées par plusieurs structures et spatialisées séparément. Ainsi, on peut se focaliser sur le comportement de la variable régionalisée à une échelle donnée en ignorant les autres structures.

3.2. Indices statistiques pour la validation

Pour valider les modèles créés, nous utilisons le jeu de données de validation (N5). Nous obtenons pour chacun des points de N5, une épaisseur de sol estimée par chaque modèle établi. Les épaisseurs estimées sont comparées aux épaisseurs de sol mesurées sur le terrain à travers le diagramme de dispersion entre valeurs mesurées et valeurs estimées et grâce aussi à un certain nombre d'indices statistiques :

- L'erreur moyenne (**ME**) qui doit être proche de zéro, permet de définir une sous-estimation ou une surestimation du modèle,

$$\boldsymbol{ME} = \frac{1}{\boldsymbol{n}} \sum_{k=1}^{n} (Ep.estimée - Ep.mesurée)$$

- La racine carrée de l'erreur quadratique moyenne (**RMSE**), qui doit être la plus faible possible pour une estimation précise et sans biais,

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{\boldsymbol{n}} \sum_{k=1}^{n} (Ep.estimée - Ep.mesurée)2}$$

- L'erreur moyenne absolue (**MAE**), indique l'ampleur moyenne des erreurs. Elle peut être inférieur ou égale à la RMSE, mais toujours positive.

$$\boldsymbol{MAE} = \frac{1}{\boldsymbol{n}} \sum_{k=1}^{n} |Ep.estimée - Ep.mesurée|$$

- Le coefficient de corrélation linéaire (**r**), indique le dégrée du lien linéaire entre deux variables, il est compris entre -1 et 1. Il représente la covariance des deux variable *Cov(X, Y)* sur le produit de leur écart-type σx et σy

$$\boldsymbol{r} = \frac{Cov(X,Y)}{\sigma x \sigma y}$$

3.3. Qualité prédictive du modèle en fonction de la taille du jeu de données de prédiction

Le modèle de prédiction de l'épaisseur de sol en fonction de la résistivité électrique a été obtenu grâce à un seul et unique jeu de donnée de prédiction N4. Pour tester la qualité prédictive de ce modèle, en fonction du nombre d'individus dans le jeu de données de prédiction, plusieurs tirages aléatoires, avec remises, ont été réalisés à partir du jeu de données des 80 points de sondages (N2). Nous obtenons plusieurs jeux de données de prédiction avec 'N' points de sondages.

Pour chacun de ces nouveaux jeux de données, une régression linéaire permet d'obtenir une variante du modèle originel obtenue avec le jeu de prédiction N4.

Chaque modèle possède son jeu de validation avec 80-N points qui nous permet de calculer les indices statistiques présentés dans le paragraphe 3.2. L'évolution de la qualité prédictive du modèle est réalisée en comparant la variation de ces indices en fonction de la taille de l'échantillon du jeu de donnée de prédiction.

3.4. Taille de la maille de spatialisation de la variable explicative

Dans les deux scénarios de spatialisation de la variable cible (Ep), l'organisation de la variable explicative sur une grille régulière est nécessaire. En interpolant la variable explicative sur des grilles de résolution (taille de la maille élémentaire de la grille) variable, la taille optimale de la maille d'interpolation peut-être déterminée.

En effet, la variable cible (Ep) peut-être, dans un premier temps, spatialisée avec différentes résolutions et, dans un second temps, la taille optimale de la maille d'interpolation est déterminée en comparant les indices statistiques, présentés dans le paragraphe 3.2., qui seront obtenus grâce le jeu de donnée de validation (N5) et les différentes cartes de la variable cible (Ep dans notre cas).

4. Résultats et discussions

4.1. Détermination du modèle prédictif par régression linéaire simple

Les tableaux 2 et 3 résument les statistiques descriptives des jeux de données de prédiction et de validation.

*Tableau 2 : Statistiques descriptives du jeu de données de prédiction **N4***

	Effectif	Minimum	Maximum	Moyenne	Variance	Ecart-type
R1 (Ohm.m)	45	22,55	69,25	35,03	123,6	11,12
R2 (Ohm.m)	45	16,28	80,61	37,70	306,6	17,51
R3 (Ohm.m)	45	7,89	112,3	50,15	620,91	24,92
Epaisseur de sol (cm)	45	24	220	62,51	1949,18	44,15

*Tableau 3 : Statistiques descriptives du jeu de données de validation **N5***

	Effectif	Minimum	Maximum	Moyenne	Variance	Ecart-type
R1 (Ohm.m)	35	17,53	60,43	34,83	102,35	10,12
R2 (Ohm.m)	35	13,28	79,33	36,06	262,94	16,22
R3 (Ohm.m)	35	12,46	117,49	48,3	556,73	23,60
Epaisseur de sol (cm)	35	23	224	71,63	2470	49,70

Ces deux jeux de données possèdent des statistiques similaires permettant ainsi d'effectuer la validation du modèle prédictif qui sera établie à partir de N4**.** Les corrélations entre l'épaisseur du sol (Ep) et les différentes valeurs de résistivités (R1, R2 et R3) du sol, présentent dans le tableau 4, montrent un lien linéaire fort entre les valeurs Ep et R3.

Tableau 4 : Corrélation entre la variable d'intérêt (épaisseur de sol) et les valeurs de résistivités

Variables	Coefficient de corrélation
Ep	1
R1	-0,57
R2	-0,79
R3	-0,84

Le tableau 5 indique les résultats de la régression linéaire simple entre ces deux variables et permet d'obtenir l'équation du modèle entre l'épaisseur du sol et la résistivité de la voie 3.

Tableau 5 : Résultats de la régression linéaire simple entre Ep et R3

Source	Valeur	Erreur standard	t	Pr > |t|	Borne inférieure (95%)	Borne supérieure (95%)
Constante	0,000	0,080	0,000	1,000	-0,162	0,162
R3	-0,840	0,083	-10,152	< 0,0001	-1,007	-0,673

Ce modèle explique 70% de la variabilité de l'épaisseur de sol ($R^2 = 0.7$, $p<0.0001$). Le coefficient qui pondère R3 dans le modèle de régression (-0,84) est diffèrent de zéro.

La validation du modèle établi est réalisée grâce à N5. Les résultats montrent que ce modèle a tendance à sous-estimé des épaisseurs de sol (*Figure 4)*.

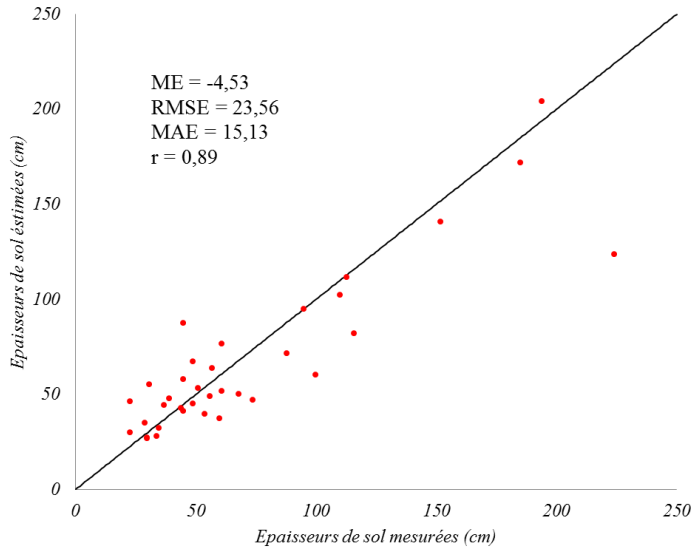

Figure 4 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimée avec le modèle : avant spatialisation

Avec une erreur moyenne proche de -4,5 cm et un coefficient de corrélation élevé, on peut dire que ce modèle permet d'obtenir une estimation proche des valeurs mesurées. Cependant, il ne parvient pas à reproduire certaines valeurs, qu'elles soient extrêmes ou non, et estime des valeurs inférieurs à celles rencontrées sur le terrain. Les valeurs de RMSE et de MAE attestent également de cette dernière remarque en prenant des valeurs assez élevées.

L'analyse des écarts entre les épaisseurs mesurées et les épaisseurs estimés (*Figure 5)* permet de constater que 77% des individus du jeu de validation (N5) possèdent moins de 20 cm d'écart avec la réalité. Ce modèle estime des valeurs proches des valeurs mesurées mais mérite d'être amélioré dans le but de mieux appréhender le bilan hydrique et azoté des sols.

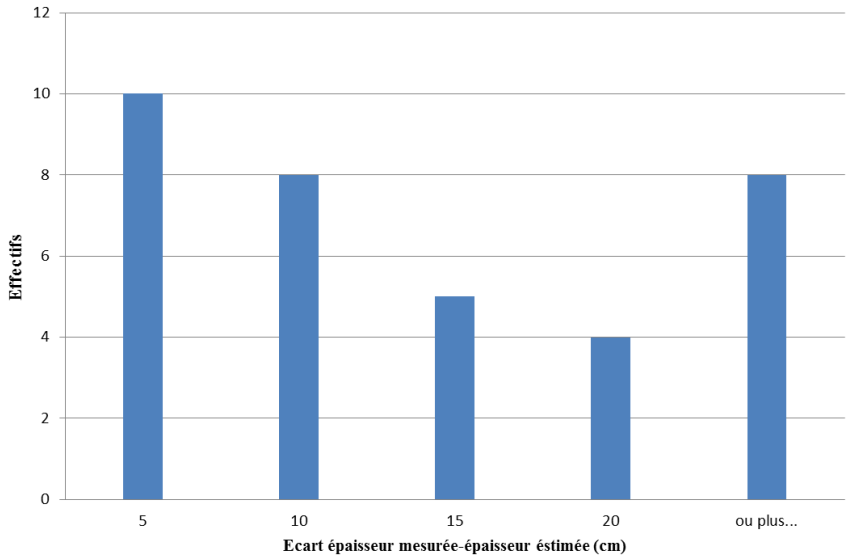

Figure 5: Ecarts entre épaisseurs mesurées et épaisseurs estimées par le modèle : statistiques sur la base du jeu de données de validation (N5)

4.2. Spatialisation de l'épaisseur de sol par régression linéaire et par régression-krigeage

4.2.1. Spatialisation par régression linéaire

Les spatialisations de l'épaisseur de sol ont été réalisées en tenant compte des variations à courte et longue distance de la variable explicative (R3). Pour déterminer ces variations nous réalisons un variogramme expérimentale. Il est estimé sur une distance de 1,5km avec un pas de 5m (*Figure 6)*. L'ajustement du variogramme expérimental a été effectué à l'aide de trois structures : un effet de pépite (**S0**) et deux modèles exponentiels (**S1** et **S2**) de portées respectives 140 et 170m.

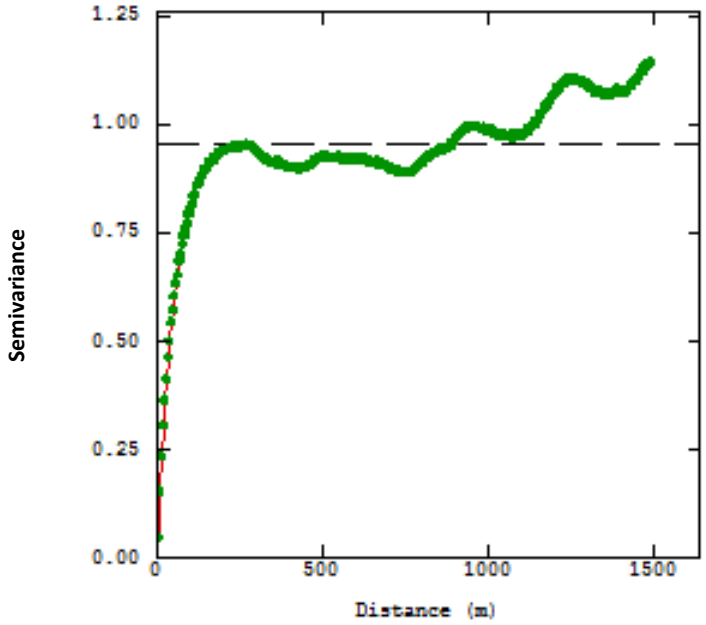

Figure 6 : Variogramme expérimental (points verts) et modèle théorique ajusté (courbe rouge) sur les valeurs Gaussiennes de R3

En Utilisant les données R3 et le variogramme présenté ci-dessus une carte (Figure 7) de la variable explicative (résistivité de la voie 3) est obtenue sur le domaine d'étude. Elle qui décrit le signal globale issue de la voie 3 de l'ARP qui spatialisé (R3*).

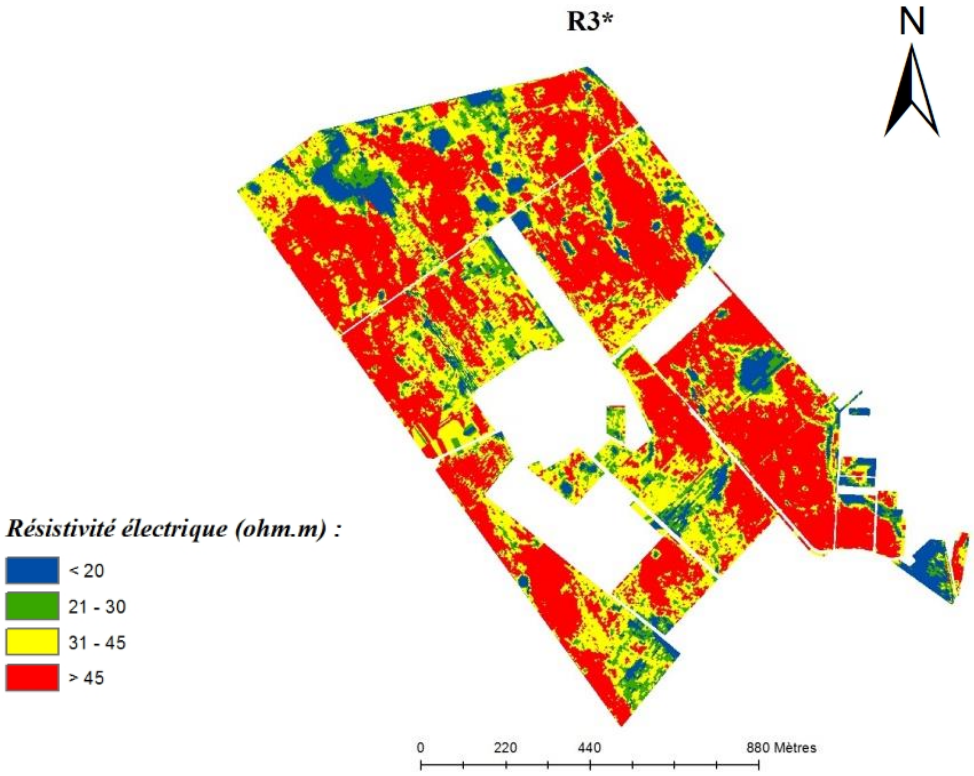

Figure 7 : Cartes de spatialisation de la résistivité des sols issues des données de la voie 3 de l'ARP.

La spatialisation de l'épaisseur par régression linéaire simple (*Figure 8*) a été réalisée en appliquant le modèle, établi dans la section 4.1, à R3* qui décrit le signal globale issue de la voie 3 de l'ARP. Cette variable est spatialisée par krigeage ordinaire sur une maille de 1m² en tenant compte du variogramme précédemment cité.

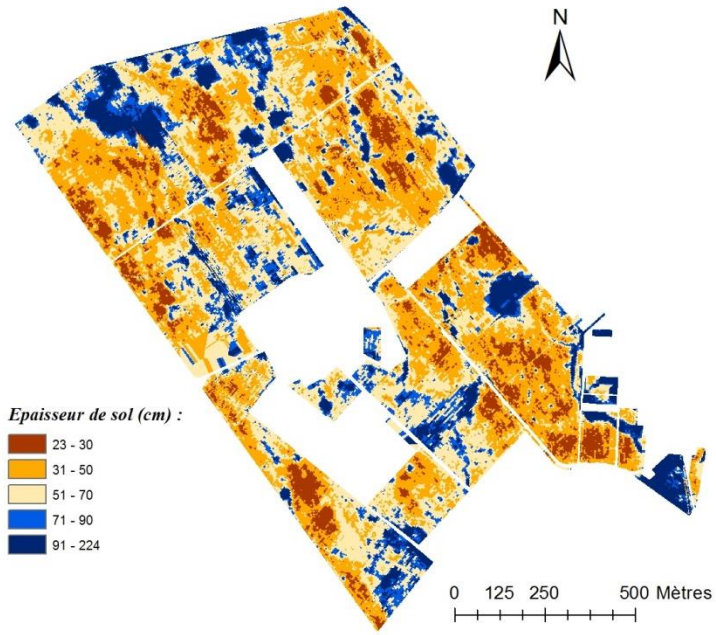

Figure 8 : Carte de l'épaisseur de sol par régression linéaire simple

La validation de la carte à partir de N5 (*Figure 9*) montre que cette spatialisation permet d'obtenir une estimation des épaisseurs de sol assez proche de la réalité observée dans le jeu de données de validation.

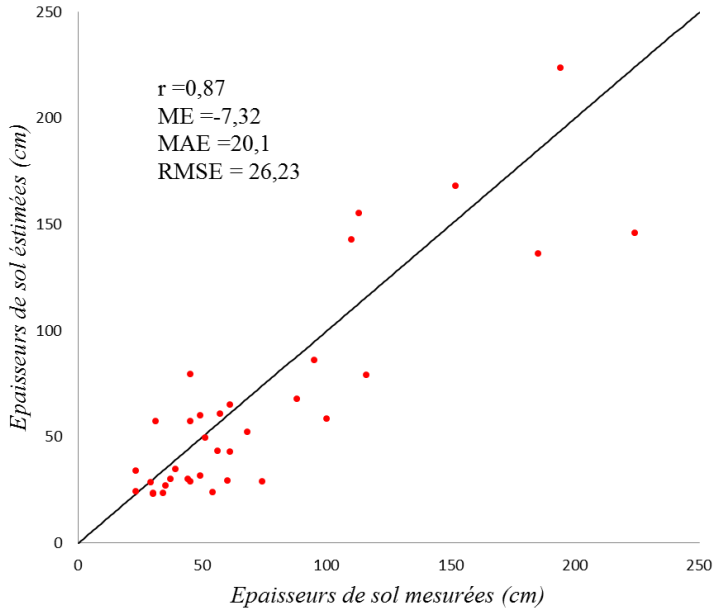

Figure 9 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimée avec le modèle : après spatialisation

La validation après spatialisation (*Figure 9*) révèle aussi des valeurs des indices statistiques de validation en moyenne quasi-similaires à celles présentées en figure 4. Cela étant, l'examen des écarts (*Figure 10*), selon que la validation soit réalisée avant ou après spatialisation, montre que la validation avant spatialisation peut engendrer des résultats très optimistes. En effet, nous constatons qu'un écart à la réalité de plus de 20cm peut être observé chez quasiment un individu sur deux du jeu de données de validation après spatialisation. Aussi, on peut dire que la spatialisation par régression linéaire simple donne une image globale de la variation de l'épaisseur des sols sur le site d'étude. Il nous semble cependant qu'une meilleure précision de la carte de l'épaisseur des sols est indispensable pour répondre aux objectifs du projet Carex.

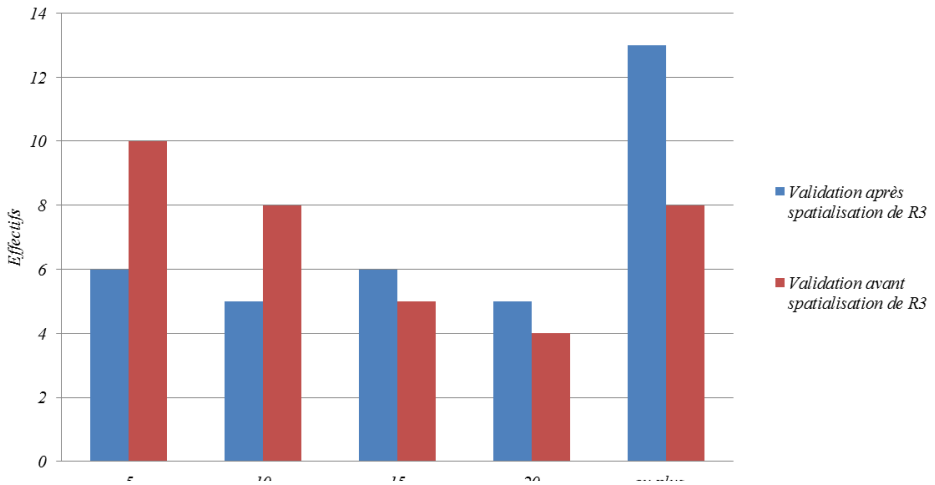

Figure 10 : Ecarts entre épaisseurs mesurées et épaisseurs estimées par le modèle : statistiques sur la base du jeu de données N5 avec validation avant et après spatialisation

4.2.2. Spatialisation par régression-krigeage

Dans cette partie, nous utilisons pour la spatialisation de l'épaisseur de sol le variogramme issue de R3* estimée dans le paragraphe précédent. Il est estimé sur une distance de 1,5km avec un pas de 5m. Il est ajusté par un variogramme théorique (*Figure 11*) composé d'un effet de pépite S0 et de deux modèles exponentiels S1 et S2 de portée respective 240 et 300 m.

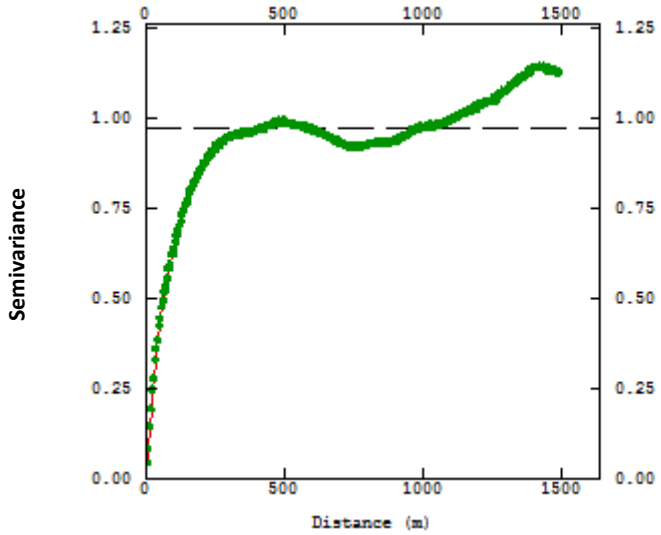

Figure 11 : Variogramme expérimental (points verts) et modèle théorique ajusté (courbe rouge) sur les valeurs de R3.*

En utilisant les données R3* et le variogramme présenté ci-dessus, trois cartes (*Figure 12*) de la variable explicative spatialisé sont obtenues pour le domaine étudié. Une première (S0- R3*) qui décrit le bruit contenu dans les données et deux autres (S1-R3* et S2-R3*) qui représentent respectivement les variations de R3* à courtes distances et à longues distances.

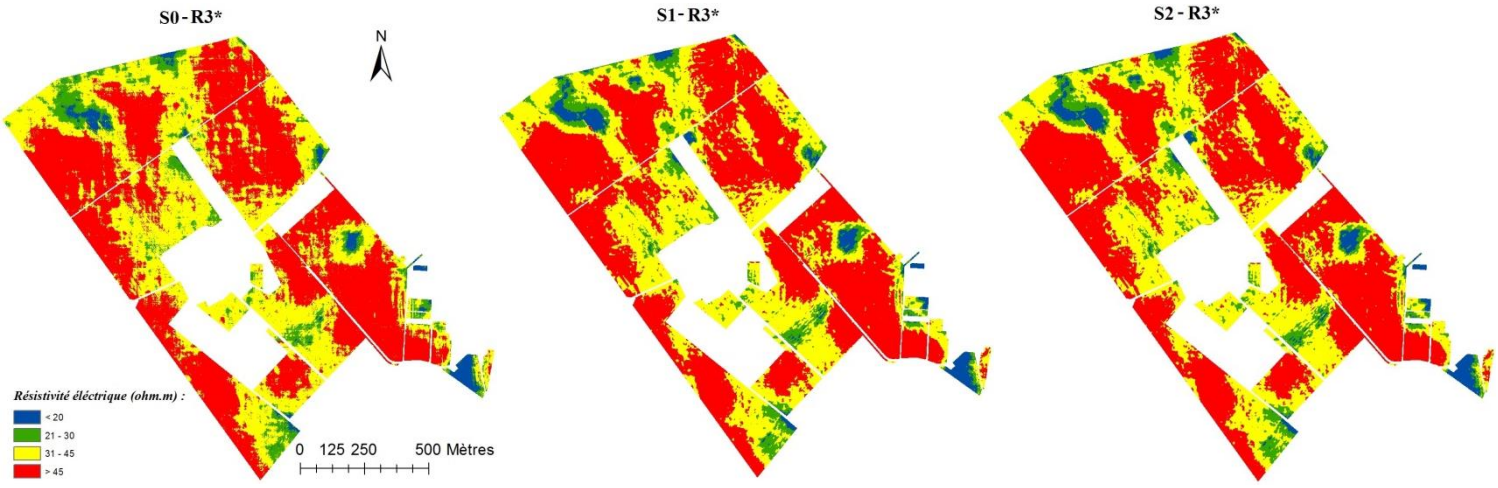

Figure 12 : Cartes de la résistivité des sols et de ses composantes spatiales issues des données de R3.*

L'utilisation des composantes du variogramme modifie la spatialisation de R3. En effet certaines zones de faibles résistivités sont harmonisées avec les zones de plus fortes résistivités qui les entourent.

La spatialisation des composantes de R3* (S0-R3*, S1-R3* et S2-R3*) et leur projection sur N2 permet d'identifier la variable explicative la plus pertinente pour la spatialisation de Ep (*Tableau 6)*.

Tableau 6 : Corrélation entre la variable d'intérêt (Ep) et R3 et ses composantes*

Variables	Coefficient de corrélation
Ep	1
R3*	-0,81
S0-R3*	-0,63
S1-R3*	-0,8
S2-R3*	-0,8

Mis à part pour les valeurs de R3*-S0, les coefficients de corrélation sont très similaires. Aussi pour la spatialisation de Ep nous retenons la composante qui représente les grandes tendances de la variable R3 c'est-à-dire S2-R3*.

Par ailleurs dans la spatialisation de Ep nous tenons compte des résidus entre Ep et R3*-S2. Pour cela, on projette S2-R3* sur N2 puis on calcule la différence entre S2-R3* et Ep. Nous obtenons ainsi les résidus présentés dans le tableau 7.

Tableau 7 : Statistiques descriptives des résidus entre R3-S2 et Ep : valeurs Gaussiennes*

	Effectif	Minimum	Maximum	Ecart-type
Ep	80	-2,33	2,33	0,99
S2-R3*	80	-2,46	2,56	1,07
Résidus	80	-4.33	4,15	1,95

Afin de pouvoir intégrer nos résidus dans le système d'équations du krigeage, nous nous assurons que ces résidus présentent une structure spatiale. Pour ce faire un variogramme a été estimé sur une distance de 1080 m avec un pas de 54 m. (*Figure 13)*.

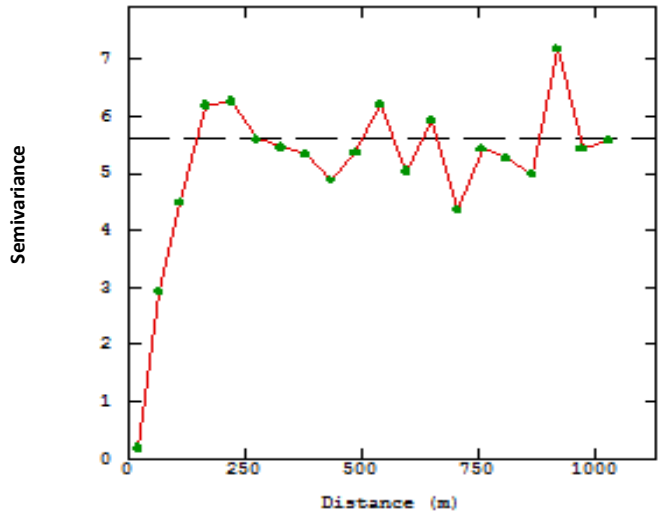

Figure 13 : Variogramme expérimental des résidus obtenus avec le jeu de données N2.

La courbe des dissemblances montre une évolution rapide aux faibles distances, ainsi qu'une absence de structuration spatiale aux grandes distances. Ces résidus sont donc fonction d'une distance de l'ordre de 200m et peuvent donc être spatialisés par krigeage ordinaire. Ainsi, la spatialisation de Ep (Figure 14) a été réalisée à partir des mesures de l'épaisseur (N2 = 45) en tenant compte de deux variables dans le systèmes d'équations du krigeage des variations spatiales des résidus et de la variable explicative S2-R3*. Ces deux variables sont connues en tout nœud de la grille d'estimation.

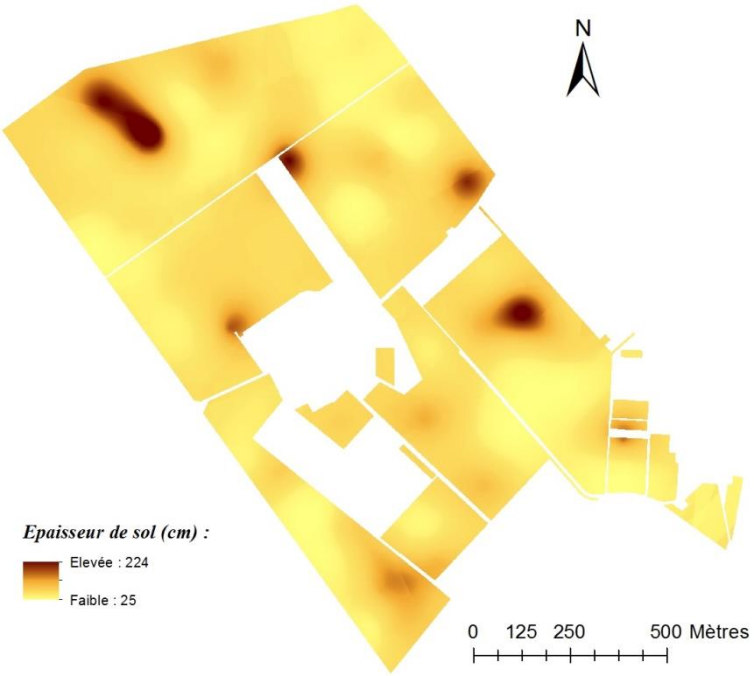

Figure 14 : Carte de l'épaisseur de sol par régression-krigeage..

La validation à partir de N5 (*Figure 15*) montre une forte corrélation entre les épaisseurs mesurées et les épaisseurs estimées. Le biais est faible (ME = -2.87cm) et la précision moyenne (RMSE) est « convenable ».

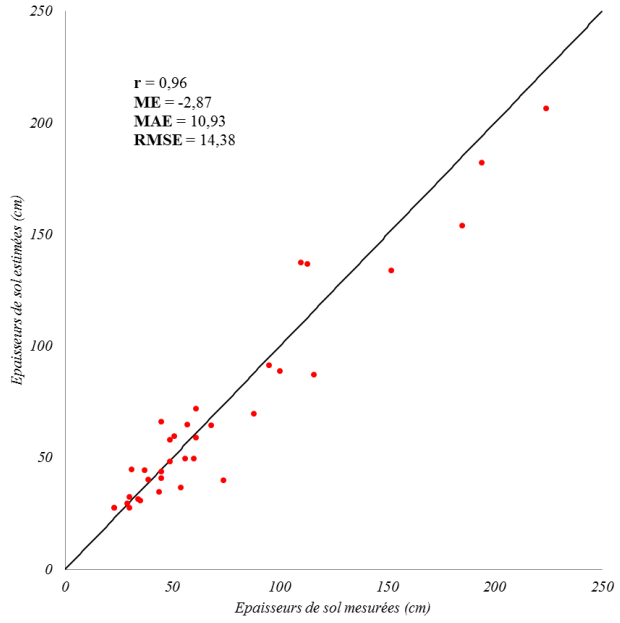

*Figure 15 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimées avec le modèle : **régression-krigeage***

Par ailleurs, les écarts entre les valeurs estimées et les valeurs mesurées en considérant N5 (Figure 16) sont moins importants dans l'estimation par régression-krigeage que dans celle par régression linéaire simple. Les écarts inférieurs à 5 cm représentent 40% de nos résultats pour la méthode par régression-krigeage et seulement 17% pour la méthode par régression linéaire simple.

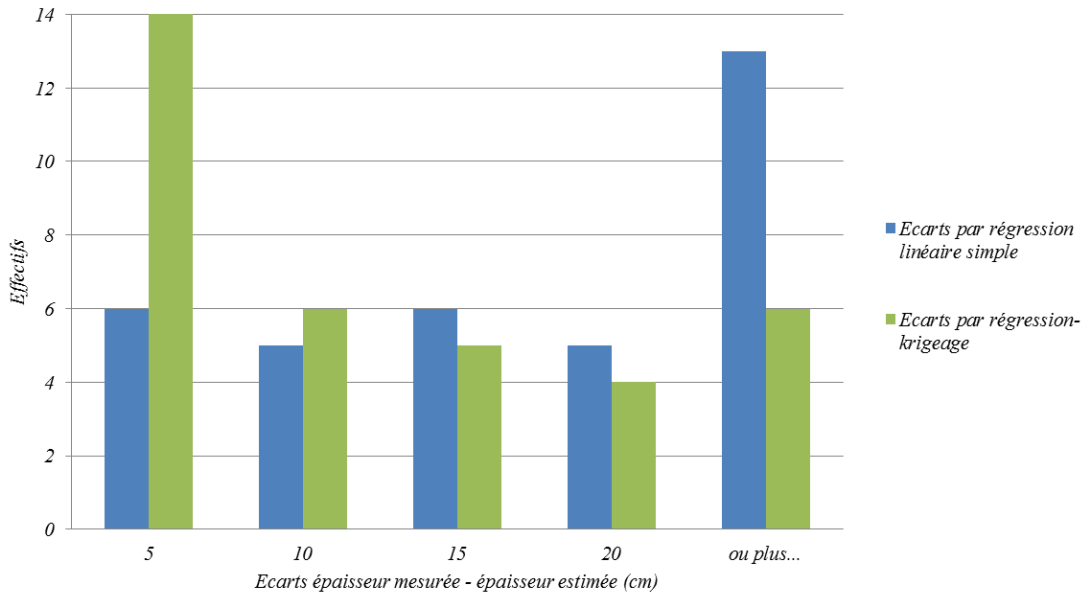

Figure 16 : Ecarts entre épaisseurs mesurées et épaisseurs estimées par régression-krigeage et régression linéaire simple : statistiques sur la base du jeu de données N5

L'ensemble de ces résultats montre que la spatialisation par régression-krigeage donne de meilleurs résultats et plaide en faveur de la : (1) prise en compte des résidus entre la variable d'intérêt et la variable explicative ; (2) sélection de la composante de la variable explicative la moins bruité (basses fréquences) lors dans la spatialisation de la variable d'intérêt.

4.3. Evolution de la qualité prédictive

4.3.1. Sensibilité du modèle à la taille du jeu de donnée de prédiction

Cas de la régression linéaire simple :

Pour examiner la sensibilité du modèle de régression à la taille du jeu de donnée de prédiction, des tirages aléatoires avec remises (Tableau 8) ont été effectués à partir de N2. Ainsi, pour chaque taille, nous disposons de 10 jeux de données de prédiction et leurs corollaires de validation.

Tableau 8 : Récapitulatif des tirages aléatoires

Jeux de prédiction (N=)	15	20	30	35	45	55	60	65	70
Jeux de validation (N=)	65	60	50	45	35	25	20	15	10

Avec chaque jeu de donnée de prédiction, à 'N' valeurs, nous effectuons dix régressions linéaires, entre Ep et R3, afin d'obtenir une amplitude des indices statistiques de validation retenus, et de déterminer l'évolution de la qualité prédictive du modèle.

Pour une taille du jeu de donnée de prédiction (par exemple pour N = 15), la figure 17 montre l'amplitude de variation des 10 valeurs du coefficient de corrélation **r** entre les valeurs mesurées des jeux de validation (10 jeux de validation avec 65 individus chacun) et les valeurs estimées par les dix modèles de régression.

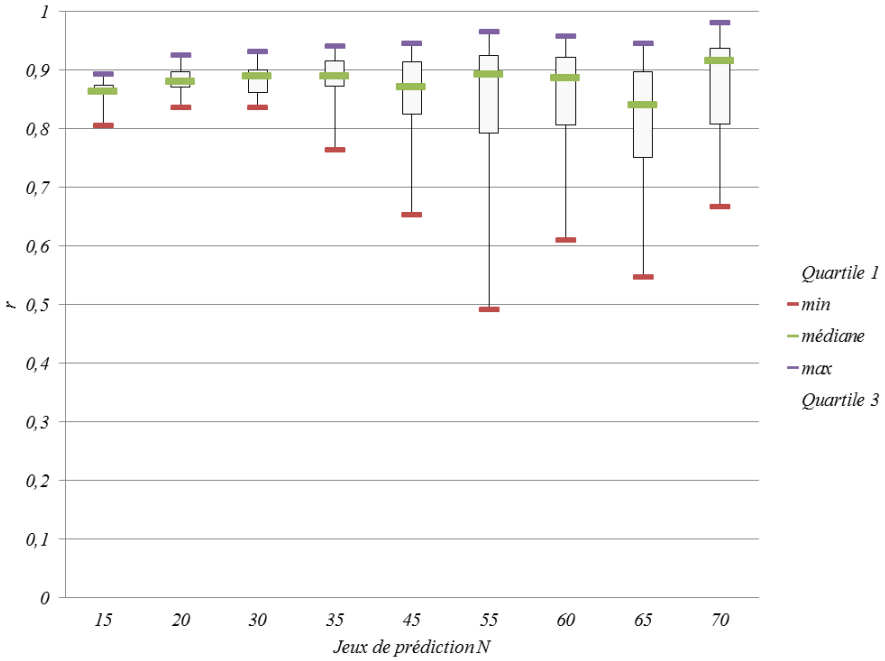

Figure 17 : Evolution du coefficient de corrélation en fonction du jeu de prédiction (N) par régression linéaire simple

Nous constatons que **r** varie peu pour des échantillons de prédiction de petites tailles. Cependant à partir d'un jeu de données dépassant les 30 individus, l'amplitude de ce coefficient augmente de manière significative. A ce stade, nous pouvons préconiser de prévoir la validation à partir d'un jeu de données plus important que le jeu de donnée de prédiction.

Pour confirmer ce résultat, nous examinons en plus l'évolution de la MAE de la RMSE pour les différentes tailles d'échantillon de prédiction (*Figure 18*).

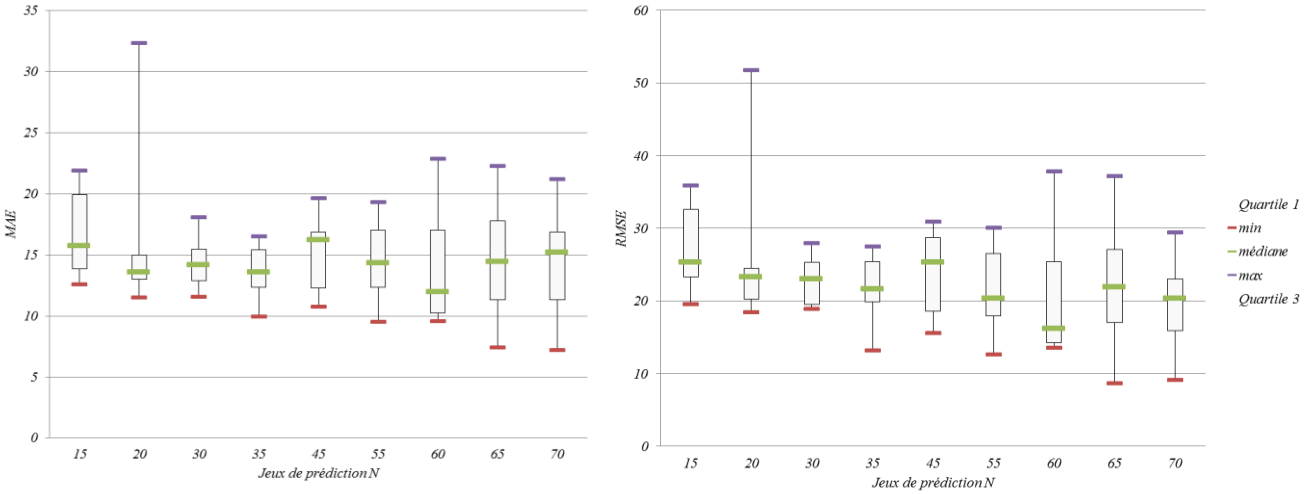

Figure 18 : Evolution du RMSE et du MAE en fonction du jeu de prédiction (N) par régression linéaire simple

Nous pouvons constater que la MAE et la RMSE suivent la même évolution. Les amplitudes de ces deux indices statistiques sont les plus importantes pour des jeux de prédictions de grandes et de petites tailles. Elles diminuent, jusqu'à des jeux de prédiction avoisinant les 30 individus, et augmentent par la suite, pour des jeux de prédictions de tailles supérieurs. Cependant ces deux indices statistiques semblent atteindre un optimum pour des jeux de données de prédiction proche de N=30.

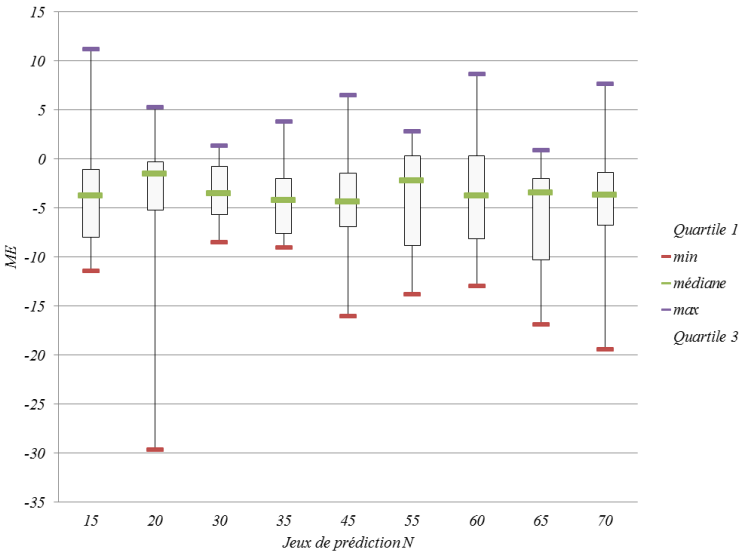

Figure 19 : Evolution de l'erreur moyenne en fonction du jeu de prédiction (N) par régréssion linéaire simple

L'évolution de l'erreur moyenne en fonction de l'évolution de la taille de l'échantillon de prédiction (*Figure 19)* est similaire à celles de la RMSE et de la MAE. Nous retrouvons de fortes amplitudes pour des jeux de prédiction de petites et de grandes tailles, ainsi qu'une amplitude minimum pour les jeux de prédiction avoisinant N=30. Une tendance à la sous-estimation de l'épaisseur de sol par notre modèle semble néanmoins présente pour l'ensemble des jeux de prédictions.

D’après ces constations, nous pouvons supposer que dans le contexte de ce travail une estimation spatiale à partir d’un jeu de donnée de l’ordre de 30 individus serait optimal pour la méthode par régression linéaire simple.

Cas de la régression-krigeage

De la même manière que dans la partie régression linéaire simple, nous validons nos résultats avec les jeux de données de validation à 80-N points associés aux jeux de prédictions.

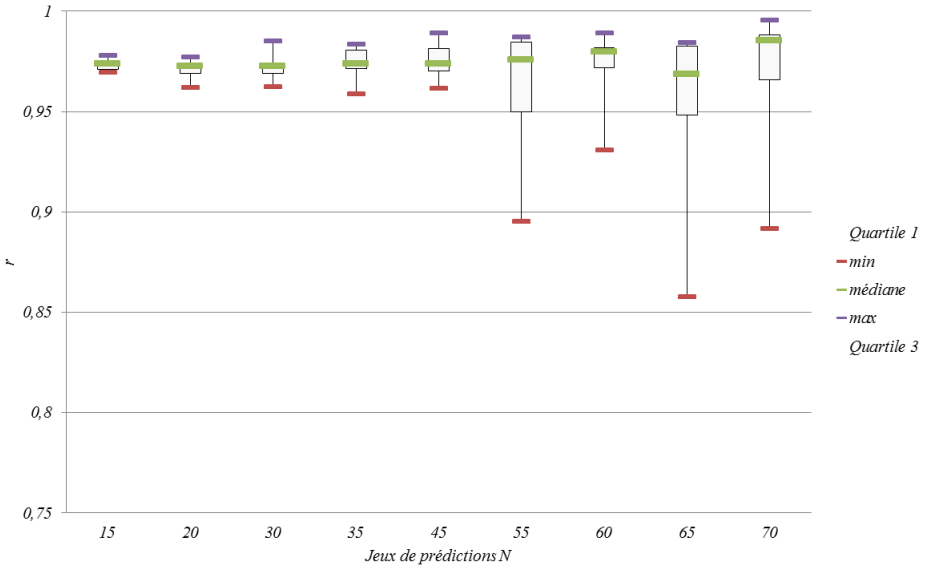

Figure 20 : Evolution du coefficient de corrélation en fonction du jeu de prédiction (N) par régression-krigeage

En augmentant la taille du jeu de donnée de prédiction, nous constatons (Figure 20) que **r** présente une amplitude de variation plus importante à partir de N = 55. Cependant ses valeurs moyennes restent élevées et quasi-stables (>0,95) et attestent d’une bonne corrélation entre nos estimations par régression-krigeage et les mesures de terrain. La comparaison des valeurs de RMSE et de MAE (*Figure 21*) montre des amplitudes de variations plus importantes et qu’un jeu de données de prédiction de l’ordre de 30 individus semble optimal.

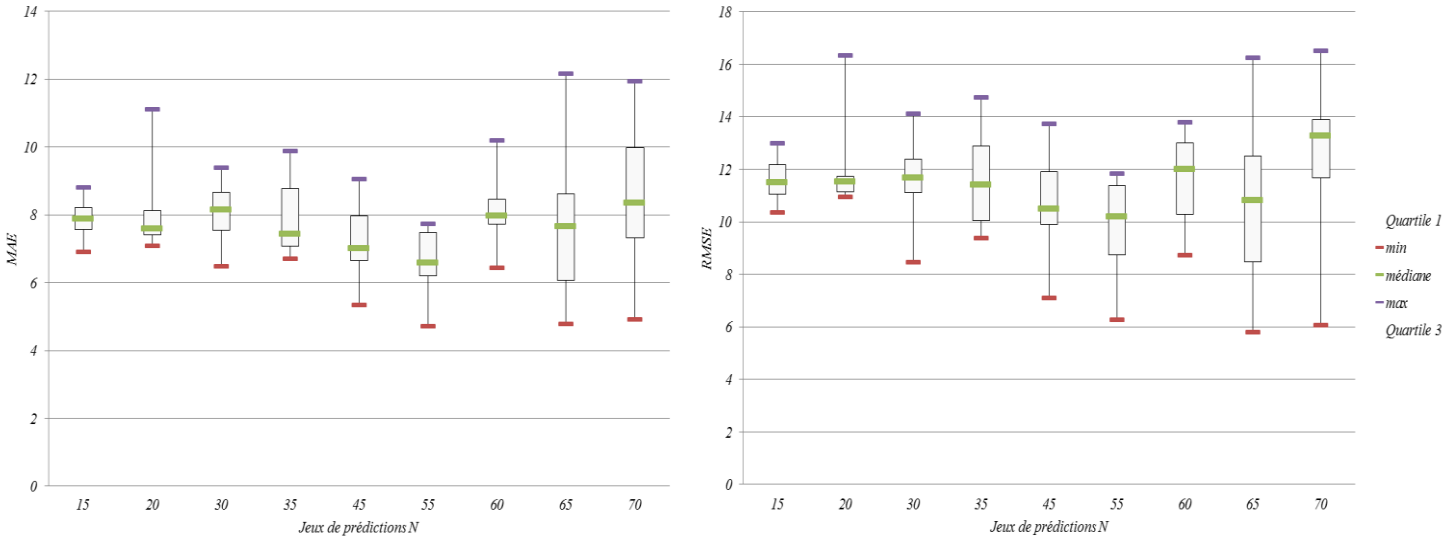

Figure 21 : Evolution du RMSE et du MAE en fonction du jeu de prédiction (N) par régression-krigeage

En effet, au cours de l'augmentation du jeu de prédiction, les résultats de MAE et de RMSE restent « convenables » pour l'ensemble des jeux de prédiction. L'amplitude de ces indices est légèrement plus faible pour un jeu de prédiction N=30. En étudiant l'évolution de l'erreur moyenne (*Figure 21)*, nous pouvons observer un phénomène similaire.

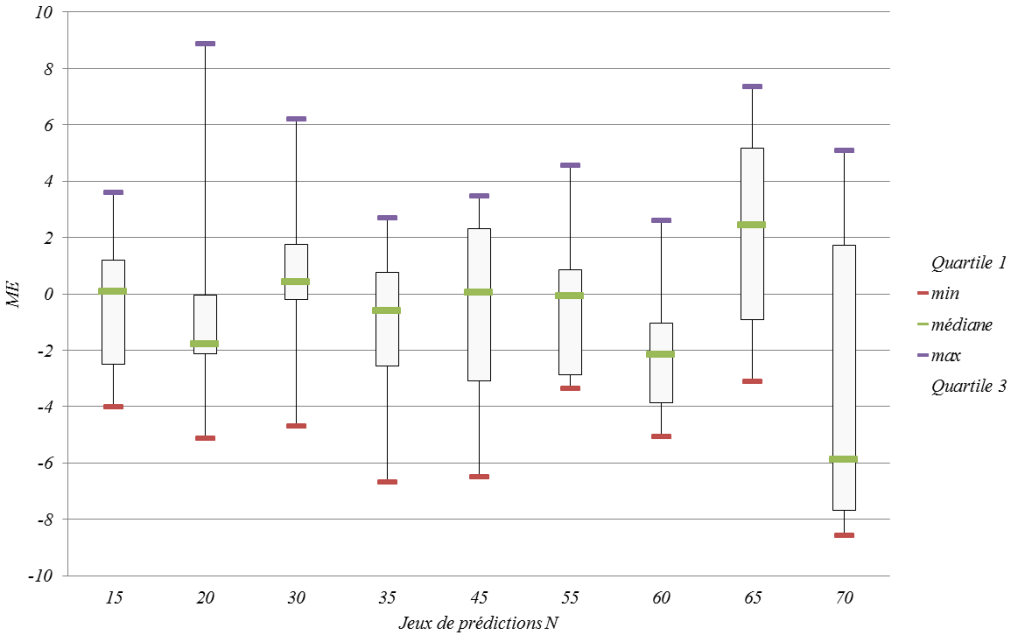

Figure 22 : Evolution de l'erreur moyenne en fonction du jeu de prédiction (N) par régréssion-krigeage

Nous retrouvons des valeurs optimales pour des jeux de prédictions avoisinant les N=30, cependant les erreurs moyennes compris entre -2 et 2 restent également présent pour un nombre important de cas.

L'évolution de ces indices statistique, en fonction de la taille du jeu de prédiction, confirme que l'approche par régression-krigeage avec un jeu de données de prédiction de l'ordre de 30 individus semble optimale dans le contexte de cette étude.

4.3.2. Sensibilité du modèle de spatialisation à la résolution de la variable explicative

La spatialisation de l'épaisseur de sol est effectuée selon deux approches, par régression linéaire simple et par régression-krigeage. La maille que nous avons utilisée précédemment est d'une résolution de 1m^2. Afin de comparer la sensibilité de l'approche de spatialisation, nous avons fait varier cette résolution selon des mailles de 1, 4, 9, 16 et 25 m^2. Les indices statistiques qui en résultent proviennent de la validation par le jeu de donnée N5. Ils nous permettent de déterminer si les approches utilisées sont stables ou non selon la taille de la maille de la variable explicative.

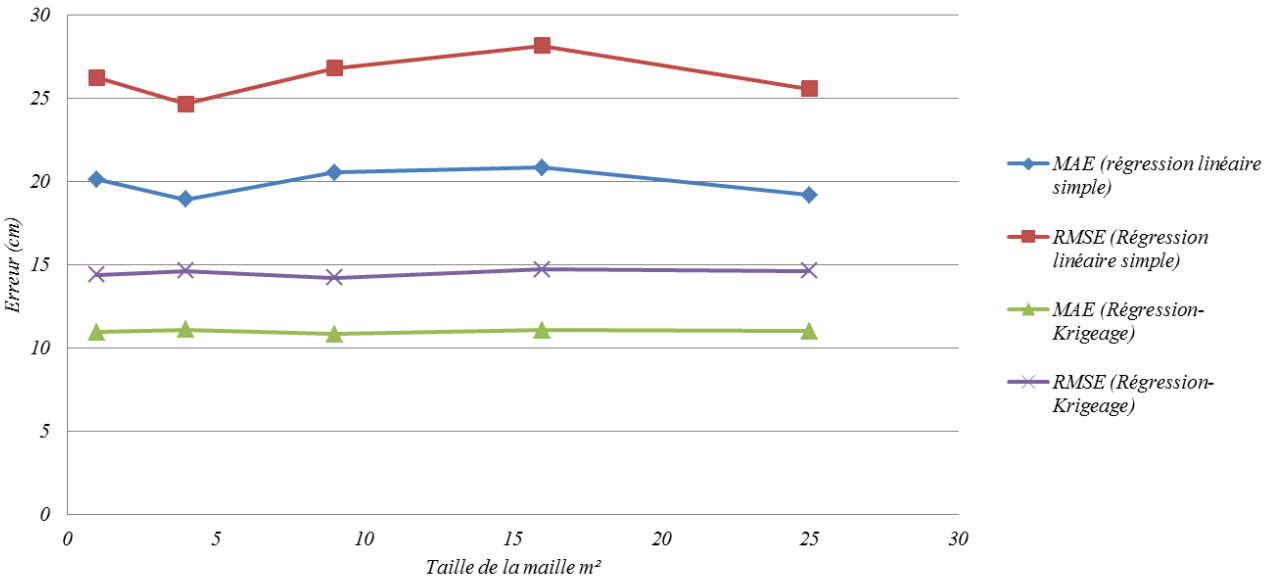

Figure 23: Evolution de la RMSE et de MAE en fonction de la maille pour les deux approches de spatialisation : régression-krigeage et régression linéaire simple

En observant l'évolution du RMSE et du MAE (*Figure 23)* en fonction de la taille de la maille, nous pouvons constater une légère différence entre la méthode par régression linéaire simple et par régression-krigeage. En effet la MAE reste comprise entre 18,9 et 20, 8 cm et la RMSE comprise entre 24,6 et 28,1 cm pour la première méthode. L'approche par régression linéaire simple semble exprimer une légère sensibilité face à la variation de la maille.

Dans le cas de la régression-krigeage, nous observont une MAE comprise entre 10,8 et 11 cm et une RMSE comprise entre 14,2 et 14,6 cm. Cette approche semble produire des estimations semblables pour n'importe qu'elle taille de maille utilisé pour la spatialisation de la variable explicative. Pour confirmer ce résultat nous étudions l'évolution de l'erreur moyenne pour ces deux méthodes (*Figure 24).*

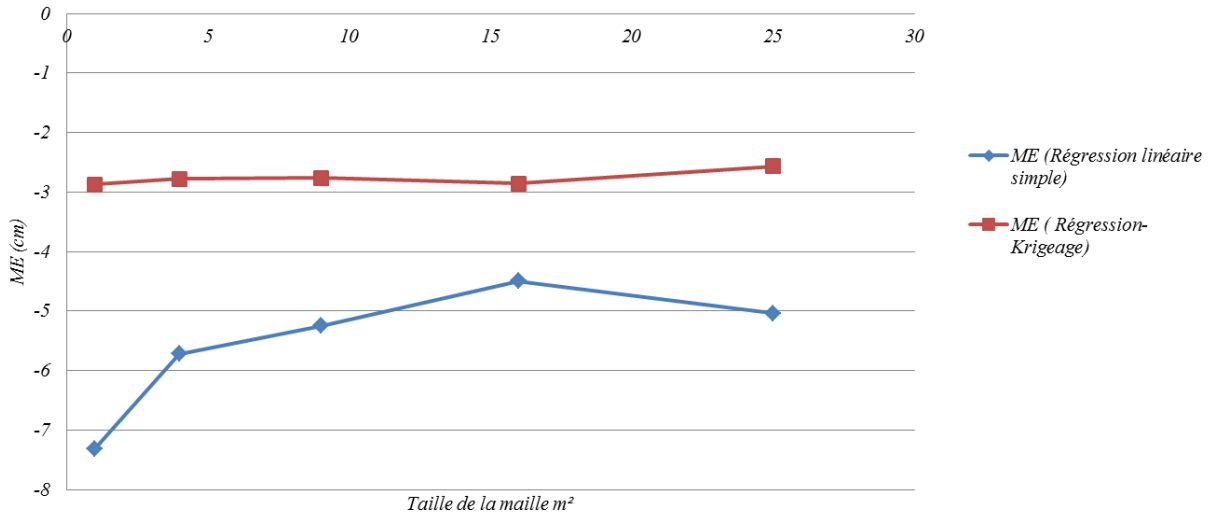

Figure 24 : Evolution de l'erreur moyenne en fonction de la maille pour la méthode régression-krigeage et régression linéaire simple.

Nous constatons que l'erreur moyenne, obtenue par la régression linéaire simple, semble atteindre un palier et n'évolue que très peu lors de l'augmentation de la taille de la maille. Elle diminue la sous-estimation de l'épaisseur de sol, lorsque l'on augmente la taille de la maille et atteint un point d'inflexion pour une résolution de 9m^2. En comparaison la méthode par régression-krigeage atteint la même estimation pour toutes les tailles de mailles

Cette dernière remarque atteste de la stabilité de la méthode par régression-krigeage face à un changement dans la résolution de la maille de spatialisation. Nous choisissons d'étudier le coefficient de corrélation afin de comprendre cette variation entre ces deux méthodes (*Figure 25).*

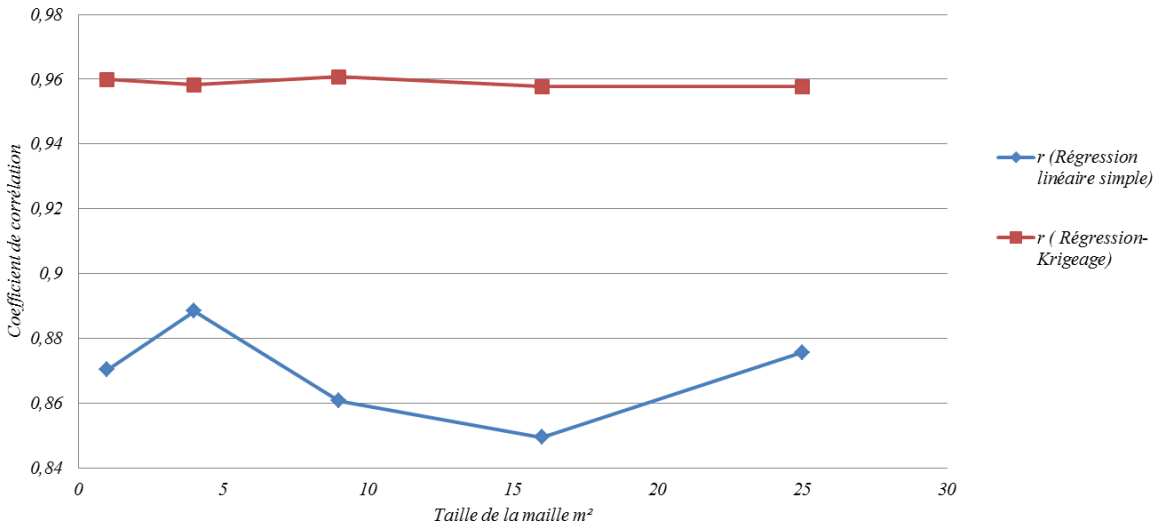

Figure 25 : Evolution du coefficient de corrélation en fonction de la maille pour la méthode KED et régression linéaire.

Cette dernière comparaison, nous permet de définir que la méthode par régression-krigeage permet d'obtenir des résultats plus corrélé avec la réalité et plus stable, en restant proche de 0,96, tandis que la méthode par régression linéaire simple présentent certaines variations tout en produisant des estimations moins réaliste lors de ces changements.

En résumé les résultats (*Figures 23 à 25*) montrent : (1) que l'approche de spatialisation par régression-krigeage est plus performante que celle de spatialisation par régression linéaire simple et ceci quelle que soit la taille de maille de spatialisation de la variable explicative ; (2) l'approche de spatialisation par régression-krigeage n'est pas sensible à la variation de la taille de maille de spatialisation de la variable explicative contrairement à la spatialisation par régression linéaire simple ou les variations des valeurs de r et ME, en particulier, présentent des points d'inflexions. Ce résultat permet donc de dire que la spatialisation de la variable explicative peut être réalisée sur une grille de faible résolution sans pour autant affecter la qualité prédictive de l'approche de spatialisation de la variable d'intérêt par régression-krigeage.

5. Conclusion

L'objectif de ce travail était de modéliser et spatialiser l'épaisseur du sol (variable d'intérêt) à partir de mesures ponctuelles de l'épaisseur du sol et de la résistivité électrique des sols (variable auxiliaire) mesurée de façon quasi-exhaustive sur le domaine d'étude. Il s'agissait aussi de valider le modèle établi et d'étudier l'évolution de sa qualité prédictive en fonction : (1) de la taille des échantillons de mesure de la variable d'intérêt ; (2) de la résolution spatiale de la variable auxiliaire.

Deux scénarios de spatialisation de la variable d'intérêt ont été proposés dans ce travail : spatialisation par régression linéaire simple et spatialisation par régression-krigeage. Les résultats ont montré que la prise en compte, de la composante spatiale de la variable auxiliaire la moins bruitée et des résidus spatialisés entre cette composante et la variable d'intérêt, dans le système d'équations du krigeage (spatialisation par régression-krigeage), permet d'obtenir des estimations spatiales très fiables.

L'analyse de sensibilité du modèle à la taille de l'échantillon de prédiction a principalement révélé que l'approche de spatialisation par régression-krigeage est légèrement sensible à la taille de l'échantillon de prédiction et suggère un échantillon de validation dont le nombre d'individus est quasiment le double de celui de l'échantillon de prédiction.

Quant à la résolution spatiale de la variable auxiliaire, les résultats ont montré qu'il n'est pas nécessaire de spatialiser la variable auxiliaire avec une forte résolution spatiale. En d'autres termes le scénario de spatialisation en particulier par régression-krigeage n'est pas sensible à la résolution spatiale de la variable auxiliaire.

Références bibliographiques

Association française pour l'étude du sol, 2008, *Référentiel pédologique*, Quae.

Bourennane, H., 1997. *Etude des lois de distribution spatiale des sols de petite Beauce - Application à la cartographie d'un horizon par couplage de méthodes morphométriques et Géostatistiques*, Thèse, Université d'Orléans.

Chrétien, J., 1976. *Carte pédologique de la France à 1/100.000*. Feuille de Dijon. INRA. Editions carte et notice. 218p

Ndjiba, M., D., 2013. *Valorisation d'une prospection électrique pour établir une carte de l'épaisseur du sol,* Mémoire Master 2 GPRE Sols, eaux et environnement, Université de Lorraine, INRA Orléans.

Sun, W., Minasny, B., Mcbratney, A., 2012. *Analysis and prediction of soil properties using local regression-kriging*, Geoderma 171-172, 16-23.

Wackernagel, H., 1995. *Multivariate geostatistics.* Springer-Verlag Berlin Heidelberg. Berlin.

Watt, S., M., Palmer, J., D., 2012. *Use of regression kriging to develop a Carbon:Nitrogen ratio surface for New Zealand*, Geoderma 183-184, 49-57.

Liste des figures

Figure 1 : Localisation de l'UE d'Epoisses 6

Figure 2 : Représentation du système ARP, modifié à partir de 3D Inversion of Automated Resistivity Profiling (ARP) 7

Figure 3 : Représentation de la résistivité R2 avec les points de sondages 8

Figure 4 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimée avec le modèle : avant spatialisation.....14

Figure 5 : Ecarts entre épaisseurs mesurées et épaisseurs estimées par le modèle : statistiques sur la base du jeu de données de validation (N5). 14

Figure 6 : Variogramme expérimental (points verts) et modèle théorique ajusté (courbe rouge) sur les valeurs Gaussiennes de R3 15

Figure 7 : Cartes de spatialisation de la résistivité des sols issues des données de la voie 3 de l'ARP. 16

Figure 8 : Carte de l'épaisseur de sol par régression linéaire simple 16

Figure 9 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimée avec le modèle : après spatialisation 17

Figure 10 : Ecarts entre épaisseurs mesurées et épaisseurs estimées par le modèle : statistiques sur la base du jeu de données N5 avec validation avant et après spatialisation ... 17

Figure 11 : Variogramme expérimental (points verts) et modèle théorique ajusté (courbe rouge) sur les valeurs de R3*. 18

Figure 12 : Cartes de la résistivité des sols et de ses composantes spatiales issues des données de R3*. 18

Figure 13 : Variogramme expérimentale des résidus obtenue avec le jeu de données **N2** 19

Figure 14 : Carte de l'épaisseur de sol par régression-krigeage**.**..... 20

Figure 15 : Comparaison entre épaisseurs de sol mesurées et épaisseurs de sol estimée avec le modèle : régression-krigeage 20

Figure 16 : Ecarts entre épaisseurs mesurées et épaisseurs estimées par régression-krigeage et régression linéaire simple : statistiques sur la base du jeu de données N5 21

Figure 17 : Evolution du coefficient de corrélation en fonction du jeu de prédiction (N) par régression linéaire simple 22

Figure 18 : Evolution du RMSE et du MAE en fonction du jeu de prédiction (N) par régression linéaire simple 23

Figure 19 : Evolution de l'erreur moyenne en fonction du jeu de prédiction (N) par régréssion linéaire simple 23

Figure 20 : Evolution du coefficient de corrélation en fonction du jeu de prédiction (N) par régression-krigeage 24

Figure 21 : Evolution du RMSE et du MAE en fonction du jeu de prédiction (N) par régression-krigeage 24

Figure 22 : Evolution de l'erreur moyenne en fonction du jeu de prédiction (N) par régréssion-krigeage 25

Figure 23 : Evolution de la RMSE et de MAE en fonction de la maille pour les deux approches de spatailisation : régression-krigeage et régression linéaire simple 26

Figure 24 : Evolution de l'erreur moyenne en fonction de la maille pour la méthode régression-krigeage et régression linéaire simple. 27

Figure 25 : Evolution du coefficient de corrélation en fonction de la maille pour la méthode KED et régression linéaire. 28

Liste des tableaux

Tableau 1 : Statistiques descriptives des données de résistivités N1. 7

Tableau 2 : Statistiques descriptives du jeu de données de prédiction N4. 13

Tableau 3 : Statistiques descriptives du jeu de données de validation N5. 13

Tableau 4 : Corrélation entre la variable d'intérêt (épaisseur de sol) et les valeurs de résistivités 13

Tableau 5 : Résultats de la régression linéaire simple entre Ep et R3 13

Tableau 6 : Corrélation entre la variable d'intérêt (Ep) et R3* et ses composantes 19

Tableau 7 : Statistiques descriptives des résidus entre R3*-S2 et Ep : valeurs Gaussiennes 19

Tableau 8 : Récapitulatif des tirages aléatoires 22

Liste des annexes

Annexe 1 *:* Tableau récapitulatifs des points de sondages pédologiques 33

Annexe 2 *:* Jeu de données de prédiction N4 utilisé pour l'établissement du modèle 34

Annexe 3 *:* Jeu de données de validation N5 utilisé pour la validation du modèle. 35

Annexe 1 : Tableau récapitulatif des points de sondages pédologiques

Numero	X	Y	Année de mesure	Epaisseur de sol (cm)	Numero	X	Y	Année de mesure	Epaisseur de sol (cm)
1	858127,609	6684604,485	2013	152	41	858576,8876	6683592,8551	2015	80
2	859213,971	6684008,445	2013	192	42	858512,9220	6683616,9136	2015	45
3	858832,025	6683382,094	2013	100	43	858736,9964	6683720,8105	2015	70
4	859473,473	6683673,431	2013	113	44	858961,6458	6683728,8016	2014	90
5	858531,705	6684289,125	2013	54	45	859049,2684	6683751,5237	2014	68
6	858446,028	6684658,609	2013	116	46	859343,8989	6683807,0988	2015	30
7	858986,867	6683684,028	2013	61	47	859221,6905	6683843,9135	2015	30
8	858847,641	6683285,093	2013	82	48	859096,6472	6683896,7575	2015	60
9	858952,060	6684165,964	2013	57	49	858401,6739	6683944,2547	2014	51
10	858842,500	6683781,558	2013	50	50	859093,7715	6683966,0773	2015	95
11	858541,256	6684044,121	2013	67	51	859247,7962	6683966,2256	2015	110
12	858322,866	6684329,287	2013	31	52	858291,8448	6683985,5754	2014	54
13	858478,157	6684532,771	2013	46	53	858315,3661	6684088,3552	2014	48
14	858783,383	6684281,093	2013	23	54	858852,5880	6684101,5067	2014	51
15	858529,028	6683764,351	2013	37	55	859124,5184	6684112,5536	2015	60
16	858976,157	6683911,609	2013	39	56	858227,9892	6684127,2045	2014	31
17	858981,512	6683504,641	2013	49	57	858455,9066	6684156,4969	2014	71
18	859308,157	6683895,545	2013	61	58	858574,1285	6684203,7552	2014	49
19	858743,221	6684650,577	2013	44	59	858171,2530	6684208,2287	2014	39
20	859474,157	6683587,641	2013	26	60	858114,6392	6684307,5462	2014	59
21	858668,254	6683477,867	2013	26	61	858720,1419	6684311,2749	2014	49
22	858887,797	6684461,780	2013	28	62	858174,5503	6684311,9568	2014	56
23	858411,495	6684489,561	2013	74	63	859070,9167	6684343,1897	2014	194
24	859188,005	6683708,289	2013	24	64	858610,6503	6684399,5325	2014	220
25	858732,088	6684559,820	2013	34	65	857959,7374	6684406,9122	2014	61
26	859334,196	6683817,593	2013	23	66	858851,8058	6684416,8005	2014	99
27	858114,072	6684179,833	2013	25	67	858085,0501	6684430,7518	2014	55
28	858456,620	6683964,552	2013	185	68	858415,5668	6684453,8372	2014	29
29	858874,3001	6683188,2569	2015	45	69	858256,5315	6684471,6412	2014	220
30	858828,0249	6683189,2799	2015	35	70	858662,5590	6684481,9870	2014	61
31	858957,9224	6683247,1083	2015	80	71	858905,7250	6684513,8481	2014	37
32	858794,6977	6683324,0181	2014	50	72	858656,5539	6684523,4502	2014	65
33	858986,4918	6683368,9206	2014	80	73	858346,9816	6684547,5139	2014	35
34	858711,0510	6683375,8301	2015	35	74	858126,0227	6684549,7085	2014	224
35	859065,3216	6683428,9562	2014	53	75	858399,4282	6684599,6891	2014	63
36	858737,5233	6683450,6117	2015	60	76	858211,8383	6684618,7114	2014	88
37	858950,7820	6683455,0104	2014	30	77	858848,2448	6684665,1623	2014	29
38	858674,3818	6683457,3411	2015	30	78	858702,6704	6684676,6082	2014	33
39	859116,9319	6683559,9925	2014	80	79	858793,8358	6684703,2433	2014	45
40	859394,6760	6683587,2667	2015	35	80	858722,7517	6684728,7163	2014	34

Annexe 2 : Jeu de données de prédiction N4 utilisé pour l'établissement du modèle

Numero	X	Y	Date de mesure	Epaisseur de sol (cm)
2	859213,971	6684008,45	2013	192
5	858531,705	6684289,13	2013	54
8	858847,641	6683285,09	2013	82
10	858842,5	6683781,56	2013	50
11	858541,256	6684044,12	2013	67
13	858478,157	6684532,77	2013	46
20	859474,157	6683587,64	2013	26
21	858668,254	6683477,87	2013	26
22	858887,797	6684461,78	2013	28
24	859188,005	6683708,29	2013	24
27	858114,072	6684179,83	2013	25
30	858828,025	6683189,28	2015	35
31	858957,922	6683247,11	2015	80
32	858794,698	6683324,02	2014	50
33	858986,492	6683368,92	2014	80
34	858711,051	6683375,83	2015	35
35	859065,322	6683428,96	2014	53
36	858737,523	6683450,61	2015	60
39	859116,932	6683559,99	2014	80
40	859394,676	6683587,27	2015	35
41	858576,888	6683592,86	2015	80
43	858736,996	6683720,81	2015	70
44	858961,646	6683728,8	2014	90
46	859343,899	6683807,1	2015	30
47	859221,691	6683843,91	2015	30
48	859096,647	6683896,76	2015	60
53	858315,366	6684088,36	2014	48
54	858852,588	6684101,51	2014	51
56	858227,989	6684127,2	2014	31
57	858455,907	6684156,5	2014	71
59	858171,253	6684208,23	2014	39
60	858114,639	6684307,55	2014	59
61	858720,142	6684311,27	2014	49
64	858610,65	6684399,53	2014	220
65	857959,737	6684406,91	2014	61
66	858851,806	6684416,8	2014	99
67	858085,05	6684430,75	2014	55
69	858256,532	6684471,64	2014	220
70	858662,559	6684481,99	2014	61
71	858905,725	6684513,85	2014	37
72	858656,554	6684523,45	2014	65
75	858399,428	6684599,69	2014	63
77	858848,245	6684665,16	2014	29
78	858702,67	6684676,61	2014	33
80	858722,752	6684728,72	2014	34

Annexe 3 : Jeu de données de validation N5 utilisé pour la validation du modèle

Numero	X	Y	Année de mesure	Epaisseur de sol (cm)
1	858127,609	6684604,485	2013	152
3	858832,025	6683382,094	2013	100
4	859473,473	6683673,431	2013	113
6	858446,028	6684658,609	2013	116
7	858986,867	6683684,028	2013	61
9	858952,060	6684165,964	2013	57
12	858322,866	6684329,287	2013	31
14	858783,383	6684281,093	2013	23
15	858529,028	6683764,351	2013	37
16	858976,157	6683911,609	2013	39
17	858981,512	6683504,641	2013	49
18	859308,157	6683895,545	2013	61
19	858743,221	6684650,577	2013	44
23	858411,495	6684489,561	2013	74
25	858732,088	6684559,820	2013	34
26	859334,196	6683817,593	2013	23
28	858456,620	6683964,552	2013	185
29	858874,3001	6683188,2569	2015	45
37	858950,7820	6683455,0104	2014	30
38	858674,3818	6683457,3411	2015	30
42	858512,9220	6683616,9136	2015	45
45	859049,2684	6683751,5237	2014	68
49	858401,6739	6683944,2547	2014	51
50	859093,7715	6683966,0773	2015	95
51	859247,7962	6683966,2256	2015	110
52	858291,8448	6683985,5754	2014	54
55	859124,5184	6684112,5536	2015	60
58	858574,1285	6684203,7552	2014	49
62	858174,5503	6684311,9568	2014	56
63	859070,9167	6684343,1897	2014	194
68	858415,5668	6684453,8372	2014	29
73	858346,9816	6684547,5139	2014	35
74	858126,0227	6684549,7085	2014	224
76	858211,8383	6684618,7114	2014	88
79	858793,8358	6684703,2433	2014	45