Chapitre 2.2.6
Modélisation du sous-sol

Amélie LAURENT-DEHECQ(1), Xavier RODIER(2)

(1) CITERES-LAT UMR 7324, CNRS, Université de Tours, Tours (2) CITERES-LAT UMR 7324, MSH Val de Loire UAR 3501, CNRS, Université de Tours, Université d'Orléans, Tours

Que cela soit à l'échelle d'un site ou d'un territoire, les archéologues sont confrontés aux questions d'évaluation du potentiel archéologique à partir de données ponctuelles (Laurent, Fondrillon 2010 ; Laurent-Dehecq 2014 ; Rodier 1994 ; Garmy 2012). Il s'agit, d'une part d'appréhender les effets de l'anthropisation sur un espace considéré et de mesurer l'épaisseur du sous-sol produit par les sociétés, d'autre part de définir les moyens à mettre en œuvre pour les étudier lors des prescriptions d'opérations archéologiques préventives. Pour y répondre, les chercheurs ont accès à la carte archéologique nationale (Patriarche, MCC) ou, pour les villes, à des bases de données plus élaborées, sur le modèle des documents d'évaluation du potentiel archéologique en milieu urbain édités par le CNAU[1] (Cnau 1989). Dans ces outils d'aide à la décision, les données sont généralement cartographiées sous forme de points. Dans la majorité des cas, elles sont de nature hétérogène et sont réparties irrégulièrement dans l'espace. Pour modéliser le sous-sol, deux questions se posent : comment produire des surfaces à partir de ces semis de points ? Et comment harmoniser et critiquer les informations disponibles ?

Les modèles d'estimation de l'épaisseur du dépôt archéologique et du toit du substrat naturel sont les plus couramment mis en œuvre pour tenter d'apporter des réponses (Dominique *et al.* 2010 ; Laurent 2007 ; Morin *et al.* 2013 ; Fondrillon *et al.* 2013). Ces modèles sont établis à l'aide de méthodes d'interpolation spatiale, qui consistent à estimer des valeurs entre les points renseignés initialement afin de créer une information continue. Pour utiliser une méthode d'interpolation, la première question à se poser porte donc sur la nature du phénomène étudié dont la continuité dans l'espace

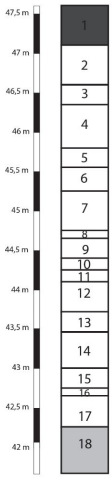
[1] Jusqu'à sa fermeture en 2016.

doit être établie comme c'est le cas pour la surface du substrat. Le dépôt archéologique n'est par définition pas continu. Toutefois, il peut être considéré comme tel, par exemple, à l'échelle d'un espace urbanisé ancien. En revanche, l'interpolation n'est pas adaptée pour analyser la répartition spatiale d'artéfacts ou de sépultures par exemple. Pour ces informations par nature discrètes, il convient de recourir, par exemple, aux méthodes d'analyse spatiale* par maillage régulier décrites précédemment (*cf.* 2.2.3).

1 Quelles sont les données utiles à la modélisation du sous-sol ?

Plusieurs sources informent les archéologues sur l'épaisseur du sous-sol et la nature des dépôts anthropiques et géologiques : les archives du sols (opérations archéologiques, des carottages géologiques et des sondages géotechniques (sondages aux pénétromètres lourds et légers de type PANDA®, sondages destructifs, tarière, pressiomètre) (Laurent 2007 : 73-78).

Log n° GEOx111_2



Opérateur : Cebtp
Appellation origine opérateur : SC4
Année de réalisation : 2005
Méthode employée : Sondage carotté
Qualité lithostratigraphique : A+

Coordonnées X,Y (Lambert Zone II) : 477514, 2264164
Altitude tête de sondage z : 47,6 m NGF
Altitude fond de sondage z : 41,6 m NGF
Qualité de l'information altimétrique : 1 (d'après levé topographique)

Numero Couche	Coordonnee Z toit Couche	Coordonnee Z mur Couche	Constituant sédimentaire 1	Constituant sédimentaire 2	Constituant sédimentaire 3	Texture Couche	Couleur Couche 1	Description Couche	Nature Couche
1	47,6	47,1						remblais	Depot anthropique
2	47,1	46,6	silt_argileux			silto_argileuse	marron	argile legerement limoneuse marron bariolee noir	Alluvions
3	46,6	46,35	sable			sableuse	marron	sable legerement graveleux (silex) marron beige	Alluvions
4	46,35	45,8	silt_argileux			silto_argileuse	marron	argile legerement limoneuse marron bariolee rouille	Alluvions
5	45,8	45,55	silt_argileux			silto_argileuse	gris	argile grise bleute	Alluvions
6	45,55	45,25	silt_argileux			silto_argileuse	marron	argile legerement limoneuse	Alluvions
7	45,25	44,7	silt_argileux			silto_argileuse	gris	argile grise bleute	Alluvions
8	44,7	44,6	silt_argileux			silto_argileuse	marron	argile limoneuse marron et quelques graves siliceuses	Alluvions
9	44,6	44,35	silt_argileux			silto_argileuse	marron	argile limoneuse marron bariolee rouille	Alluvions
10	44,35	44,2	silt_argileux			silto_argileuse	gris	argile grise bleute et quelques graves siliceuses	Alluvions
11	44,2	44,05	silt_argileux	sable		silto_argilo_sableuse	marron	limon sablo argileux marron et vegetaux	Alluvions
12	44,05	43,65	silt_argileux			silto_argileuse	marron	argile limoneuse marron bariolee rouille	Alluvions
13	43,65	43,4	silt_argileux	gravier	bloc	silto_argilo_graveleuse	gris	argile marneuse gris bleue et marron, a blocs et cailloutis	Alluvions
14	43,4	42,95	silt_argileux	sable		silto_argilo_sableuse	marron	argile marno sableuse marron bariolee gris bleu	Alluvions
15	42,95	42,7	sable	silt_argileux		sableuse_a_matrice	gris	sable marneux gris a beige	Alluvions
16	42,7	42,6	sable	silt_argileux		sableuse_a_matrice	vert	sable marneux vert jaune	Alluvions
17	42,6	42,4	sable	gravier	silt_argileux	sablo_graveleuse_a_matrice	gris	sable marneux gris a beige jaune + silex	Alluvions
18	42,4	41,6	calcaire					tuffeau tendre jaune	Substrat

Extrait de la base de données (BDS Tours) / E. Morin, X. Rodier, Base de données sédimentaires de la plaine alluviale de Tours (BDS Tours), plate-forme Rehcherche, données , information du Plan Loire Grandeur Nature 2007-2013

Fig. 1 : Exemple de log stratigraphique.

Le niveau d'information obtenu est hétérogène selon le type d'investigation et la nature des dépôts, c'est pourquoi il est nécessaire de l'harmoniser sous la forme de logs stratigraphiques[2], soit des coupes ponctuelles qui peuvent ensuite être reliées entre elles. Il convient donc de réunir les compétences disciplinaires propres à chacune des approches, mentionnées ci-dessus, pour constituer un jeu de données robuste et homogène du point de vue de l'information à traiter. Des critères de qualité de localisation,

[2]Un log stratigraphique est une représentation schématique verticale d'une série de couches stratigraphiques observées en un lieu donné.

de **complétude**[3] et d'interprétation stratigraphique des couches ou séquences stratigraphiques doivent ainsi être retenus afin de trier les informations (Laurent 2007 ; Morin *et al.* 2013). En milieu urbain, il est recommandé d'établir un à trois logs stratigraphiques complets, représentatifs pour chaque site archéologique et significatifs à l'échelle de la ville (Figure 1). Les données sont enregistrées dans un SGBD[4] relationnelle de façon à les sélectionner suivant leur niveau de qualité (défini selon la complétude et la précision*[5]) afin d'éviter les biais statistiques dans la modélisation.

2 Quelle méthode d'interpolation spatiale choisir ?

L'interpolation spatiale* permet d'estimer la valeur d'une variable* en un point à partir des valeurs de la même variable mesurée en des points voisins (Arnaud, Emery 2000 ; Zaninetti 2005 : 213). Plusieurs méthodes permettent de faire cette estimation mais elles comportent chacune des prérequis qu'il convient de connaître afin de déterminer la plus adaptée à la modélisation du sous-sol et être en mesure d'interpréter correctement les résultats. En outre, la plupart d'entre elles sont disponibles dans les logiciels de SIG et peuvent être appliquées de manière plus ou moins automatique. Une bonne compréhension de ces prérequis est donc indispensable pour paramétrer correctement l'analyse.

D'une façon générale, la valeur en un point particulier est estimée sous la forme d'une moyenne pondérée des valeurs mesurées dans le voisinage*, soit globalement soit localement (*cf.* 11). Les méthodes de calcul de la pondération sont regroupées en deux approches principales (Tableau 1) :

- une approche utilisant un cadre déterministe où la pondération est définie sur la seule base des positions des mesures déjà effectuées (méthode Spline, inverse de la distance / IDW, méthodes des plus proches-voisins, polygones de Thiessen…).

- une approche utilisant un cadre probabiliste où la pondération est dérivée à la fois des positions relatives entre les points (distance qui les sépare), mais aussi des valeurs de points de mesures déjà effectuées : ce sont les méthodes de **krigeage**[6] (simple, ordinaire…). L'objectif du krigeage est d'estimer linéairement des valeurs en des positions non observées, en minimisant les écarts à la moyenne (écart type* ou écart quadratique moyen*), soit en minimisant la covariance* des valeurs attributaires des

[3] Signifie ici qu'un log stratigraphique est au moins complet depuis la surface actuelle jusqu'au toit du substrat géologique. De façon générale, la complétude des données des données fait référence à l'exhaustivité ou à l'intégralité des données sans lacune ou information manquante.

[4] Système de Gestion de Base de données.

[5] La précision des données correspond d'une part à la précision de la localisation géographique et d'autre part à la qualité de l'information selon le type de sondage réalisé. Par exemple, les informations issues de l'observation directe du sous-sol par un archéologue ou d'un sondage carotté sont jugées de très bonne qualité tandis que celles issues de sondages géotechniques destructifs ne sont pas très fiables.

[6] Le krigeage porte le nom de son précurseur, D.G. Krige, ingénieur minier, qui a développé des méthodes statistiques empiriques afin de déterminer la distribution spatiale de minerais à partir de forages à partir des années 1950 (Krige 1951). C'est G. Matheron (Matheron 1970) qui formalise l'approche « en utilisant les corrélations entre les forages pour en estimer la répartition spatiale » (Gratton 2002 : 1). Il met au point la théorie des variables régionalisées, connue sous le nom de géostatistique. Elle est l'application des probabilités aux phénomènes naturels qui se développent dans l'espace et dans le temps. Le krigeage est un estimateur géostatistique qui devient alors une référence auprès des géographes, des sciences de la terre et des sciences humaines (Zaninetti 2005 : 233-234). Ce modèle probabiliste donne également un indicateur de la précision de l'estimation.

points (caractéristiques thématiques) et des valeurs de distance entre les points dans un espace (caractéristique de position relative).

Ainsi, le krigeage se différencie des méthodes déterministes par une analyse variographique préalable à l'interpolation, qui consiste à élaborer un modèle de covariance (variogramme*) à partir de données et d'informations disponibles (Matheron 1970 ; Arnaud, Emery 2000 : 179 ; Zaninetti 2005 : 234). L'intérêt du variogramme est de pouvoir analyser et modéliser le degré de ressemblance (autocorrélation spatiale*) entre deux mesures selon leur positionnement respectif. Le variogramme sert ainsi à identifier la distance maximale entre deux points de mesures au-delà de laquelle, il n'y a plus de corrélation*.

Dans toutes les méthodes d'interpolation spatiale, l'estimation d'une valeur est calculée suivant une fonction F où chaque valeur est pondérée. Le problème consiste à déterminer la pondération de chacun des points environnants. Par exemple, dans les approches utilisant un cadre déterministe, il peut s'agir d'une simple pondération en fonction de la distance qui sépare chaque point d'enregistrement à tous les autres ou à un nombre limité de points voisins (*cf.* 11). Dans le cas du krigeage, la pondération (ou poids) correspond au degré de ressemblance (estimateur) obtenu par l'analyse variographique (Gratton 2002 : 1). Pour mettre en œuvre le krigeage, il faut donc répondre à différentes conditions (Arnaud, Emery 2000) :

- la moyenne* et la variance* de la fonction F doivent être stationnaires[7], c'est-à-dire qu'elles ne dépendent pas de la localisation des points, mais de la distance qui les sépare ;
- les données voisines du point à estimer, appelé voisinage, doivent être en nombre suffisant pour réaliser une estimation précise et peu sensible à l'ajout ou au retrait d'une donnée.

La condition de stationnarité* est souvent difficile à remplir sur les sites ou territoires étudiés, notamment en ville, où la production de sol, par définition anthropique, n'est pas continue intra-muros et extra-muros par exemple. Toutefois, d'un point de vue exploratoire et avec toutes les précautions que cela suppose en termes d'interprétation, il intéressant d'utiliser le krigeage car il permet de s'affranchir du bruit produit par les mesures ponctuelles réalisées à l'échelle locale. On évite les effets de « halos » visibles avec les méthodes déterministes qui auront tendance à survaloriser une mesure particulière (éloignée de la moyenne des valeurs voisines). À contrario, les méthodes du cadre déterministe pourront mettre en évidence de manière plus ou moins forte ces valeurs de mesure particulières par rapport au voisinage, y compris les grosses erreurs de mesure.

[7] On parle de **stationnarité** quand il y a invariance du semis de point par translation, c'est-à-dire si les propriétés du semis de points ne varient pas d'une position à l'autre de l'espace. Ainsi, leur moyenne et leur variance ne varient pas dans l'espace.

TAB. 1 : Description de méthodes d'interpolation spatiale.

		Méthode déterministe			Méthode probabiliste				
					Géostatistiques				
Méthodes locales (voisinage)	IDW : inverse de la distance pondérée	Pondération par la distance inverse élevée à une puissance (calcul d'une moyenne mobile spatiale pondérée) -> grille avec cellule associés à un poids Wavec n éléments + 2 paramètres frottement de la distance + rayon de recherche à définir (ajustement linéaire)	Plus les points sont proches, plus la pondération affectée est forte (“effet oeil de bœuf”) Pb pour déterminer les 2 paramètres. Estime que les points plus proches de l'emplacement à calculer auront plus d'influence.	Krigeage simple	Variable stationnaire de moyenne connue	Méthode globale qui ne prend pas en compte les variations locales ce la tendance		
	NN : Voisin Naturel	Cherche le sous-ensemble d'échantillons le plus proche à un point et applique une pondération en fonction de la zone où ils se trouvent.	Ne produit pas de pics, de fosses, de crêtes ou vallées qui ne sont pas déjà présents dans les échantillons en entrée et s'adapte localement à la structure des données en entrée. Ne nécessite pas de paramétrage par l'utilisateur et fonctionne aussi bien pour des données réparties régulièrement ainsi qu'irrégulièrement.	Krigeage ordinaire	Variable stationnaire avec moyenne inconnue	Prend en compte les variations locales ce la tendance.		
	Splines : base radiales	Utilise une fonction mathématique afin de minimiser la courbure de la surface et produit une surface lisse qui correspond exactement aux points d'entrée (ajustement de polynômes)	Adapte une surface courbe minimale en passant par les points d'entrée. Elle préserve les tendances dans les données de l'échantillon et s'adapte aux changements rapides de gradient ou de pente.	Krigeage universel ou krigeage avec dérive externe	Variable non-stationnaire avec une moyenne inconnue	La tendance n'est pas constante et est modélisée par régression. On découpe alors une dérive et des résidus considérés stationnaires que l'on analysera (variogramme)	3 Etapes : 1-Analyse structuration spatiale (variogramme). dépendra aussi à partir d'une distance de similarité des points, dépend de la distance entre les points et non pas leur position 2-Interpolation (krigeage) 3- Validation croisée calcul d'écart-type entre valeurs mesurées et valeurs simulées	
	ANUDEM (topo vers Raster ArcGIS 3D analyst)	Utilise une technique d'interpolation spécialement conçue pour créer une surface qui représente au mieux une surface de drainage naturel et préserve à la fois les lignes de crête ainsi que les réseaux de cours d'eau.	Bonnes performances quand il s'agit du calcul de crêtes et de zones de flux d'écoulement	Krigeage sous contrainte d'inégalité	Variable stationnaire	Associe mesure “vraie” et mesure “simulée” à partir de bornes définies, utile pour des données incomplètes		
Méthode globales (tous les points)	Fonctions polynomiales ex : Surface de tendance (approximation des moindres carrés)	3 ETAPES PREALABLES à l'interpolation *Statistique descriptive générale *Statistique inductive (test stationnarité, indice de dispersion, méthode du voisin le plus proche, méthode des quadrats) *Etude des agrégats	Méthode approximative, algorithmes lourds	Krigeage d'indicatrices, Méthodes gaussiennes, Krigeage disjonctif	méthode non-linéaire	Détermine des indicateurs comme une probabilité d'excéder un seuil		

D'après Arnaud Emery 2006 / Zaninetti 2005 / François 2016

3 L'exemple d'une méthode utilisant un cadre déterministe : les TIN (Triangulated irregular network)

La méthode des réseaux de triangles irréguliers est souvent utilisée pour créer des MNT (Modèle Numérique de Terrain) car sa mise en œuvre est simple et rapide et parce qu'elle peut s'avérer être efficace quand le jeu de données initial est assez dense. Elle consiste à relier les points sous forme de segments et à générer des facettes triangulaires qui portent une valeur interpolée de la variable étudiée (altimétrie, épaisseur, *cf.* Figure 2). En pratique, les SIG proposent le plus souvent des outils reposant sur la méthode de triangulation de Delaunay[8][9] qui s'assure qu'aucun sommet ne se trouve à l'intérieur des cercles dans lesquels les triangles du réseau sont circonscrits[10]. Cette méthode évite la construction de facettes allongées du fait qu'elle maximise le minimum des angles des triangles.

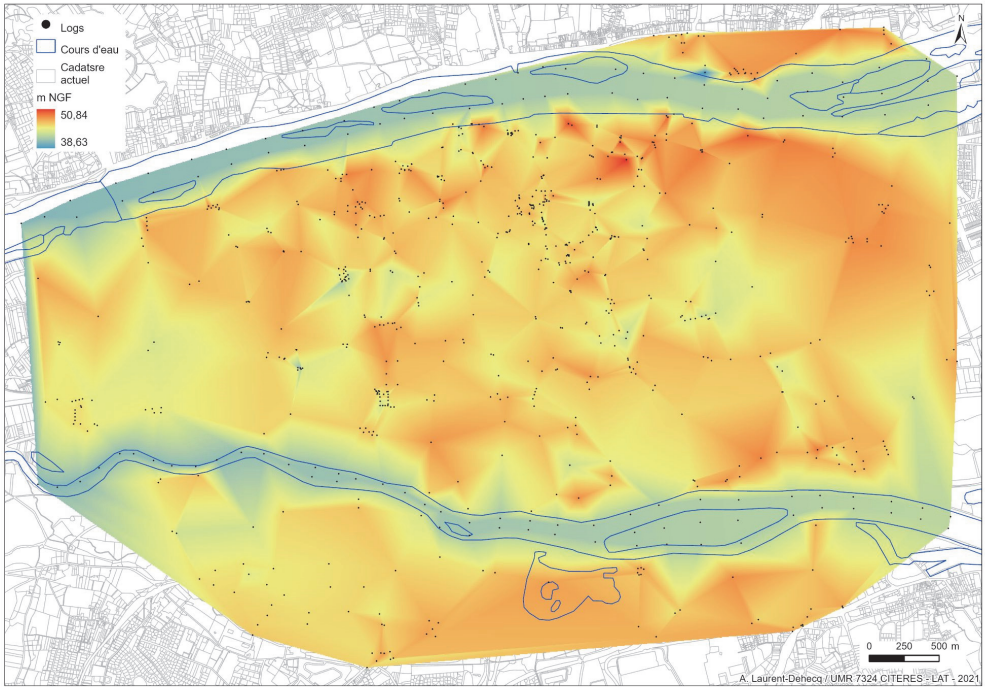


Fig. 2 : TIN du toit des alluvions de la plaine alluviale de Tours.

[8] La méthode de Delaunay et les calculs de l'interpolation sont présentés en détail dans la bibliographie (Arnaud, Emery 2000 : 61-67).

[9] Pour créer un TIN, il est conseillé de travailler dans un système de coordonnées projetées (Ex. Lambert 93). L'option de la triangulation de Delaunay sous contraintes consiste à ne pas densifier les segments de triangles le long des lignes de fractures.

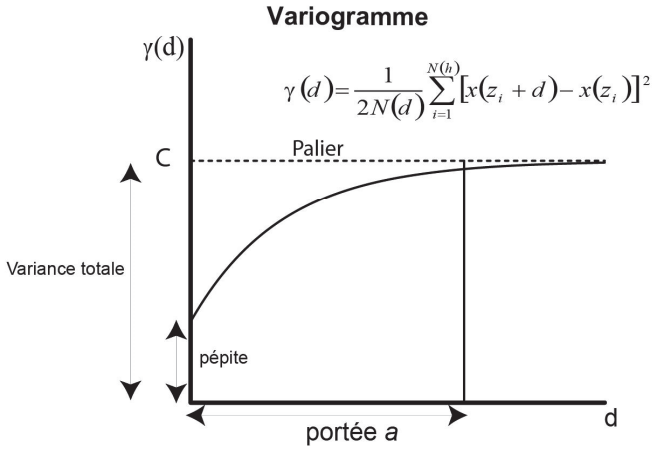
[10] La triangulation de Delaunay est complémentaire du diagramme de Voronoï dont les sommets des triangles sont au centre des cercles : la triangulation de Delaunay est le graphe dual du diagramme de Voronoï.

4 L'exemple d'une méthode de géostatistique : le krigeage ordinaire

L'interpolation par krigeage nécessite de suivre trois étapes (Laurent 2011).

4.1 Le variogramme

Dans une phase exploratoire, il s'agit d'abord d'analyser la variabilité spatiale de la valeur à estimer à l'aide du **variogramme**, soit le graphe qui représente l'écart quadratique moyen de tous les couples de points en fonction de la distance qui les sépare (Figure 3).



z_i et $z_i + d$ sont les positions des mesures et x(zi) est la valeur étudiée correspondant au point considéré. La variable *d* correspond à la distance entre deux mesures. La sommation est étendue à tous les couples de points (z_i, $z_i + d$) distants de *d* et N(d) représentant l'effectif des couples de points. La fonction $\gamma(d)$ quantifie la perte progressive de corrélation spatiale entre les données. Elle représente la perte de corrélation moyenne entre deux points pris au hasard quand la distance *d* entre ces points augmente. Le variogramme est souvent une fonction croissante de *d* et peut se stabiliser autour d'une valeur finie pour les grandes valeurs de *d*. Elle correspond alors à la variance du signal ou « palier » *C. A priori*, la variance des données croît avec la distance entre les points jusqu'à une distance *a* appelée « portée ». Au-delà de cette valeur, les valeurs sont complètement décorrélées. Souvent, la valeur que prend le variogramme quand *d* tend vers zéro n'est pas nulle du fait en particulier des bruits de mesure (variance de mesure) et est nommée « effet de pépite ».

Fig. 3 : Variogramme théorique

Le but est de déterminer la distance de corrélation spatiale entre les points d'observation, soit la portée *a* du variogramme. Pour réaliser le variogramme expérimental, il faut définir plusieurs paramètres (Arnaud, Emery 2000 : 128-132 ; Zaninetti 2005 : 237) :

- la portée limite qui définit la limite de la recherche (en général, la moitié de la fenêtre d'étude) ;
- le nombre de pas de recherche. Pour chaque pas de recherche, on obtient le nombre de paires de points qui ont contribué à calculer la variance correspondante. Il est conseillé d'avoir au minimum 20 paires de points de mesure ; à partir de 50 paires, les résultats sont considérés comme robustes.

- la distance de voisinage. Elle s'exprime dans l'unité de la mesure des coordonnées cartésiennes du site. Elle correspond à la division de la portée limite par le nombre de pas.

- la tolérance. Elle permet de prendre en compte les voisins qui ne se trouvent pas exactement à la distance de voisinage définie. Elle doit être inférieure à la distance de voisinage. Elle peut être estimée à partir de la densité de points théorique[11] ;

- la direction du variogramme, omnidirectionnel (isotrope) ou directionnel (anisotrope). Une carte variographique représente la répartition spatiale de la variance par pas de recherche. La variance est calculée en tout point de l'espace étudié selon la résolution choisie (distance de voisinage et avec un nombre de paires de points minimum pour le calcul de la variance, appelé aussi le seuil). Si la variance se concentre suivant un axe directionnel, il faut considérer que le phénomène est anisotrope (Figure 4) ;

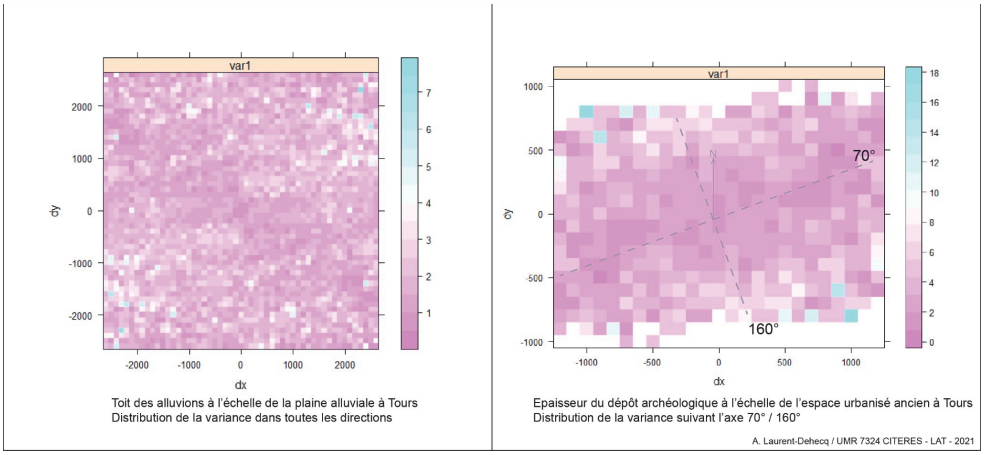


FIG. 4 : Cartes variographiques (variogramme omnidirectionnel à gauche, directionnel à droite)

- l'orientation. Elle définit l'orientation de l'axe de recherche en degré par rapport au nord (0°) suivant le sens des aiguilles d'une montre. L'angle doit être compris entre 0 ° et 180 ° .

- le champ. Il définit l'étendue du cône de recherche à partir de l'orientation dans le sens des aiguilles d'une montre (entre 0° et 90°).

[11] La densité de point théorique correspond à l'estimation de la distance théorique entre les points. Par exemple, à l'échelle de l'espace urbanisé ancien (Laurent 2007), on dénombre 143 points sur 235,27 ha, soit 1 point sur 16433,5 m² donc 1 point tous les 128 m. Ici, la tolérance retenue pour le modèle est de 100 m.

Il faut ensuite ajuster un variogramme théorique (modèle)[12] sur le variogramme expérimental (Figure 5).

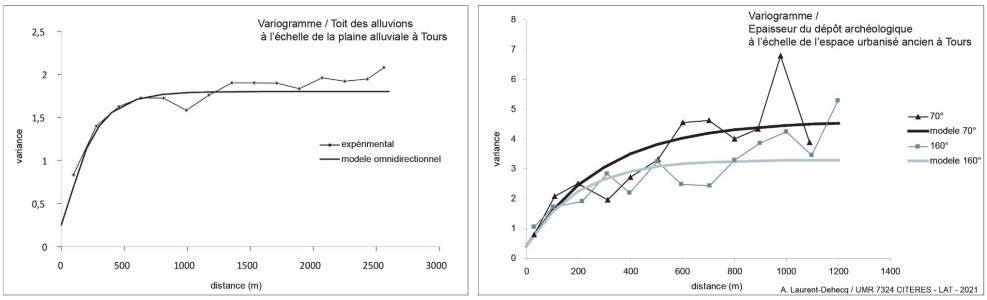


Fig. 5 : Ajustement du variogramme théorique sur le variogramme expérimental (variogramme omnidirectionnel à gauche, directionnel à droite)

L'opération d'ajustement est délicate puisqu'elle dépend de l'objectivité de l'utilisateur[13]. Lorsque x = 0, la valeur de la variance correspond à l'effet de pépite[14] (Figure 3) qui est interprété comme une micro-structuration des données que le variogramme ne peut mesurer. Il est important de réduire cet effet pépite au maximum par rapport à la variance totale afin que les micro-variations locales aient peu d'effet sur le modèle. Dans l'exemple sur la plaine alluviale de Tours (Figure 5), l'effet de pépite est de 0,25 alors que la variance totale est d'1,75. Il représente donc 14 % de la variance totale, ce qui est empiriquement considéré comme acceptable en contexte archéologique. La portée du variogramme, soit la distance de corrélation entre les points de mesure, est déterminée également par l'utilisateur lors de l'ajustement. La portée prise en compte dans l'exemple est de 200 m : au-delà de cette distance, on considère qu'il n'y a pas de corrélation spatiale entre les données. En d'autres termes, les points distants de plus de 200 m au maximum ne sont pas comparables entre eux[15]. La connaissance de cette distance limite est intéressante pour mieux comprendre le phénomène étudié. Il permet en outre d'orienter des investigations futures du sous-sol par la mise en place d'un maillage représentatif de l'espace étudié[16].

4.2 L'interpolation des données par krigeage ordinaire

Cette deuxième étape est le résultat cartographique de l'interpolation par la méthode du krigeage ordinaire (Figure 6) qui utilise le modèle ajusté à l'aide du variogramme

[12] Les modèles théoriques (sphérique, exponentiel, gaussien…) sont présentés dans la bibliographie (Arnaud, Emery 2000 : 133-138 ; Zaninetti 2005 : 241-242).

[13] Cette opération est présentée en détail dans la bibliographie (Arnaud, Emery 2000 : 132-163 ; Zaninetti 2005 : 239-240).

[14] Représente la variation entre deux mesures effectuées à des emplacements infiniment proches, et peut donc provenir de deux effets. Le premier représente une variabilité de l'instrument de mesure : la pépite mesure donc en partie l'erreur statistique de l'instrument de mesure. Le deuxième est un réel effet pépite qui correspond à une variation brutale du paramètre mesuré ; le cas historique est le passage sans transition d'une pépite d'or à un sol ne contenant quasiment pas d'or (Matheron 1965).

[15] La portée est définie au moment du calcul du variogramme. Le logiciel R propose un ajustement automatique dans un premier temps utile à la réflexion, puis dans un second temps l'ajustement manuel des valeurs de la portée et de l'effet de pépite permet d'obtenir le modèle présentant le meilleur R^2.

[16] En théorie, dans le cas de la plaine alluviale de Tours, un maillage de points distants de 200 m serait suffisant pour restituer l'altimétrie du toit des alluvions.

(Figures 4 et 5). Pour réaliser l'interpolation, il faut déterminer un voisinage : il peut être unique si on veut prendre en compte tous les points de l'espace étudié dans le calcul de l'estimation, ou bien, glissant si on choisit une distance limite de recherche autour de chaque mesure initiale.

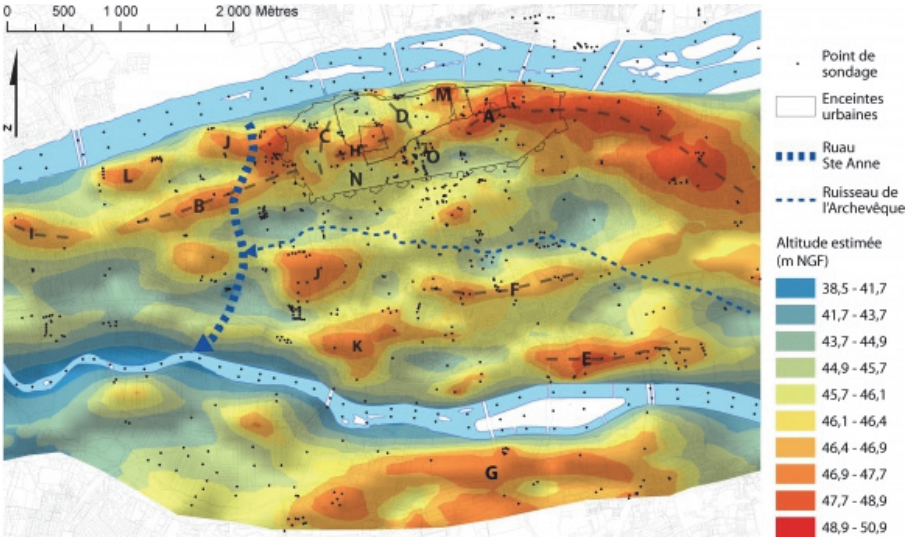


Fig. 6 : Carte de l'estimation du toit des alluvions dans la plaine alluviale de Tours (Morin *et al.* 2013)

Dans le cas de Tours, le modèle a été ajusté, dans un premier temps, sur des données mesurées (722 points) et a été appliqué lors de l'interpolation, dans un second temps, sur le jeu de données complétés de données extrapolées (depuis des levées bathymétriques des cours d'eau (la Loire et le Cher), soit 838 points au total localisés avec des coordonnées géographiques x et y, et une valeur attributaire correspondant à l'altitude z du toit des alluvions. L'interpolation par krigeage ordinaire a été utilisée pour estimer la valeur z des points avec les valeurs du variogramme décrites supra. Cette interpolation permet de passer d'un semis de points d'altitudes à un modèle de surface altimétrique[17] mettant en évidence des variations modérées de la surface des alluvions et délimitant des secteurs plus ou moins submersibles lors des inondations (présence de montilles et de levées naturelles : formations A à M) (Morin *et al.* 2013).

4.3 La critique des modèles par la validation croisée et la carte des erreurs standards

La phase de validation croisée permet d'évaluer la fiabilité de l'interpolation spatiale en analysant la relation entre les valeurs mesurées (initiales) et les valeurs estimées de chaque point du jeu de données à partir du modèle. La relation entre ces deux valeurs (mesurée et estimée) en un même point est mesurée par un coefficient de détermination* (R^2) (*cf.* 5 et Figure 7). La moyenne des écarts entre les valeurs

[17] Parmi les modèles de surface altimétrique, on citera par exemple le MNT (modèle numérique de terrain)

mesurées et estimées est également calculée. Enfin une carte des erreurs standards est produite (Figure 8).

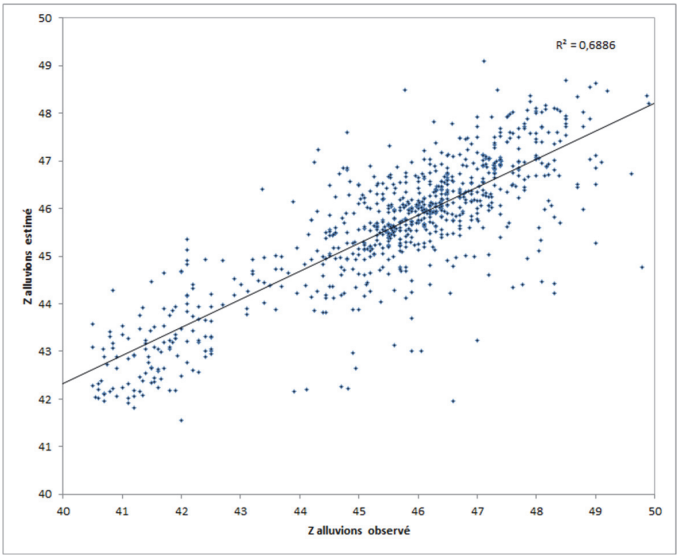


Fig. 7 : Validation croisée des valeurs estimées et mesurées du toit des alluvions dans la plaine alluviale de Tours ($R^2 = 0{,}68$).

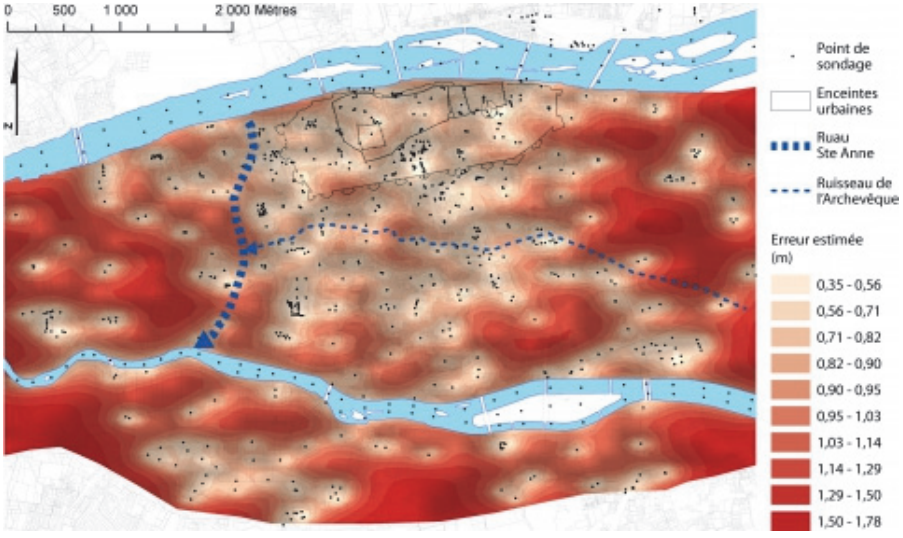


Fig. 8 : Carte des erreurs standards du toit des alluvions dans la plaine alluviale de Tours (Morin *et al.* 2013).

5 Lecture et interprétation des modèles du sous-sol

Les modèles du sous-sol sont généralement de 2 types, des modèles de restitution de surface (toit du substrat, toit de la surface actuelle) ou d'épaisseur sédimentaire (formation géologique, dépôt anthropique). Ils se présentent sous la forme d'un raster*. Les secteurs dont les estimations sont moins fiables concernent ceux qui sont les moins documentés, généralement dans les extrémités de l'emprise étudiée. Dans le cas où le krigeage est utilisé, l'interprétation archéologique des modèles repose sur une lecture conjointe des cartes d'interpolation et des cartes des erreurs standard.

Pour modéliser les épaisseurs sédimentaires, il est nécessaire d'étudier la structuration spatiale des mesures dédiées. Ainsi, la soustraction des rasters des estimations du toit et du fond de la variable analysée n'est pas appropriée. Par exemple, la modélisation de l'épaisseur du dépôt anthropique se fait directement par l'analyse variographique de cette variable et ne peut être déduite de la soustraction de deux rasters issus de la modélisation de l'altimétrie en surface et l'altimétrie de l'apparition du substrat.

L'exemple de la modélisation de l'apparition des alluvions dans la plaine alluviale de Tours par quatre méthodes met en évidence des différences (Figure 9) :

- Les méthodes sont basées sur différents modèles mathématiques dans le calcul de l'interpolation et expliquent les différences entre les bornes inférieures et supérieures des valeurs interpolées[18]. Les valeurs du jeu de données au départ sont comprises entre 38,5 et 50,8 m NGF. Les méthodes TIN et de l'inverse à la distance (IDW) estiment des fourchettes similaires. La méthode Spline interpole des valeurs au-delà de la fourchette des données mesurées tandis que le krigeage réduit les valeurs extrêmes. Ainsi, pour chaque pixel, les valeurs estimées pourront varier selon la méthode employée.

- La méthode TIN fait bien apparaître les valeurs interpolées suivant des facettes triangulaires alors que la méthode Spline a tendance à créer des halos autour des points. Les méthodes IDW et du krigeage génèrent des résultats assez proches au premier abord et, à petite échelle, permettent de mettre en évidence des ensembles sédimentaires similaires comme la présence des hauts fonds interprétée comme des montilles. À plus grande échelle, on observe notamment des différences dans les secteurs où il y a peu de points. Sans compter que les fourchettes de valeurs sont différentes (*cf.* point précédent), il faut garder en tête que ces méthodes ont des modes de calculs très différents (*cf.* Tableau 1). Le krigeage a l'avantage de générer des indicateurs pour quantifier la robustesse du modèle (validation-croisée des données, coefficient de détermination).

[18] La classification des données au sein d'un SIG est primordiale pour représenter les données. Pour cet exemple, dans un premier temps, des rasters ont été générés à l'issue de chaque modélisation (avec ArcgisPro ou R). Pour assurer la comparaison, ces rasters ont été exportés suivant un maillage de 5 m (taille de la cellule à régler). Dans un second temps, la représentation graphique a été réglée avec un affichage des données compris entre le « minimum et le maximum des valeurs ». Il faut être attentif à la méthode proposée par le logiciel utilisé qui peut rééchantillonner les données suivant un « écart-type » par exemple (réglage de l'étirement des valeurs).

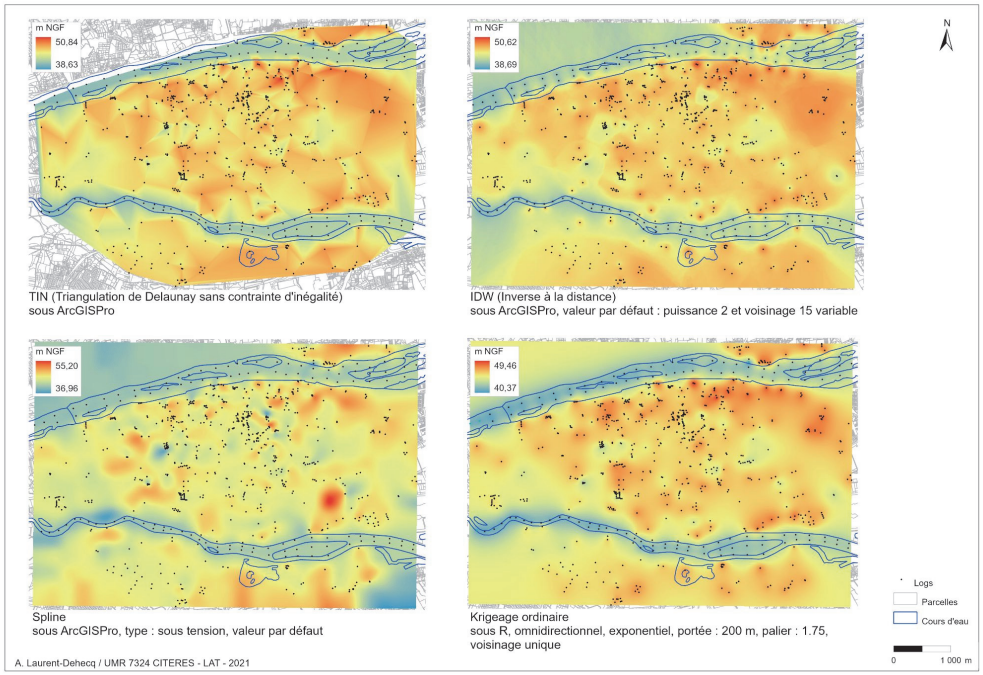


FIG. 9 : Comparaison de cartes de l'estimation du toit des alluvions dans la plaine alluviale de Tours suivant différents modèles.

En résumé :

- Modéliser le sous-sol à partir de données ponctuelles est possible à condition de suivre des étapes préalables à la cartographie du phénomène étudié : c'est-à-dire produire un jeu de données harmonisé et robuste et utiliser une méthode d'interpolation adaptée. L'interpolation est mise en œuvre en partant du postulat que le phénomène est continu dans l'espace.
- Les outils d'interpolation spatiale sont présents dans les logiciels de SIG et sur R. Il faut néanmoins veiller à bien régler les paramètres (*cf.* 11).
- Dans le cas du choix d'une méthode probabiliste tel que le krigeage ordinaire, il est nécessaire d'étudier, en amont, la structuration spatiale des données et, en aval, la validation croisée des mesures, étapes indispensables à la critique des données. La mise en œuvre du krigeage est plus longue que celle des méthodes déterministes mais elle a l'avantage de produire un modèle qui n'est pas dépendant de la position des mesures et d'éva-

luer la représentativité statistique de ce modèle. Si certaines données sont incomplètes, le krigeage sous contrainte d'inégalité sera adapté. Si les mesures sont contraintes par une autre variable continue bien échantillonnée (exemple : relief), le cokrigeage pourra alors être mis en œuvre.

Outils :

- Dans Arcgis, privilégier le module Geostatistical wizard afin de rentrer les paramètres nécessaires.
- Dans R, les packages…pour le krigeage. Des scripts sont disponibles ici : https://github.com/archeomatic/AA2020.1/