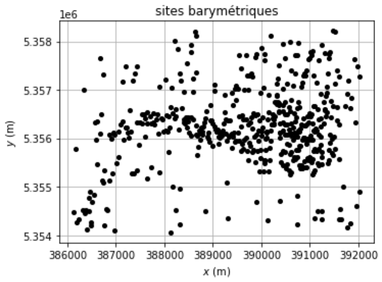
Léonie LEROUX et Laure GREGOIRE

Rapport tp de Géostatistiques

Validation d’un levé bathymétrique

Points d’observations:

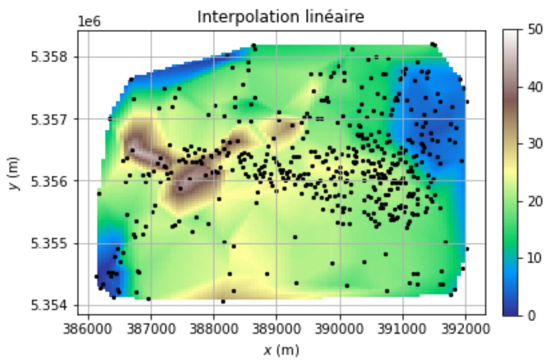


On peut déjà souligner que la répartition spatiale des sites du levé est, à première vue, loin d’être homogène, ce qui nous donne une première idée des méthodes d’interpolation que nous pouvons utiliser. Par exemple, la méthode linéaire pourrait exclure un certain nombre de points sur les bords du graphique qui ne possèdent pas beaucoup de données.

Cartes issues des différentes méthodes d’interpolation:

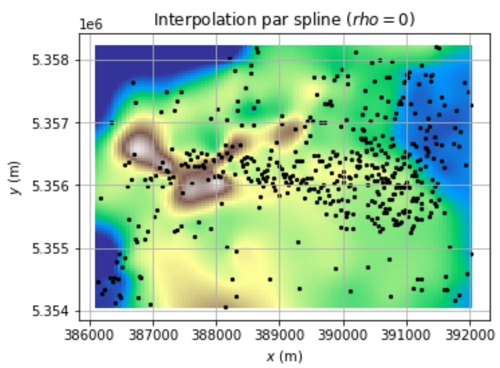
Interpolation linéaire:

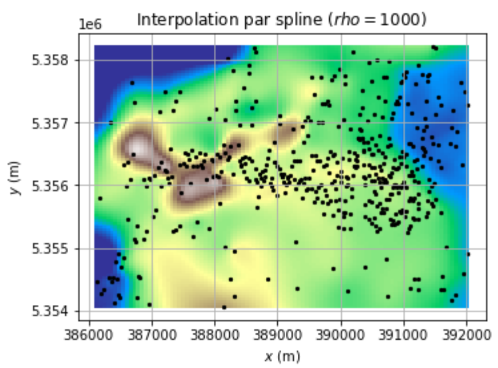
Comme on peut le remarquer sur l’image ci-contre, il y a des zones sur les bords dont on n’a pas réussi à interpoler la profondeur, notamment sur les points les plus excentrés (points en dehors de l’enveloppe convexe). Ces premières observations rejoignent notre hypothèse de départ: la méthode d’interpolation linéaire convient pour des levés ayant des observations à peu près régulières et homogènes (surtout sur les bords), ce qui n’est pas notre cas. Ainsi, les points les plus éloignés sont exclus du calcul.



Interpolation par splines:

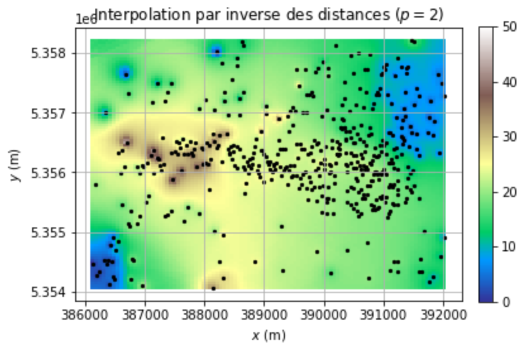
L’interpolation par spline permet d’avoir un résultat qui suit le modèle observé, les profondeurs calculées sont donc homogènes, la structure globale est cohérente. Mais, la carte obtenue semble plus lisse que celle des autres méthodes ce qui laisse à penser que certains résultats notamment dans des zones à fortes variations ont été adoucis.

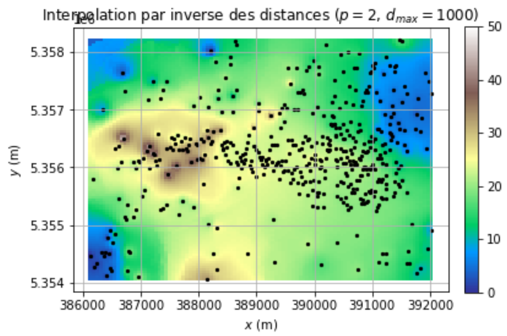




Le changement entre spline de lissage et spline d'interpolation n'est pas flagrant mais on peut tout de même souligner qu'en passant le paramètre rho à 1000 cela permet d'atténuer les valeurs extrêmes et de réguler nos résultats.

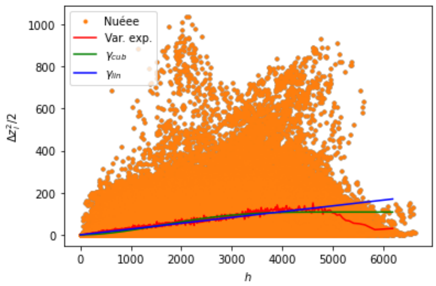
Interpolation par inverse des distances





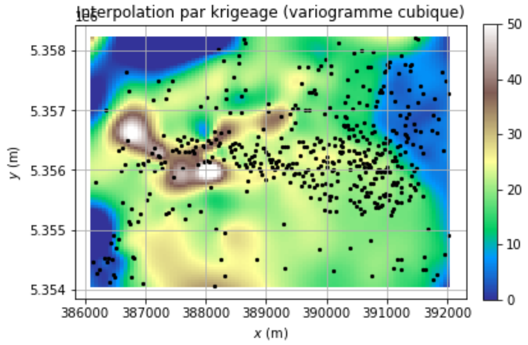
Cette méthode prend en compte la distance du point à interpoler avec les observations ce qui nous permet ici d'obtenir des résultats plus précis dans les zones avec des variations locales importantes. Mais pour les zones où il y a beaucoup d'observations avec des valeurs similaires, elles peuvent prendre un poids trop important dans le calcul. Ainsi, après différentes valeurs tests nous avons choisi de mettre en paramètre une distance maximale dans le calcul de 1000 mètres. Cela nous a permis de conserver une carte homogène, tout en ayant des valeurs plus précises selon l'emplacement des points. En effet, on peut voir que les zones excentrées sont ajustées lorsque l'on met une distance maximale.

Interpolation par krigeage



La nuée variographique met en avant cette disparité de la répartition des observations, avec notamment deux pics principaux. Nous avons donc choisi de nous tourner vers un variogramme cubique puisqu'il permet de prendre en compte la tendance globale de la zone étudiée en prenant en compte la répartition non linéaire de nos données.

Contrairement aux résultats précédents on peut noter que les zones à grandes profondeurs ont tendance à atteindre des valeurs plus extrêmes. De la même manière, on peut observer en haut à gauche de la carte que les profondeurs sont plus importantes que dans les autres interpolations. Donc, le krigeage permet d'avoir des données précises et homogènes mais il peut laisser place à des valeurs aberrantes.



Comparaison par validation croisée

Pour la validation croisée, nous avons choisi d'avoir un nombre d'observations plus important pour notre entraînement que pour le jeu test, afin que le modèle soit le plus précis possible. Ainsi, sur un ensemble de 493 points nous avons utilisé 300 observations pour l'entraînement (tirées de manière aléatoire), et 193 pour les tests. Cette répartition nous permet de garantir l'obtention d'un modèle suffisamment entraîné tout en conservant un ensemble de tests significatifs pour évaluer nos modèles. Pour le modèle du krigeage, nous avons décidé de conserver les regroupements de 500 points bien que nous ayons plus de valeurs.

La matrice de comparaison obtenue est la suivante:

	moyenne	écart type	moy quadratique	Nash-Sutcliffe
linéaire	-0.12438167	3.55603769	3.55821231	0.78720848
inverse distance	0.0434176	4.14097683	4.14120444	0.72179902
inverse d = 1000	-0.13228234	3.70728119	3.70964047	0.77676157
spline rho = 0	0.05059055	2.96966712	2.97009802	0.85689761
spline rho = 1000	0.04266186	2.97024483	2.97055119	0.85685394
krigeage	0.11089324	3.0192444	3.0212802	0.8519231

L'interpolation par spline avec un rho de 1000 est celle qui a les meilleurs écarts. Mais, globalement l'ensemble de nos résultats sont très rapprochés: nous avons donc fait une validation croisée pour différentes tailles de jeu modèle (200, 300 et 400) afin de déterminer quel modèle d'interpolation est le meilleur et le plus constant, d'autant plus que nos jeux de données modèles changent à chaque fois puisque qu'ils sont tirés de manière aléatoire.

Modèle de taille 200:

	moyenne	écart type	moy quadratique	Nash-Sutcliffe
linéaire	0.05035962	2.18021174	2.18079328	0.90446687
inverse distance	0.43577938	3.65024146	3.67616191	0.74672352
inverse d = 1000	0.23609224	3.36112682	3.36940841	0.78465984
spline rho = 0	-0.26521671	3.17666625	3.18771836	0.80955679
spline rho = 1000	-0.2588263	3.15005181	3.16066725	0.8127753
krigeage	-0.81017075	5.56962935	5.62824552	0.40632102

Modèle de taille 400:

	moyenne	écart type	moy quadratique	Nash-Sutcliffe
linéaire	-0.47941766	3.06940072	3.10661585	0.8471467
inverse distance	0.01172089	3.89764794	3.89766556	0.76734249
inverse d = 1000	-0.23270581	3.35406359	3.3621265	0.82688447
spline rho = 0	-0.43000106	2.83916033	2.87153831	0.87371925
spline rho = 1000	-0.42780537	2.81328123	2.84562273	0.87598832
krigeage	-0.40566382	2.95537295	2.98308438	0.86371784

On peut remarquer que la méthode d'interpolation linéaire paraît être plutôt bien adaptée à notre étude mais il faut souligner que les statistiques de cette méthode ne prennent pas en compte les points externes à l'enveloppe convexe (qui n'ont donc pas de valeurs interpolées). Bien que ces points ne soient pas considérés comme une erreur ils représentent une grande perte d'informations, c'est pour cela que nous avons décidé d'écarter cette méthode.

Le calcul par l'inverse des distances avec une distance maximale imposée semblait être une méthode envisageable. Mais on observe que l'écart-type et la moyenne quadratique dépassent les trois mètres pour l'ensemble de nos comparaisons. C'est pour cela que nous n'avons pas conservé cette méthode. De la même manière, les méthodes par l'inverse des distances et du krigeage ont des résultats inférieurs à ceux attendus. On peut noter que le krigeage varie de façon conséquente entre les différents jeux de données modèles.

Ainsi, la méthode du spline d'interpolation (rho = 1000) nous semble être celle qui s'accorde le mieux à notre levé bathymétrique. Les écarts sont constants et satisfaisants, notamment le Nash-Sutcliffe proche de 1 indique une bonne correspondance entre nos profondeurs calculées et celles observées.

Evaluation du levé avec les données de contrôle

évaluation	moyenne	écart type	moy quadratique
spline rho = 1000	-0.17381684	2.20893709	2.21576518

Les écarts obtenus avec les données de contrôle paraissent satisfaisants étant donné qu'ils sont inférieurs à ceux de la validation croisée. Le fait que l'écart moyen est relativement faible (environ 0,17 m) va dans l'idée que notre choix de méthode d'interpolation concorde avec les données étudiées. De plus, si l'écart-type ainsi que la moyenne quadratique ont des valeurs nettement supérieures cela peut s'expliquer par la disparité importante de profondeur dans des zones réduites.

Ainsi, une erreur sur un site peut prendre rapidement du poids lors de l'évaluation. Par ailleurs, il faut noter que les profondeurs les plus grandes sont d'environ 30 mètres, donc un écart-type de 2 mètres ne paraît pas aberrant. Selon nous, notre interpolation est donc valide.