Rapport

TP 2
GÉOSTATISTIQUE

Réalisé par:

Abdelaziz Babioui

Encadré par:

Mr Ahmed DOUAIK

المدرسة الحسنية للأشغال العمومية
ECOLE HASSANIA DES TRAVAUX PUBLICS

I-Introduction :

Interpolation Spatiale: :

L'interpolation spatiale est une technique utilisée pour estimer des valeurs entre des points d'observation ou des échantillons dispersés dans l'espace. Cela permet de créer des surfaces continues représentant des phénomènes ou des variables qui varient dans l'espace, tels que la température, l'humidité, la pollution, l'altitude, etc.

Parmi les méthodes couramment utilisées pour l'interpolation spatiale, on trouve le krigeage (ou kriging), le TIN (Triangulated Irregular Network) et l'IDW (Inverse Distance Weighting).

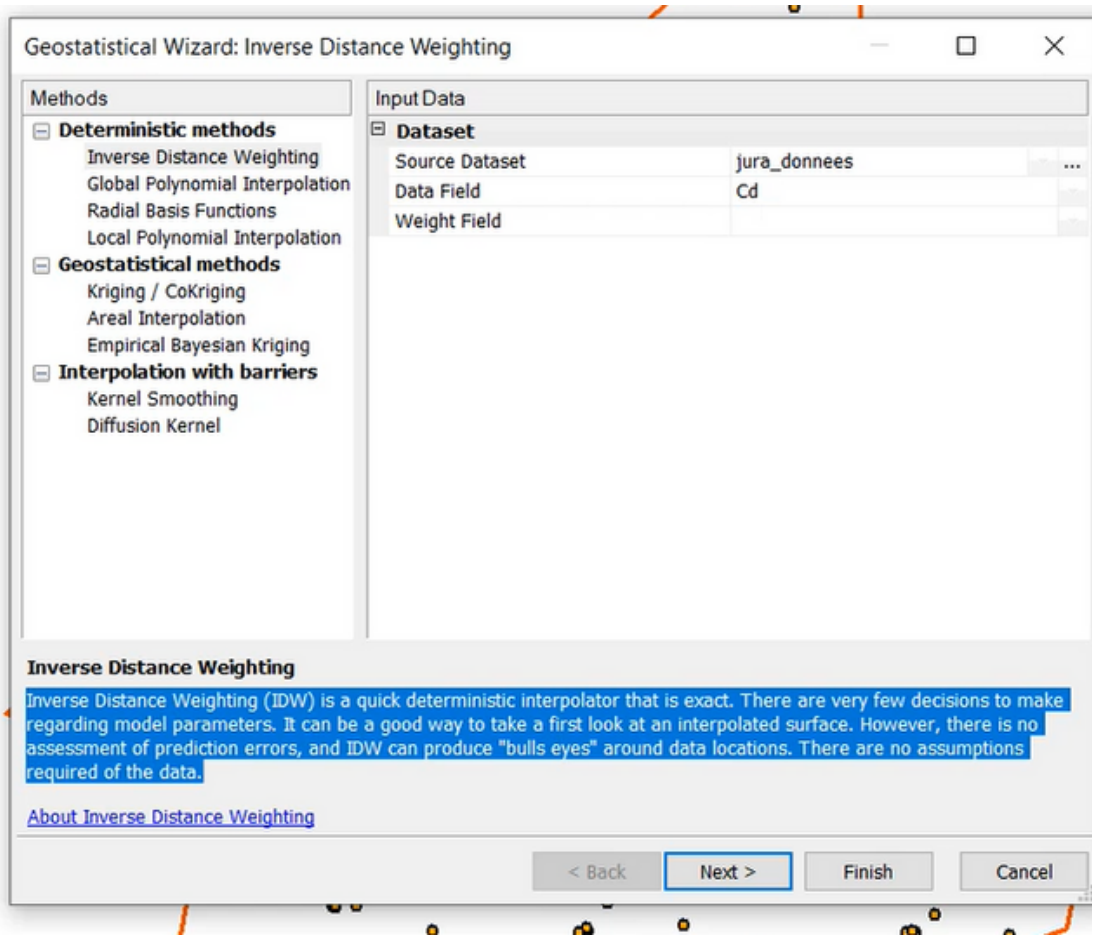
1. **Le krigeage : est une méthode d'interpolation basée sur la théorie des processus gaussiens. Elle repose sur l'hypothèse que les valeurs mesurées à proximité les unes des autres sont corrélées spatialement. Le krigeage estime les valeurs inconnues en utilisant les valeurs observées pondérées en fonction de la distance et de la corrélation spatiale. Il fournit également une estimation de l'erreur d'estimation associée à chaque point interpolé.**
2. **Le TIN :est une méthode d'interpolation qui divise l'espace en une série de triangles connectés. Chaque triangle est défini par trois points d'observation voisins. Les valeurs inconnues sont ensuite estimées en interpolant linéairement les valeurs des points d'observation à l'intérieur du triangle correspondant. Cette méthode est particulièrement adaptée aux jeux de données irréguliers ou avec des données manquantes, car elle ne nécessite pas de grille régulière.**
3. **L'IDW : est une méthode d'interpolation qui repose sur le principe que les valeurs inconnues sont influencées de manière inversement proportionnelle à leur distance par rapport aux points d'observation. Autrement dit, les valeurs les plus proches ont plus de poids dans l'interpolation que les valeurs plus éloignées. L'IDW attribue des poids aux valeurs observées en fonction de la distance et calcule la moyenne pondérée pour estimer les valeurs inconnues.**

Ces méthodes d'interpolation spatiale sont largement utilisées dans de nombreux domaines, tels que la cartographie, l'environnement, l'agriculture, la géologie, etc.

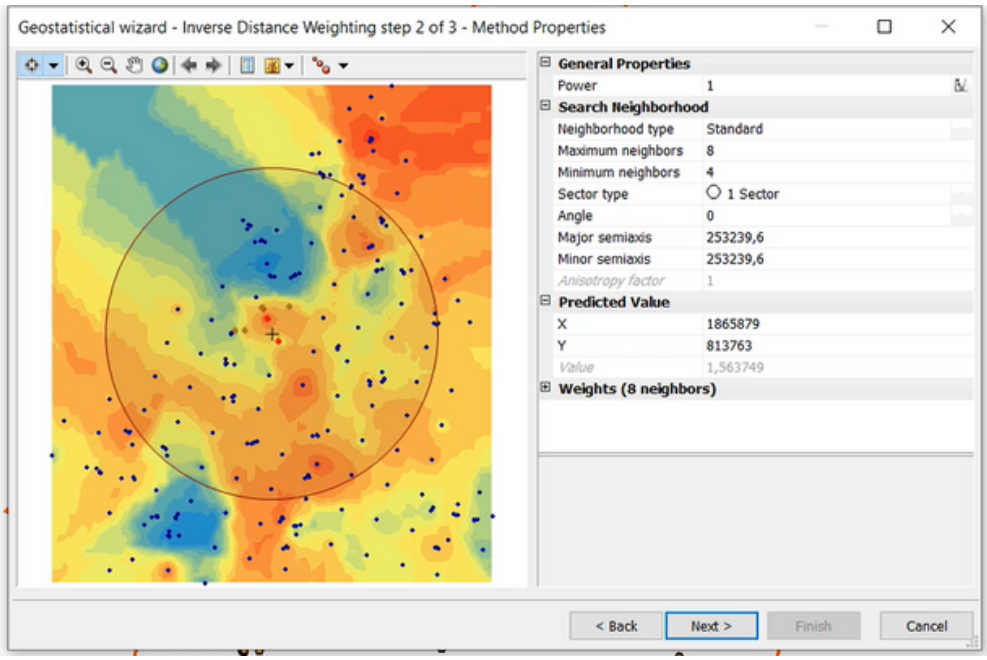
II-Réalisation du TP sur Arcgis:

en utilisant l'outil "Geostatistical wizard" sur ArcGis , on peux exploiter les différentes méthode disponibles à savoir les méthodes détérministiques et les méthodes géostatistiques

1-En un premier lieu , on va opter pour IDW (distance inverse pondérée avec power =1)



2ème etape :



on change les propriétés de la méthode , en changeant l'exposant p en 1 , ainsi que le minimum et le maximum voisins (4 et 8)

en changeant le point , le poids des voisins change également

Weights (8 neighbors)

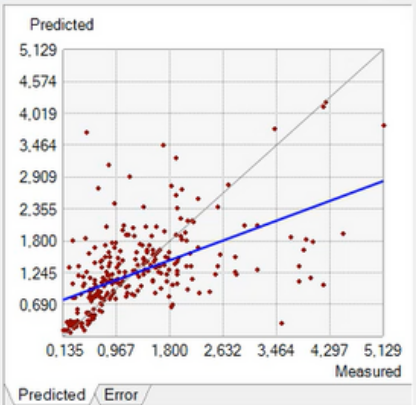
OID	Weight
0	0,28531
40	0,13223
214	0,10467
163	0,10169
174	0,09698
112	0,09616
218	0,09276
110	0,0902

Sector type
Allows a choice of 1, 4, 4 with an offset of 45º, or 8 sectors.

La validation croisée : voici le résultat de la 1ère méthode, avec la valeur mesurée observée de chaque point , la valeur prédis et l'erreur calculée (la différence entre les deux)

Geostatistical wizard - Inverse Distance Weighting step 3 of 3 - Cross Validation

Source ID	Included	Measured	Predicted	Error
0	Yes	1.74	1.0967...	-0.6432153210528517
1	Yes	1.335	1.7533...	0.4183226656067267
2	Yes	1.61	1.2638...	-0.3461036745636217
3	Yes	2.15	1.1484...	-1.0015333820952073
4	Yes	1.565	1.4828...	-0.08213219179082598
5	Yes	1.145	1.0274...	-0.11754651591214227
6	Yes	0.894	1.0609...	0.1669830971432117
7	Yes	0.525	0.8309...	0.30596214942530364
8	Yes	0.24	0.2243...	-0.015644855880887876
9	Yes	0.625	1.0690...	0.44406380359236564
10	Yes	3.873	1.6472...	-2.225773793096419
11	Yes	1.425	2.0508...	0.6258441611623915
12	Yes	0.455	1.0608...	0.6058077493577592
13	Yes	1.025	1.0440...	0.01902014932070517
14	Yes	0.775	0.7904...	0.015454449392339087
15	Yes	0.62	0.5240...	-0.0959346874209408
16	Yes	0.4	0.3831...	-0.016885064526275084
17	Yes	0.625	1.4253...	0.8003795246108909
18	Yes	0.355	0.6282...	0.2732189371172592
19	Yes	1.585	1.3727...	-0.2122333992608838
20	Yes	1.885	1.9050...	0.020032967351441444
21	Yes	1.3	0.9468...	-0.35316719749935643



Predicted / Error

Regression function	0,41346600312269 * x + 0,...
Prediction Errors	
Samples	259 of 259
Mean	0,006895521
Root-Mean-Square	0,8080416
Export Result Table	

< Back | Next > | Finish | Cancel

Interpretation : en cas parfait , chaque point estimé doit avoir la meme valeur predite . Donc toute éloignement de la ligne montre une moins bonne estimation

la fonction utilisée : ÿ=a+by , tels que ÿ est la valeur estimée et y est la valeur observée . en espère avoir a = 0 , et b =1 pour que ÿ=y . Mais dans notre cas voici la fonction de la droite

Regression function	0,41346600312269 * x + 0,71985289459726

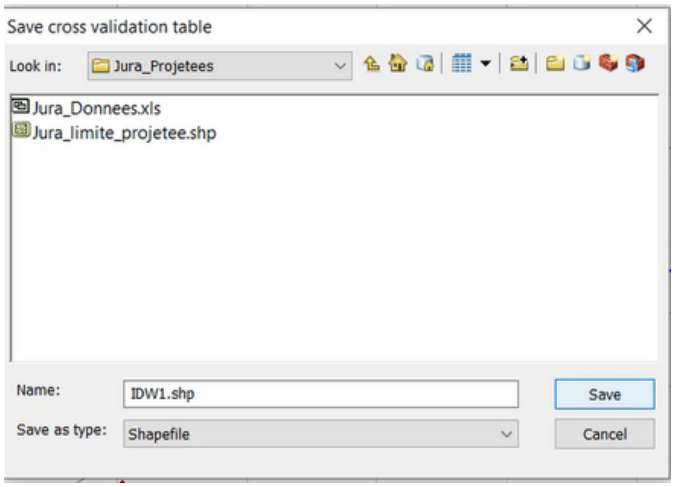
et voici la synthèse de la validation croisée avec le nombre de prédiction totale 259 , l'erreur moyen de ME=0.00689 et RMSE=0.8080

Predicted / Error

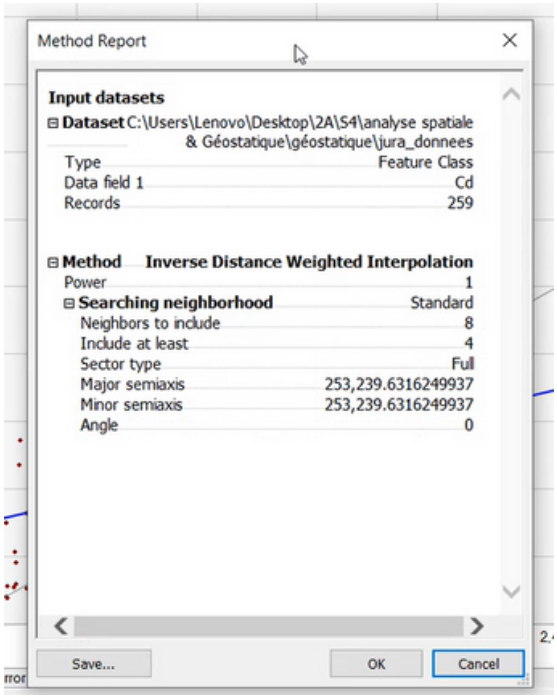
Regression function	0,41346600312269 * x + 0,...
Prediction Errors	
Samples	259 of 259
Mean	0,006895521
Root-Mean-Square	0,8080416
Export Result Table	

< Back | Next > | Finish | Cancel

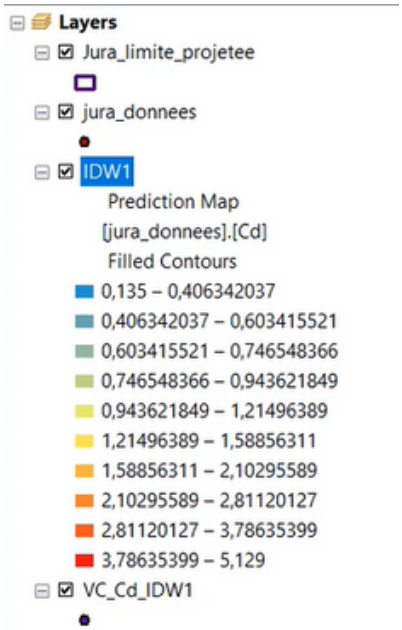
l'exportation des résultats en format shp pour les exploiter plus tard



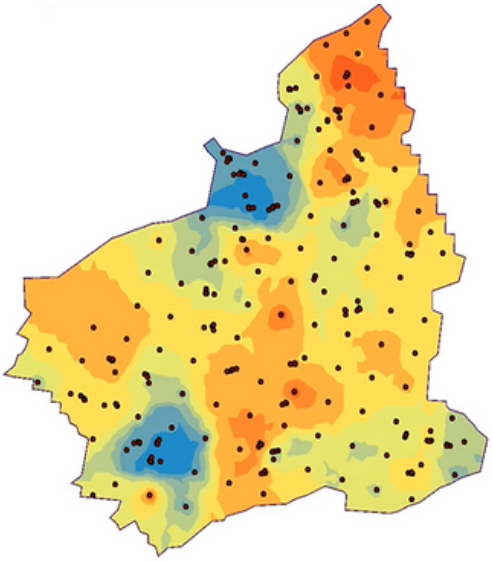
Voici un résumé de ce qui a été fait pour la 1ère méthode :



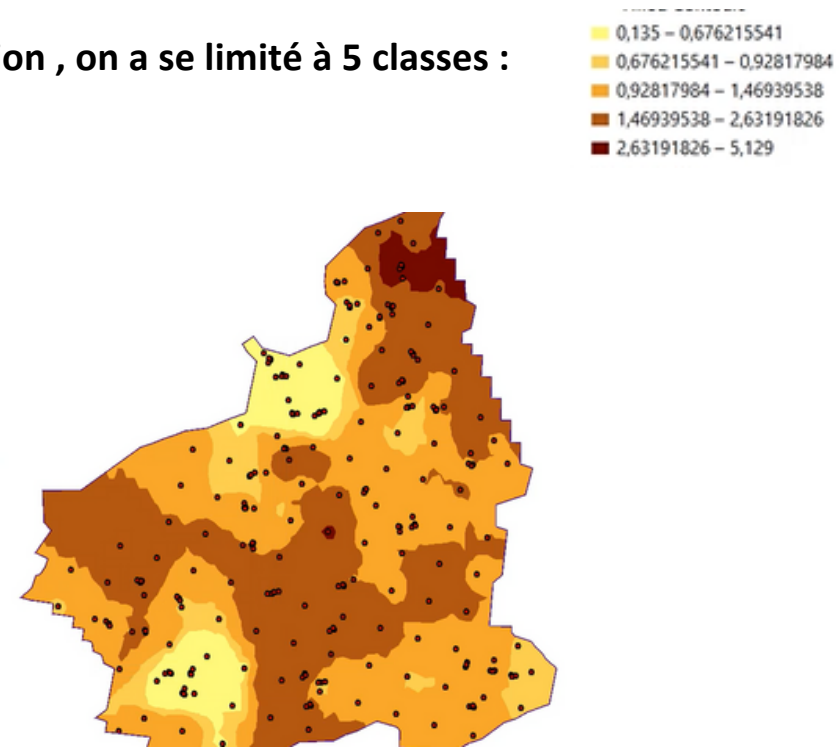
par la suite , on peux exploiter la couche exportée:



voici la carte de toute la zone d'étude après qu'on a fait un CLIP



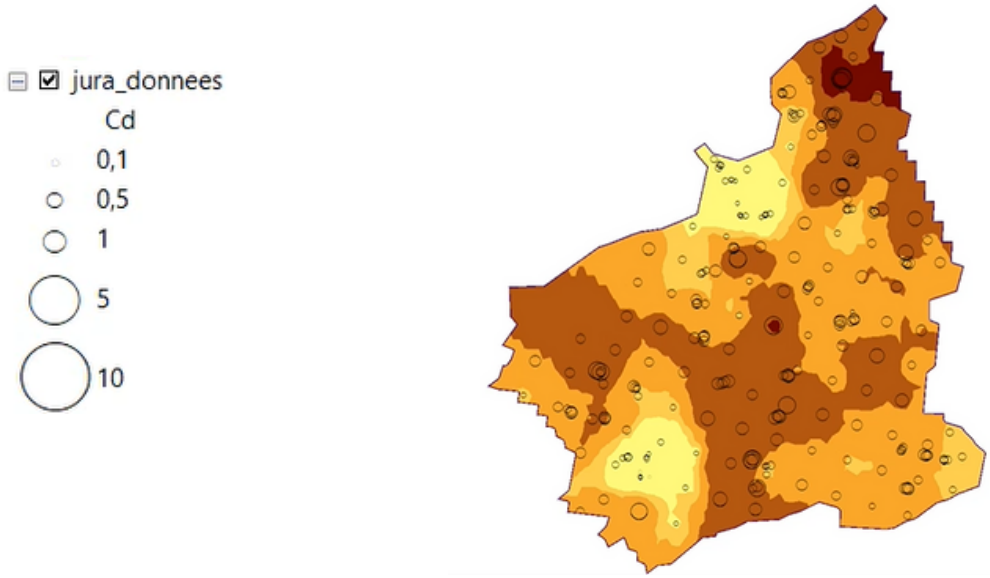
pour le but de simplification , on a se limité à 5 classes :



Interpretation :

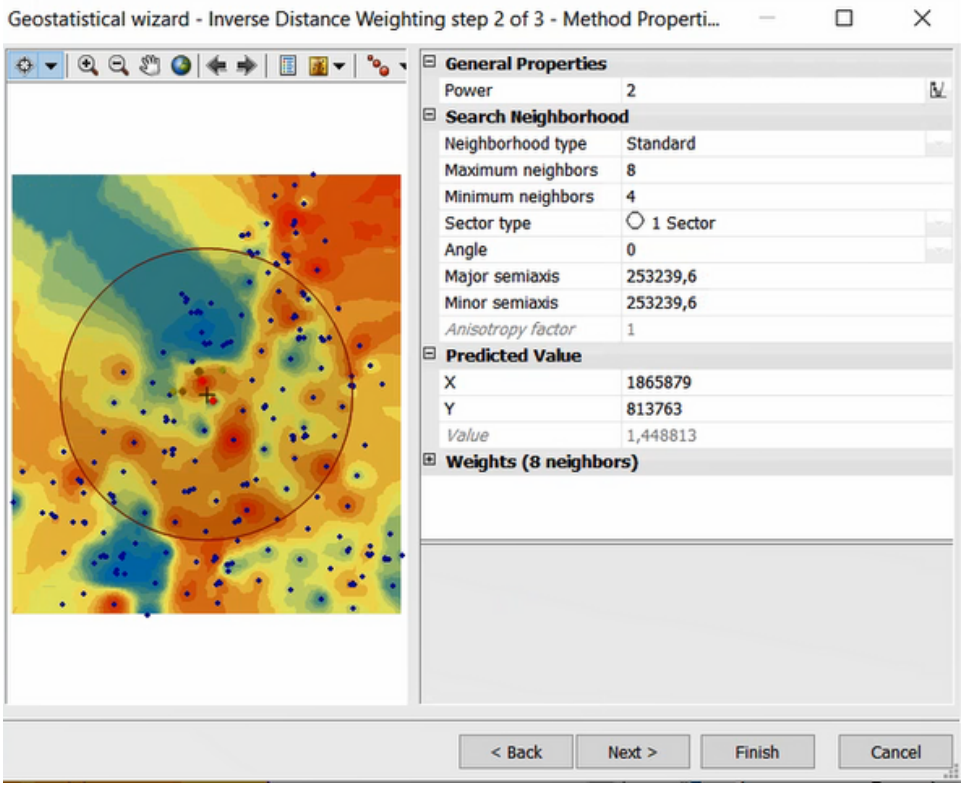
les régions jaunes sont lesquelles où il y a le moins de Cadmium , et les régions foncées c'est là où il ya le plus de Cadmium dans notre zone d'étude

->afin de comparer les prédictions avec les mesures qu'on a en données , on va utiliser les cartes de position (soit avec des couleurs ou bien des cercles)

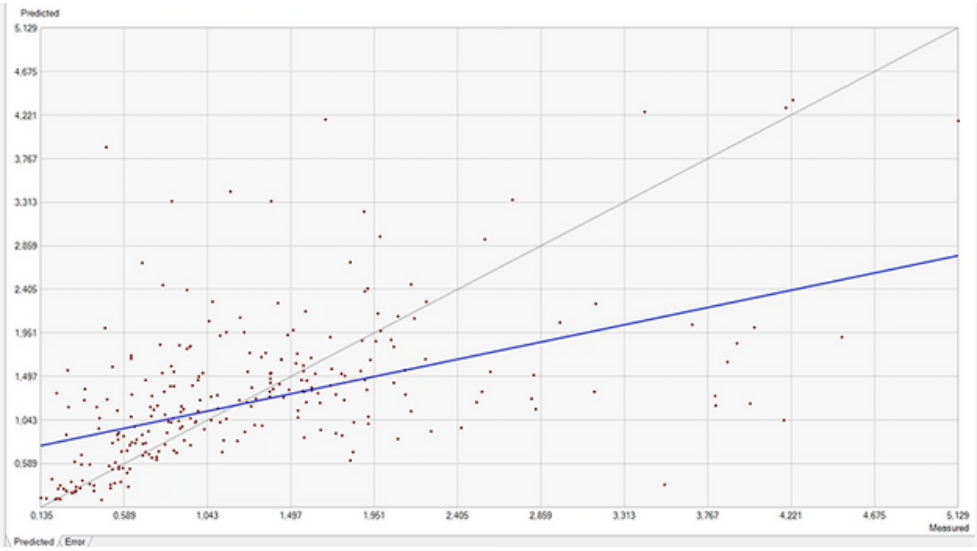


2-on va opter de meme pour IDW (mais la distance inverse pondérée de power =2)

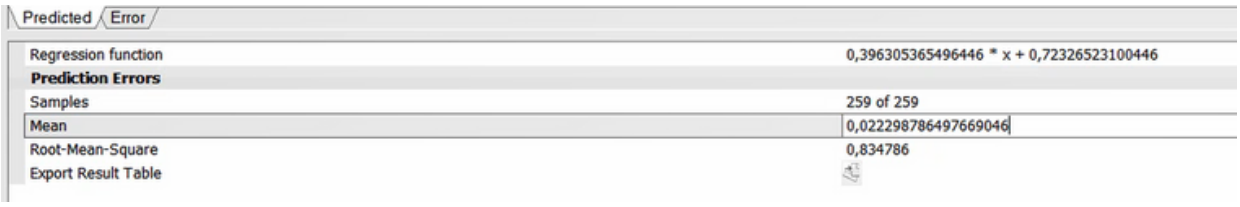
on va fixer les memes valeurs des propriétés de la méthode précédente (sauf power =2) afin de pouvoir les comparer



voici le résultat de IDW2



et voici la synthèse de la validation croisée avec le nombre de prédiction totale 259 , l'erreur moyen de ME=0.022298 et RMSE=0.834786

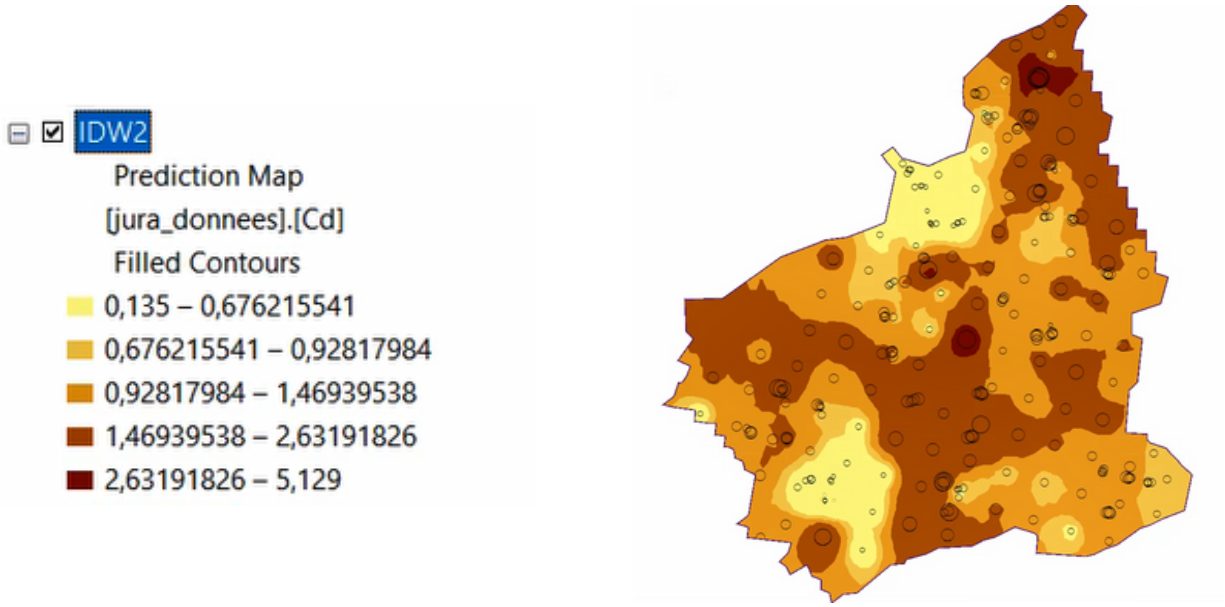


Regression function	0,396305365496446 * x + 0,72326523100446
Prediction Errors	
Samples	259 of 259
Mean	0,022298786497669046
Root-Mean-Square	0,834786
Export Result Table	

Conclusion:

IDW1 est meilleure que IDW2 car les erreurs ME et RMSE de la 1ère approche sont inférieures que celles de la 2ème approche , donc IDW1 va donner plus de précision lors des prédictions

on fait de meme pour IDW2 , en exportant le résultat et l'afficher afin de le comparer graphiquement avec le résultat de IDW1 et les données mesurées



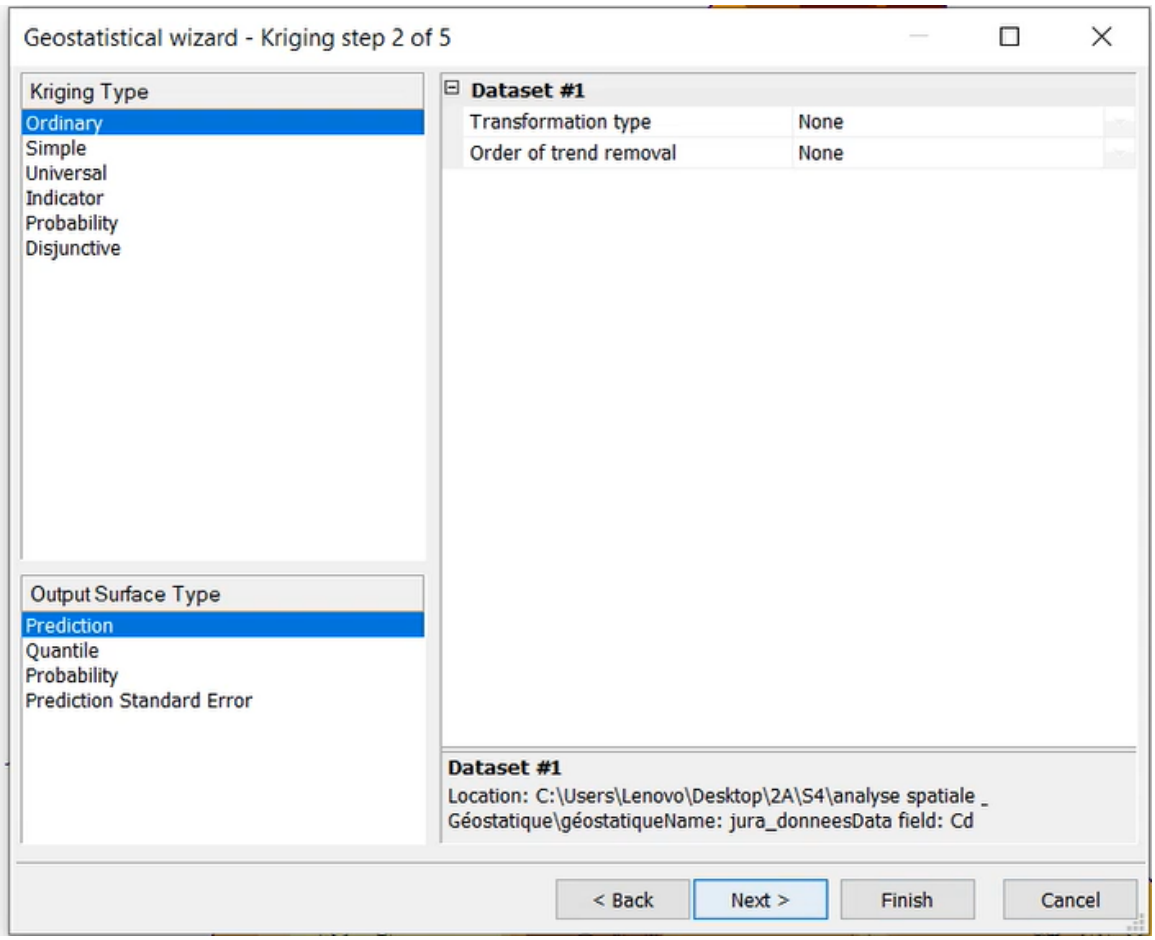
on remarque que les 2 cartes (IDW1 et IDW2) diffèrent en plusieurs zones

Remarque : l'intérêt pratique de l'étude de savoir les régions où le Cd est loin des seuils réglementaires du pays
pour le sol , les plantes , et l'eau afin proteger la nature et l'etre hum proteger la nature et l'etre humain.
Pour notre cas , si le Cadmium est faible (les régions jaunes) ça ne pose pas un problème pour la santé humaine , mais pour les régions où le niveau du Cd est élévée il y a un risque dangereux de polution.

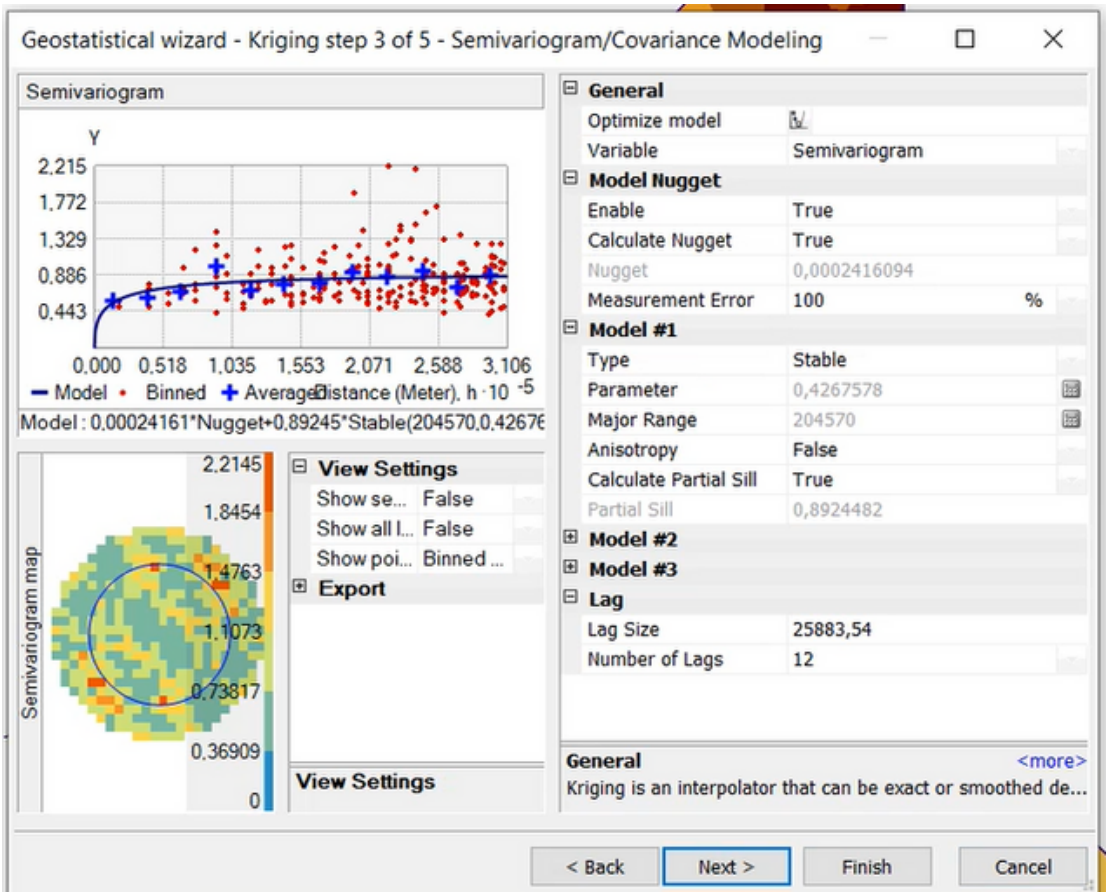
la solution est d'éliminer le Cd en excès par un type de plante qui permet de le diminuer

3-La 3ème méthode : Krigeage

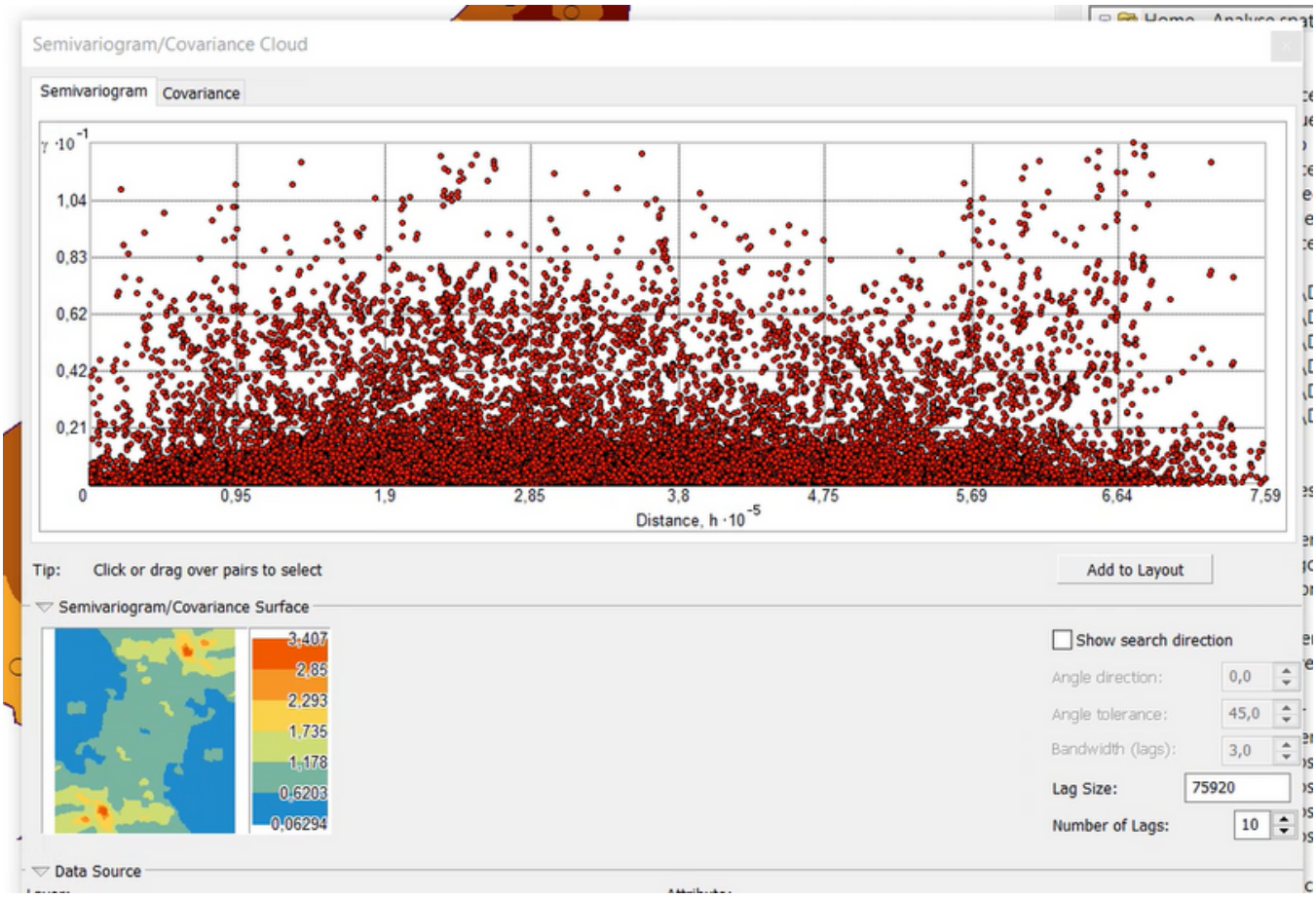
après l'utilisation des méthodes déterministes comme IDW , on passe à une méthode géostatisique comme Krigeage



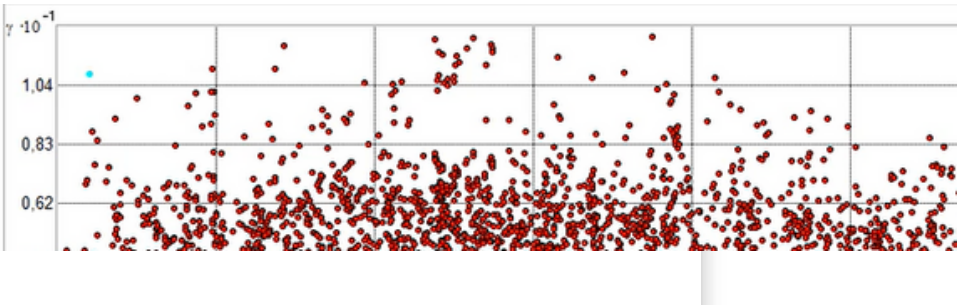
il y a deux étapes supplémentaire par rapport à IDW , il s'agit du variogramme , qui est l'une des caractéristiques des méthodes géostatistiques



voici le nuage du variogramme



en se basant sur ce nuage du variogramme , on peut détécter les points les plus proches dans l'éspace et très éloignés en valeurs de Cd . afin de détérminer s'il y a des erreurs de mesures (les valeurs les plus extrêmes)

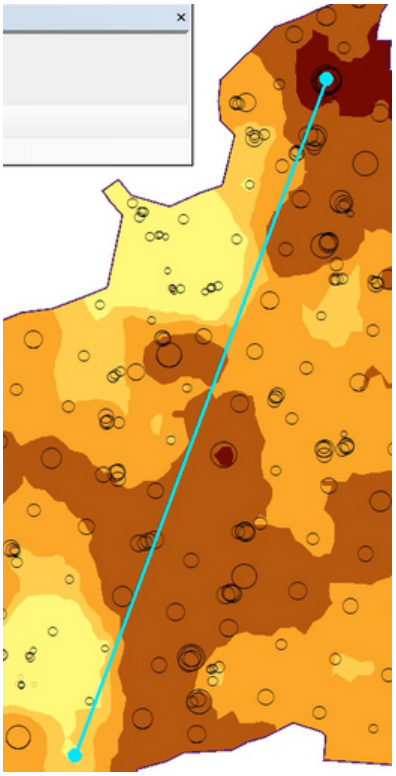


FID	Shape *	LU	Rock	Cd	Co	Cr	Cu	Ni	Pb	Zn	x	y	sample
255	Point ZM	1	1	0,49	7,44	34,52	4,64	16,76	31,64	56,8	2019218	1093261	256
173	Point ZM	2	1	5,12	8,6	39,24	14,76	18,28	36,8	135,6	2016665	1112849	174

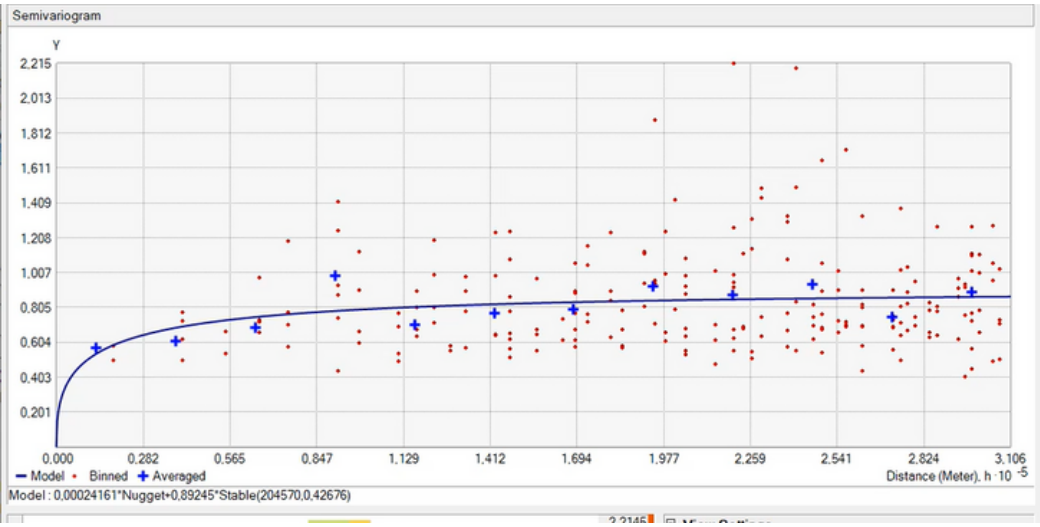
de meme on peut détecter les points les plus éloignés dans l'éspace et éloignés en valeurs de Cd.

Ces valeurs semblent logiques

FID	Shape *	LU	Rock	Cd	Co	Cr	Cu	Ni	Pb	Zn	x	y	sample
92	Point ZM	3	2	0,28	9,6	38	14,16	24,52	44	76	1763598	434163	93
173	Point ZM	2	1	5,12	8,6	39,24	14,76	18,28	36,8	135,6	2016665	1112849	174



->Afin d'optimiser l'estimation , on joue sur plusieurs paramètres ayant un effet sur le variogramme :



Ces paramètres sont :

- le modèle du variogramme (sphérique ,gaussien , exponentiel ,...)
- les 3 caractèristiques du variogramme (l'effet pépite ,la portée , le palier)

Conclusion :

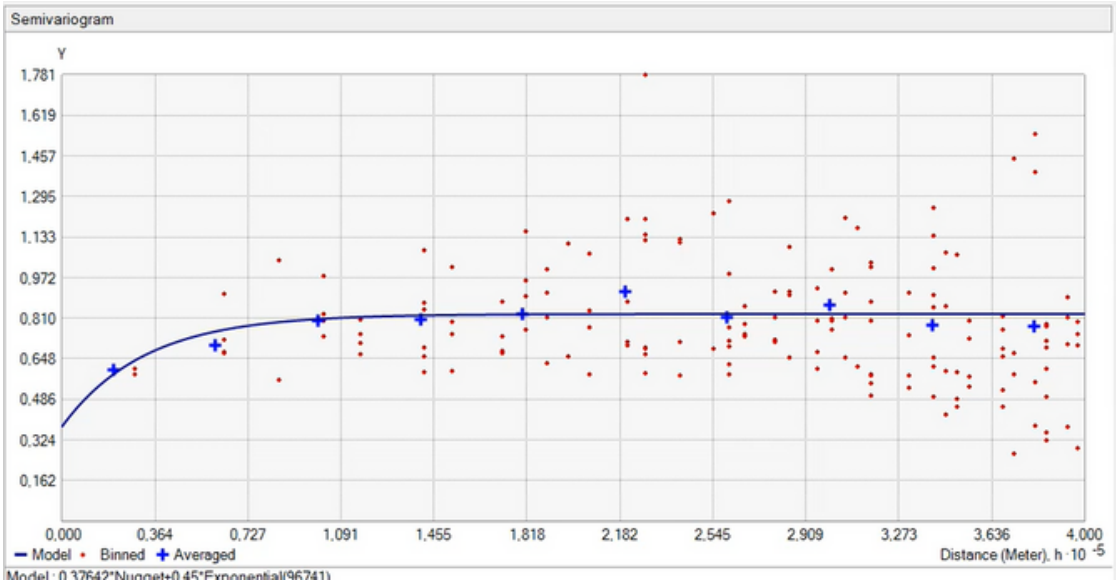
-si on augmente l'effet pépite , l'erreur augmente . De meme pour le palier ,mais si on augmente la portée l'erreur diminue

-le changement du modèle affecte les valeurs estimées

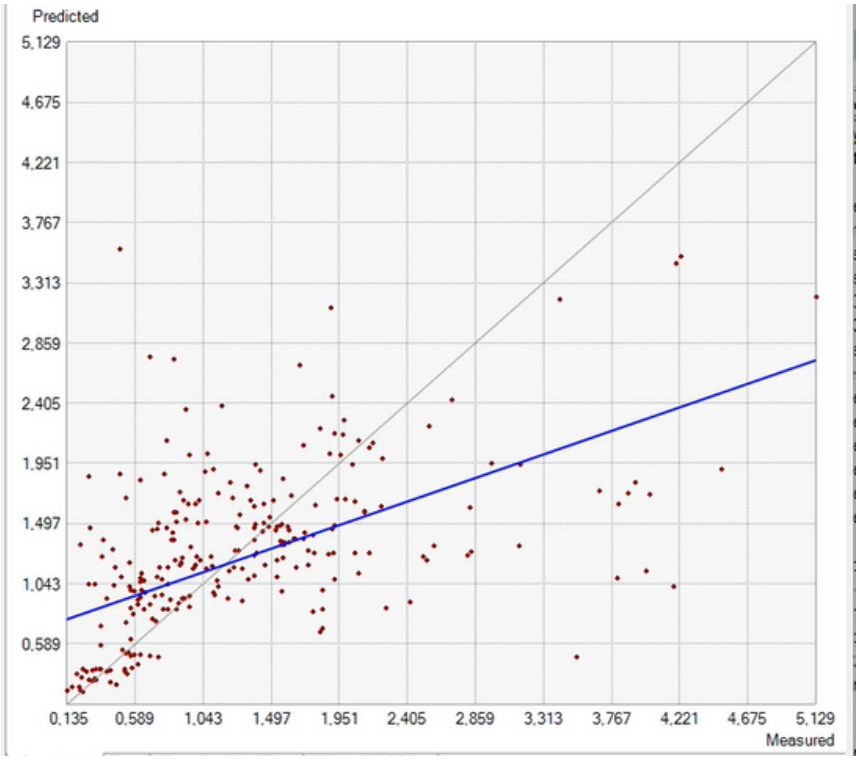
enfin , on a opté pour un modèle exponentiel avec les valeurs des caractéristiques suivantes :

General		
Optimize model		
Variable	Semivariogram	
Model Nugget		
Enable	True	
Calculate Nugget	True	
Nugget	0,3764163	
Measurement Error	100	%
Model #1		
Type	Exponential	
Major Range	96740,65028864906	
Anisotropy	False	
Calculate Partial Sill	False	
Partial Sill	0,45	
Model #2		
Model #3		
Lag		
Lag Size	40000	
Number of Lags	10	

il est clair que cette variogramme est plus optimale , car les moyennes des mesures sont tres proches de la droite du modèle



voici le résultat de Krigeage:



et voici la synthèse de la validation croisée avec le nombre de prédiction totale 259 , l'erreur moyen de ME=0.006430 et RMSE=0.7970

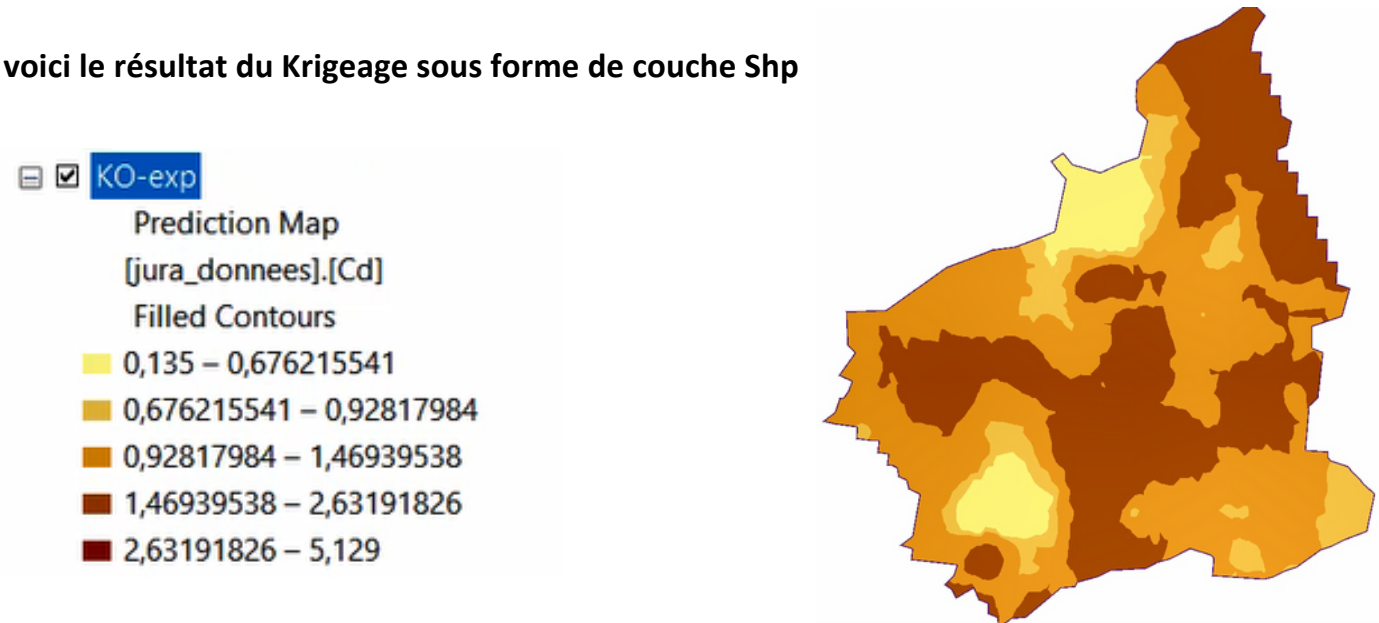
Predicted | Error | Standardized Error | Normal QQPlot

Regression function	0,390990118670086 * x + 0,723969214848473
Prediction Errors	
Samples	259 of 259
Mean	-0,00643053
Root-Mean-Square	0,7970568
Mean Standardized	-0,0136783
Root-Mean-Square Standardized	0,9723413

Conclusion :

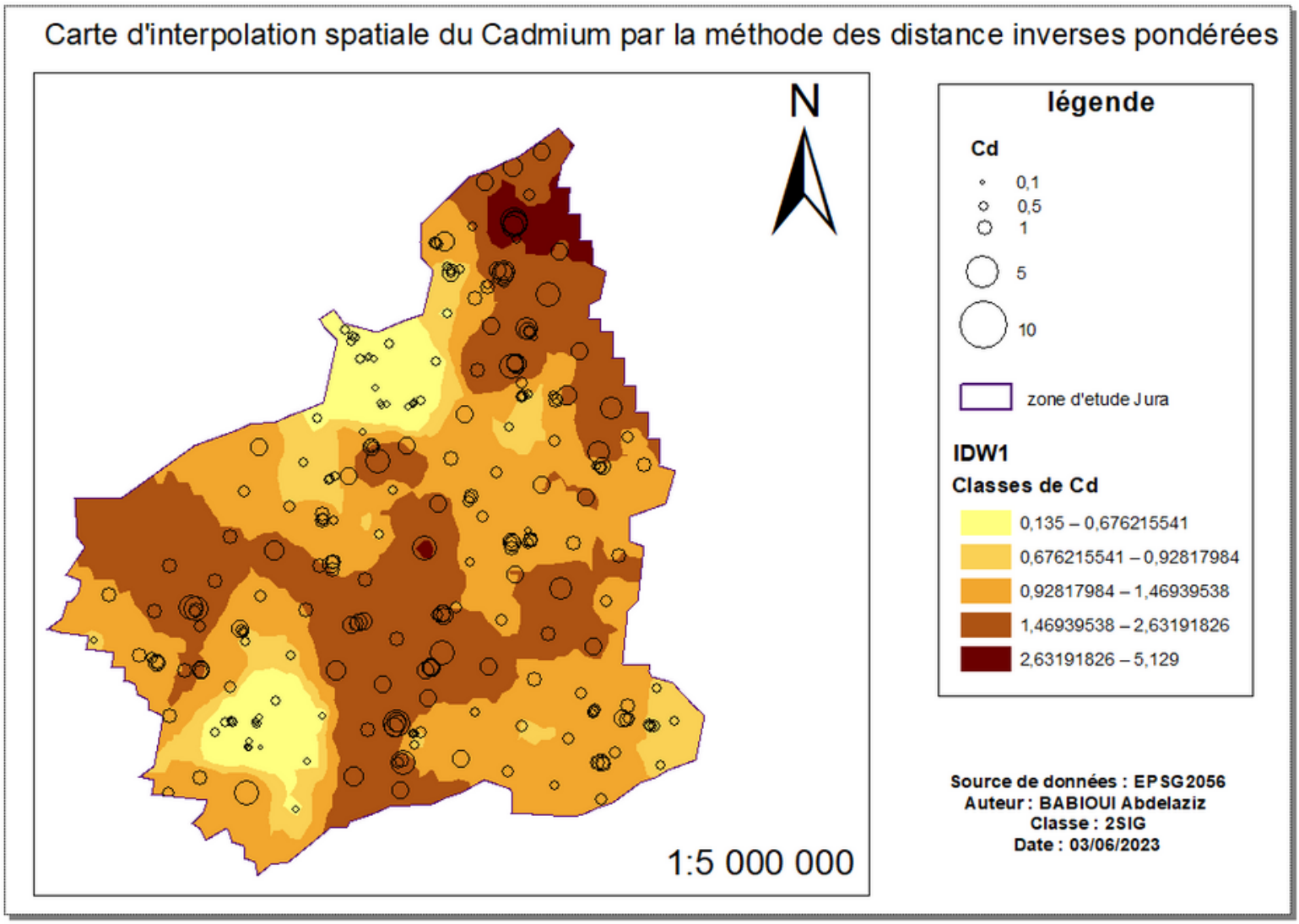
D'après la comparaison des erreurs ME et RMSE des trois methodes , on peux conclure que Krigeage Exp est meilleure que IDW2 mais, IDW1 est le meilleure entre les trois .

voici le résultat du Krigeage sous forme de couche Shp



en comparant les 3 cartes , on remarque que krigeage a fait le lissage des valeurs. (augmenter les valeurs faibles et diminuer les valeurs élévées)

Conclusion : on va opter pour le modèle IDW1 pour créer une carte du Cd dans notre zone d'étude



Remarque : On peux faire les memes étapes pour le Zn afin d'aboutir à créer une carte similaire