**Objectifs:** Apprendre à interpréter les résultats d'un KI et en comprendre les principales étapes.

**Données:** Contamination par le plomb dans un quartier résidentiel de Dallas (données de EPA, Environmental Protection Agency).

Une fonderie de plomb se trouve au centre d'un quartier résidentiel. Afin de répondre aux inquiétudes des résidents, une vaste campagne d'échantillonnage a été menée pour caractériser les sols de la région. Après vérification des données, 5 données ont été éliminées pour cause de pollution locale (batterie, rebuts de métaux, etc...). Le fichier nettoyé renferme 175 échantillons. Vous noterez au nord-est un "corridor" moins échantillonné correspondant à l'emplacement d'une rivière et de sa plaine d'inondation.

Les coordonnées x et y sont en pieds, les valeurs de plomb sont en ppm. La fonderie est située aux coordonnées (6200, 6200). Le fichier de données plomb.dat contient 4 variables, dans l'ordre: un numéro d’échantillon, x, y, et la concentration de plomb.

Après discussion avec les représentants de la municipalité de Dallas, vous déterminez, entre autres que les points suivants sont d'intérêt pour eux:

1. Déterminer la probabilité, en chaque point, d'excéder le niveau 500 ppm de plomb considéré comme dangereux pour la santé par le gouvernement concerné.
2. Définir un intervalle de confiance (disons de niveau 70%), en chaque point, pour la concentration en Pb et éventuellement suggérer des zones à échantillonner davantage.
3. Estimer le volume de sol à traiter et le coût selon divers scénarios. (On suppose une masse spécifique de 2.1t/m3 et un coût de traitement de 100$/t)

* Copier le contenu de k:\geostat\glq3401\2020\tp9 sur \usager
* Lire les données et initialiser certaines variables: **>>go**

Observez les seuils et le codage pour quelques points choisis.

* Afficher les statistiques élémentaires, localisation et carte préliminaire avec griddata: **>>statelem**

Après vérification par le calcul des variogrammes directionnels, le champ étudié ne présente pas d’anisotropie importante. Pour la suite, seuls des variogrammes omnidirectionnels et des modèles isotropes sont considérés.

* Calculer et modéliser les variogrammes du plomb et des indicatrices. Sept seuils sont sélectionnés: 50 100 150 200 300 500 et 700 ppm. Les valeurs représentatives de chaque classe sont les centres des classes (25, 75, 125, 175, 250, 400, 600) et 1735 pour la dernière classe (cette valeur a été choisie égale à la médiane des valeurs dont la concentration est supérieure à 700ppm) **: >>variogrammes**

Note: vous pouvez enlever l'affichage des paires sur les variogrammes à l'aide du menu "affichage".

*Q1.* **(1pts)** *La figure 11 montre le variogramme du plomb et ceux des différentes indicatrices. Décrivez l'évolution de la structure spatiale en fonction du seuil considéré et commentez. Pourquoi les variogrammes des indicatrices permettent-ils d’exclure la possibilité que le phénomène étudié soit gaussien?* (Aide : cela n’a rien à voir avec le modèle gaussien de variogramme)

* Effectuer le krigeage d'indicatrices: **>>kriger.** Le krigeage est effectué en prenant les 16 données les plus près du point à estimer (recherche circulaire).

*Q2.* **(1pts)** *La figure 21 montre les résultats pour les distributions conditionnelles estimées par KI en quatre points distincts. Identifiez visuellement celle présentant:*

1. *La plus grande variance conditionnelle*
2. *L'espérance conditionnelle la plus élevée*
3. *La plus forte probabilité d'excéder le seuil 500ppm*
4. *Un problème de relation d'ordre*
5. *Le 70e percentile le plus élevé*

* Corriger les relations d'ordre: **>>corriger**

*Q3.* **(1pts)** *La figure 22 montre quatre fonctions de répartition estimées par KI avant et après correction. Ces fonctions sont représentatives des violations des relations d'ordre susceptibles d'être rencontrées. Les corrections apportées ont-elles modifié substantiellement les probabilités calculées ?*

* Visualiser les résultats: **>>visualisation**

*Q4.* **(1pts)** *La figure 31 montre la carte des espérances conditionnelles, la figure 32 celle des écarts-types conditionnels. Quel lien observez-vous entre ces 2 cartes ? Expliquez.*

*Q5.* **(1pts)** *La figure 33 montre la valeur krigée du plomb par krigeage ordinaire et la figure 34 l'écart-type de krigeage ordinaire. Comparez la figure 33 à la figure 31 et la figure 34 à la figure 32.*

La figure 35 montre la probabilité, en chaque point, d'excéder le seuil 500ppm. La figure 36 montre le 15e percentile, la figure 37 le 85e percentile. La figure 38 montre la largeur de l’intervalle de confiance 85e-15e percentiles. La carte est obtenue en soustrayant la carte figure 37 de la carte figure 36.

*Q6.* **(1pts)**  *Est-ce que la largeur de l'intervalle de confiance obtenu (fig. 38) montre un lien particulier avec la carte des écarts-types conditionnels (fig. 32)? Confirmez ou infirmez votre impression en calculant la corrélation ente les deux cartes :* >>corrcoef(p85(:)-p15( :),vcond(:))

* Calculer les volumes à traiter (excavation sur 6'') et les coûts ($100/t) en fonction de différents scénarios. Pour cela taper au prompt matlab **>> scenario=i; volume**; où i vaut 1, 2, 3 ou 4 et représente :

i=1: traiter si Prob(Z(x0)>500ppm)>0.5

i=2: traiter si valeur krigée>500ppm

i=3: traiter si espérance conditionnelle Z(x0)\*>500ppm

i=4: traiter si 85e percentile>500ppm

*Q7.* **(1pts)** *Discutez les résultats obtenus par chaque scénario (similitudes et différences, scénario le plus coûteux, le moins coûteux). Que pourriez-vous recommander, en termes d’échantillonnage, avant d’appliquer le scénario 4?*

*Q8.* **(1pts)** *Sur la carte 37 du 85e percentile que pouvez-vous dire de la probabilité que la teneur excède 500 ppm dans les zones non-traitées (i.e. zones où 85e percentile<500ppm) ?*

* Pour les 2 questions suivantes, on suppose que l'on a décidé de traiter suivant le scénario 1 indiqué précédemment, i.e. traiter si Prob(Z(x0)>500ppm)>0.5. x0 contient les coordonnées des points et fzc contient les fonctions de distribution conditionnelle. Les seuils utilisés sont indiqués dans la variable "seuils".

*Q.9* **(1pts)** *Tapez >>[x0(1100,:) fzc(1100,:)]. Quelle décision doit-on prendre en ce point ? Quelle est la probabilité que l'on ait pris la mauvaise décision?*

*Q.10* **(1pts)** *Tapez >>[x0(1000,:) fzc(1000,:)]. Quelle décision doit-on prendre en ce point ? Quelle est la probabilité que l'on ait pris la mauvaise décision?*