ANNEXE A

Méthodologie d'évaluation des intervalles de confiance et de l'incertitude spatiale

Etape 1 : analyse exploratoire des données et estimation

- Analyse exploratoire des données et modélisation
- Validation croisée du (ou des) modèle(s) ajusté(s):
  - calcul des statistiques d'erreur,
  - calcul des erreurs relatives de validation croisée → appréciation **qualitative** de l'incertitude dans la zone d'échantillonnage
- Krigeage :
  - calcul des estimations
  - calcul des écarts-types de krigeage associés→ appréciation **qualitative** de la précision de l'estimation (au sens de l'écart quadratique entre estimation et observation)

Etape 2 : calcul d'intervalles de confiances

Géostatistique linéaire : estimation d'intervalles de confiance, **en faisant une hypothèse sur la loi de distribution des erreurs.**

- Hypothèse d'une distribution gaussienne des erreurs
  - Borne inférieure de l'intervalle de confiance à X%:
    $a = Z^*(x) - 2\sigma_K(x)$
  - Borne supérieure de l'intervalle de confiance à X%:
    $b = Z^*(x) + 2\sigma_K(x)$

  G : fonction de répartition associée à la loi normale

  X : taux de confiance choisi

- Hypothèse d'une distribution continue et unimodale des erreurs
  - Borne inférieure de l'intervalle de confiance à 95% :
    $a = Z^*(x) - 3\sigma_K(x)$
  - Borne supérieure de l'intervalle de confiance à 95% :
    $b = Z^*(x) + 3\sigma_K(x)$

Ces hypothèses ne sont pas vérifiables par le calcul. L'usage de la géostatistique non linéaire est préférable.

Géostatistique non linéaire : estimation d'intervalles au taux de confiance X

Option n°1 : utilisation de l'espérance conditionnelle

Pour des estimations ponctuelles

- Transformation de la variable initiale Z en variable gaussienne grâce aux polynômes d'Hermite → nouvelle variable Y et fonction (modèle) d'anamorphose.

  Le logiciel ISATIS possède une interface (Gaussian Anamorphosis Modelling) permettant de réaliser cette transformation de façon simple et de vérifier les résultats obtenus:

  - Au préalable, il faut définir le nombre de polynômes : quelques dizaines suffisent.
  - L'interface du logiciel ISATIS propose trois méthodes pour transformer l'histogramme brut en histogramme gaussien (*linear interpolation inversion, empirical inversion*, *frequency inversion*). Il est conseillé de tester les trois méthodes et de choisir celle qui, dans les vérifications ci-après, conduit aux meilleurs résultats (dans l'étude présentée en annexe, la méthode employée a été « empirical inversion ») :
    - Apres transformation, il convient d'examiner les statistiques de la nouvelle variable afin de s'assurer qu'elle est bien de loi gaussienne (histogramme, statistiques élémentaires, nuage de corrélation avec une variable gaussienne parfaite, etc.).
    - Contrôle du modèle d'anamorphose : vérifier les coefficients des polynômes calculés. Le premier coefficient doit être égal à la moyenne de la variable brute ; la somme des coefficients des autres polynômes, élevés au carré, doit être proche de la variance des données expérimentales brutes.
    - Réalisation du calcul inverse : calculer la variable brute à partir de la variable gaussienne en utilisant l'inverse de la fonction d'anamorphose et comparer la nouvelle variable brute avec la variable d'origine (nuage de corrélation, histogrammes, statistiques).

- Analyse variographique de Y

  Il s'agit de vérifier que la variable gaussienne présente la même structure que la variable brute - mêmes anisotropies, même stationnarité, etc-. On construit à cette fin le variogramme expérimental de Y ; les paramètres de calcul sont ceux du variogramme de Z.

- Modélisation variographique de Y

  Le variogramme de Y est modélisé en prenant pour structures de base celles qui composent le modèle de Z :

  - Charger le modèle de la variable brute, conserver les portées et modifier uniquement les paliers car ceux-ci doivent être proportionnels à la variance de la nouvelle variable gaussienne (variance=1).

- Une fois les paliers définis, réajuster si nécessaire les portées afin d'affiner le modèle.

- Calcul du krigeage simple ponctuel de Y, noté Y^{KS}, et de l'écart-type de krigeage associé.

- Calcul de l'intervalle de confiance à X% de la variable gaussienne :

  - Borne inférieure de l'intervalle de confiance à X %, pour Y:

$$a^{Y} = Y(x)^{KS} + G^{-1}\left\{1-\left(\frac{X+\left(\frac{(100-X)}{2}\right)}{100}\right)\right\} * \sigma_{KS}$$

  - Borne supérieure de l'intervalle de confiance à X %, pour Y:

$$b^{Y} = Y(x)^{KS} + G^{-1}\left\{1-\left(\frac{\left(\frac{(100-X)}{2}\right)}{100}\right)\right\} * \sigma_{KS}$$

- Transformation des bornes de l'intervalle de confiance de Y par l'inverse de l'anamorphose gaussienne → bornes inférieure et supérieure de l'intervalle de confiance pour Z : a^Z, b^Z.

<u>Pour des estimations par bloc</u>

Le processus est le même qu'en estimation ponctuelle avec, de surcroît, l'application des formules de changement de support .

- Transformation de la variable initiale Z en variable gaussienne de bloc grâce aux polynômes d'Hermite → nouvelle variable Y et fonction (modèle) d'anamorphose de bloc:

  - Calcul du coefficient de changement de support par la formule:

$$r = \sqrt{\frac{Var[Z(x)] - \dddot{\gamma}(v,v)}{Var[Z(x)]}}$$

  - Le variogramme moyen de bloc (γ_{vv}) est calculé par la méthode de discrétisation comme cela est expliqué dans l'étude présentée en annexe.

Si la somme des paliers du variogramme n'est pas égale à la variance expérimentale des données, le variogramme moyen des blocs doit être normalisé :

$$\sum(paliers) \neq Var[Z(x)], \qquad alors:$$
$$\dddot{\gamma}(v,v)_{normalise} = \dddot{\gamma}(v,v)\frac{Var[Z(x)]}{\sum(paliers)}$$

- Les coefficients des polynômes d'Hermite correspondant à la fonction d'Hermite de bloc sont égaux à ceux de la fonction d'Hermite ponctuelle, multipliés par le coefficient de changement de support élevé à la puissance n (n étant le degré du polynôme).

- Après avoir établi le modèle d'anamorphose de bloc, procéder de la même façon que pour des estimations ponctuelles (choisir l'option de bloc pour le calcul du krigeage simple de bloc de la variable gaussienne).

Option n°2 : utilisation des simulations

L'idée de l'algorithme est de transformer la variable brute en variable gaussienne, de simuler la variable dans l'espace gaussien et de transformer les valeurs gaussiennes en valeurs brutes par l'inverse de la fonction d'anamorphose.

Une fois que les résultats de toutes les simulations conditionnelles ont été stockés, le calcul d'un intervalle de confiance à X% pour un point ou un bloc consiste à déterminer les quantiles d'ordre $\frac{1-(X/100)}{2}$ et $\frac{1+(X/100)}{2}$ dans la distribution de fréquence des concentrations simulées en ce point ou ce bloc.

La méthode de simulation choisie est celle des bandes tournantes du menu d'Isatis. L'interface « Cond. Turning Bands » réalise de façon automatique ces procédures. La démarche est la suivante :

- Définition des paramètres de la simulation:
  - Anamorphose gaussienne ponctuelle (même étape que dans l'espérance conditionnelle).
  - Nombre de simulations : plus le nombre de simulations est grand, plus les résultats sont précis. Il faut cependant trouver un compromis entre la qualité des résultats et le temps de calcul informatique. Une centaine de simulations ont été réalisées dans l'exemple traité.
  - Nombre de bandes tournantes: le même critère s'applique. Une cinquantaine de bandes tournantes a été utilisée.

- Le calcul est de type ponctuel ou de bloc selon les cas. Pour un calcul de bloc, le programme discrétise le bloc en points puis réalise les simulations en chacun de ces points dans l'espace gaussien. En conséquence, le temps de calcul augmente notablement.

- Calcul en tout point x ou pour tout bloc v des bornes a^Y et b^Y de l'intervalle de confiance pour Y : calcul des centiles correspondant à chaque borne.
- Transformation par l'inverse de l'anamorphose → bornes inférieure et supérieure de l'intervalle de confiance pour Z : a^Z, b^Z.

<u>Etape 3 : Comparaison avec les objectifs de qualité des directives</u>

Soit d_{max}, une incertitude maximale autorisée.

Au taux de confiance choisi, l'objectif $\frac{|Z^* - Z|}{Z} \leq d_{\max}$ est respecté si :

$$\boxed{(1 - d_{\max}) \cdot b \leq Z^* \leq (1 + d_{\max}) \cdot a}$$