

Calcul des incertitudes liées aux stockages des déchets nucléaires

Anca Badea*, Olivier Gipouloux**

Dans le cadre de la gestion des déchets radioactifs nous nous intéressons à la possibilité du stockage des déchets à vie longue en formation géologique profonde avec une sécurité suffisante et durable. Plus précisément, nous nous intéressons à une situation où un stockage de polluant en milieu géologique profond (plus de 200 m sous terre) serait amené à fuir (suite à une détérioration par exemple du confinement). Les questions qui se posent alors sont: va-t-il y avoir remontée à la surface de polluant (par diffusion à travers les couches géologiques et par convection grâce aux écoulements souterrains) ?, quel est le taux de polluant qui apparaîtra à la surface ? et quand ? Pour faire ceci, nous considérons les équations du transport des radionucléides (le polluant) dans un milieu poreux, nous calculons les concentrations de ces radionucléides pour différents moments à l'exutoire (i.e. dans un voisinage d'un point de la surface), puis nous calculons la probabilité que ces concentrations dépassent un certain seuil admissible de toxicité.

Les phénomènes en relation avec l'environnement sont d'une grande complexité, ils dépendent de données (géologiques, physiques, chimiques) qu'il est souvent difficile de connaître parfaitement, en un lieu donné, en un temps donné, etc. De plus, les observations peuvent être entachées de subjectivité, les mesures affectées d'imprécision, et enfin des caractéristiques pourront être de nature aléatoire.

Dans le cas d'un stockage de déchets, les différentes caractéristiques du stockage ne sont jamais connues de manière exacte (diffusivité, perméabilité et porosité des roches...): les informations disponibles, issues de mesures ponctuelles sont loin de conduire à une description précise des différents champs. L'incertitude sur le stockage se traduit donc en une incertitude sur les variables d'intérêt, i. e. les concentrations des radionucléides. C'est pourquoi le but de cet article est de montrer comment on peut cependant fournir une mesure quantitative de ces incertitudes.

* ENSM. SE, 3MI, 158 cours Fauriel, 42023 Saint-Étienne cedex 2, France

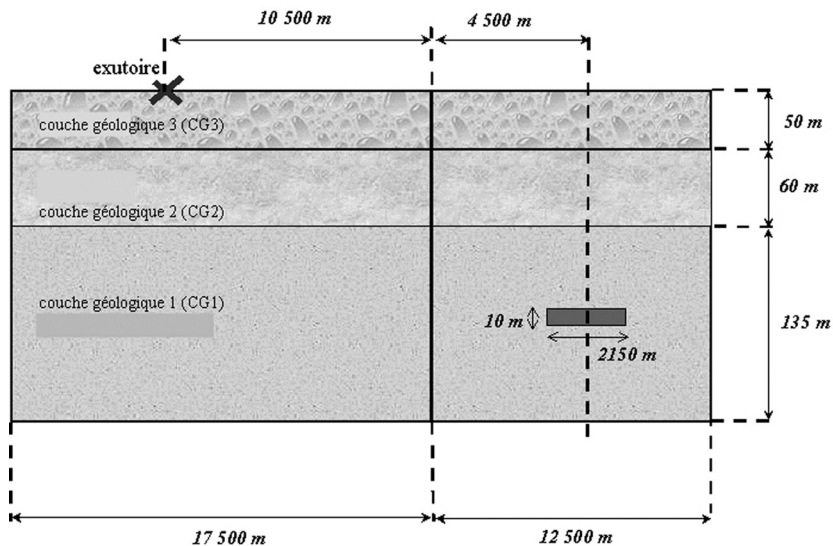
** Institut Camille Jordan UMR 5208, Bâtiment ISTIL, 15 Boulevard Latarjet, 69622 Villeurbanne cedex, France

Nous mettons donc en œuvre la méthodologie des plans d'expériences numériques sur un exemple simplifié mais synthétique de stockage en utilisant les résultats d'un simulateur numérique construit à partir du code CAST3M (Cast3m). Il s'agit d'un code de calcul permettant de résoudre les équations aux dérivées partielles régissant le transport des polluants par des méthodes de type volume ou éléments finis. Pour chaque jeu de données de paramètres physiques, ce simulateur fournit les concentrations aux lieux et aux temps choisis.

Description du problème

La géométrie simplifiée de stockage que nous avons choisi, représentée dans la Figure 1, est conforme au benchmark Couplex. Il s'agit d'un exemple bidimensionnel (une coupe verticale d'un domaine volumique), avec trois couches géologiques différentes et une seule zone de stockage (le rectangle situé à droite dans la couche géologique 1).

Figure 1 : La géométrie du stockage.



Le modèle est décrit par le système (1), dont les deux premières équations représentent la partie hydrodynamique (loi de Darcy et conservation de la masse) régissant les écoulements souterrains et la troisième équation représente la partie transport (modèle de convection dispersion) décrivant l'évolution dans le temps et l'espace du radionucléide. À ce système se rajoutent des conditions aux limites, au bord du domaine, et initiales (au temps de début de la fuite).

La quantité d'intérêt est la concentration du radionucléide I129 (notée c) à un certain endroit appelé exutoire.

$$\begin{aligned} u &= -K^{(i)} \nabla p && \text{dans } CGi, i = 1,2,3 \\ \nabla \cdot u &= 0 && \text{dans } CGi, i = 1,2,3 \\ R\omega \frac{\partial c}{\partial t} - \nabla \cdot (d\nabla c - uc) + \lambda R\omega c &= f && \text{dans } CGi, i = 1,2,3 \end{aligned}$$

Les notations utilisées ici sont :

u : la vitesse de Darcy

p : la pression du fluide

c : la concentration du radionucléide dans le fluide

$K^{(i)}$: la perméabilité absolue (horizontale, verticale ou régionale) de la couche géologique numéro i , CGi

R : le facteur de retard

ω : la porosité effective des anions

d : le coefficient de diffusion effectif

$\lambda = \frac{\log 2}{T}$: où T est le temps de demi-vie du I129, le radio nucléide considéré

f : le terme source (localisé dans la zone de stockage)

t : le temps

CGi : couche géologique numéro i , $i=1, 2, 3$.

Pour plus de détails concernant la modélisation d'un stockage, voir par exemple Bourgeat et al. (2002) et Bourgeat et al. (2004).

La résolution numérique de ces équations est obtenue par le code de calcul numérique CAST3M développé par le Commissariat à l'Énergie Atomique (Cast3m). Il permet de considérer une discrétisation de type éléments finis ou volumes finis pour la partie spatiale et des schémas implicites, explicites ou semi-explicites pour la discrétisation en temps. Les difficultés de résolution sont de deux types : tout d'abord le domaine de calcul est très grand (30 km de long sur près de 250 m de haut) avec en même temps des niveaux de détail très petits (zone de 2 km de long sur seulement 10 m de haut) ce qui nécessite une discrétisation spatiale très fine et donc très importante en terme de nombre d'inconnues ; et ensuite la durée sur laquelle on doit étudier l'évolution de la concentration du radionucléide est extrêmement longue (du même ordre que la période de demi-vie du radionucléide) concrètement 10 millions d'années à partir du moment où la fuite est constatée. Notre approche a été de type volume fini (qui permet de considérer des anisotropies de maillage) et de type discrétisation implicite (qui nous a permis de considérer des pas de temps relativement importants). Le temps moyen d'un calcul (i.e. résolution complète du système pour un seul jeu de donnée de paramètres physiques) est de l'ordre de l'heure et demie.

Dans les cas que nous présentons, les incertitudes se situent au niveau des paramètres intervenant dans le système (1). Ainsi, nous avons retenu, pour cet exemple, six paramètres incertains :

- la perméabilité horizontale de la première couche géologique (notée $Kh1$)
- la perméabilité verticale de la première couche géologique (notée $Kv1$)
- la perméabilité régionale de la deuxième couche géologique (notée $K2$)
- la perméabilité régionale de la troisième couche géologique (notée $K3$)
- la porosité effective des anions (considérée comme étant la même dans les trois couches géologiques) (notée *poro*)
- le coefficient de diffusion effectif (considéré comme étant le même dans les trois couches géologiques) (noté *de*).

Les lois de probabilité de ces paramètres sont supposées connues : lois log normales pour les perméabilités, uniforme pour la porosité et normale pour le coefficient de diffusion.

La méthodologie utilisée

Pour quantifier l'incertitude affectant la prévision de la concentration de I129 à l'exutoire, nous utilisons la méthodologie des plans d'expériences numériques (on remplace les vraies expériences ou mesures par des résultats du code de calcul).

D'une façon générale, les plans d'expériences consistent en un ensemble de techniques mathématiques et statistiques utilisées pour modéliser et analyser un processus dans lequel une variable d'intérêt (appelée la réponse) est influencée par plusieurs autres variables (appelées les facteurs). Quant aux plans d'expériences numériques, ils sont apparus au tout début des années 1990, avec des deux articles fondateurs de Sacks *et al.* (1989), et Sacks *et al.* (1989), puis celui de Welch *et al.* (1992). Dans cet article nous suivons la démarche du livre de Santner *et al.* (2003).

L'idée est d'identifier et de quantifier les sources de variations de la réponse par rapport aux facteurs et ensuite ajuster un modèle « simple » liant la réponse aux facteurs retenus (surface de réponse) et qui par la suite remplacera le simulateur dans le calcul des incertitudes.

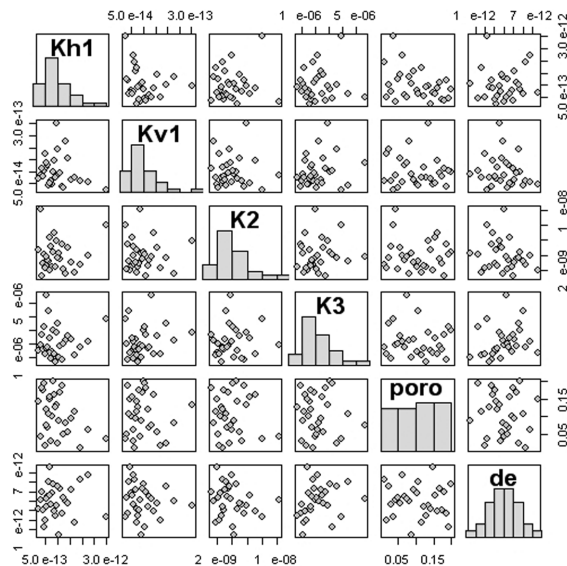
Cette méthodologie est souvent utilisée car la durée des calculs numériques pour la résolution des équations aux dérivées partielles gouvernant le transport des radionucléides provenant d'un stockage de déchets nucléaires est très longue et par conséquent il est important de pouvoir disposer de modèles « alternatifs », rapides, permettant d'analyser de nouvelles situations et surtout permettant de fournir à la fois une analyse de sensibilité de la réponse en fonction des facteurs et une mesure de la propagation des incertitudes sur les facteurs à la réponse.

Dans ce qui suit, nous utiliserons les notations suivantes :

- les facteurs : $x_i, i = 1, \dots, d$ (ici $d = 6$)
- $x = (x_1, \dots, x_d)$; $x^j = (x_1^j, \dots, x_d^j)$: la j -ème expérience numérique
- la réponse (du simulateur) : $y = f(x)$, où f désigne une fonction simple qui « modélise » le simulateur numérique (appelée modèle de simulateur) et qui doit être estimée à partir d'un nombre réduit d'expériences numériques.

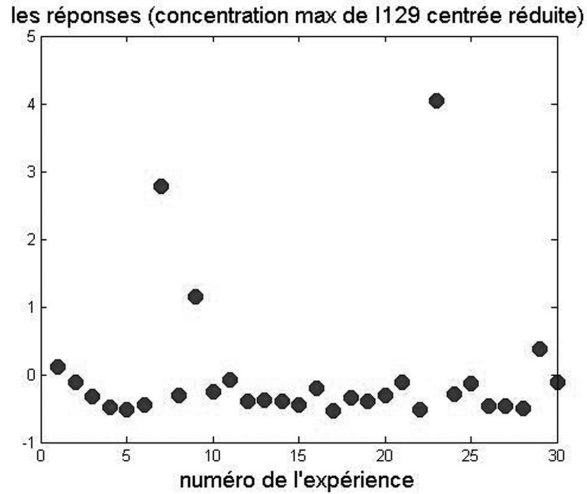
Cette fonction simple sera obtenue, dans notre cas, par interpolation, avec une méthode de krigeage, à partir de n expériences numériques. S'agissant d'expériences numériques, réalisées donc par un code numérique, il apparaît que l'utilisation d'une méthode d'interpolation est plus judicieuse que celle d'une méthode d'approximation, voir Santner et al. (2003). Le plan d'expérience utilisé est un plan de type hyper cube latin (les projections des points x du plan sur tous les axes sont distinctes) maximin (qui maximise la distance minimum entre les points) avec 30 niveaux pour les 6 facteurs. S'agissant de facteurs incertains (qui suivent des lois de probabilités), les niveaux correspondent aux quantiles $(i+0.5)/30, i=0, \dots, 29$ de la loi correspondante. Le plan d'expériences est représenté dans la figure 2 ci-dessous.

Figure 2 : Le plan d'expériences (les facteurs).



La réponse du simulateur est le maximum en temps de la concentration de 1129 à l'exutoire ; nous représentons dans la figure 3 les 30 réponses correspondant aux 30 expériences.

Figure 3 : Les réponses du simulateur



Le modèle de simulateur obtenu par krigage est constitué de deux parties: la première partie constitue la tendance, et la deuxième partie le complément de modèle représenté par la réalisation d'un processus aléatoire, Gaussien, stationnaire de moyenne nulle et fonction de covariance qu'on supposera connue :

$$f(x) = t(x) + z(x) \quad (2)$$

Comme dans la plupart des études parues dans le domaine des plans d'expériences numérique, nous avons supposé que la tendance est constante sur le domaine (krigeage ordinaire), $t(x) = \beta_0$ (constante à estimer). On se base sur le deuxième terme du modèle pour « tirer » la surface de réponse de sorte qu'elle passe par les observations, en quantifiant la corrélation entre les points voisins. La fonction de covariance du processus $Z(x)$ est supposée isotrope, c'est-à-dire :

$$\gamma(x^1, x^2) = \sigma^2 R(\|x^1 - x^2\|) \quad (3)$$

La variance du processus σ^2 est un scalaire à estimer, tout comme les deux paramètres θ, κ de la fonction de corrélation spatiale R , de type Matérn,

$$R(h) = \frac{1}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{h^\kappa}{\theta}\right) K_\kappa\left(\frac{h}{\theta}\right)$$

(où on a noté par h la variable qui représente une distance et dont dépend la fonction de corrélation). Le paramètre θ représente la portée du processus $Z(x)$ et κ représente un paramètre de lissage. On dénote par $\Gamma(\kappa)$ la fonction gamma et par K_κ la fonction de Bessel modifiée de troisième type et d'ordre κ .

Pour estimer ces quatre paramètres $\{\beta_0, \sigma^2, \theta, \kappa\}$ (représentant donc la moyenne, la variance, la portée et le paramètre de lissage) on peut utiliser soit la méthode de maximum de vraisemblance, soit une méthode de validation croisée (et qui est celle utilisée ici).

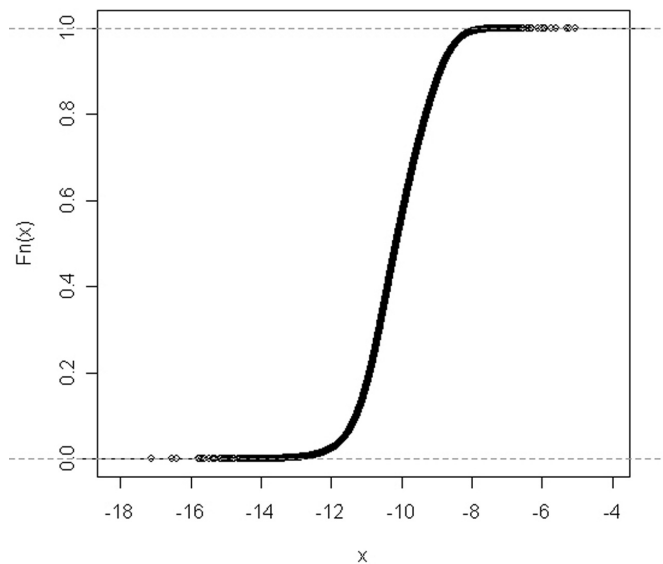
Les résultats

Les résultats ci-après ont été obtenus en utilisant le logiciel libre R (<http://www.r-project.org>) avec les packages *geoR* (voir Ribeiro et Diggle, 2001) et *fields* (voir Fields Development Team, 2004).

Après avoir estimé le modèle de simulateur (plus précisément ses paramètres), on utilise la méthode de Monte-Carlo pour approximer la fonction de répartition théorique de la réponse :

- on simule d’abord selon la loi jointe
- on calcule pour x précédemment simulé la réponse estimée du simulateur (avec le modèle de simulateur)
- on calcule la fonction de répartition empirique de cette réponse, ce qui permet de calculer la probabilité de dépassement de seuil.

Figure 4: La fonction de répartition empirique, échelle logarithmique pour la concentration.



Afin de connaître l’influence des facteurs sur la réponse, on calcule les indices de sensibilité de Sobol (du premier ordre), voir Saltelli, 2002.

Tableau 1 : Les valeurs des indices de sensibilité de Sobol correspondant à chaque facteur

Facteur	Indice Sobol 1 ^{er} ordre
<i>Kh1</i>	0.0292
<i>Kv1</i>	0.1214
<i>K2</i>	0.0167
<i>K3</i>	0.4176
<i>poro</i>	0.5221
<i>de</i>	0.0352

Discussion et interprétation des résultats.

Avant d'interpréter les résultats du tableau 1 il faut d'abord préciser qu'ils ont été obtenus pour le modèle très simplifié (1), que seulement six paramètres ont été considérés comme étant incertains et nous les avons fait varier. De plus des hypothèses restrictives ont été considérées, comme celle d'indépendance statistique des facteurs et celle d'isotropie de la fonction de corrélation. Et enfin, le plan d'expériences considéré compte seulement 30 points dans un espace à 6 dimensions, ce qui signifie que l'espace est presque vide.

Dans ces conditions il n'est pas étonnant que les indices de sensibilité de Sobol indiquent comme facteurs les plus influents la porosité effective des anions (*poro*) puis la perméabilité régionale de la troisième couche géologique (*K3*). En effet, les indices de sensibilité de Sobol sont liés à la décomposition de la variance de la réponse (ici la concentration du radio nucléide à l'exutoire) en termes de chaque facteur pris individuellement; plus sa valeur est grande, plus l'influence du facteur correspondant sur la réponse est grande. En travaillant dans des conditions plus réalistes nous nous attendons à obtenir des résultats indiquant que les paramètres associés à la première couche géologique (celle correspondante à la zone de stockage) sont les plus influents.

Conclusion

Ce travail constitue une première étape, car nous avons utilisé ici un modèle de stockage très simplifié (ne faisant intervenir qu'un faible nombre de paramètres), et nous avons fait des hypothèses très restrictives (paramètres indépendants, isotropie au niveau de la fonction de corrélation).

Il sera continué dans plusieurs directions. Tout d'abord la modélisation du stockage sera complexifiée, de sorte que tous les paramètres considérés « intéressants » par les spécialistes soient pris en compte. Cela conduira évidemment à une augmentation conséquente du temps de calcul du code CAST3M pour chaque simulation. Un plan

d'expériences plus grand (un minimum de 5 expériences par facteur étant nécessaires pour obtenir des résultats corrects, d'après Santner *et al.*, 2003), sera donc construit.

Parallèlement, ayant considéré un modèle de krigeage comme modèle de simulateur, le choix du modèle de covariance utilisé est important. Dans l'exemple présenté ci-dessus, nous n'avons pas fait d'étude particulière du variogramme expérimental afin de retenir le modèle de covariance le plus pertinent. La prise en compte de l'anisotropie de la covariance est une des étapes futures de notre travail.

D'autre part nous nous intéressons à comparer des résultats obtenus par krigeage ordinaire (qui est le modèle utilisé jusqu'à présent) et par krigeage universel afin d'améliorer à la fois l'estimation des paramètres $\{\beta, \sigma^2, \theta\}$ et la variance des prédictions obtenues par le modèle de simulateur, voir l'article de Martin, Simpson, 2005.

Ce travail a été réalisé dans le cadre du GDR CNRS, Modélisations mathématiques et simulations numériques liées aux problèmes de gestion des déchets nucléaires (MoMaS).

Bibliographie

BOURGEAT A., GIPOULOUX O., MARUSIC-PALOKA E., 2002. Mathematical modelling of an array of underground waste containers, *C. R. Acad. Sc. Mécanique*, 330, pp. 371-376.

BOURGEAT A., GIPOULOUX O., MARUSIC-PALOKA E., 2004. Modeling of an underground waste disposal site by upscaling, *Mathematical Modelling in Applied Sciences*, 27 (4), pp. 381-403.

CAST3M, <http://www-cast3m.cea.fr/cast3m/index.jsp>

COUPLEX, <http://www.andra.fr/Couplex/>

FIELDS DEVELOPMENT TEAM, 2004. fields: Tools for Spatial Data. National Center for Atmospheric Research, Boulder, CO. <http://www.cgd.ucar.edu/Software/Fields>.

MARTIN J.-D., SIMPSON T.W., 2005. Use of Kriging Models to Approximate Deterministic Computer Models. *AIAA Journal*, 43 (4), pp. 853-863.

RIBEIRO JR. P.J., DIGGLE P.J., 2001. geoR: A package for geostatistical analysis, *R-NEWS*, vol. 1, n° 2. ISSN 1609-3631. <http://www.est.ufpr.br/geoR>.

SACKS J., WELCH W.J., MITCHELL T.J., WYNN H.P., 1989. Design and Analysis of Computer Experiments, *Statistical Science*, 4 (4), pp. 409-435.

SACKS J., SCHILLER S.B., WELCH W.J., 1989. Designs for Computer Experiments, *Technometrics*, 31, n° 1, pp. 41-47.

SALTELLI A., 2002. Making best use of model evaluations to compute sensitivity indices, *Computer Physics Communications* 145, pp. 280-297.

SANTNER T.J., WILLIAMS B.J., NOTZ W.I., 2003. *The design and analysis of computer experiments*, Springer.

WELCH W. J., BUCK R. J., SACKS J., WYNN H. P., MITCHELL T. J., MORRIS M. D., 1992. Screening, predicting, and computer experiments, *Technometrics*, 34, n° 1, 15-25.