

Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol

D. Guyonnet⁽¹⁾, D. Dubois⁽²⁾, B. Bourguine⁽¹⁾, H. Fargier⁽²⁾, B. Côme⁽³⁾ et J.-P. Chilès⁽⁴⁾

(1) BRGM, 3 avenue C. Guillemin, BP 6009, 45060 Orléans Cedex

(2) IRIT, Université Paul Sabatier, 31063 Toulouse

(3) ANTEA, 3 avenue Claude Guillemin, BP 6119, 45161 Orléans Cedex

(4) Centre de géostatistique, 35 rue Saint-Honoré, 77305 Fontainebleau Cedex

RÉSUMÉ

Les paramètres qui influencent le risque d'exposition aux polluants du sol sont souvent entachés d'incertitude. Une méthode classiquement utilisée pour prendre en compte cette incertitude est la méthode dite Monte Carlo, qui applique la théorie des probabilités et s'appuie sur une représentation statistique de l'information disponible. Depuis quelques années, d'autres théories de l'incertitude ont été proposées pour traiter le cas où l'information dont on dispose n'est pas suffisante pour identifier des distributions de probabilité statistiquement représentatives, en raison notamment du faible nombre de données mesurées. La théorie la plus simple est la théorie des possibilités, qui utilise la notion de distribution de possibilité (aussi appelée nombre flou) pour représenter l'incertitude. Dans la pratique, il peut arriver que certains paramètres puissent effectivement être représentés par des distributions de probabilité (en raison d'un nombre suffisant de données), tandis que d'autres soient mieux représentés par des nombres flous (en raison d'une information trop incomplète). La question se pose alors de savoir comment ces deux modes de représentation de l'incertitude peuvent être combinés dans l'estimation du risque. C'est l'objectif de la méthode dite « hybride » qui combine calcul Monte Carlo et calcul en nombres flous. La méthode est expliquée dans son principe et appliquée au calcul d'une dose d'élément trace métallique absorbée par une cible humaine par le biais de la consommation de légumes. L'exemple d'application illustre le potentiel de la méthode hybride, qui permet de représenter l'incertitude liée aux paramètres qui influencent le risque d'une manière qui est cohérente avec l'information dont on dispose réellement.

Mots clés

Incertitude, risques, probabilités, possibilités, Monte Carlo, Méthode hybride, nombres flous, cadmium, sols pollués.

SUMMARY**ACCOUNTING FOR UNCERTAINTY IN ESTIMATING THE RISK OF EXPOSURE TO SOIL POLLUTANTS**

Parameters that influence the risk of exposure to soil pollutants are generally fraught with uncertainty. A method that is commonly used to address this uncertainty is the Monte Carlo method, which applies probability theory and relies on a statistical representation of available information. In recent years, other uncertainty theories have been proposed as alternative approaches to address model parameter uncertainty in situations where available information is insufficient to identify statistically representative probability distributions, due in particular to data scarcity. The simplest such theory is possibility theory, which uses so-called fuzzy numbers to represent model parameter uncertainty. In practice, it may occur that certain model parameters can be reasonably represented by probability distributions, because there is sufficient data available to substantiate such distributions, while others are better represented by fuzzy numbers (due to data scarcity). The question then arises as to how these two modes of representation of model parameter uncertainty can be combined for the purpose of estimating the risk of exposure. This is the objective of the proposed « hybrid » method which combines Monte Carlo random sampling of probability distribution functions with fuzzy calculus. The method is explained and applied to the calculation of a dose of cadmium absorbed by a human receptor via vegetable consumption. The application illustrates the potential of the proposed method, that allows the uncertainty affecting model parameters to be represented in a way which is consistent with the information at hand.

Key-words

Uncertainty, risk, probability, possibility, Monte Carlo, Hybrid method, fuzzy numbers, cadmium, contaminated soil

RESUMEN**CONSIDERACION DE LA INCERTIDUMBRE EN LA EVALUACION DEL RIESGO DE EXPOSICION A LOS CONTAMINANTES DEL SUELO**

Los parámetros que influyen en el riesgo de exposición a los agentes contaminantes del suelo, se ven a menudo rodeados de incertidumbre. Un método clásico que se utiliza para tener en cuenta esta incertidumbre es el llamado método Monte Carlo, que aplica la teoría de las probabilidades basándose en una representación estadística de la información disponible. Desde hace unos años, se proponen otras teorías de la incertidumbre para tratar los casos en que no se dispone de la suficiente información para identificar una distribución de probabilidades estadísticamente representativa, especialmente en caso de un número insuficiente de datos medidos. La teoría más sencilla es la teoría de las posibilidades, que define la noción de distribución de posibilidad (también conocida como número borroso) para representar la incertidumbre. En la práctica, puede ocurrir que determinados parámetros puedan ser efectivamente representados por distribuciones de probabilidad (gracias a un número suficiente de datos), mientras que otros se representan mejor con números borrosos (debido a una información demasiado incompleta). Así pues se plantea la cuestión de saber cómo pueden combinarse estas dos formas de representación de la incertidumbre, para hacer una estimación del riesgo. Este es el objetivo del llamado método "híbrido", que combina el cálculo Monte Carlo con el cálculo en números borrosos. Se explica el principio del método y se aplica al cálculo de una dosis de elemento en forma de trazas metálicas absorbidas por un individuo al consumir verduras. El ejemplo de aplicación ilustra el potencial del método híbrido que permite la representación de la incertidumbre, ligada a los parámetros que influyen en el riesgo, de forma coherente con respecto a la información de que se dispone realmente.

Palabras clave

Incetidumbre, riesgos, probabilidades, posibilidades, Monte Carlo, método híbrido, números borrosos, cadmio, suelos contaminados.

Parmi les objectifs exprimés dans l'appel à proposition de recherche du programme GESSOL, piloté par l'INRA pour le Ministère de l'Écologie et du Développement Durable (MEDD), figure celui de définir des critères de qualité des sols qui ne soient pas uniquement liés à des concepts de fertilité physico-chimique et biologique, mais qui intègrent également un souci de préservation de la santé humaine. Lorsqu'il est confronté à une mesure de la teneur en polluant d'un sol, l'investigateur peut envisager plusieurs approches pour appréhender la dimension sanitaire. Il peut, par exemple, comparer cette teneur à une valeur de référence (si elle existe) établie le cas échéant par une instance sanitaire. Il s'agit là d'une approche que l'on peut qualifier de « générique », en ce qu'elle ne tient compte ni des spécificités de l'environnement du point où la teneur a été mesurée, ni de la manière dont une cible humaine pourrait être exposée à cette pollution suivant l'usage qu'elle pourrait faire de ce sol. La question peut se poser, en effet, de savoir s'il convient d'appliquer les mêmes critères de qualité pour un site destiné à une utilisation strictement industrielle que pour un site destiné à l'implantation de résidences ou à une utilisation agricole. On peut également considérer la teneur mesurée en se référant à un fond pédogéochimique local (voir Baize et Tercé, 2002). Pour utile qu'elle soit, lorsqu'il s'agit de mettre en évidence le caractère anthropique de la teneur mesurée, cette démarche n'intègre pas de manière explicite la dimension sanitaire.

Une approche qui peut trouver des applications dans certains contextes est l'approche dite « spécifique » où l'appréciation de la qualité d'un sol s'appuie sur une évaluation du risque d'exposition de populations humaines aux polluants de ce sol, évaluation qui tient compte d'un usage actuel ou futur. C'est l'approche qui a été retenue par le MEDD pour ce qui concerne la gestion des sites potentiellement pollués (voir MEDD, 2000). Ce choix a été guidé par un certain pragmatisme et en particulier par le constat que la France doit gérer un passé industriel lourd et que les ressources humaines et financières qui peuvent être consacrées à cette gestion sont forcément limitées. L'objectif est donc de concentrer prioritairement ces ressources sur les sites les plus susceptibles de provoquer des nuisances. Mais cette approche, si elle présente certains avantages, a aussi comme inconvénient de devoir s'appuyer dans une certaine mesure sur de la modélisation, avec les conséquences que cela entraîne en terme d'incertitude. En effet, s'il est incontestable que la mesure directe doit être privilégiée chaque fois que cela est techniquement et économiquement possible, de nombreux paramètres qui interviennent dans l'exposition de populations aux polluants du sol sont difficilement mesurables, ne serait-ce que parce que cette exposition peut couvrir des périodes de temps de plusieurs décennies et que les mécanismes de migration des polluants depuis la source de pollution vers les cibles potentielles sont souvent très lents.

Deux démarches sont traditionnellement utilisées pour tenir compte de l'incertitude dans l'évaluation de l'exposition d'une cible humaine aux polluants d'un sol. Une approche que l'on qualifiera de « déterministe » et l'approche probabiliste. La démarche déterministe consiste à attribuer des valeurs dites « sécuritaires et majo-

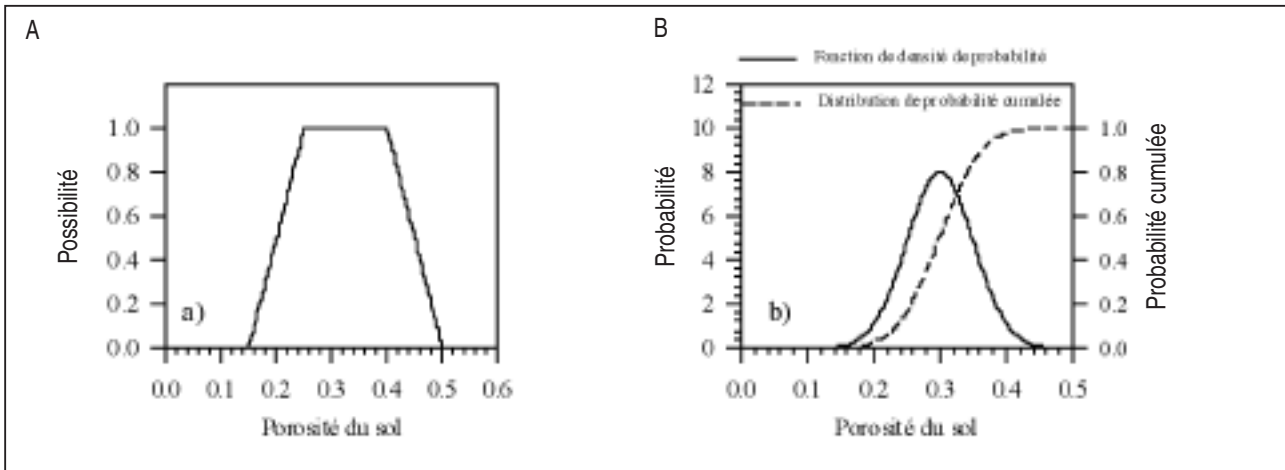
rantes » aux paramètres incertains qui influencent le risque d'exposition. Si le risque calculé demeure en deçà d'un seuil maximal admissible fixé par l'autorité sanitaire, on estime alors que le sol ne pose pas de problème d'ordre sanitaire. Mais cette démarche peut parfois constituer une application trop stricte du « principe de précaution » et se traduire, soit par des objectifs de dépollution trop ambitieux (et donc coûteux), ou alors par des restrictions d'usage sur des zones trop étendues (au travers de servitudes; voir MEDD, 2000).

La démarche probabiliste consiste à représenter les paramètres incertains par des distributions de probabilité et à propager l'incertitude relative à ces paramètres sur celle du risque encouru par la cible, en appliquant notamment la technique dite Monte Carlo (voir Vose, 1996). Si cette technique est bien connue, toute la difficulté tient à une définition cohérente des distributions de probabilité. En effet, dans un contexte d'évaluation des risques liés à l'exposition aux polluants du sol, l'information dont on dispose concernant certains paramètres ne permet généralement pas une analyse fréquentielle et le calage de lois de probabilité sur des distributions de fréquences observées. Pour pallier cette lacune, certains chercheurs proposent d'utiliser des distributions de probabilité dites subjectives, définies à partir de jugement d'expert dans un cadre « Bayésien » (Bonano *et al.*, 1990; Cullen et Frey, 1999). Plusieurs chercheurs (voir par exemple Wonneberger *et al.*, 1995) ont néanmoins critiqué cette approche sur la base qu'elle introduit artificiellement une information qu'on ne possède pas réellement et génère un degré de confiance dans le résultat de l'évaluation qui n'est pas justifié par l'information disponible. Ferson et Ginzburg (1996) insistent notamment sur le fait qu'on ne peut pas aborder la variabilité (le hasard) et l'ignorance (l'imprécision) avec les mêmes vecteurs d'information (voir aussi Kaufmann et Gupta, 1985).

Lorsqu'on est confronté à une information qui est incomplète ou imprécise, il peut être préférable de parler de « possibilités » plutôt que de « probabilités ». La théorie des possibilités (Zadeh, 1965, 1978, Dubois et Prade, 1988, Dubois *et al.*, 2000) fournit un cadre méthodologique pour traiter ce type d'information. Au lieu d'être représentées par des distributions de probabilité, les variables incertaines sont représentées par des distributions de possibilité (aussi appelées des nombres « flous »). De manière schématique, on peut se représenter une distribution de possibilité comme une superposition d'intervalles emboîtés de valeurs du paramètre incertain, chaque intervalle correspondant à un certain degré de confiance. Ce degré de possibilité, qui traduit la notion d'absence de surprise (Shackle, 1961), correspond d'un point de vue mathématique à une borne supérieure de probabilité (Dubois et Prade, 1992). La *figure 1a* illustre la notion de distribution de possibilité pour le cas de la porosité d'un sol. Cette distribution nous informe que la porosité de ce sol est vraisemblablement située entre 0.25 et 0.4 (on appelle cet intervalle le noyau), sans qu'il y ait de préférence exprimée à l'intérieur de cet intervalle, mais que des valeurs aussi faibles que 0.15 ou aussi élevées que 0.5 ne peuvent pas être totalement exclues (l'intervalle 0.15 - 0.5 est

Figure 1 - Illustration de (a) : un nombre flou représentant l'incertitude sur la porosité d'un sol et (b) : une fonction de densité de probabilité et sa fonction associée de distribution de probabilité cumulée.

Figure 1 - Illustration of (a) : a fuzzy number representing the uncertainty with respect to the porosity of a soil and (b) : a probability density function and its associated cumulative probability distribution function.



appelé le support). Des valeurs respectivement supérieures ou inférieures à ces valeurs sont considérées comme n'étant pas possibles.

A noter que ce type de représentation constitue, dans bien des cas, une traduction bien plus « fidèle » de l'information disponible que ne le fait une distribution de probabilité définie par exemple par une moyenne et un écart type (*figure 1b*). Lorsqu'on est en présence d'une telle information, on pourrait envisager de travailler non pas avec une seule, mais avec toute une famille de distributions de probabilité. Mais cette approche est particulièrement fastidieuse et la théorie des possibilités fournit un cadre beaucoup plus simple et pratique. A noter également que la distribution de possibilité la plus simple (et la moins informative) est une distribution rectangulaire définie uniquement par une borne minimum et maximum. On retombe alors sur un classique intervalle min-max et la forme rectangulaire du nombre flou correspondant traduit simplement la certitude que la valeur se trouve à l'intérieur de cet intervalle. La forme trapézoïdale de la *figure 1a* traduit le fait que l'investigateur possède plus d'information que celle véhiculée par un simple intervalle min-max, une certaine plage de valeurs lui paraissant plus vraisemblable qu'une autre. Bien que linéaires dans la *figure 1a*, les branches qui relient les valeurs limites du noyau et du support pourraient être courbes, une forme concave exprimant l'idée que les valeurs situées en dehors du noyau sont possibles mais jugées peu vraisemblables et une forme convexe exprimant l'inverse.

De nombreux chercheurs ont utilisé soit l'un soit l'autre de ces modes de représentation de l'incertitude. Par exemple Labieniec *et al.* (1997) applique la méthode Monte Carlo à l'échantillonnage de distributions de probabilité pour estimer les risques liés à la présence de sols pollués. Freissinet *et al.* (1998) ont appliqué la théorie des possibilités à la modélisation de la pollution des eaux souterraines.

Cazemier (1999) et Lagacherie *et al.* (2000) l'ont utilisée pour cartographier des paramètres du sol imprécis. Par contre peu de chercheurs se sont intéressés à la question de savoir comment combiner des informations de type probabiliste et possibiliste dans une même analyse. Il s'agit pourtant là d'une question importante car dans la pratique on peut être confronté à des paramètres pour lesquels l'information disponible justifie l'utilisation d'une représentation probabiliste, tandis que pour d'autres l'insuffisance d'information milite en faveur d'une représentation possibiliste. Cette combinaison de différents modes de représentation de l'incertitude liée aux paramètres influençant l'exposition aux polluants du sol est au centre du projet qui a été réalisé dans le cadre du programme GESSOL. Un des principaux objectifs de ce projet a été d'identifier une méthodologie d'estimation du risque permettant de représenter l'incertitude d'une manière qui soit cohérente par rapport à l'information dont on dispose dans la pratique. On notera que cette question n'est pas uniquement d'ordre scientifique mais, comme il sera montré plus loin, peut avoir des conséquences importantes en terme d'orientation des choix en matière de gestion des sites potentiellement pollués. La méthodologie proposée peut trouver également des applications dans d'autres domaines pour lesquels se pose le problème de l'incertitude.

LA MÉTHODE « HYBRIDE »

Considérons un « modèle », noté M, qui fait intervenir des paramètres dont certains sont entachés d'incertitude. Ce modèle pourrait être, par exemple, une simple équation ou formule de calcul. L'approche purement « probabiliste » consiste à représenter les paramètres incertains par des distributions de probabilité (*figure 1b*). Lors

d'un calcul Monte Carlo classique on génère des nombres aléatoires compris entre 0 et 1 et on « échantillonne » les distributions de probabilité pour constituer un jeu de valeurs qui sert ensuite pour le calcul à partir du modèle. En répétant cette procédure un grand nombre de fois on obtient plusieurs résultats du modèle qui peuvent être représentés sous la forme d'une distribution de fréquences relatives. Cette distribution peut ensuite être comparée à un critère de tolérance pour un certain niveau de probabilité. Par exemple: le résultat calculé a-t-il une probabilité d'au moins 95 % d'être inférieur au

critère ?

L'approche purement « possibiliste » consiste à représenter les paramètres incertains par des distributions de possibilité (*figure 1a*). Le calcul à partir du modèle peut être effectué par la méthode dites des α -coupes (Dubois et Prade, 1988). Cette méthode revient schématiquement à effectuer de simples calculs d'intervalle sur des plages de valeurs correspondant à différents degrés de vraisemblance. On choisit tout d'abord une valeur de possibilité (notée α ; par exemple 0.4) et on sélectionne toutes les valeurs des paramètres flous dont la possibilité est d'au moins 0.4 (cet ensemble de valeurs est appelé α -coupe 0.4; *figure 2*). On cherche ensuite les valeurs minimum et maximum du modèle, pour toutes les combinaisons possibles de valeurs et on attribue une possibilité de 0.4 à ces valeurs minimum et maximum. En répétant ce calcul pour différentes valeurs de possibilité α (de 0 à 1 par pas de 0.1) on reconstitue la distribution de possibilité du résultat du modèle. Pour revenir au cas simple du calcul d'intervalle on notera que si les paramètres incertains étaient représentés par des intervalles min-max (nombres flous rectangulaires), alors la méthode des α -coupes se réduit à un calcul d'intervalle unique où on cherche les valeurs min et max du modèle sur toute la gamme de valeurs des paramètres.

Une illustration très simple de la différence fondamentale entre les approches probabiliste et possibiliste est obtenue en considérant le problème suivant (voir Ferson, 1996). Supposons que l'on ait deux paramètres non reliés, A et B, dont on ne sait que peu de choses si ce n'est que A est situé entre 0 et 1, tandis que B est situé entre 1 et 2. On s'intéresse à la somme de A et B et plus particulièrement à la valeur moyenne de cette somme. Si on se place dans un cadre possibiliste, les paramètres A et B sont des intervalles de type min-max et leurs nombres flous res-

Figure 2 - Illustration d'une α -coupe pour une valeur $\alpha = 0.4$: toutes les valeurs de porosité situées dans l'intervalle grisé.

Figure 2 - Illustration of an α -cut for a value of $\alpha = 0.4$: all the values of porosity located within the shaded area.

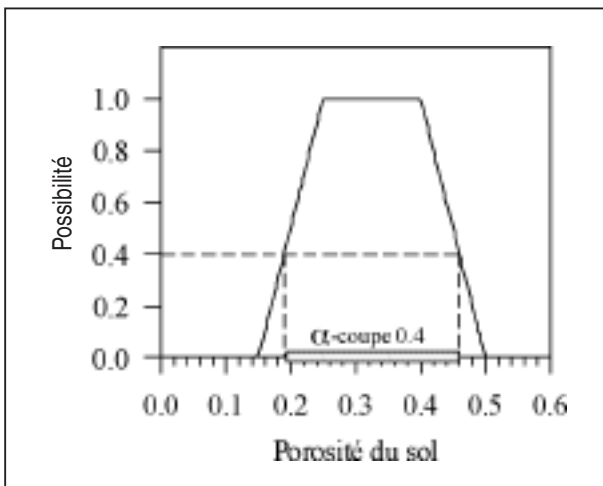


Figure 3 - Illustration de la différence entre un calcul possibiliste (a) et probabiliste (b).

Figure 3 - Illustration of the difference between a possibilistic (a) and a probabilistic (b) calculation.

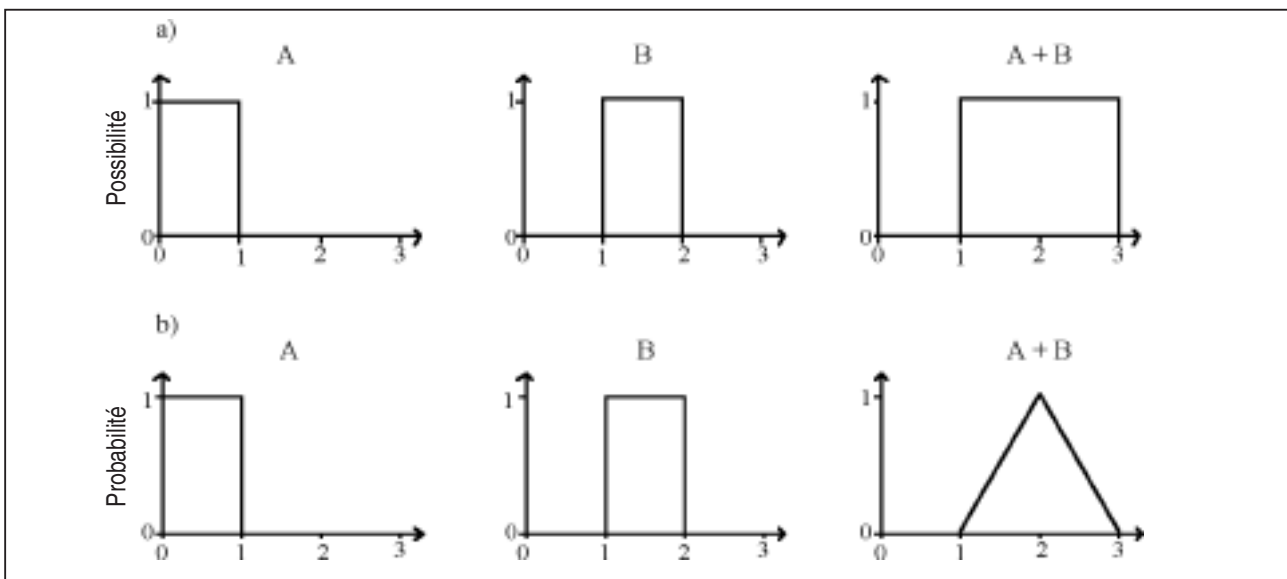
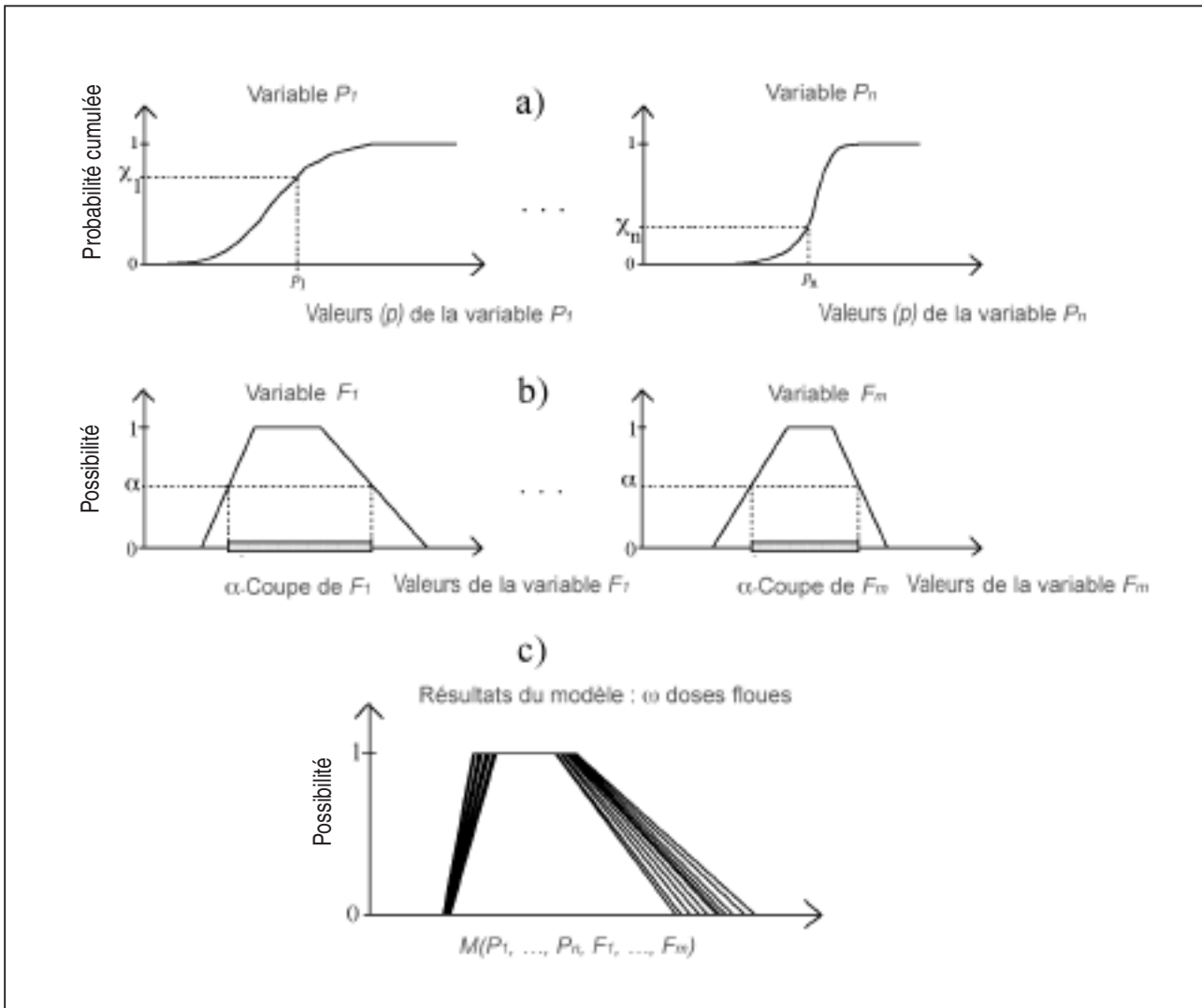


Figure 4 - Illustration de la méthode dite « hybride » (Guyonnet *et al.*, 2003).**Figure 4** - Illustration of the "hybrid" method (Guyonnet *et al.*, 2003).

ectifs sont représentés dans la *figure 3a*. La somme de A et B est simplement l'intervalle 1-3. La valeur moyenne est l'intervalle tout entier; sur la base de l'information disponible on ne peut pas définir de préférence à l'intérieur de cet intervalle. En adoptant maintenant une approche probabiliste que l'on pourrait qualifier de « naïve », on supposera que les paramètres A et B sont représentés par des fonctions de densité de probabilité uniformes (*figure 3b*). Dans ce cas la somme de A et B est une fonction de densité de probabilité triangulaire de valeur moyenne 2 (ici égal au mode) et de valeurs minimum et maximum 1 et 3 respectivement. On constate donc que le choix *a priori* de distributions de probabilité aboutit à un résultat beaucoup plus « précis » que dans le cas possibiliste. Mais cette précision est artificielle car on a supposé l'équi-répartition des probabilités alors

que cette hypothèse n'était pas justifiée par l'information disponible. Dans un cadre probabiliste il aurait fallu travailler avec toutes les lois de probabilité envisageables et prendre l'enveloppe pour définir l'intervalle de probabilité sur A + B. On aboutirait au même résultat qu'avec l'approche possibiliste, mais de manière moins directe.

La méthode dite « hybride » (voir Guyonnet *et al.* 2002, 2003, Ferson et Ginzburg, 1995) pour combiner des informations de type probabiliste et possibiliste dans un même calcul du risque d'exposition aux polluants du sol, associe la méthode Monte Carlo d'échantillonnage aléatoire des distributions de probabilité au calcul en nombres flous. Son principe général est décrit ci-après. On suppose que le modèle M est fonction de paramètres dont certains peuvent être représentés par des distributions de probabi-

lité, tandis que d'autres sont jugés mieux représentés par des nombres flous (l'information étant incomplète ou imprécise). Pour l'illustration on considérera le calcul d'une dose d'élément trace métallique absorbée par une cible humaine. On a : Dose = $M(P_1, \dots, P_n, F_1, \dots, F_m)$ où P_1 à P_n sont les n variables pouvant être représentées par des distributions de probabilité tandis que F_1 à F_m sont les m variables représentées par des nombres flous. On suppose que les n variables P sont indépendantes. Le calcul hybride consiste à échantillonner de manière aléatoire les distributions de probabilité (figure 4a), puis à effectuer le calcul en nombres flous à partir des distributions de possibilité (figure 4b). On obtient une dose calculée sous la forme d'un nombre flou (figure 4c). En répétant la procédure ω fois (on ré-échantillonne les distributions de probabilité) on obtient ω doses floues (soit un nombre flou aléatoire ; Gil, 2001).

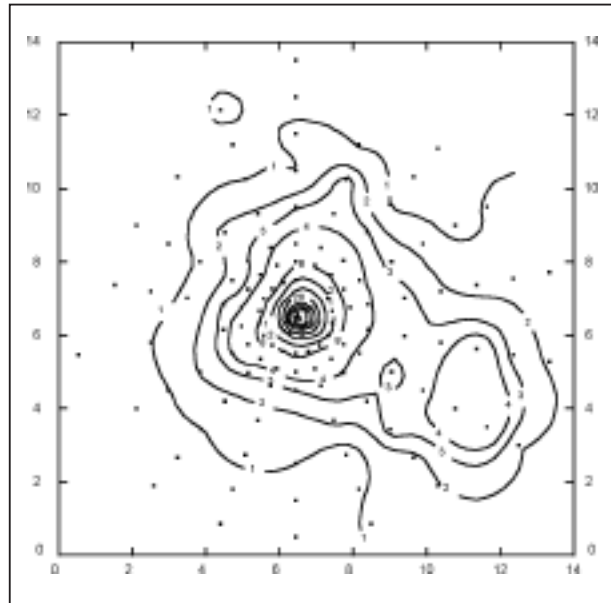
Pour ce qui concerne la comparaison du résultat avec un critère de tolérance (par exemple une dose jugée tolérable), Guyonnet *et al.* (2003) ont proposé d'extraire tout d'abord un résultat flou final, en sélectionnant les valeurs inf et min (pour chaque α -coupe) telles qu'il y ait 5 % de probabilité d'être respectivement en dessous ou au dessus. Ce résultat flou final est ensuite comparé au critère de tolérance (voir le chapitre suivant). Une autre alternative consiste à d'abord comparer chaque résultat flou obtenu au critère de tolérance et ensuite à éliminer les 5 % de valeurs extrêmes. A noter que différentes méthodes de post-traitement de la méthode hybride sont actuellement en cours d'évaluation dans le cadre d'une thèse (Baudrit *et al.*, 2003)

EXEMPLE D'APPLICATION

L'approche « hybride » est appliquée à l'estimation d'une dose de cadmium absorbée par une cible humaine par le biais de la consommation de légumes sur un site industriel. Le site sélectionné pour cette application est le site métallurgique de Noyelles Godault (Nord-Pas-de-Calais), dont les sols superficiels sont contaminés par des éléments traces métalliques liés aux fumées rejetées par une cheminée d'usine. A noter que l'objectif de cette application n'est en aucun cas une détermination définitive du risque que représente la présence de cadmium sur ce site, mais uniquement l'illustration de l'approche hybride. Aussi, pour les besoins de l'illustration, des hypothèses simplificatrices ont-elles été adoptées, comme cela est indiqué dans le texte. Par rapport aux objectifs du projet, le jeu de données disponible pour ce site présente un double avantage. D'une part, le nombre de mesures de teneurs totales en cadmium dans les sols superficiels est suffisant pour permettre une approche statistique de la répartition spatiale de la teneur. D'autre part, les mesures de l'INRA (Luttringer et de Cormis, 1979) fournissent des informations directes sur le transfert de cadmium depuis le sol vers la plante sur ce site. Pour une étude récente et plus détaillée de la répartition d'éléments métalliques dans des végétaux cultivés dans l'environnement de ce site, on se référera à Douay et Sterckeman

Figure 5 - Carte de répartition des teneurs en Cd (ppm) des sols superficiels. (triangle = cheminée d'usine, points = points de prélèvement, graduations en km)

Figure 5 - Map of the spatial distribution of Cd (ppm) in superficial soil. (triangle = chimney location, points = sampling points, graduation in km).



(2002).

Dans un premier temps, les outils de la géostatistique (Chilès et Delfiner, 1999) ont fourni une carte de la répartition de la teneur en cadmium dans les sols superficiels (tranche 0-10 cm). Un variogramme expérimental a été calé à l'aide d'un modèle de variogramme linéaire qui a ensuite permis d'interpoler une carte des teneurs par krigeage (figure 5). Cette méthode présente l'avantage de fournir l'écart-type de krigeage, c'est-à-dire un ordre de grandeur de la précision avec laquelle l'estimation est réalisée. Cet écart-type permet d'aborder l'incertitude sur la teneur dans un cadre probabiliste.

Les travaux de l'INRA relatifs au transfert sol-plante mettent en évidence la multiplicité des facteurs d'influence (spéciation du métal dans le sol, pH du sol, compétition entre métaux, type de plante, etc.). Il est par conséquent préférable de disposer de mesures directes de ce transfert plutôt que de faire référence par exemple à des facteurs de bioconcentration génériques. Les mesures de Luttringer et de Cormis (1979) effectuées notamment sur des poireaux, bien que peu nombreuses et ayant été réalisées à des dates différentes de la plupart des prélèvements de sols représentés en figure 5, fournissent néanmoins une information directe sur ce transfert, information qui est spécifique au site en question. Etant peu nombreuses, ces données ne se prêtent pas à une analyse statistique, mais permettent par contre une analyse corrélatrice floue.

A partir des valeurs mesurées indiquées dans la figure 6, on peut

tenter d'établir une corrélation entre la teneur en cadmium dans les poireaux et celle du sol, corrélation spécifique à ce site :

$$Cd_{pl} = Or + (As - Or) (1 - \exp(-k Cd_s))$$

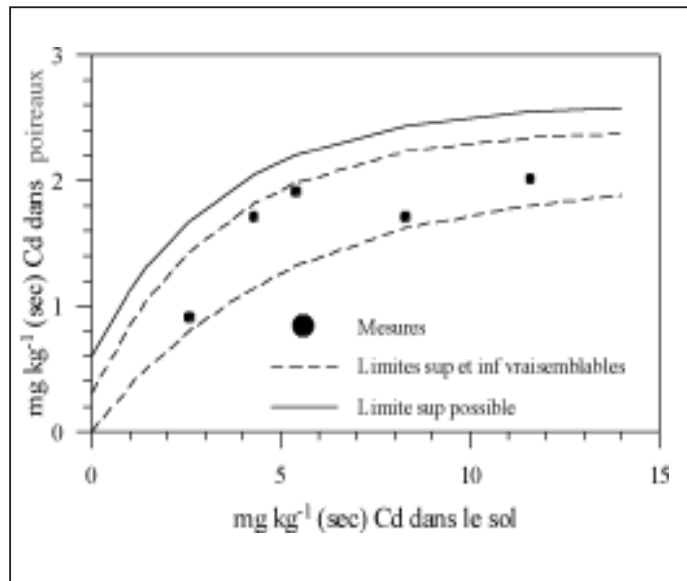
où : Cd_{pl} = concentration en cadmium dans la partie comestible de la plante ($mg\ Cd\ kg^{-1}$ plante sèche) ; Cd_s = concentration en cadmium mesurée dans le sol ($mg\ Cd\ kg^{-1}$ sol sec) ; Or = valeur de Cd_{pl} à l'origine ($Cd_s = 0$) ; As = Asymptote (Cd_{pl} pour des valeurs élevées de Cd_s) ; k = paramètre contrôlant la forme de la courbure dans la *figure 6*. A noter que l'expression proposée est un exemple mais qu'il peut en exister d'autres (voir Douay et Sterckemann, 2002).

Cette équation permet de cadrer les valeurs mesurées comme indiqué par les courbes en pointillé de la *figure 6*. Ces courbes en pointillé suggèrent qu'une mesure de la concentration en cadmium dans des poireaux cultivés sur ce site donnera « vraisemblablement » une valeur située entre ces deux courbes. Par contre, l'expert agronome peut juger que compte tenu du faible nombre de mesures, il ne peut exclure des valeurs plus élevées et de proposer par conséquent une courbe haute jugée possible et représentée par la courbe en trait plein de la *figure 6*. Tandis qu'on peut débattre sur la position précise de cette courbe haute, on notera que ce mode de représentation est au moins cohérent avec les quelques valeurs mesurées disponibles (5 valeurs) et qu'il peut être facilement adapté pour tenir compte de l'opinion de l'expert. Les paramètres Or , As et k de l'équation de corrélation présentée plus haut sont donc représentés par des nombres flous pour tenir compte de l'imprécision sur la corrélation, tandis que les valeurs de Cd_s sont représentées par des fonctions de distribution de probabilité (lois normales) qui découlent de l'analyse géostatistique. A partir de la *figure 6*, on identifie les nombres flous des paramètres qui interviennent dans l'équation de corrélation et qui sont représentés en *figure 7* ainsi qu'un autre paramètre décrit plus bas. Les nombres flous des *figures 7a à 7c* se déduisent directement de la *figure 6*. Par exemple, le nombre flou en *figure 7a* signifie que la concentration en cadmium dans la plante à l'origine (valeur de concentration nulle dans le sol) est vraisemblablement située entre 0 et $0,3\ mg\ kg^{-1}$, mais qu'on n'exclut pas une valeur haute possible de $0,6\ mg\ kg^{-1}$. A noter que les valeurs non nulles pourraient traduire une concentration dans la plante indépendante de celle du sol, liée par exemple à un transfert par le biais de poussières déposées sur les feuilles. Le nombre flou en *figure 7b* signifie que la valeur maximale de cadmium dans la plante (asymptote) est vraisemblablement située entre 2 et $2,4\ mg\ kg^{-1}$, tandis qu'une valeur maximale de $2,6\ mg\ kg^{-1}$ est considérée comme étant possible. L'information sur le paramètre k (*figure 7c*) est représentée par un simple intervalle min-max ($0,2 - 0,3$), les valeurs étant obtenues par ajustement de la courbe de corrélation aux données mesurées.

Le calcul de dose est effectué à partir de l'équation suivante

Figure 6 - Concentrations en Cd dans les poireaux en fonction de la teneur totale en Cd du sol superficiel et corrélation floue

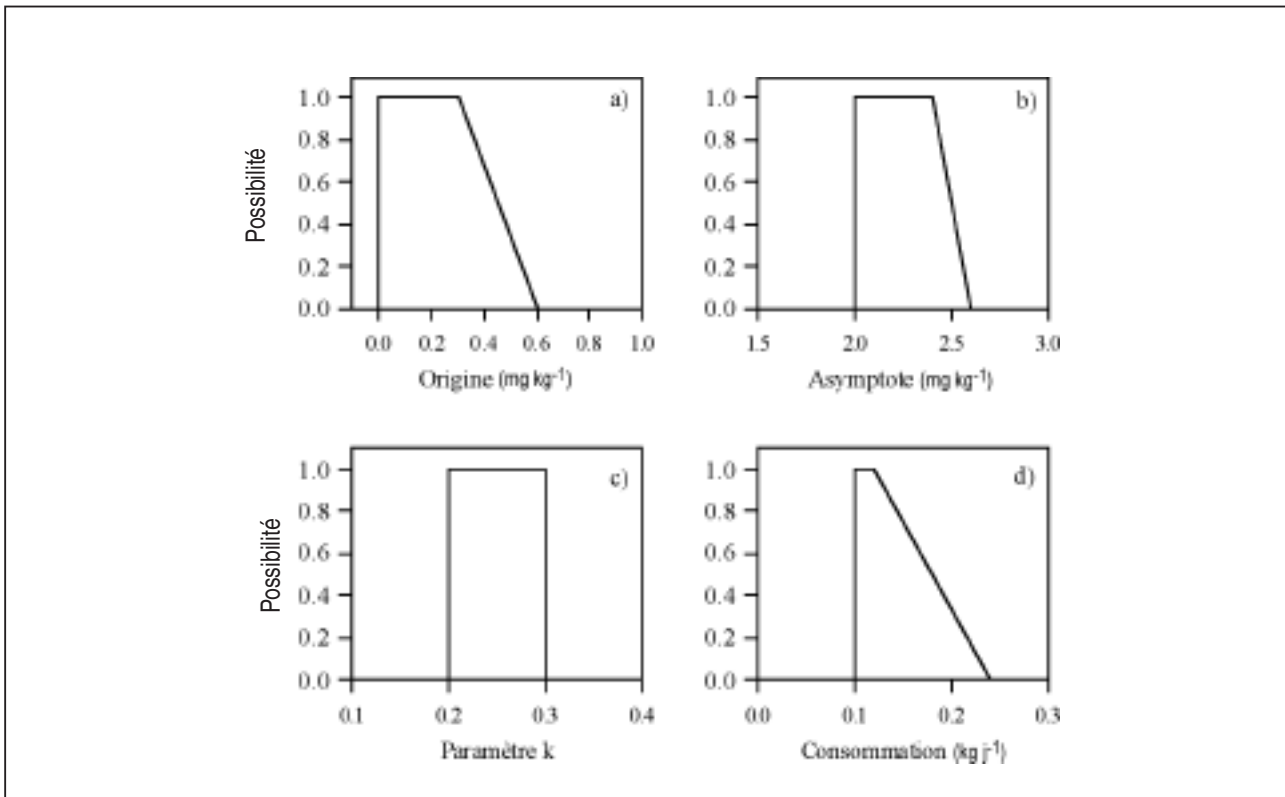
Figure 6 - Cd concentrations in leeks as a function of total soil Cd concentration and fuzzy correlation



$$Dose = \frac{Cd_{pl} \cdot 1000 \cdot Con \cdot TMS}{PC}$$

ou : Dose = dose absorbée par la cible humaine ($mg\ Cd$ par jour et par kg de poids corporel) ; Con = consommation journalière de poireaux (kg par jour) ; TMS = teneur en matière sèche des poireaux (pourcentage pondéral) ; PC = Poids corporel de la cible humaine (kg).

La teneur en matière sèche des poireaux tient compte du fait que la consommation journalière est définie par rapport à la masse totale humide, tandis que la concentration en cadmium dans les poireaux est définie par rapport à la masse sèche. La concentration en cadmium dans les poireaux (Cd_{pl}) est obtenue à l'aide de l'équation de corrélation floue. Il est supposé (hypothèse simplificatrice) que les poireaux sont représentatifs des légumes pour ce qui concerne l'exposition de l'homme au cadmium par le biais de la consommation de légumes. A noter que cette hypothèse n'est ni exacte ni nécessairement majorante, les mesures de Douay et Sterckeman (2002) suggérant en effet que le transfert du sol vers la salade serait supérieur à celui observé pour les poireaux. La consommation journalière de légumes (*figure 7d*) et la teneur en matière sèche sont sélectionnées sur la base de valeurs proposées dans INERIS (1999). La consommation vraisemblable est prise entre 100 et $120\ g/jour$, tandis qu'une valeur haute possible est prise comme étant égale à deux fois la valeur haute ($240\ g/jour$). La teneur en matière sèche (TMS) et le poids corporel sont considérés comme étant constants : 15 % et $70\ kg$ respectivement. Un exemple de résultat du calcul hybride pour $Cd_s = 7.97\ ppm$ et écart-type = 1.15, est présenté en *figure 8*.

Figure 7 - Nombres flous pour plusieurs paramètres**Figure 7** - Fuzzy numbers for several parameters

= 7.97 ppm et écart-type = 1.15, est présenté en *figure 8*.

L'étape suivante consiste à examiner l'acceptabilité des doses calculées. Cette acceptabilité est considérée en faisant référence à la dose journalière maximale admissible préconisée par l'OMS (1994) ($1 \text{ mg j}^{-1} \text{ kg}^{-1}$). A l'aide de l'indicateur de possibilité (noté Π ; Dubois et Prade, 1988), on obtient une mesure de la véracité de la proposition "la dose calculée dépasse la dose jugée tolérable". Cet indicateur est illustré dans la *figure 9*. Tant que la dose calculée floue est en deçà de la dose de référence (notée D_0), la possibilité de dépassement est jugée nulle ($\Pi = 0$; *figure 9a*). Dès que le nombre flou recoupe le critère, la possibilité de dépassement croît de 0 à 1 (*figures 9b et c*). A noter qu'un autre indicateur de la véracité de la proposition; la mesure dite de « nécessité » (voir Dubois et Prade, 1988), traduit le degré de certitude que la dose calculée dépasse la dose de référence. Cet indicateur s'appliquerait dans le cas de la *figure 9d*, où le noyau de la dose floue calculée (les valeurs dont la possibilité est l'unité) passe de l'autre côté de la dose de référence. En appliquant ce critère à tous les points d'un maillage couvrant le site, on arrive à une carte de la possibilité de dépassement de la dose jugée tolérable (*figure 10*). On distingue des zones où il y a une forte possibilité de dépas-

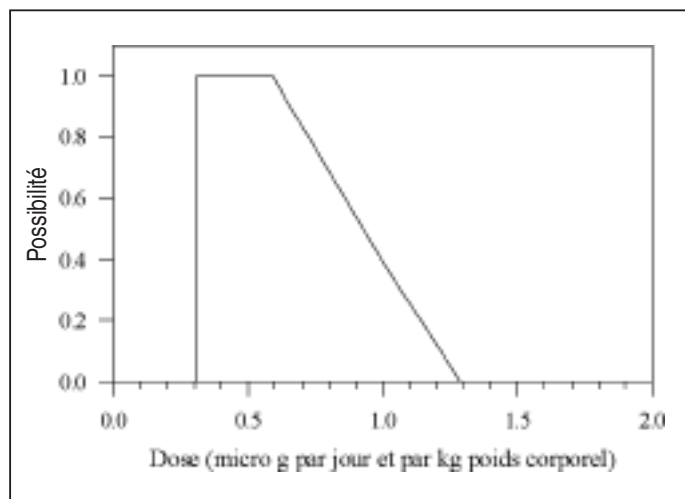
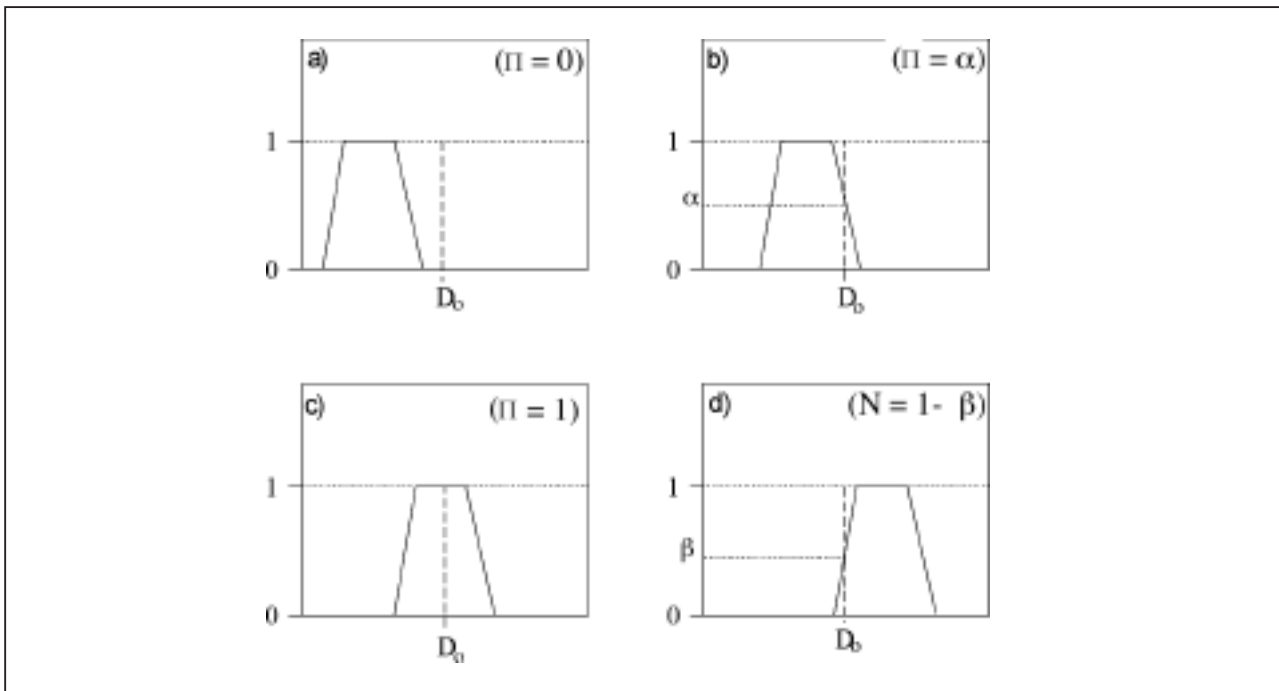
Figure 8 - Résultat de calcul de dose pour $\text{Cd}_s = 7.97 \text{ ppm}$ et écart-type = 1.15**Figure 8** - Result of the dose calculation for $\text{Cd}_s = 7.97 \text{ ppm}$ and standard deviation = 1.15

Figure 9 - Degré de « possibilité » (Π) que la dose calculée dépasse la dose de référence D_0 (a) - (c). Degré de « nécessité » (N) (d).
Figure 9 - Degree of « possibility » (Π) that the calculated dose should exceed a reference dose D_0 (a) - (c). Degree of « necessity » (N) (d)

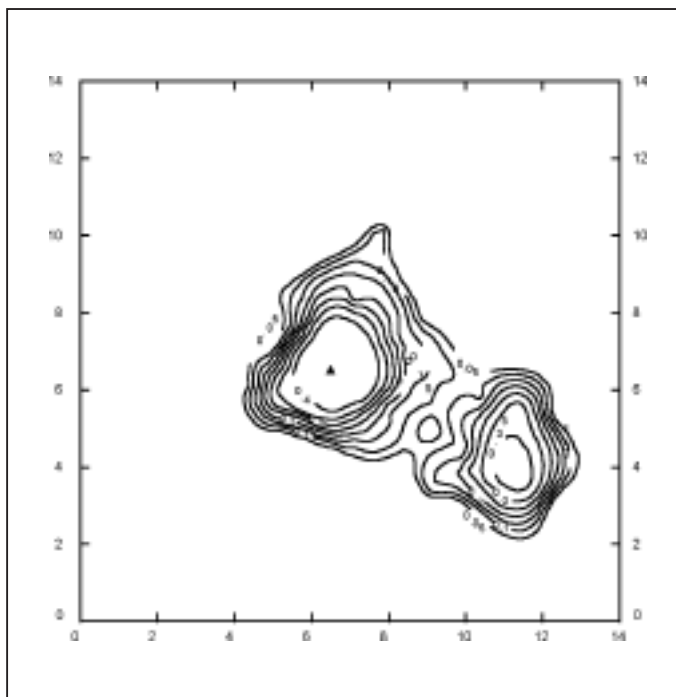


sement (les zones proches de la cheminée d'usine au centre) et des zones où cette possibilité est faible voire nulle.

Afin d'illustrer le caractère non sécuritaire de l'approche Monte Carlo classique dans ce contexte, on a effectué le même calcul en supposant que les nombres flous définis précédemment étaient en fait des fonctions de densité de probabilité (en conservant exactement les mêmes formes de fonctions). On a sélectionné une teneur en cadmium du sol élevée qui correspond à un prélèvement effectué dans le voisinage immédiat de la cheminée d'usine. Le calcul Monte Carlo aboutit aux résultats de la *figure 11*, qui suggèrent que le risque serait acceptable même dans le voisinage immédiat de la cheminée (la dose calculée a environ 95 % de chances d'être inférieure à la dose admissible). Mais cette conclusion est entièrement imputable au choix *a priori* d'une représentation de l'incertitude par des distributions de probabilité uniques pour chaque paramètre, ainsi qu'à l'hypothèse d'indépendance entre les paramètres. Il est rappelé que la probabilité que deux événements indépendants se réalisent est le produit de leurs probabilités respectives. Par conséquent lors de simulations Monte Carlo, les scénarios associant plusieurs événements peu probables ont d'autant moins de chances d'être réalisés. L'approche floue, par contre, s'intéressant aux possibilités et non aux probabilités, ne répercute pas de manière multiplicative les degrés de possibilité respectifs des différentes variables sur celui du scénario tout entier.

Figure 10 - Carte de la « possibilité » de dépassement de la dose jugée tolérable ($1 \mu\text{g j}^{-1} \text{kg}^{-1}$).

Figure 10 - Map of the « possibility » of exceeding the tolerable dose ($1 \mu\text{g j}^{-1} \text{kg}^{-1}$).



CONCLUSIONS ET PERSPECTIVES

Cet article propose une approche alternative pour prendre en compte l'incertitude relative aux paramètres des modèles communément utilisés en évaluation des risques d'exposition aux polluants du sol. L'objectif de cette approche est de permettre une représentation de cette incertitude qui soit cohérente avec l'information réellement disponible. Un nombre croissant de chercheurs dans le domaine de l'évaluation des risques prennent conscience du fait que des vecteurs d'information différents sont nécessaires pour représenter la variabilité (le hasard) et l'ignorance (l'imprécision). Lorsque des distributions de probabilité uniques sont simplement supposées (souvent pour des raisons de confort méthodologique), l'investigateur biaise le résultat de l'évaluation et introduit un degré de confiance dans ce résultat qui n'est pas cohérent avec l'information disponible.

La méthode hybride évite cet écueil en permettant de combiner probabilités et possibilités. Une carte comme celle de la *figure 10* peut servir d'outil d'aide à la décision en matière de gestion d'un site pollué. A noter toutefois qu'il appartient à l'autorité sanitaire compétente de fixer une limite pour le niveau de possibilité de dépassement de la dose tolérable. Des valeurs de 0.1 ou 0.2, par exemple, se traduiraient par une restriction sur la culture de légumes destinée à la consommation humaine, sur une superficie d'environ 20 kilomètres carrés. Exiger une possibilité de zéro peut être trop strict et se traduire par des restrictions d'usage sur des superficies de terrain trop importantes, ou encore par des coûts de réhabilitation trop élevés. Comme le calcul en nombres flous est plus majorant que le calcul Monte Carlo, les critères de tolérance en terme de possibilités ne devraient pas être aussi stricts qu'en terme de probabilités résultant d'une analyse Bayésienne.

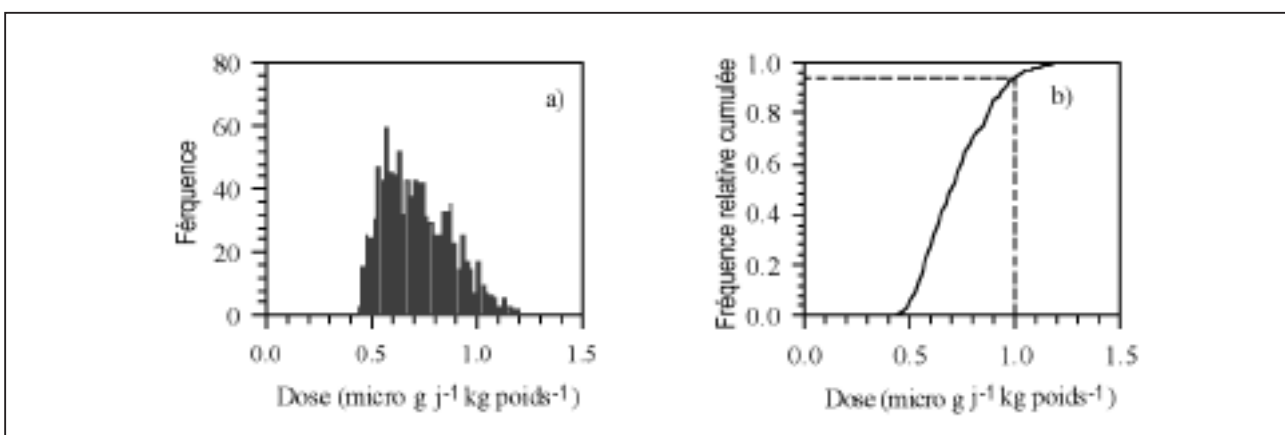
Malgré ses atouts on peut citer deux inconvénients de la méthode proposée. Le premier est d'ordre opérationnel. La méthode « hybride » combine deux techniques qui sont gourmandes en

temps de calcul: le calcul Monte Carlo et le calcul en nombres flous. Le modèle utilisé dans l'exemple traité était simple. Dans le cas de modèles plus complexes, analytiques ou numériques, il est à craindre qu'une application « brutale » de la méthode ne soit limitée par la puissance des machines disponibles. L'autre inconvénient que l'on peut citer est d'ordre sociologique. Les décideurs en matière de sols pollués et de risques en général commencent seulement à prendre en compte la notion de « probabilité d'apparition de dommages ». Celle de « possibilité d'apparition de dommages » pourrait rencontrer une certaine réticence de leur part, ou de la part du public ou des médias. Une restitution des résultats possibilistes en terme de valeurs hautes et basses de probabilités (Dubois et Prade, 1992) pourrait être une solution naturelle à ce problème. A noter que des travaux sont actuellement en cours visant à identifier comment intégrer la perception sociale du risque dans la manière de le représenter.

Pour conclure il convient de souligner que la méthode hybride proposée est une alternative parmi d'autres et que le travail présenté ici représente une image partielle de travaux de recherche en cours. Une thèse, cofinancée par le BRGM, l'INERIS et l'IRSN et avec également la participation de l'ADEME, est actuellement en cours à l'Université Paul Sabatier de Toulouse. Il a déjà été mentionné que dans le cadre de cette thèse, le post-traitement de la méthode hybride était évaluée. Par ailleurs, d'autres méthodes de propagation de l'incertitude sont à l'étude et en particulier une méthode qui fait appel à la théorie des fonctions de croyance de Dempster-Shafer (Shafer, 1976), théorie qui fournit un cadre commun à la fois aux probabilités et aux possibilités. Cette approche permettrait également d'appréhender un aspect important de la propagation de l'incertitude à savoir les éventuelles dépendances qui peuvent exister entre les paramètres.

Figure 11 - Résultat d'une simulation Monte Carlo (1000 itérations) (a) et comparaison du résultat avec le critère de tolérance (b).

Figure 11 - Result of a Monte Carlo simulation (1000 iterations) (a) and comparison between the result and the acceptance criterion (b).



BIBLIOGRAPHIE

- Baize D. et Tercé M., (Coord.) 2002 - Un point sur : les éléments traces métalliques dans les sols. Approches fonctionnelles et spatiales. Publication INRA, Paris.
- Baudrit C., Dubois D., Guyonnet D. et Fargier H., 2003 - Joint treatment of imprecision and randomness in uncertainty propagation. Dans: *Information Processing in Management of Uncertainty in Knowledge-Based Systems*, 4-9 Juillet 2003, Perugia, Italie.
- Bonano E., Hora S., Keeney R. et von Winterfeldt D., 1990 - Elicitation and use of expert judgement in performance assessments for high-level radioactive waste repositories. Sandia National Laboratories Report SAND89-1821, Albuquerque, New Mexico.
- Cazemier D., 1999 - Utilisation de l'information incertaine dérivée d'une base de données sols. Thèse de l'Ecole Nationale d'Agronomie, Montpellier (France). 170 p.
- Chilès J.-P. et Delfiner P., 1999 - *Geostatistics : Modeling Spatial Uncertainty*. Wiley, New York.
- Cullen A.C. et Frey H.C., 1999 - Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs. New York Plenum Press, 333 pp.
- Douay F. et Sterckeman T., 2002 - Teneurs en Pb, Cd et Zn dans les végétaux cultivés aux alentours d'usines métallurgiques. Dans: *Un point sur : les éléments traces métalliques dans les sols. Approches fonctionnelles et spatiales*. D. Baize et M. Tercé (Coord.). Publication INRA, Paris.
- Dubois D., Nguyen H.T., Prade H., 2000 - Possibility theory, probability and fuzzy sets : misunderstandings, bridges and gaps. Dans: *Fundamentals of fuzzy sets*. Dubois, D. et Prade, H. (Eds.). Kluwer Academic, Boston, USA, pp.343-438.
- Dubois D. and Prade H., 1992 - When upper probabilities are possibility measures. *Fuzzy Sets and Systems*, 49, 95-74.
- Dubois D. et Prade H., 1988 - Possibility theory. New York Plenum Press, 263 pp.
- Ferson D., 1996 - What Monte Carlo Methods Cannot Do. *Human and Environmental Risk Assessment*, 2, 990-1007.
- Ferson S., Ginzburg L., 1996 - Different methods are needed to propagate ignorance and variability. *Reliability and System Safety*, 54, 133-144.
- Ferson S., Ginzburg L., 1995 - Hybrid Arithmetic. Dans: *ISUMA-NAFIPS'95*, IEEE Computer Society Press, Los Alamitos, Californie, 619-613.
- Freissinet C., Erlich M., Vauclin M., 1998 - A fuzzy logic-based approach to assess imprecision of soil water contamination modelling. *Soil & Tillage Research*, 47, 1-17.
- Gil M.A., 2001 - Fuzzy Random Variables. Edition spéciale de "Information Sciences", 133, nos. 1-2.
- Guyonnet D., Bourguine B., Chilès J.-P., 2002 - Prise en compte de l'incertitude dans l'évaluation du risque d'exposition aux polluants du sol. Programme GES-SOL. Rapport final. Rapport BRGM RP-51683-FR.
- Guyonnet D., Bourguine B., Dubois D., Fargier H., Côme B., Chilès J.-P., 2003 - Hybrid approach for addressing uncertainty in risk assessments. *Journal of Environmental Engineering*, 129, 68-78.
- INERIS, 1999 - Méthode de calcul des valeurs de constat d'impact dans les sols. Rapport INERIS, Avril 1999, Verneuil-en Halatte, France.
- Kaufmann, A., Gupta, M., 1985 - Introduction to fuzzy arithmetic: theory and applications. Van Nostrand Reinhold, New York, 1985.
- Lagacherie P., Cazemier D., Martin-Clouaire R., Wassenaar, T., 2000 - A spatial approach using imprecise soil data for modelling crop yields over vast areas. *Agriculture, Ecosystems and Environment*, 81 :5-16.
- Labieniec P. Dzombak D., Siegrist R., 1997 - Evaluation of uncertainty in a site-specific risk assessment. *Journal of Environmental Engineering*, 123(3), 234-243.
- Luttringer M. et de Cormis L., 1979 - La pollution par les métaux lourds à Noyelles-Godault et ses environs (Pas de Calais). Rapport de l'Institut National de la Recherche Agronomique, Montfavet (France), 12 pp.
- MEDD, 2000 - Gestion des sites (potentiellement) pollués. Editions BRGM, Orléans.
- OM, 1994 - Directives de qualité pour l'eau de boisson. Volume 1 : Recommandations. 2^e édition. Organisation Mondiale de la Santé, Genève Suisse.
- Shackle G., 1961 - *Decision Order and Time in Human Affairs*. Cambridge University Press, UK.
- Shafer G., 1976 - *A mathematical theory of evidence*. Princeton University Press.
- Vose D., 1996 - *Quantitative risk analysis - A guide to Monte-Carlo simulation modeling*. Wiley, New York.
- Wonneberger S., Kisting S., Deckert A., 1995 - Unbiased guess, a concept to cope with fuzzy and random parameters ? Rapport de la Commission Européenne EUR 16199 EN. Bureau des Publications Officielles des Communautés Européennes, Luxembourg, 42 pp.
- Zadeh L., 1965 - Fuzzy Sets. *Information and Control*, 8, 338-353.
- Zadeh L., 1978 - Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1, 3-28.

