



Analyse critique d'un logiciel d'évaluation des coûts
pour la dépollution de sols

Société de Calcul Mathématique SA

Juillet 2008

Le logiciel "RACER" est destiné à évaluer les coûts d'une dépollution des sols. Nos remarques sont les suivantes :

1. Commentaires généraux

Tout comme la plupart des logiciels scientifiques supposés servir à la prise de décision, celui-ci n'est pas approprié à cet objet. En effet, il prétend décrire la réalité de manière fine dans les détails, mais ignore les variantes majeures. Du fait de l'impression de sécurité qu'il procure (mise en œuvre de techniques mathématiques à base de probabilités), il y a là un véritable danger.

Il n'y a pas d'interface de haut niveau (prenant en compte la géographie), et de ce fait l'utilisateur est obligé d'entrer, un par un, les paramètres liés à chaque site pris en compte.

Si un événement imprévu survient (par exemple découverte d'un nouveau polluant, retards, fouilles archéologiques, grèves, accidents, etc.), les indications fournies par le logiciel deviennent absolument fausses. Or, dans la réalité, il y a toujours des impondérables. Il faudra donc, en ce cas, reprogrammer entièrement le logiciel.

2. Sur l'aspect probabiliste

Les incertitudes sont supposées prises en compte sous forme de lois de probabilités, que l'on peut mettre en place pour chaque paramètre. Or cet aspect est également tout à fait trompeur :

- Les lois de probabilités choisies sont absolument factices (on met ce qu'on veut) ;
- Les divers paramètres sont considérés comme indépendants, ce qui n'est absolument pas le cas dans la réalité : la découverte d'une pollution plus importante, une augmentation de tarifs, l'inflation, tout ceci affecte de nombreux paramètres simultanément.

L'hypothèse d'indépendance conduit à une loi de probabilité sur le coût total qui est absolument fautive : les extrêmes paraissent avoir une très faible probabilité. En présence d'indépendance, la loi de probabilité paraît très concentrée autour de sa valeur moyenne.

Il n'est pas même assuré que l'intervalle min - max affiché pour le résultat global soit correct. Il ne l'est pas si le nombre de paramètres est élevé. Pour faire comprendre ceci, imaginons qu'il y ait 15 paramètres, prenant chacun la valeur 0 ou 1 (deux valeurs seulement), et que le coût maximal soit obtenu lorsque tous valent 1. La probabilité que tous les paramètres valent 1 est 2^{-15} et 5000 runs (comme c'était le cas dans la démonstration que nous avons vue) ne mettront jamais ce cas en évidence.

Pour mettre en évidence les pires cas, il est impératif d'abandonner l'approche aléatoire : il faut prendre soin de mettre individuellement, et à la main, chaque paramètre sur sa pire valeur.