Available online at www.sciencedirect.com
ScienceDirect
Journal of Environmental Sciences 19(2007) 1272–1280
JOURNAL OF ENVIRONMENTAL SCIENCES
ISSN 1001-0742
CN 11-2629/X
www.jesc.ac.cn

Sensitivity of key factors and uncertainties in health risk assessment of benzene pollutant

LIU Zhi-quan, ZHANG Ying-hua*, LI Guang-he, ZHANG Xu

Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China. E-mail: liu.zhiquan@sepa.gov.cn

Received 21 December 2006; Revised 18 March 2007; Accepted 15 May 2007

Abstract

Predicting long-term potential human health risks from contaminants in the multimedia environment requires the use of models. However, there is uncertainty associated with these predictions of many parameters which can be represented by ranges or probability distributions rather than single value. Based on a case study with information from an actual site contaminated with benzene, this study describes the application of MMSOILS model to predict health risk and distributions of those predictions generated using Monte Carlo techniques. A sensitivity analysis was performed to evaluate which of the random variables are most important in producing the predicted distributions of health risks. The sensitivity analysis shows that the predicted distributions can be accurately reproduced using a small subset of the random variables. The practical implication of this analysis is the ability to distinguish between important versus unimportant random variables in terms of their sensitivity to selected endpoints. This directly translates into a reduction in data collection and modeling effort. It was demonstrated that how correlation coefficient could be used to evaluate contributions to overall uncertainty from each parameter. The integrated uncertainty analysis shows that although drinking groundwater risk is similar with inhalation air risk, uncertainties of total risk come dominantly from drinking groundwater route. Most percent of the variance of total risk comes from four random variables.

Key words: Monte Carlo; sensitivity; uncertainty; MMSOILS models; risk assessment

Introduction

The assessment of risk in complex systems is almost inevitably affected by uncertainties. In the case of assessing potential health risks from different exposure routes, there are many contributing factors. The importance of adequately characterizing variability and uncertainty in fate, transport, exposure, and dose-response assessments for human health and ecological risk assessments has been emphasized in several U. S. Environmental Protection Agency documents and activities (USEPA, 1986, 1992, 1996c, 1997), which also recommend that a sensitivity analysis be used to help determine factors of importance to the assessment. Although the USEPA's guidance does not dictate an approach to sensitivity, examples provided in the guidelines including use of rank correlation coefficients, standardized rank regression coefficients, and scatter plots. The importance of systematically distinguishing and evaluating various kinds of uncertainties was too identified by some researchers (Bennett *et al*., 1998; Fewtrell *et al*., 2001; Havelaar, 1998).

There are many articles on performing sensitivity and uncertainty analyses for environmental numerical simulation based on some actual case study of the transport of contaminants (Hamby, 1994; James and Oldenburg, 1997; Helton *et al*., 1995; Pelmulder *et al*., 1996; Manteufel, 1996). Approaches to sensitivity and uncertainty analyses cited in these works include: differential sensitivity analysis, one-at-a-time sensitivity measures, sensitivity indices, qualitative analysis, Pearson's ρ and Spearman's ρ, regression and standardized regression techniques, variance analysis, tests involving segmented input distributions, and the propagation of uncertainty owing to variance in major parameter values in model. However, no consensus exists as to a "best" method. Rather, it is indicated that the method and the objectives of the analysis should be jointly evaluated (Mills *et al*., 1999).

An approach was presented in this study, using the multimedia models of MMSOILS (USEPA, 1996b), to address the question posed–that is, how many variables need to be treated as random and the parameter uncertainty in the human health risk model for a particular case study. The first objective of this paper was to implement an approach to sensitivity analysis, within a multimedia setting, to identify the random variables most important in contributing to the predicted variability of selected model endpoints, such as lifetime potential risk owing to exposure to benzene in drinking groundwater. The second objective was to compare the sensitivity analysis results

Project supported by the National Natural Science Foundation of China (No. 40572141). *Corresponding author. E-mail: yinghuazhang8@gmail.com.

for the models evaluated to determine their degree of similarity, despite differences of exposure route. The final objective was to demonstrate model integration including uncertainty analysis. Parameters used in health risk assessment were uncertain and thus were represented by distributions. By considering the combined uncertainties attributable to both the transport and exposure/risk input parameters, the authors obtained total human health risk as a distribution of values. The integration of uncertainty analyses also allows comparison of the relative importance of uncertainty arising from drinking groundwater route, inhalational air route and every random variable.

1 Methods

The MMSOILS model is Multimedia Contaminant Fate, Transport, and Exposure model for estimating exposure and risk resulting from the multimedia release of contaminants from hazardous waste units. The model addresses the transport of a chemical in groundwater, surface water, soil erosion, the atmosphere, and accumulation in the foodchain. The human exposure pathways considered in the model include: soil ingestion, air inhalation of volatiles and particulates, dermal contact, ingestion of drinking water, consumption of fish, consumption of plants grown in contaminated soil, and consumption of animals grazing on contaminated pasture. For multimedia exposures, the methodology provides estimates of human exposure through individual pathways and combined exposure through all pathways considered. The risk associated with the total exposure dose is calculated based on chemical-specific toxicity data. With a Monte Carlo simulation capability MMSOILS model allows the user to specify statistical distributions and associated characteristics (e.g., mean, standard deviation) of model input parameters, thus, enabling the user to quantify uncertainty in model estimates as a function of input parameter uncertainty (USEPA, 1996b).

The USEPA has made clear that there are a number of situations in which a Monte Carlo analysis can be useful. For example, a Monte Carlo may be useful in performing risk assessments that addresses sensitivity and uncertainty analysis issues. MMSOILS uses the Monte Carlo method. Given a set of deterministic values for each of the input variables, X_1, X_2 ... X_n, the composite model computes results such as exposure concentration C, i.e.:

$$C = \text{function}(X_1, X_2, X_3 ... X_n) \qquad (1)$$

Application of the Monte Carlo simulation procedure requires that at least one of the input variables, X_1 ... X_n, be uncertain, with the uncertainty represented by a cumulative probability distribution (USEPA, 1996b). The method involves the repeated generation of pseudo-random values of the uncertain input variable(s). The pseudo-random values are drawn from the specified distribution and are within the range of any imposed bounds. Then the model is applied, using these values, to generate a series of model responses. These responses are statistically analyzed to yield the cumulative probability distribution of the output. Optionally, for each Monte Carlo realization, a vector of information unique to that realization can be saved. By doing this, all the information generated by the model during all Monte Carlo simulations is preserved, and can be used for post-processing purposes, such as for sensitivity and uncertainty analysis.

2 Case study and model descriptions

2.1 Case study

The case study modeled in this article is based on an actual site in which groundwater has become the sole source of urban and industrial water supplies due to the lack of surface water. However, a petrochemical factory was built in 1984 on the recharge area which supplies the drinking water well field. The potential health risk of the drinking groundwater well field is of particular concern with approximately 3700 m away from the factory because well water is pumped for domestic direct uses. The foundations of oil tanks and waste water pipes were placed directly on the aquifer. Because of the easily corroded pipes, leakage from pipe lines and oil tanks has resulted in petroleum contamination of the soils. The thickness of the contaminated soils is about 3 m. With precipitation recharge into groundwater the petroleum byproducts continuously leached from the contaminated soils are threatening the drinking groundwater. The mean infiltration through the soil to the water table is 4.12 cm/a. The sources of benzene pollutant in daily life of human being and environment were drinking groundwater and air inhalation resulting from contaminant soil. High and significant acute myeloid leukemia risks with positive benzene dose response relationships were identified across published studies and especially in more highly exposed workers in benzene-related industries. Although risks for chronic myeloid leukemia and acute lymphocytic leukemia are sparse and inconclusive, risks for chronic lymphocytic leukemia tended to show possible dose response relationships (Robert *et al.*, 2005). Because benzene with 138.64 mg/kg in soil is a typical carcinogenic contaminant of the petroleum byproducts, it was chosen as the chemical of concern for this study. The contamination source areas and location of the receptor wells are shown in Fig.1. For the multimedia scenario examined, it is assumed to be initially

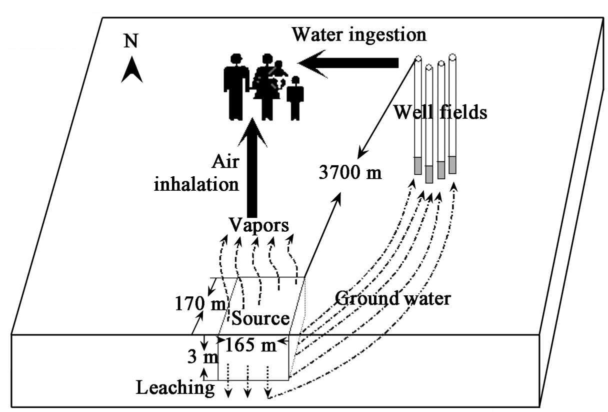

Fig. 1 Environmental setting and exposure routes of case study

present in the contaminated soils but not initially present in any other media (i.e., at the beginning of the problem all other media are uncontaminated). Finally, it is important to note that these are model projections assuming no future remedial actions and no biodegradation of benzene.

The variables that are treated as random for simulations are shown in Table 1. The variables listed are grouped by source geometry, contaminant-related variables, infiltration and leaching data, groundwater data, and exposure factor data. Food ingestion or dermal contact scenarios were not considered because the exposure routes just were drinking groundwater from the well field and inhalation air working on the contaminated soils in the study case. For each variable, the distribution type and the parameters of the distribution are provided (minimum, maximum, mean, and standard deviation). Cross-correlations between variables are not considered, and the random variables are deliberately chosen to eliminate any cross-correlations. The distributions for the random variables are chosen to reflect a site that is generally well characterized; that is, the distributions are chosen to have narrow ranges. The exposure factor data come from EPA's statistical data (U S EPA, 1996a). The remaining data are deterministic and are the same as previously used to benchmark three multimedia models in a deterministic setting (Mills *et al*., 1997).

2.2 MMSOILS model descriptions

The model uses analytical or quasi-analytical techniques to solve transport equations, and simulate contaminant migration from a source to the unsaturated zone, saturated zone, surface soil, atmosphere, and surface water. The four basic functions of the multimedia methodology are developed as a tool to estimate exposures and health risks associated with the release and subsequent fate and transport of chemicals from contaminated soils and various hazardous waste sites. They are: (1) based on chemical properties and land use at the site, estimate the chemical release rate from the soil into each environmental media; (2) based on the chemical release rate and the proximity to exposed populations, estimate the chemical concentration at exposure points in each environmental media considered; (3) based on the chemical concentration at exposure points and assumptions regarding human intake levels, estimate the human exposure through inhalation, ingestion and absorption; 4) based on the estimated human exposures at exposure points, estimate the potential health risk based on toxicity data for the specific chemical (USEPA, 1996a). Although the model predicts risks from exposure to chemicals only and assumes the well used by the receptor is screened near the surface of the aquifer and concentrations in the water withdrawn are reflective of water in the aquifer at that location.

Often the most basic parameters, such as contaminant concentration in soil, vary significantly over a given site and the distribution may be poorly understood. These uncertainties, coupled with approximations that were used to streamline the modeling process, lead to the results that may differ from reality by orders of magnitude. To evaluate these uncertainties, MMSOILS utilizes the Monte Carlo simulation method. By representing input parameters in terms of a probability distribution rather than a single deterministic value, this method allows quantitative estimation of the uncertainty in the predicted concentrations and human health risks. The Monte Carlo simulator can be used to quantify the uncertainty in the three broad

Table 1 Random variables for Monte Carlo uncertainty analysis scenario for benzene contaminant

Variable	Distribution type	Mean	Minimum	Maximum	Standard deviation
Source geometry					
W: source width (m)	Uniform	165	27	303	79.7
L: source length (m)	Uniform	170	85	255	49
Contaminant-related variables					
K_d: partition coefficient for contaminated soil (ml/g)	Uniform	1.37	0	2.74	0.79
K_h: henry's law constant (atm m^3/mol)	Truncated normal	5.55×10^{-3}	4.7×10^{-3}	6.4×10^{-3}	4.20×10^{-4}
Infiltration and leaching					
I: annual recharge to groundwater (cm/a)	Truncated normal	4.12	0.21	8.24	2.04
F_c^1: field capacity for 1st unsaturated zone (cm^3/cm^3)	Truncated normal	0.27	0.198	0.342	0.036
F_c^2: field capacity for 2nd unsaturated zone (cm^3/cm^3)	Truncated normal	0.091	0.07	0.243	0.01
K_S^1: hydraulic conductivity for 1st unsaturated zone (cm/h)	Truncated normal	1.32	0.06	5.75	2.65
K_S^2: hydraulic conductivity for 2nd unsaturated zone (cm/h)	Truncated normal	21	0.91	181	21
ρ: bulk density of the contaminated soil (mg/cm^3)	Truncated normal	1.4	1.1	1.7	0.14
T_s: thickness of contaminated soil (m)	Uniform	3	2.29	3.71	0.41
Groundwater data					
S: hydraulic gradient (m/m)	Truncated normal	0.007	0.004	0.011	0.002
θ: effective porosity (cm^3/cm^3)	Truncated normal	0.3	0.17	0.42	0.058
a_x: longitudinal dispersivity coefficient (dimensionless)	Truncated normal	0.1	0.05	0.15	0.025
H: thickness of aquifer (m)	Uniform	80	66.7	93.3	7.68
K_{gw}: groundwater hydraulic conductivity (m/d)	Truncated normal	36.6	17.5	111.84	9.15
X: distance to groundwater receptor along centerline (m)	Uniform	3700	2982	4418	414
Y: distance to groundwater receptor in cross-gradient direction (m)	Uniform	0	–50	50	28.9
Exposure factor					
I_w: water ingestion (L/d)	lg-normal	1.24	0.3	3.5	0.65
SF_w: carcinogen potency factor for oral route (mg kg^{-1} d^{-1})$^{-1}$	Uniform	0.035	0.015	0.055	0.012
I_a: inhalation rate (m^3/d)	Triangular	19.0	6.0	32.0	7.5
SF_a: carcinogen potency factor for inhalation (mg kg^{-1} d^{-1})$^{-1}$	Uniform	0.0174	0.0077	0.027	0.0056

classes of input data that exhibit different uncertainty characteristics: chemical properties, environmental media properties, and human health exposure/intake parameters, so all MMSOILS parameters in each category can be evaluated for uncertainty and summary statistics and distributions of the input and output parameters can be provided (USEPA, 1996b; Mills *et al*., 1997).

In the study of human exposure and risk analysis, the tail of the total risk distribution is of particular interest. Exposure and risk analyses often use the probability of the upper 5% tail (the 95th percentile) of their distribution to provide a conservative estimate of the possible effects of environmental contamination on a human population. Because the Monte Carlo sample size to be used in the analysis must be large enough to obtain a reasonable statistic for the 95th percentile, ten thousand Monte Carlo simulations were performed with 95% percentile confidence level for this study. This number is large enough to provide a statistically representative sampling for each sample size.

3 Results and discussion

3.1 Simple regression for sensitivity analysis

A question of practical concern in performing probabilistic risk assessments is the importance of specific random variables in influencing an endpoint of importance, such as distribution of risks from exposure to contaminated groundwater. These, along with parameter values randomly selected from their respective probability distributions were input to MMSOILS and were used to calculate risk with Monte Carlo methods. The process results in a distribution of risk based on the uncertainties in model parameters. Issues of importance in the sensitivity analysis are to determine how sensitive every random variable is by comparing correlation coefficient between input parameters and risks.

A correlation coefficient is a number between –1 and 1 which measures the degree to which two variables are linearly related. There are a number of different correlation coefficients that might be appropriate depending on the kinds of variables being studied. Pearson's product moment correlation coefficient, usually denoted by r, is one example of a correlation coefficient. It is a measure of the linear association between two variables that have been measured on interval or ratio scales. However, it can be misleadingly small when there is a relationship between the variables but it is a non-linear one. Especially the implicit assumption is made that the two variables are jointly normally distributed. When this assumption is not justified, a non-parametric measure such as the Spearman rank correlation coefficient (RCC) might be more appropriate. The Spearman rank correlation coefficient may also be a better indicator that a relationship exists between two variables when the relationship is non-linear (Howell, 1987). A set of data that monotonically increases in a highly non-linear manner might have a small r, but the RCC might be high near ±1.

The sensitivity analysis is performed for three endpoints including inhalation air risk, drinking groundwater risk and total risk for exposure to benzene using MMSOILS. Results are shown in Table 2 and indices are used in the table to evaluate sensitivity. Because the model response is nonlinear, to reduce the possibility that some important random variables might have been missed in the analysis, the Spearman's rank correlation coefficients were also calculated. The column labeled "rank" in Table 2 is based on decreasing values of R^2. Note the difference among the rankings of drinking groundwater, inhalation air and total risks in Table 2. The most important variables for all three endpoints are source geometries. The remaining variables

Table 2 Results of simple regression analysis for human health risk based on modeling results using full suite of random variables

Parameter	Drinking water risk			Air inhalation risk			Total risk		
	R^2	RCC[a]	Rank[b]	R^2	RCC[a]	Rank[b]	R^2	RCC[a]	Rank[b]
W	0.15	0.5022	1	0.0223	–0.3342	5	0.064	0.2710	4
I_w	0.146	0.4032	2	NA	NA	NA	0.111	0.3267	1
I	0.138	0.4828	3	0.0001	–0.0137	15	0.103	0.3651	2
SF_w	0.068	0.3019	4	NA	NA	NA	0.05	0.2378	5
L	0.053	0.2733	5	0.0238	–0.2766	4	0.012	0.1132	10
K_{gw}	0.033	–0.1879	6	2.77E–05	0.0136	16	0.025	–0.1508	7
S	0.027	–0.1910	7	0.0004	0.0162	8	0.017	–0.1442	9
a_x	0.01	–0.0986	8	0.0003	–0.0117	9	0.01	–0.0957	11
T_s	0.01	0.1081	9	1.53E–05	–0.0102	17	0.008	0.0759	12
X	0.009	–0.0958	10	NA	0.0139	NA	0.006	–0.0734	14
H	0.0087	–0.0936	11	0.00017	–0.0094	12	0.008	–0.0879	13
ρ	0.0028	0.0570	12	0.0029	0.0399	6	0.006	0.0757	15
K_d	0.001	0.1071	13	0.2611	–0.4849	1	0.079	–0.3575	3
K_s^2	0.0008	0.0312	14	9.7E–07	0.0088	18	0.0007	0.0270	17
θ	0.0003	–0.0120	15	0.00015	0.0173	14	4.4E–05	–0.0014	21
Y	0.0001	–0.0057	16	NA	0.0196	NA	2.25E–06	0.0059	22
K_h	4.65E–05	0.0065	17	0.0015	0.0401	7	0.0009	0.0329	16
K_s^1	1.8E–05	–0.0188	18	0.00016	0.0113	13	0.0001	–0.0090	18
F_c^2	9.15E–06	0.0060	19	0.00025	–0.0167	11	4.78E–05	–0.0051	20
F_c^1	2.29E–06	0.0074	20	0.00028	0.0109	10	7.76E–05	0.0163	19
I_a	NA	NA	NA	0.0584	0.2566	3	0.019	0.1714	8
SF_a	NA	NA	NA	0.079	0.2976	2	0.029	0.2105	6

[a] Spearman rank correlation coefficient; [b] rank is based on R^2 value, the square of Pearson's product moment correlation coefficient.

are mostly different shown by the order of selection.

To more fully explore the above results, three scatter plots of correlations generated by MMSOILS are shown in Fig.2, respectively. All the plots show untransformed data, the probability distribution used for each independent variable, and simple R^2 values. The independent variables chosen for the scatter plots are source geometry width, infiltration rate and thickness of aquifer. Some variables like source geometry width produce higher R^2 and RCC values for all the three endpoints (Table 2); on the contrary, some variables like thickness of aquifer have low values of all indices. The scatter plot shows that infiltration rate has a high correlation with drinking groundwater risk, however, has a very low correlation with inhalation air risk. There are some variables like infiltration rate shown by Table 2. Increasing absolute values of the indicators corresponds to increasing sensitivity to endpoints, as defined by the ratio in the Table 2, but the results in Table 2 and Fig.2 are interpretable only with respect to a specific problem. For these results, R^2 and RCC both provide comparable indication of this sensitivity, both in absolute value of the index and possession of the appropriate positive or negative sign to indicate the direction of sensitivity. To provide a more sensitivity analysis, the information previously discussed and presented is based on for farther stepwise regression analysis.

3.2 Stepwise multiple regression for sensitivity analysis

One practical issue inherent in the probabilistic approach is the number of variables that have to be treated as random and the justification for the choice of random variables. Suppose, $M+N$ variables are required as input to an analysis, where M equals the number of variables treated as random and N equals the number of variables treated as deterministic. By performing a Monte Carlo analysis using such a model as MMSOILS, a distribution of endpoints is predicted. Now, the question is whether a comparable distribution can be generated where less than M random variables are used and more than N deterministic variables are used ($M+N$ remains the same). Mills *et al.* (1999) has shown the implicit procedures of the approach. Here, a stepwise multiple regression approach is used to perform this task as simple regression approach only could test sensitivity for every random variable.

One of the same endpoints used previously is also selected for the analysis. Based on the correlation coefficient of the simple regression analysis, the procedure to produce higher and higher R^2 is continued until R^2 is calculated for all random variables, is automated within the software used. The results are shown in Fig.3a. The number of sensitive random variables for three endpoints is different. The plots in Fig.3a show that the asymptotes for R^2 are approximated 0.641 with eleven variables, 0.447 with six variables and 0.541 with fourteen variables for drinking groundwater risk, inhalation air risk and total risk, respectively. The R^2 with maximum possible 1.0 is not achieved by either model, and less than 1.0 is indicative of the nonlinear relationships between variables. However, many of the sensitive variables remain the same. As shown in Fig.3a, only two of the six variables selected as most important for inhalation air risk are the same with those for drinking groundwater risk; however, some variables like the source K_d are selected by inhalation air risk but not by drinking groundwater risk, and some like

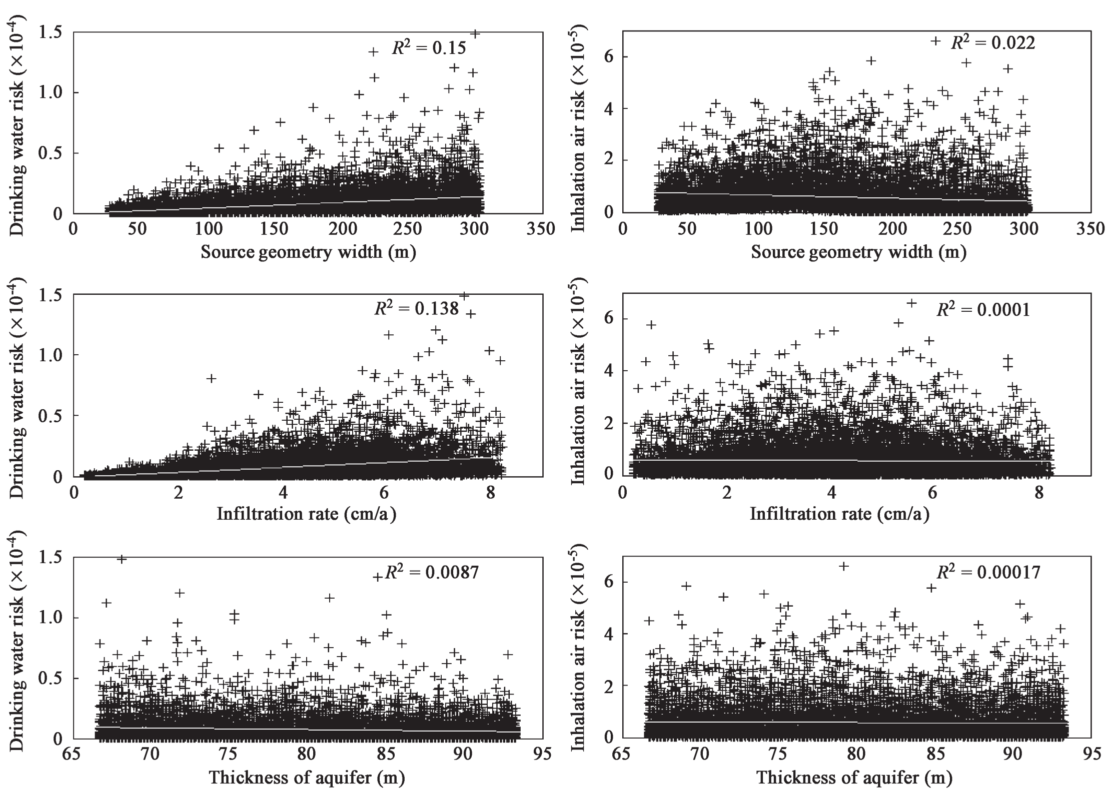

Fig. 2 Scatter plots created from MMSOILS simulations of drinking water and inhalation air risks.

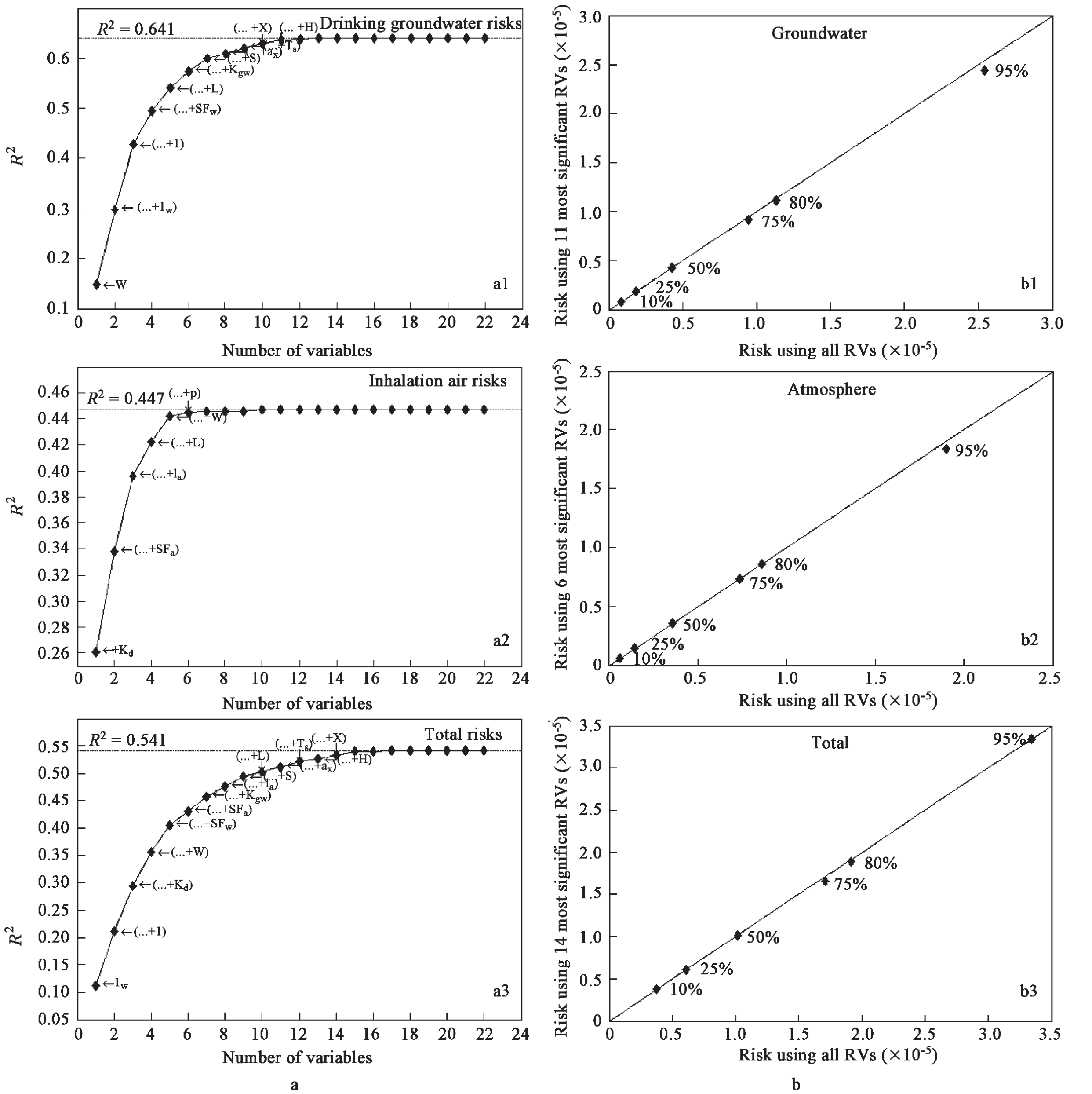

Fig. 3 Stepwise multiple linear regression approach to sensitivity analysis for selected endpoints (a) and comparisons of CDFs for selected model endpoints using all random variables (RVs) to generate the CDFs and using a sensitivity subset of random variables to generate the sumset (b).

the x-direction dispersivity coefficient, a_x, are selected by drinking groundwater risk but not by inhalation air risk. A similar result occurs for different risks as endpoints as shown in Fig.3a. The most important random variables selected by each endpoint are very different. Considering the plots along with Table 2, the choice of these variables is plausible.

Using the random variables identified as most important in producing the distribution of concentrations in Fig.3b, each model is rerun and treats only those 6, 11 and 14 variables as random according respectively to three different endpoints; the remaining former random variables are treated deterministically, using mean values from their distributions. The sensitivity analysis is performed for groundwater, atmospheric and total risk endpoints for exposure to benzene. The comparison is shown in Fig.3b. The 45-degree line shown in Fig.3b is the line of perfect correlation; if the two distributions being compared are identical, all points would fall on those lines. The result for atmospheric risk fully agrees to within 80%, but approximately agrees to within 80% for groundwater risk. However, the result for total risk is the same apart from difference of 75% percentile values. Relative to the percentile values themselves for all endpoints, the confidence intervals are small. According to the results and those in the simple regression analysis too, it is conceivable that these random variables above specified are enough to indicate the specific sensitivity.

3.3 Integrated uncertainty analysis

To determine the relative influence of random variables attributable to different endpoints is done using sensitivity analysis. As previously described, a complete Monte Carlo method was performed by varying input variables, yielding a range of exposure risk values. This, in turn, may indicate whether or not an uncertainty analysis is required or if the uncertainty analysis can be simplified and performed on one endpoint using only some variables.

The following exercise illustrates how we determined in which form the data of uncertainty are most appropriately expressed. Consider different input random variables with known distributions, X_1, X_2 ... X_n, who were transformed by a model to yield an output Y. The proportion of the variance in the predicted output Y attributable to the variance from the input parameter X_i is a function as follows (Bevington, 1969):

$$Y = \text{function}(\delta_{X_1}\Big/\overline{X}_1, \delta_{X_2}\Big/\overline{X}_2, \cdots \delta_{X_n}\Big/\overline{X}_n) \tag{2}$$

The resulting influence of the X_i distribution can be calculated using the following equation:

$$p_i = \frac{\omega_i \delta_{X_i} / \overline{X}_i}{\sum_{i=1}^{n} \omega_i \delta_{X_i} / \overline{X}_i} \times 100\% \tag{3}$$

Where, p_i is the exact analytic formulation of the proportion of variance in the output Y attributable to the variance from input X_i; δ_{X_i} is standard deviation of variable X_i; $\overline{X}_i$ is mean value of variable X_i; $\delta_{X_n}/\overline{X}_n$ is the coefficient of variation for X_n; ω_i is the weight for X_i, and here equals to the correlation coefficient, because the more sensitive X_i is to Y with the higher correlation coefficient value between X_i and Y. As discussed in an earlier section, when the absolute value of Pearson's product moment correlation coefficient is very small, but that of Spearman rank correlation coefficient is high, the former is more accurately selected as ω_i instead of the later avoiding spurious correlation.

Uncertainty analysis through the model yields a range of human health risk results. Fig.4 illustrates projections of the total risk, the risk from soil gas, and the risk from the corresponding groundwater. The uncertainty analyses with three different distributions are illustrated in the small figure, which is a box-and-whisker plot included in Fig.4. The "whisker" is line with 95th percentile extending right the box. They show the extent of the rest of the sample. The more left and right lines of the "box" are the 25th and 75th percentiles of the sample. The middle line is mean value and the distance between the left and right of the box is the inter-quartile range. It is clear that the risk from groundwater is higher than that from soil gas from 60th to 95th percentile of risk, and both are almost the same at higher and lower percentiles of risk. The drinking groundwater risk has a more dispersive distribution than inhalation air risk while the distribution of total risk is most dispersive. Note that the probability of total risk higher than 10^{-6}, the USEPA advised health risk value (USEPA, 1996d), is larger than 95%, that was also shown by Fig.3b. Although the percent of risk from soil gas and groundwater is little difference, the percent of uncertainty from them is not able to be distinguished for total risk. The following discussion demonstrates ways to identify the percent uncertainty of random variables and the relative influence of soil gas and groundwater on the total risk.

Because the total risk is always concerned with health, here it is only analyzed for uncertainty. With the results from regression analysis and Eq. (3), the findings of the uncertainty analysis, presented in Fig.5, are based on the calculation of correlation coefficient indicating the contribution of the variance of fourteen different random variables and two exposure routes on total risk. These two exposure routes are drinking groundwater and inhalation air. The results indicate that uncertainty in drinking groundwater accounts for approximately 74% of the variance in the total risk indicating its potential importance to risk uncertainty, while uncertainty stemming from soil gas exposure accounts for approximately 26%.

An analysis of the importance of variance of individual

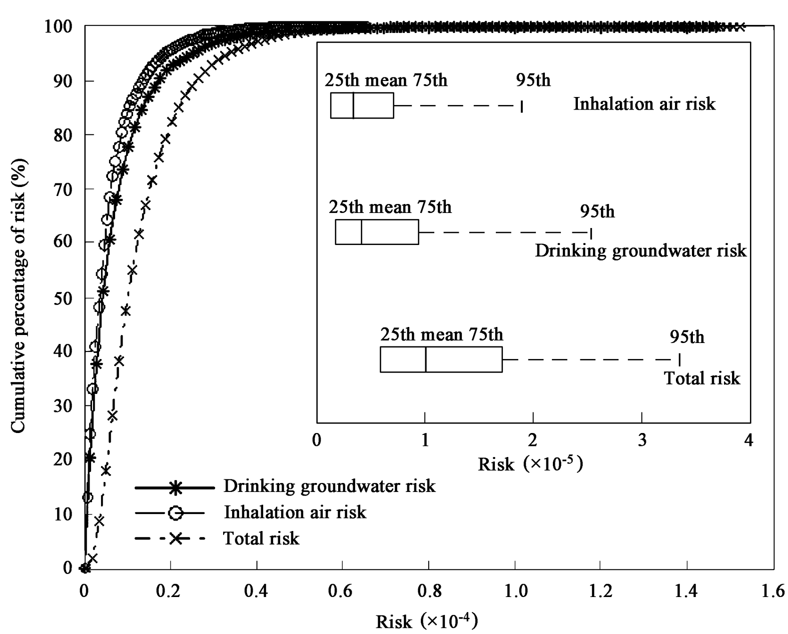

Fig. 4 Cumulative percentile of three different risk endpoints ant the box-and-whisker plot.

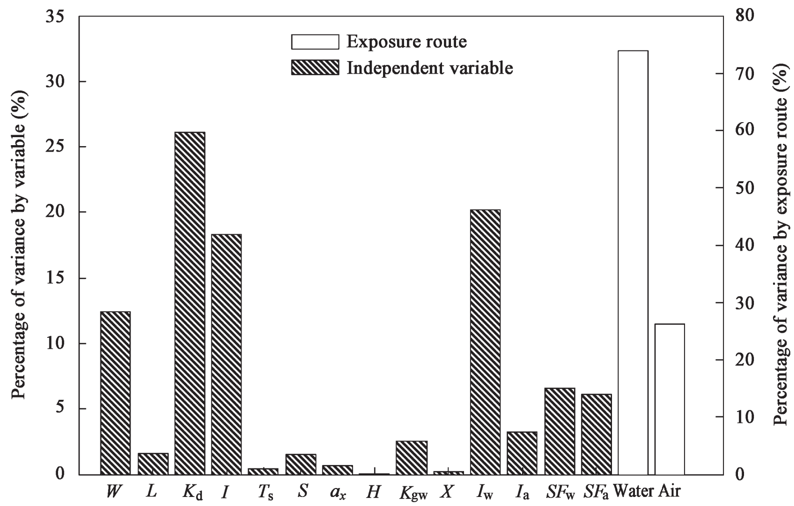

Fig. 5 Percentage describing the variance of total risk as a result of variance of variable and exposure route.

parameters reveals that four random variables are the dominant sources of variance in total risk as illustrated in Fig.5. They are in sequence contaminated soil K_d, water ingestion, annual recharge and source geometry width. In the case studied here, uncertainty in these four variables contributes roughly 77% of the variance in the total risk prediction. With knowledge such as this, resources could be focused on obtaining a more accurate representation of these variables thus decreasing uncertainty in total risk. Other variables in Fig.5 show contributions at the 1%–7% level, indicating that greater knowledge of these distributions will reduce variance but to a lesser degree. However, the summed contribution of these variables is not trivial relative to the variance attributed to other four variables, accounting for 23% of the variance in total risk. The influence of each model varies with the case study being considered. The results provided here serve to illustrate the analysis techniques and the type of information they can yield. More accurate characterization of variance importance for these parameters requires methods with better resolution at low variance such as regional sensitivity analysis (Spear *et al*., 1994).

4 Conclusions

This paper describes the application of MMSOILS model with Monte Carlo simulation, to a contaminated soils containing a typical organic chemical (Benzene) resulting from petrochemical factory. The actual risk calculation is hypothetical because of our assumptions to include no future remedial action and no biodegradation of benzene. Three endpoints drinking groundwater risk, inhalation air risk and total risk are predicted. A sensitivity analysis, using simple and stepwise multiple regression, is constructed to evaluate which of the random variables are most important in producing the predicted distributions of health risks. Pearson's product moment correlation coefficient and Spearman rank correlation coefficient give comparable results which have practical implications, in that a method is provided to distinguish between important versus unimportant random variables in terms of sensitivity of selected endpoints. The numbers of the random variables accurately reproduce the probability distribution originally developed using 22 random variables are six for inhalation air risk, eleven for drinking groundwater risk and fourteen for total risk while the remaining random variables are set to their mean values, an affirmation that none of the remaining variables is important. This directly translates into a reduction in data collection and modeling effort.

The integrated uncertainty analyses have also enabled the evaluation of which parameters from random variables and both exposure routes (groundwater and air) exert the greater influence on final risk variance for a particular total risk endpoint. It was observed that for this particular contamination scenario with the chosen parameter distributions, uncertainty introduced by drinking groundwater route was dominant of total risk uncertainty. Further analysis indicated that most percent of the variance of total risk in this scenario comes from four random variables which are in sequence contaminated soil K_d, water ingestion, annual recharge and source geometry. The uncertainty resulting from the four variables is almost 77% percent of the variance in total prediction, so it should be to get more accurate data charactering these variables. It is important to remark that the influence of individual parameters will always be site specific and that by developing the methodology and framework with which to perform the uncertainty analysis, site risk assessment will be more robust.

References

Bennett D H, James A L, Mckone T E, 1998. On uncertainty in remediation analysis: variance propagation from subsurface transport to exposure modeling[J]. Reliability Engineering and System Safety, 62: 117–129.

Bevington P R, 1969. Data reduction and error analysis for the physical sciences[M]. New York, NY: McGraw-Hill.

Fewtrell L, Macgill S M, Kay D *et al*., 2001. Uncertainties in risk assessment for the determination of drinking water

pollutant concentrations: cryptosporidium case study[J]. Wat Res, 35(2): 441–447.
Hamby D M, 1994. A review of techniques for parameter sensitivity analysis of environmental models[J]. Environmental Monitoring and Assessment, 32(2): 135–154.
Havelaar A H, 1998. Emerging microbiological concerns in drinking water quality[R]. In: Drinking Water Research 2000. Papers from a Research Conference, held at the University of Warwick. Drinking Water Inspectorate, London.
Helton J C, Johnson J D, Mckay M D *et al*., 1995. Robustness of an uncertainty and sensitivity analysis of early exposure results with the MACCS reactor accident consequence model[J]. Reliability Engineering and System Safety, 48: 129–148.
Howell D C, 1987. Statistical methods for psychology[M]. Boston, MA: Daxbury Press. 172–222.
James A L, Oldenburg C M, 1997. Linear and Monte Carlo uncertainty analysis for subsurface contaminant transport simulation[J]. Water Resour Res, 33(11): 2495–2508.
Kalagnanam J, Kandlikar M, Linville C, 1998. Importance ranking of model uncertainties: A robust approach[R], IBM Research Report, RC 21130(94479), T. J. Watson Research Center, Yorktown Heights, New York.
Manteufel R D, 1996. Variance-based importance analysis applied to a complex probabilistic performance assessment[J]. Risk Analysis, 16(4): 587–598.
Mills W B, Cheng J J, Droppo J G *et al*., 1997. Multimedia benchmarking analysis for three risk assessment models: RESRAD, MMSOILS, and MEPAS[J]. Risk Analysis, 17(2): 187–201.
Mills W B, Lew C S, Hung C Y, 1999. Sensitivity of concentration and risk predictions in the PRESTO and MMSOILS Multimedia Models: Regression technique assessment[J]. Risk Analysis, 19(3): 511–525.
Pelmulder S D, Yeh W W-G, Kastenberg W E, 1996. Regional scale framework for modeling water resources and health risk problems[J]. Water Resour Res, 32: 1851–1861.
Robert S A, Rosamilia K, Wojcik N C, 2005. Review of the literature on benzene exposure and leukemia subtypes[J]. Chemico-Biological Interactions, 153: 9–21.
Spear R C, Grieb T M, Shang N, 1994. Parameter uncertainty and interaction in complex environmental models[J]. Water Resour Res, 30(11): 3159–3169.
USEPA, 1986. Guidelines for carcinogenic risk assessment[J]. Federal Register, 51(185): 33992–34003.
USEPA, 1992. Guidelines for exposure assessment[J]. Federal Register, 57(104): 22888–22938.
USEPA, 1996a. Exposure factors handbook[S]. Washington, DC: EPA/600/P-96/002A.
USEPA, 1996b. MMSOILS: Multimedia contaminant fate, transport, and exposure model[S]. Documentation and User's Manual Version 4.0, Draft, Athens, Georgia: Prepared for Environmental Research Laboratory U.S. Environmental Protection Agency.
USEPA, 1996c. Summary report for the workshop on Monte Carlo analysis[R]. U.S. EPA report EPA/630/R-96/010.
USEPA, 1996d. Soil screening guidance: technical background document[R]. 2cd ed. Washington, DC: Office of Emergency and Remedial Response. EPA/540/R-95/128.
USEPA, 1997. Policy for use of probabilistic analysis in risk assessment at the U.S. Environmental Protection Agency[R]. Risk Assessment Forum, EPA/630/R-97/001.