SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 22, No. 12, pp. 3081–3087, 2003

Printed in the USA
0730-7268/03 $12.00 + .00

ISSUES OF REPLICABILITY IN MONTE CARLO MODELING: A CASE STUDY WITH A PESTICIDE LEACHING MODEL

IGOR G. DUBUS*† and PETER H.M. JANSSEN‡
†Cranfield Centre for Ecochemistry, Cranfield University, Silsoe, Beds MK45 4DI, United Kingdom
‡RIVM, Netherlands Environmental Assessment Agency, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

(*Received* 4 *October* 2002; *Accepted* 22 *April* 2003)

Abstract—Sensitivity and uncertainty analyses based on Monte Carlo sampling were undertaken for various numbers of runs of the pesticide leaching model (PELMO). Analyses were repeated 10 times with different seed numbers. The ranking of PELMO input parameters according to their influence on predictions for leaching was stable for the most influential parameters. For less influential parameters, the sensitivity ranking was severely influenced by the seed number used. For uncertainty analyses, probabilities of exceeding a particular concentration were significantly influenced by the seed number used in the random sampling of values for the two parameters considered, even for those cases in which 5,000 model runs were undertaken (coefficient of variation of 10 replicated analyses, 5%). A decrease in the variability of exceedance probabilities could be achieved by further increasing the number of model runs. However, this may prove to be impractical when complex deterministic models with a relatively long running time are used. Attention should be paid to replicability aspects by modelers when devising their approach to assessing the uncertainty associated with the modeling and by decision makers when examining the results of probabilistic approaches.

Keywords—Monte Carlo Sensitivity Uncertainty Latin hypercube sampling Model

INTRODUCTION

Modeling techniques based on Monte Carlo (MC) sampling are increasingly used in environmental disciplines [1]. The two main applications are typically the study of the relationship between model input and output (sensitivity analysis [2]) and the assessment of the variability/uncertainty in modeling predictions resulting from the variability/uncertainty in model input parameters (uncertainty analysis, also referred to as MC modeling [3]). An MC uncertainty analysis for a given model typically involves six stages: The selection of outputs of interest on which the analysis will be performed; the selection of model input parameters, the associated uncertainties of which will be considered; the attribution of ranges and distributions to each parameter selected and specification of their dependence/correlation; the generation of random samples from the joint distributions assigned to parameters; the running of the model for each of the sample elements; and the examination of model predictions in statistical terms (e.g., estimation of the mean and variance, construction of cumulative distribution charts). In a sensitivity analysis, the sixth step consists of examining changes in model predictions resulting from changes in model input parameters using scatterplots, regression analysis, or correlation measures [4]. Monte Carlo sensitivity and uncertainty analyses have been reported for pesticide leaching models [5–9].

Reasons for the widespread adoption of MC techniques in the study of model sensitivity and uncertainty in modeling include the following: They are transparent, conceptually simple, easy to explain, and can therefore be readily understood by decision makers; they are generic and can be applied to virtually any model (although this might involve a significant amount of work for models not amenable to an easy automation of modeling tasks); they have been applied on numerous occasions in the past; they are relatively easy to implement when using appropriate software; and they normally do not require modifications to the model code. Various sampling procedures are used in MC studies, and these include random sampling, stratified sampling, and quasirandom sampling [4]. Although sampling schemes are termed random, the numerical processes involved are deterministic and, therefore, are not random in the strictest sense. Random number generation typically requires the specification of a seed number that will be used to initiate the random sequence. Different seed numbers will result in different random samples. Although MC simulation is widely used in many fields of science, little research has been conducted on what influence the generation of a random sample with a particular seed number may have on the overall outcome of MC modeling (replicability issue).

The present paper reports on investigations to assess the robustness of MC applications in pesticide fate modeling (MC-based sensitivity and uncertainty analyses) and, hence, the confidence that should be attributed to results obtained through these techniques. The research looked at the influence on results of sensitivity and uncertainty analyses of the seed number used to generate random values for model input parameters and the size of the random sample (equivalent to the number of model runs undertaken). Sensitivity and uncertainty analyses were undertaken for the pesticide leaching model (PELMO; Staatliche Lehr- und Forschunganstalt für Landwirtschaft, Weinbau und Gartenbau, Neustadt, Germany) with varying numbers of model runs, and all analyses were repeated 10 times with different seed numbers.

* To whom correspondence may be addressed (i.dubus@brgm.fr). The current address of I. G. Dubus is BRGM, Avenue Claude Guillemin, BP 6009, 45060 Orléans Cedex 2, France.

Table 1. Model input parameters included in the sensitivity analysis and parameterization of probability density functions

Parameter[a]	Description	Nominal value	Lower value[b]	Upper value[b]	Distribution	Variance
ANET	Depth of evapotranspiration computation (cm)	10	5	15	Normal	6.51
CINT	Maximum interception storage (cm)	0.28	0.21	0.35	Normal	1.28×10^{-3}
AMXD	Maximum active rooting depth (cm)	110	82.5	137.5	Normal	1.97×10^{2}
COVM	Maximum soil cover by the crop (%)	100	80	100	Uniform	—
BUD[c]	Bulk density (g/cm^3)	1.5	1.35	1.65	Normal	5.86×10^{-3}
WC[c]	Initial soil water content (%)	0.2	0.15	0.25	Normal	6.51×10^{-4}
SA1	Sand content, first horizon (%)	68.3	64.9	71.7	Normal	3.04
CL1	Clay content, first horizon (%)	7.2	6.8	7.6	Normal	3.37×10^{-2}
OC[c]	Organic carbon (%)	1.5	1.35	1.65	Normal	5.86×10^{-3}
SA2	Sand content, second horizon (%)	67	63.65	70.35	Normal	2.92
CL2	Clay content, second horizon (%)	6.7	6.365	7.035	Normal	2.92×10^{-2}
SA3	Sand content, third horizon (%)	96.2	92.4	100.0[d]	Normal	3.66
CL3	Clay content, third horizon (%)	0.9	0.855	0.945	Normal	5.27×10^{-4}
APPL	Application rate (kg/ha)	0.015	0.01125	0.01875	Normal	3.66×10^{-6}
DT50	Laboratory half-life (d)	58.5	29.25	117	Log normal	3.85×10^{2}
TEMP	Temperature of laboratory incubation (°C)	20	19	21	Normal	2.60×10^{-1}
QTEN	Increase in degradation for a temperature increase of 10°C (−)	2.2	1.76	2.64	Normal	5.03×10^{-2}
MOIS	Soil moisture content during incubation experiment (%)	40	30	50	Normal	2.60×10^{1}
MEXP	Exponent for moisture correction (−)	0.7	0.56	0.84	Normal	5.10×10^{-3}
KOC	Sorption coefficient normalized to organic carbon (ml/g)	91.45	45.725	182.9	Log normal	9.41×10^{2}
NF	Freundlich exponent (−)	0.895	0.716	1.074	Normal	8.34×10^{-3}

[a] Variations of values in the subsoil were linked to that in the topsoil.
[b] Lower and upper values were assumed to correspond to the 2.5th and 97.5th percentiles of the normal and log-normal distributions.
[c] Depth-dependent parameter.
[d] A truncation at 100% was integrated into the sampling to avoid the use of unrealistic values.

MATERIALS AND METHODS

Model selection, scenario simulated, and automation of modeling tasks

The pesticide leaching model used was PELMO version 3.00 SP2. This is a one-dimensional leaching model that integrates descriptions of water movement and pesticide transfer through the soil column [10]. Hydrology is described using a tipping-bucket approach, whereas solute transport is simulated using the convection–dispersion equation. Subroutines describing the fate of reactive solutes include the description of sorption, degradation, pesticide losses in leaching, runoff, soil erosion, and volatilization.

The leaching of a pesticide in the German soil Borstel was simulated using replicated climatic data from Hamburg [10]. The weather dataset used was that known as the Hamburg normal, which consists of a number of replicated years of weather data for the year 1978 for Hamburg (annual rainfall, 777 mm). Properties of the pesticide were selected from within the range of those of registered compounds to give a reasonable likelihood of leaching to groundwater [11]. A laboratory time for a 50% decline of the initial amount of pesticide value (DT50) of 58.5 days (assumed to have been determined at 20°C and 40% field capacity) was initially considered in the modeling. The sorption distribution coefficient normalized to organic carbon (K_{oc}) and the Freundlich exponent were set to 91.45 ml/g and 0.895, respectively. The pesticide was assumed to be applied to the soil on May 15 in each of the 20 years simulated. No guidelines are yet available from pesticide regulatory authorities regarding an acceptable level of exceedance of a particular threshold concentration, but it is anticipated that within the context of probabilistic modeling, the acceptance level will be less than 5% exceedance. Accordingly, an application rate of 15 g active substance/ha was selected to provide, on average, a probability of just less than 5% that the annual average concentration in the 20th year simulated exceeds the European Union threshold for drinking water (0.1 μg/L). Values attributed to the main PELMO parameters are presented in Table 1. The output variable of interest in the present exercise was the annual average pesticide concentration in leachate in the 20th year simulated.

Replicability in MC sensitivity analysis

The aim of the sensitivity analyses was to establish a classification of PELMO input parameters according to their influence on PELMO predictions for leaching (i.e., to identify the parameters to which the model is sensitive or insensitive). The sensitivity of PELMO was studied using MC sampling combined with regression analysis. This popular methodology [2,4] has recently been used to investigate the sensitivity of the pesticide leaching models MACRO [8] and PRZM [9]. In the present instance, 21 PELMO input parameters were attributed normal (18 parameters), log-normal (two parameters), or uniform (one parameter) distributions on the basis of expert judgement (Table 1). For the normal and log-normal probability density functions, lower and upper values for variation were arbitrarily selected (Table 1), and these were assumed to correspond, respectively, to the 2.5th and 97.5th percentiles of the distributions. For organic carbon content, bulk density, and water capacity, only the parameter in the top horizon was included in the analysis, and values for the parameter in the deeper horizons were modified by the same variation applied to the value in the top horizon.

Latin hypercube sampling (LHS) [2,4] was used to sample values for the 21 parameters from their probability density functions. The two sampling packages used in the present research were Crystal Ball 2000 (Ver 5.1 [12]) and @RISK (Ver

4.0.5 [13]), which both implement random LHS. Four different sample sizes were considered (250, 1,000, 2,500, and 5,000 elements), and for each sample size, 10 different replicated samples were generated by randomly selecting the seed number used in the sampling. The maximum number of elements (5,000) was determined by limitations with regard to computational resources.

The PELMO was run automatically and independently for each element of each replicated sample, and the annual average pesticide concentration in leachate in the 20th year simulated was automatically extracted. The total number of PELMO runs was 10(250 + 1,000 + 2,500 + 5,000) = 87,500.

The sensitivity of PELMO to changes in input parameters was derived by regression analysis with raw (i.e., nonranked) and rank-transformed data. The rank transformation consisted in replacing each value for inputs and outputs by their rank in the sample. For instance, the largest organic carbon content randomly selected in the sample of size 250 received the rank 1, whereas its smallest value was attributed the rank 250. Rank transformation is a popular transformation in sensitivity analysis, which is used to decrease nonlinearity/nonmonotonicity in highly nonlinear systems [14]. Both raw and ranked data were analyzed in the same way. Annual average concentrations in the 20th year simulated and model inputs were standardized (i.e., the mean of each variable was subtracted from each value, and the result was divided by the standard deviation of the variable) and linearly related through multiple linear regression using the Statistica package (Ver 6.0 [15]) as follows:

$$Y = \sum_{i=1}^{21} b_i \cdot X_i + \varepsilon \quad (1)$$

where Y is the standardized concentration, X_i is a standardized input parameter, b_i is the regression coefficient for each X_i, and ε is the regression error.

The sensitivity of PELMO to each input parameter is given by the coefficient b_i for that parameter, which is known as the standardized regression coefficient (SRC; raw data) and the standardized ranked regression coefficient (SRRC; rank transformed data) [16]. The SRC or SRRC values (S(R)RC) may be positive or negative. A positive S(R)RC means that an increase in the input parameter will generally result in an increase in the output, whereas a decrease in the output will result for a parameter with a negative S(R)RC. For each replicated sample and for the raw and ranked data, absolute values of S(R)RC were used to sort the 21 PELMO input parameters according to their influence on the prediction of the annual average concentration in the 20th year of simulation. The SRRC indices have been shown to be one of the most robust and reliable means of assessing model sensitivity [17].

Replicability in MC uncertainty assessments

The probabilistic modeling undertaken was based on MC simulations and was restricted to the variation of two parameters only: K_{oc} and DT50. The PELMO has been shown previously to be most sensitive to parameters related to sorption and degradation [18], and these two parameters play an important role within environmental risk assessments for pesticides. The number of parameters considered in the MC analysis was voluntarily limited to two in an effort to keep the modeling system as simple as possible and to ensure the general validity of the results obtained. Similar results would have been obtained if a larger number of parameters had been selected.

Data regarding the environmental properties of pesticides are expected to be limited in most instances when an environmental risk assessment is needed and not numerous enough to support a robust attribution of probability distribution functions to parameters via distribution fitting. In the present case, log-normal distributions were attributed to K_{oc} and DT50 through expert judgment. The focus of the present study was on illustrating replicability issues, and no distinction between the contributions of stochastic variability and incertitude (lack of knowledge) to the overall uncertainty was made. In our experience, a common rule of thumb is that K_{oc} and DT50 values obtained through standard laboratory experiments will vary within approximately a factor of 2 from the median values, although some consider that a factor of 4 might be more appropriate. This expert knowledge was reflected in the parameterization of the log-normal distributions by considering that 95% of the overall probability for the two parameters was contained within a range defined by $M/2$ and $M \cdot 2$, where M is the value considered in the initial scenario.

The LHS procedure was the same as that used for sensitivity analyses. Twelve different sample sizes were considered (10, 50, 100, 150, 200, 250, 500, 750, 1,000, 1,500, 2,500, and 5,000 elements), and for each sample size, 10 different samples were generated by randomly selecting the seed number used in the sampling. The maximum number of elements (5,000) was determined by limitations with regard to computational resources.

The PELMO was run for each element of each replicated sample, and the annual average pesticide concentration in leachate in the 20th year simulated was automatically extracted. For each sample, the probability p of exceeding a threshold concentration of 0.1 μg/L was calculated as follows:

$$p = \frac{n(C > 0.1\ \mu g/L)}{N} \quad (2)$$

where $n(C > 0.1\ \mu g/L)$ is the number of model runs within the sample for which the average annual concentration in the 20th year of simulation (C) exceeded 0.1 μg/L and N is the total number of model runs carried out in the sample.

The total number of PELMO runs undertaken to investigate replicability issues in uncertainty analyses was 10(10 + 50 + 100 + 150 + 200 + 250 + 500 + 750 + 1,000 + 1,500 + 2,500 + 5,000) = 315,200. Adding the number of PELMO runs for the sensitivity analysis, PELMO was run 402,700 times. Each PELMO run took between 17 to 35 s to complete, depending on the computer used.

RESULTS

Replicability in sensitivity analysis

Model sensitivity was determined by regression analysis relating the annual average pesticide concentration in leachate in the 20th year simulated to model inputs. For those instances in which regression analyses were carried out on the raw untransformed data, relatively small r^2 values were obtained ($0.31 < r^2 < 0.49$) (Table 2). The poor quality of the regressions reflected the inherent nonlinear character of the relationship between inputs and outputs in pesticide fate models and may question the validity of the sensitivity results obtained. Application of a rank transformation to the input and output data resulted in a large increase in r^2 values ($0.95 < r^2 < 0.97$), confirming the usefulness of this approach.

Results of replicability investigations are presented in Figures 1 (raw data) and 2 (ranked-transformed data), which provide a rapid and effective visual assessment concerning the

Table 2. Coefficient of determination (r^2) obtained by regression analysis on the raw and rank-transformed data for different number of model runs

Number of model runs	Raw data		Rank-transformed data	
	Mean[a]	SD	Mean[a]	SD
250	0.415A	0.040	0.958A	0.005
1,000	0.360B	0.023	0.955B	0.002
2,500	0.358B	0.012	0.954B	0.003
5,000	0.347B	0.009	0.954B	0.002

[a] Means having the same letter are not different at the 5% probability level (post-hoc Bonferroni's test).

stability of sensitivity results. The four charts in each of the two figures present results of sensitivity analyses carried out for the 10 replicated samples. Each chart corresponds to a different sample size (i.e., a different number of PELMO runs). Each of the 21 lines in the charts (one for each parameter) joins sensitivity rankings obtained for a particular parameter in the 10 replicated sensitivity analyses. A chart with only horizontal lines would be obtained if the rankings of parameters were the same whatever the seed number used. Charts reflecting poor replicability will appear to be disorganized.

Figure 1 (raw data) shows that replicability of sensitivity rankings was fairly poor except for the three most influential parameters. All charts were relatively disorganized, with only the top three parameters displaying horizontal lines. Figure 3 presents standardized regression coefficient absolute values for PELMO input parameters together with their 95% confidence

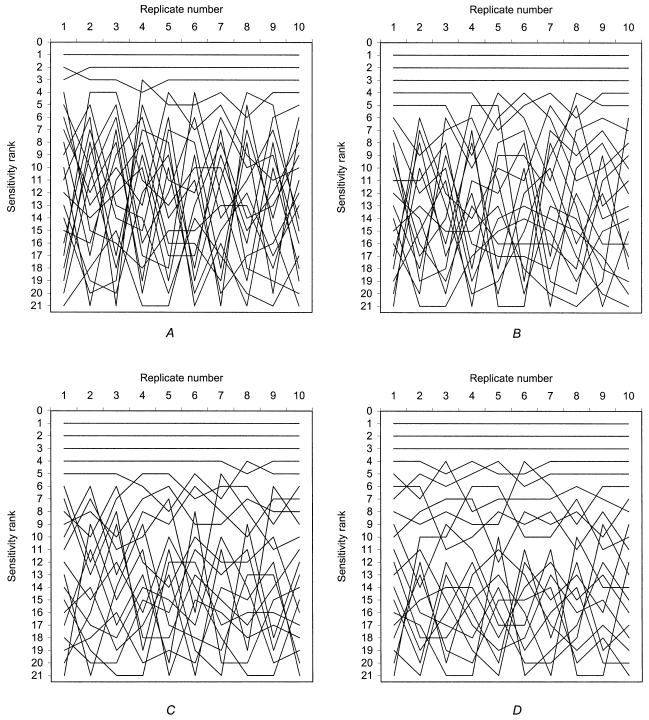

Fig. 1. Stability in sensitivity ranking obtained for 10 different seed numbers and four sample sizes (untransformed data). A larger sensitivity rank denotes a greater sensitivity of the model to this parameter. Each line joins sensitivity rankings obtained for the 10 different seed numbers for one particular input parameter. A horizontal line means that the same sensitivity ranking was obtained for the 10 replicated random samples. Charts **A, B, C,** and **D** correspond to 250, 1,000, 2,500, and 5,000 model runs, respectively.

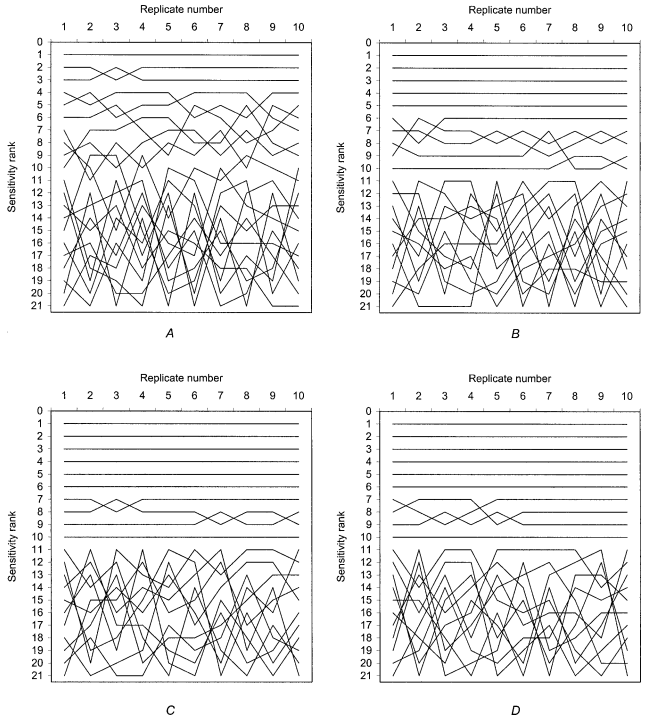

Fig. 2. Stability in sensitivity ranking obtained for 10 different seed numbers and four sample sizes (rank-transformed data). A larger sensitivity rank denotes a greater sensitivity of the model to this parameter. Each line joins sensitivity rankings obtained for the 10 different seed numbers for one particular input parameter. A horizontal line means that the same sensitivity ranking was obtained for the 10 replicated random samples. Charts **A, B, C,** and **D** correspond to 250, 1,000, 2,500, and 5,000 model runs, respectively.

intervals for the case of 5,000 runs and provides an explanation for the instability of the various replicates observed for the less influential parameters. Confidence intervals for the first three most influential parameters are clearly separated, whereas those for the remaining 18 parameters overlap. It should be noted that the strict conditions associated with derivation of these confidence intervals in terms of residue randomness are not fulfilled when applying a regression approach to a deter-

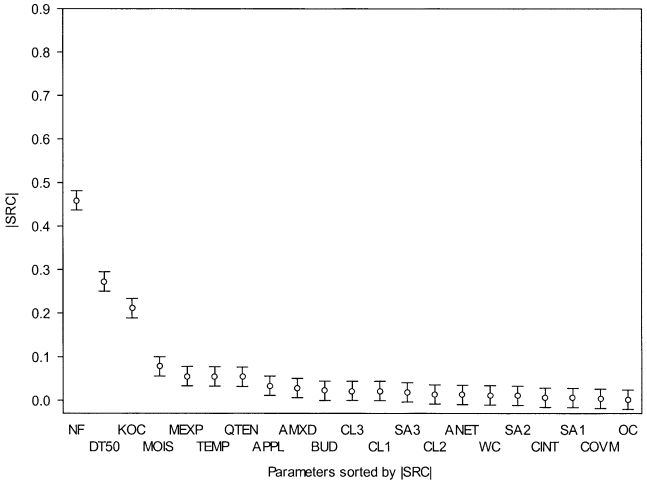

Fig. 3. Classification of pesticide leaching model input parameters according to absolute values of the standard sensitivity index (SRC; 5,000 runs). Intervals for each point represent the 95% confidence intervals. Parameter names are as described in Table 1.

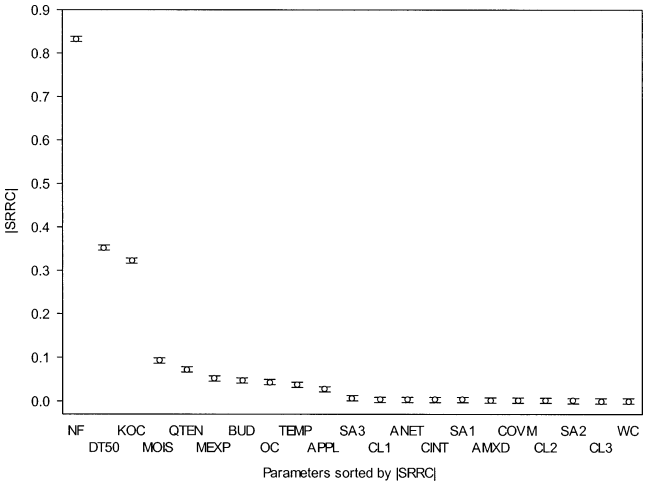

Fig. 4. Classification of pesticide leaching model input parameters according to absolute values of the standard ranked regression coefficient (SRRC) sensitivity index (5,000 runs). Intervals for each point represent the 95% confidence intervals. Parameter names are as described in Table 1.

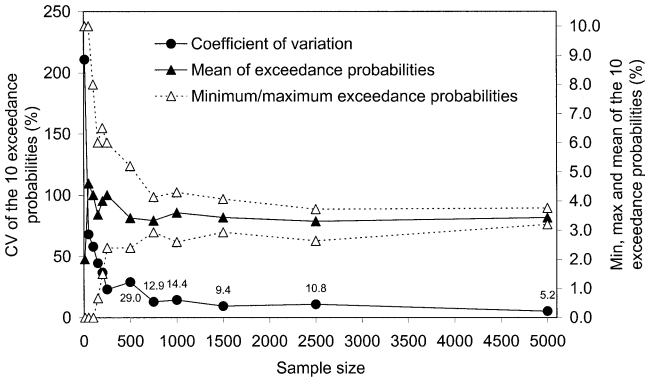

Fig. 5. Variability in the probability of exceeding 0.1 μg/L obtained when repeating Monte Carlo simulations 10 times with different seed numbers. The sample size (equivalent to the number of pesticide leaching model runs undertaken for each of the 10 replicated samples) was varied from 10 to 5,000. CV = coefficient of variation.

ministic model, but the approach provides a tentative form of guidance as to whether the model output is significantly influenced by the various model inputs [19,20].

The rank transformation of the data led to an improvement in the stability of the ranking of PELMO input parameters for different replicated random samples (Fig. 2). Also, the increased stability obtained through an increase in the number of model runs was visible. Again, the relative stability of the most influential parameters for the case of 5,000 runs was reflected, to some extent, by the absence of overlap of the confidence intervals for SRRC (Fig. 4). Based on the SRRC results obtained for 5,000 runs, the 10 input parameters (parameter names as presented in Table 1) that most influenced PELMO predictions for pesticide were

$$\text{NF} > \text{DT50} > \text{KOC} > \text{MOIS} > \text{MEXP} > \text{TEMP}$$

$$\approx \text{QTEN} \approx \text{APPL} \approx \text{BUD} \approx \text{OC}$$

Parameters that had the largest influence on PELMO predictions for leaching were those related to pesticide sorption and degradation. These sensitivity results are in agreement with those obtained in other studies investigating one-at-a-time and MC-based sensitivity analyses for the PELMO model [21]. Evaluation of the sum of squares of S(R)RC indices indicated that the 10 most influential parameters were expected to explain 95% of the variance associated with predicted concentrations on the ranked data, whereas this figure was 35% when nonranked data were used. This evaluation constitutes only a first-step assessment of the variance contribution, because it only focuses on linear aspects of the input–output relation (Eqn. 1). A complete analysis of the variance contribution of the individual parameters would require more dedicated and computationally intensive approaches [22]. Results for the ranked data suggest that the inconsistent ranking of the remaining parameters could be considered to have little overall relevance, unless the interest was specifically in the influence of these remaining parameters.

The present investigations demonstrated that the stability in the ranking of model input parameters according to their influence on model predictions is dependent on the ranking of the parameter itself. For those parameters to which the model is more sensitive, sensitivity ranking will not be influenced by the use of different seed numbers in replicated samples. In contrast, the seed number used in the sampling will severely affect the ranking of those parameters with a lesser influence on model predictions. Replicability issues in sensitivity analysis, which have also been reported by Saltelli and Homma [23] as well as by Helton et al. [24], will tend to decrease as the number of model runs increases.

Replicability in MC uncertainty assessments

Monte Carlo modeling exercises were conducted for different numbers of model runs and repeated 10 times for different seed number values used in generation of the LHS. Variability in the probability that the average annual concentration in the 20th year of simulation will exceed 0.1 μg/L is presented in Figure 5. The figure presents the minimum, average, and maximum of the exceedance probabilities as well as its coefficient of variation (CV) as estimated from the 10 replicated uncertainty analyses. Significant variabilities were obtained for all sample sizes considered, although it should be noted that variability estimates may be limited in their accuracy because of the small size of the random samples. The largest variabilities were obtained for the smallest sample sizes, and the CV for the largest number of PELMO runs considered (5,000 model runs) was 5.2%.

The experimental results can be supported by theoretical considerations on the asymptotic behavior of confidence intervals of probabilities to exceed a particular concentration. Estimating a probability of exceedance on the basis of a random sample is equivalent to estimating the probability of success in a binomial distribution. Using the results of the binomial test [25], the CV of the nonparametric exceedance probability estimator (Eqn. 2) can be expressed as follows under conditions of random sampling and asymptotic behavior

$$\text{CV} = \frac{\sqrt{\frac{1}{p} - 1}}{\sqrt{N}} \tag{3}$$

where p is the (unknown) true exceedance probability and N is the size of the random sample.

Applying Equation 3 to the results obtained experimentally (N = 5,000 runs, p = 0.0343) yields a CV of 7.5%, which agrees, to some extent, with the empirical results (estimated

CVs of 5.2%). Differences between empirical and theoretical results can be attributed to the facts that the estimations of CVs are empirical and based on 10 replications only; that empirical results are based on LHS, whereas calculations of theoretical figures assume random sampling; and that true exceedance probabilities are unknown. Iman and Helton [26] consider that satisfactory results with regard to the coverage of the range of each parameter for evaluating the uncertainty in model output can be obtained with LHS if the size of the sample (N) follows the equation $N > 4/3 \cdot p$, where p being the number of parameters to be sampled. The results obtained here suggest that this equation should not be used to infer stability in results of sensitivity analysis and in estimates of exceedance probabilities derived by uncertainty analysis. A much larger sample size will be required for these purposes.

DISCUSSION

Probabilistic modeling based on MC sampling currently receives much attention in the pesticide fate-modeling community as a possible means to account for the uncertainty and variability in model inputs (e.g., ECOFRAM, www.epa.gov/oppefed1/ecorisk; EUPRA [27]). The MC approaches to uncertainty estimation have been applied successfully in numerous fields of science [4]. The LHS is traditionally considered to be a very efficient sampling scheme, but the present research demonstrates that care should be exercised when minimizing the number of model runs for estimating exceedance probabilities. Uncertainties in the exceedance probability in excess of 5% (in terms of CVs) were noted when carrying out replicated uncertainty analyses for 5,000 PELMO runs, whereas uncertainties of approximately 20% to 30% were obtained when 500 model runs were carried out. Levels of variability reported will be of relevance to pesticide registration, especially if they are close to a threshold level used in decision making to grant authorization for placement of pesticides on the market. Although the present study focused on the uncertainty in the probability of exceeding a particular threshold concentration, the results suggest that similar issues of replicability will be encountered when considering specific percentiles in distributions.

Improvement in the robustness of exceedance estimates could be achieved by increasing the sample size (running models for a larger number of times), assessing the use of more efficient sampling schemes (e.g., importance sampling [28]), and investigating more efficient estimates of exceedance probabilities (e.g., [semi-]parametric methods, extreme value theory [29]). The former solution might be difficult to implement in practice, because pesticide fate models that integrate a complex description of hydrological and pesticide fate processes, such as the preferential flow model MACRO [8], typically require minutes to hours for one year to be simulated. The achievement of robustness in probabilistic estimates may therefore be hampered by the computational effort associated with running complex environmental models. Issues of replicability are expected to be of less importance for situations in which models are linear and amenable to a large number of runs and the focus is on estimating nontail probabilities.

CONCLUSIONS

The robustness of MC approaches for assessing model sensitivity and the associated variability/uncertainty in model predictions were investigated for the pesticide leaching model PELMO. Replicated analyses for different seed numbers and for different sample sizes suggest that sensitivity results obtained through LHS in combination with multiple linear regression are only stable for those parameters that most influence model predictions; that the estimation of exceedance probabilities from MC modeling is subject to uncertainty, which may affect subsequent decision making; and that stability of results can be improved by applying a rank transformation to model input and output (sensitivity analysis) and by increasing the size of the random sample (both sensitivity and uncertainty analyses). Although the present study focused on a pesticide leaching model, similar replicability issues can be expected for environmental models that share the same general characteristics (nonlinearity, large number of sensitive parameters, strong levels of model sensitivity, and long running times).

Further work is required to improve our understanding of the influence of using replicated random samples on MC predictions. Attention should be paid to replicability aspects by modelers when devising their approach to assessing the uncertainty associated with the modeling and by decision makers when examining the results of probabilistic approaches. The magnitude of the uncertainty associated with the use of different seed numbers in the sampling, however, should be placed in the wider context of the uncertainty associated with the overall approach to risk assessment.

Acknowledgement—The first author gratefully acknowledges the funding of this work by the United Kingdom Department for Environment, Food and Rural Affairs on behalf of the Pesticides Safety Directorate (Projects PL0532 and PL0548). Thanks are also due to the two anonymous reviewers for valuable comments and suggestions on the manuscript. Opinions expressed in this paper are those of the authors and may not reflect those of the study sponsor.

REFERENCES

1. Beven K, Binley A. 1992. The future of distributed models—Model calibration and uncertainty prediction. *Hydrol Process* 6: 279–298.
2. Helton JC. 1993. Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliability Engineering and System Safety* 42:327–367.
3. Morgan MG, Henrion M. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, New York, NY, USA.
4. Saltelli A, Chan K, Scott EM. 2000. *Sensitivity Analysis.* John Wiley, New York, NY, USA.
5. Carsel RF, Parrish RS, Jones RL, Hansen JL, Lamb RL. 1988. Characterizing the uncertainty of pesticide leaching in agricultural soils. *J Contam Hydrol* 2:111–124.
6. Soutter M, Musy A. 1999. Global sensitivity analyses of three pesticide leaching models using a Monte Carlo approach. *J Environ Qual* 28:1290–1297.
7. Ma L, Ascough JC, Ahuja LR, Shaffer MJ, Hanson JD, Rojas KW. 2000. Root zone water quality model sensitivity analysis using Monte Carlo simulation. *Trans ASAE* 43:883–895.
8. Dubus IG, Brown CD. 2002. Sensitivity and first-step uncertainty analyses for the preferential flow model MACRO. *J Environ Qual* 31:227–240.
9. Carbone JP, Havens PL, Warren-Hicks W. 2002. Validation of pesticide root zone model 3.12: Employing uncertainty analysis. *Environ Toxicol Chem* 21:1578–1590.
10. Klein M. 1995. User manual—*PELMO, Pesticide Leaching Model,* Ver 2.01. Fraunhofer-Institut für Umweltchemie und Ökotoxikologie, Schmallenberg, Germany.
11. U.S. Environmental Protection Agency. 1992. Pesticides in ground water database. A compilation of monitoring studies: 1971–1991. EPA 734R92001. National Summary. Washington, DC.
12. Decisioneering. 2000. *Crystal Ball 2000—User Manual*. Denver, CO, USA.
13. Palisade. 2000. *Guide to Using @RISK*. New York, NY, USA.

14. Iman RL, Conover WJ. 1979. The use of rank transformation in regression. *Technometrics* 21:499–509.
15. Statsoft. 2002. *Statistica System Reference.* Tulsa, OK, USA.
16. Hamby DM. 1994. A review of techniques for parameter sensitivity analysis of environmental models. *Environ Monit Assess* 32:135–154.
17. Saltelli A, Marivoet J. 1990. Nonparametric statistics in sensitivity analysis for model output—A comparison of selected techniques. *Reliability Engineering and System Safety* 28:229–253.
18. Dubus IG, Brown CD, Beulke S. 2003. Sensitivity analyses for four leaching models. *Pest Manag Sci* 59:962–982.
19. Helton JC, Davis FJ. 2000. Sampling based methods. In Saltelli A, Chan K, Scott EM, eds, *Sensitivity Analysis.* John Wiley, New York, NY, USA, pp 101–153.
20. Helton JC, Davis FJ. 2002. Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Anal* 22:591–622.
21. Dubus IG, Brown CD, Beulke S. 2000. Sensitivity analyses for leaching models used for pesticide registration in Europe. SSLRC Research Report PL0532. Cranfield University, Silsoe, UK.
22. Chan K, Tarantola S, Saltelli A. 2000. Variance-based methods. In Saltelli A, Chan K, Scott EM, eds, *Sensitivity Analysis.* John Wiley, New York, NY, USA, pp 167–197.
23. Saltelli A, Homma T. 1992. Sensitivity analysis for model output. Performance of black box techniques on three international exercises. *Computational Statistics and Data Analysis* 13:73–94.
24. Helton JC, Johnson JD, McKay MD, Shiver AW, Sprung JL. 1995. Robustness of an uncertainty and sensitivity analysis of early exposure results with the MACCS reactor accident consequence model. *Reliability Engineering and System Safety* 48:129–148.
25. Conover WJ. 1971. *Practical Nonparametric Statistics.* John Wiley, New York, NY, USA.
26. Iman RL, Helton JC. 1985. A comparison of uncertainty and sensitivity analysis techniques for computer models. NUREG/CR-3904, SAND84-1461. Sandia National Laboratories, Albuquerque, NM, USA.
27. EUPRA. 2001. Probabilistic risk assessment for pesticides in Europe. Implementation and research needs. In Report of the European Workshop on Probabilistic Risk Assessment for the Environmental Impacts of Plant Protection Products, The Netherlands, June 2001. Central Science Laboratory, York, UK.
28. Glasserman P, Heidelberger P, Shahabuddin P. 2001. Efficient Monte Carlo methods for value-at-risk. In Alexander C, ed, *Mastering Risk,* Vol 2—Applications. Financial Times-Prentice Hall, London, UK.
29. Bassi F, Embrechts P, Kafetzaki M. 1998. Risk management and quantile estimation. In Adler RJ, Feldman RE, Taqqu MS, eds, *A Practical Guide to Heavy Tails: Statistical Techniques and Heavy Tails.* Birkhaeuser Boston, Basel, Switzerland, pp 111–130.