SETAC PRESS

Environmental Toxicology and Chemistry, Vol. 25, No. 8, pp. 2227–2236, 2006

Printed in the USA
0730-7268/06 $12.00 + .00

USER SUBJECTIVITY IN MONTE CARLO MODELING OF PESTICIDE EXPOSURE

SABINE BEULKE,*† COLIN D. BROWN,†‡ IGOR G. DUBUS,§ HECTOR GALICIA,‖ NICHOLAS JARVIS,#
DIETER SCHAEFER,†† and MARCO TREVISAN‡‡
†Central Science Laboratory, Sand Hutton, York YO41 1LZ, United Kingdom
‡Environment Department, University of York, Heslington, York YO10 5DD, United Kingdom
§Bureau de recherches géologiques et minières (BRGM), Water Division, avenue Claude Guillemin, BP 6009, 45060 Orléans Cedex 2, France
‖Springborn Smithers Laboratories AG, 9326-Horn, Switzerland
#Department of Soil Science, Swedish University of Agricultural Sciences, Box 7014, 750 07 Uppsala, Sweden
††Bayer CropScience, Metabolism and Environmental Fate, 40764 Monheim, Germany
‡‡Instituto di Chimica Agraria ed Ambientale, Facoltà di Agraria, Università Cattolica del Sacro Cuore, 29100 Piacenza, Italy

(*Received* 13 *May* 2005; *Accepted* 20 *January* 2006)

Abstract—Monte Carlo techniques are increasingly used in pesticide exposure modeling to evaluate the uncertainty in predictions arising from uncertainty in input parameters and to estimate the confidence that should be assigned to the modeling results. The approach typically involves running a deterministic model repeatedly for a large number of input values sampled from statistical distributions. In the present study, six modelers made choices regarding the type and parameterization of distributions assigned to degradation and sorption data for an example pesticide, the correlation between the parameters, the tool and method used for sampling, and the number of samples generated. A leaching assessment was carried out using a single model and scenario and all data for sorption and degradation generated by the six modelers. The distributions of sampled parameters differed between the modelers, and the agreement with the measured data was variable. Large differences were found between the upper percentiles of simulated concentrations in leachate. The probability of exceeding 0.1 μg/L ranged from 0 to 35.7%. The present study demonstrated that subjective choices made in Monte Carlo modeling introduce variability into probabilistic modeling and that the results need to be interpreted with care.

Keywords—Monte Carlo User subjectivity Pesticide exposure modeling Degradation Sorption

INTRODUCTION

Current environmental risk assessments for pesticide registration in Europe rely on the comparison between a calculated exposure and an ecotoxicological endpoint (surface water) or a legal threshold concentration (groundwater). Mathematical models, including the Pesticide Leaching Model (PELMO) [1,2], Pesticide Root Zone Model (PRZM) [3], Pesticide Emission Assessment at Regional and Local Scales (PEARL) [4], and MACRO (http://www.mv.slu.se/bgf/Macrohtm/macro43b/TechMacro43.pdf), often are used to calculate exposure concentrations. Traditionally, deterministic approaches have been applied in which a single combination of model input parameters is used to calculate a single set of predicted environmental concentrations. The parameter combination often is selected to be protective of the actual range of use conditions. Such deterministic ''realistic worst-case'' approaches are useful at the lower tiers of the regulatory assessment process, because they are relatively quick to deploy and act as a screening step. However, the likelihood that the predicted risk will occur under real environmental and usage conditions is not assessed, and the degree of environmental protection provided by deterministic worst-case approaches is not very well quantified. Therefore, interest has increased concerning the use of probabilistic techniques in environmental risk assessment for pesticides to quantify better both the likelihood and the magnitude of the risk involved, such as the Ecological Committee on Federal Insecticide, Fungicide, and Rodenticide Act Risk Assessment Methods (ECOFRAM), European Workshop on Probabilistic Risk Assessment for the Environmental Impacts of Plant Protection Products (EUPRA; http://www.eupra.com/report.pdf), and European Framework for Probabilistic Risk Assessment of the Environmental Impacts of Pesticides (EUFRAM).

Uncertainty in exposure modeling arises from various sources [5], including the spatial and temporal variability in factors influencing pesticide behavior [6] and the incertitude associated with the measurement, calculation, or estimation of input parameters [7]. The most widely used technique to account for this uncertainty in pesticide fate modeling is the Monte Carlo approach [8–12], although other techniques also have been applied [13–16]. The Monte Carlo approach involves running a model iteratively for a large number of different input values or modeling scenarios followed by a statistical analysis of the model output. The parameter values are sampled from statistical distributions. A Monte Carlo analysis allows the user to evaluate the uncertainty in model predictions arising from uncertainty in the input parameters and to estimate the confidence that should be assigned to the modeling results. An end result of such probabilistic assessments often is the likelihood and frequency of exceeding a threshold environmental concentration [17,18].

Dubus et al. [19] (http://www.silsoe.cranfield.ac.uk/ecochemistry/publications/papers/pl0548.pdf) investigated the merits and shortcomings of Monte Carlo modeling for predicting pesticide exposure from a conceptual, technical, and registration point of view. Those authors found that the subjective choices made during the implementation of this technique can influence the outcome of the analysis. These choices

* To whom correspondence may be addressed (s.beulke@csl.gov.uk).

include the type of statistical distribution attributed to model parameters, the upper and lower limits of the distribution within which samples are taken (truncation), the specification of dependencies and/or correlations between parameters, the tool and method used for sampling, and the number of samples generated from the distributions. The relative importance of the various assumptions made by different users to the result of probabilistic risk assessments depends on the situation at hand. Previous work has suggested that correlation and truncation may have a significant impact [19] (http://www.silsoe.cranfield.ac.uk/ecochemistry/publications/papers/pl0548.pdf), but the possible influence of other subjective choices should not be overlooked.

The influence of user subjectivity on the results of probabilistic modeling was investigated further in the present study. User subjectivity refers to the subjective choices that are made when using an existing model in a probabilistic way. The influence of subjective assumptions made during the development of the model was not evaluated. Six pesticide fate modelers who were experienced with the implementation of Monte Carlo approaches in the regulatory context were provided with data concerning degradation and sorption of metamitron (4-amino-4,5-dihydro-3-methyl-6-phenyl-1,2,4-triazin-5-one) measured in 18 soils [20]. Each modeler was asked to analyze these data, to assign statistical distributions, and to sample values from these distributions. A leaching assessment was then carried out using a harmonized scenario and modeling protocol (e.g., simulation time, output generated) for each of the six sets of sampled degradation and sorption data.

No comparison was made between simulated concentrations in leachate and measured data from lysimeter or field studies. Metamitron was only used as an example compound to illustrate the implications of user subjectivity in Monte Carlo sampling. Results from probabilistic assessments that are carried out for regulatory purposes should, however, be verified wherever possible.

MATERIALS AND METHODS

Ring test

The aim of the present study was to investigate the influence of user subjectivity on the results of probabilistic modeling of pesticide fate. The likelihood of an example pesticide (metamitron) leaching through soil was assessed based on the variability in its sorption and degradation properties. Six modelers participated in the ring test (i.e., a study in which the results from different, independently working participants who followed the same set of instructions are compared). Each participant received data concerning degradation and sorption of metamitron measured in laboratory studies with 18 soils having a range of properties [20] (Table 1). The soils were incubated at 20°C and a moisture content equivalent to a tension of −33 kPa. The authors of the original paper [20] derived first-order degradation rate constants (k) and Freundlich sorption coefficients from their experimental results. Degradation rate constants were converted to 50% disappearance times (DT50 values; i.e., the time at which the amount of pesticide has declined to half its initial level) for the present study as $DT50 = \ln(2)/k$. Sorption coefficients normalized to organic carbon content (K_{oc}) were calculated from the Freundlich coefficients (K_f) and organic carbon contents of the soils as $K_{oc} = (K_f/\%\ \text{organic carbon}) \cdot 100$. Information regarding the Freundlich exponent was not available, and the Forum for the Co-Ordination of Pesticide Fate Models and Their Use (FOCUS) default value of 0.9 was used [21]. The variability in the laboratory values provided to the modelers originated from the use of soils with different textures, pH values, organic carbon contents, and microbial activities [20]. Although the modelers were not provided with the original paper, no explicit restrictions were given. One modeler derived relationships between pesticide and soil properties from the published data and accounted for these relationships in the sampling of modeling input parameters.

Table 1. Sorption and degradation data for metamitron provided to the participants of the ring test[a]

	K_f[b] (L/kg)	OC[b] (%)	K_{oc} (L/kg)	k[b] (/d)	DT50 (d)
Soil 1	1.5	1.4	107	0.014	49.5
Soil 2	2.1	1.2	175	0.020	34.7
Soil 3	2.4	1.9	126	0.027	25.7
Soil 4	7.6	2.0	380	0.019	36.5
Soil 5	0.77	0.7	110	0.065	10.7
Soil 6	1.0	1.4	71	0.042	16.5
Soil 7	1.3	1.7	76	0.029	23.9
Soil 8	6.6	2.2	300	0.025	27.7
Soil 9	2.2	1.3	169	0.028	24.8
Soil 10	3.3	1.5	220	0.014	49.5
Soil 11	1.4	1.2	117	0.017	40.8
Soil 12	2.9	2.4	121	0.034	20.4
Soil 13	5.0	2.0	250	0.029	23.9
Soil 14	3.5	1.4	250	0.016	43.3
Soil 15	4.5	2.2	205	0.017	40.8
Soil 16	7.5	2.2	341	0.024	28.9
Soil 17	1.5	0.6	250	0.027	25.7
Soil 18	2.2	1.5	147	0.033	21.0
Minimum	0.77	0.6	71	0.0140	10.7
Maximum	7.6	2.4	380	0.0650	49.5
Mean	3.2	1.6	190	0.0267	30.2
Median	2.3	1.5	172	0.0260	26.7
Standard deviation	2.2	0.51	91	0.0122	11.1
Coefficient of variation (%)	69	32	48	46	37

[a] K_f = Freundlich sorption coefficient; OC = organic carbon content; K_{oc} = sorption coefficients normalized to organic carbon contents; k = degradation rate constant; DT50 = 50% disappearance time.
[b] Values taken from Allen and Walker [20].

The herbicide metamitron is approved by most European Member States for use in a range of crops. Sugar beets, one of the major crops, are treated with this herbicide at rates of up to 3.5 kg active ingredient/ha. The compound was chosen for the present study because of the availability of degradation and sorption data from the open literature, and it was only used as an example to illustrate the influence of user subjectivity on probabilistic leaching assessments.

Each modeler was asked to analyze these data regarding sorption and degradation of metamitron, to assign a statistical distribution to these data, to sample sorption and degradation data from the distributions, and when necessary, to convert these data to K_{oc} values and degradation rate constants for direct use as input values for the leaching assessment. No further instructions were given regarding the assumptions, methods, or tools to be used in the present study. A leaching assessment was then carried out for each of the six sets of input data.

Model selection and scenario simulated

The potential for metamitron to leach to groundwater was investigated using the same model and scenario for each set

of input values provided by the modelers. The stand-alone version of the pesticide fate model PELMO 3.00 Service Pack 2 [1,2] was used to simulate leaching of metamitron to depth. The PELMO model is a one-dimensional leaching model that describes water movement through the soil column using a capacitance approach. Solute transport is simulated using the convection–dispersion equation. Descriptions of pesticide sorption and degradation and of pesticide losses via runoff, soil erosion, and volatilization are included in the model.

The combination of input parameters sampled by the modelers was entered into the PELMO model one by one, and the model was run deterministically every time. The model reads input parameters from a file in text format. The SENSAN (sensitivity analysis) utility provided with the PEST software (Model Independent Parameter Estimation, Watermark Computing, Corinda, QLD, Australia) was used to automatically paste the K_{oc} values and degradation rate constants provided by each modeler into the PELMO input file, run the PELMO model, process the output data, and store the annual average concentrations in leachate for each of the runs together with the input parameters. The PELMO model does not run in a probabilistic way, and all input parameters must be sampled a priori outside the model. Correlation between the parameters or truncation also was included a priori in the sampling of the input values.

Leaching of metamitron was simulated on the basis of a soil, weather, and crop scenario that is commonly used within pesticide registration in Europe and is provided with the PELMO model. The Borstel soil is a sandy loam over sand (U.S. Department of Agriculture classification) with relatively low organic carbon contents throughout the profile (1.5, 1.0, 0.2, and 0% at 0–30, 30–60, 60–75, and >75 cm, respectively). Weather data from the standard scenario Hamburg normal (year 1978; annual rainfall, 777 mm) represent average weather conditions in northern Germany [1]. These data for the single year were repeated to give a total simulation period of 20 years. Metamitron was assumed to be applied to a sugar beet crop on May 1st in each of the 20 years at a rate of 3 kg/ha. Routines for losses via runoff, erosion, and volatilization were turned off in the model, because the exercise concentrated on the simulation of leaching.

The model output of interest was the average annual concentration in leachate at a depth of 1 m for the 20th year of the simulation period. This endpoint is normally compared with a threshold concentration of 0.1 μg/L within the context of pesticide registration in Europe. Simulated average annual concentrations in leachate tend to increase from one year of the simulation to the next until a plateau concentration is reached. A relatively long simulation period of 20 years was selected to ensure that the plateau concentration was reached by the end of the model run for all combinations of input parameters. Annual average concentrations in leachate for each of the 20 years were calculated from the simulated mass of metamitron in leachate and the volume of leachate.

RESULTS AND DISCUSSION

Data analysis

All modelers analyzed the data regarding metamitron sorption and degradation for correlations and selected statistical distributions. Correlation coefficients between degradation rate constants and K_f values (−0.35), degradation rate constants and K_{oc} values (−0.38), DT50 values and K_f values (0.20), and DT50 values and K_{oc} values (0.26) were not significant at the 10% probability level.

Five of the six modelers evaluated the fit of different distributions and retained either the normal or log-normal distribution. Figure 1 shows histograms of the measured K_{oc} values, degradation rate constants, and DT50 values as well as the fitted distributions. The histograms for K_{oc} values and degradation rate constants were skewed to the left, and compared with a normal distribution, a log-normal distribution gave a somewhat better fit to these data (Fig. 1). The DT50 values for metamitron were described only slightly better by a log-normal distribution.

All modelers made a subjective decision regarding whether to truncate the distributions (i.e., to exclude values outside a certain range). Truncation is a means of avoiding the sampling of extreme values from the tails of the distribution. The modelers truncated the distribution if they considered the extreme values to be unrealistic.

Sampling of degradation and sorption parameters

Degradation and sorption parameters were sampled from the selected distributions. Table 2 summarizes the choices made by the six modelers.

Figure 2 shows cumulative frequency distributions of sampled K_{oc} values and DT50 values of metamitron (either sampled directly from a distribution or calculated from sampled degradation rate constants) produced by the six modelers. The frequency distribution of the measured values is shown for comparison. Scatter plots of sampled sorption versus degradation data are shown in Figure 3. Details of the methodology chosen by each modeler are given below.

Based on the lack of a significant correlation between K_f values and degradation rate constants or between K_{oc} values and degradation rate constants, modeler 1 sampled sorption and degradation parameters of metamitron independently (i.e., no correlation was included in the sampling). Normal and log-normal distributions were fitted to the K_{oc} values and the degradation rate constants. In both cases, the log-normal distributions gave a better fit. The log mean and log standard deviation of the log-normal distribution of the degradation rate constants were −3.71 and 0.400, respectively. The log mean and log standard deviation of the distribution of K_{oc} values were 5.13 and 0.496, respectively. The distributions were not truncated. Combinations of degradation rate constants and K_{oc} values were then sampled from the log-normal distributions with the random number generator in Microsoft Excel® 97 (Microsoft, Redmond, WA, USA). Sets of 100, 1,000, and 10,000 values were sampled from each distribution. The coefficient of variation of the 5th and 95th percentiles of the sampled values between 10 consecutive samplings was approximately 10% for samples of 1,000 values; therefore, 1,000 model runs were considered to be sufficient for a robust leaching assessment. The results are shown in Figures 2 and 3.

Modeler 2 analyzed the correlation between degradation and sorption data. Because the correlations were weak and not significant at the level of $p < 0.1$, it was decided not to include correlation in the sampling. A normal and a log-normal distribution were then fitted to the K_{oc} values and DT50 values using the Crystal Ball 2000 software (Decisioneering, Denver, CO, USA). A log-normal distribution was chosen based on the Anderson-Darling index for both parameters. The log mean and log standard deviation of the distribution of K_{oc} values were 5.13 and 0.49, respectively. The distribution was trun-

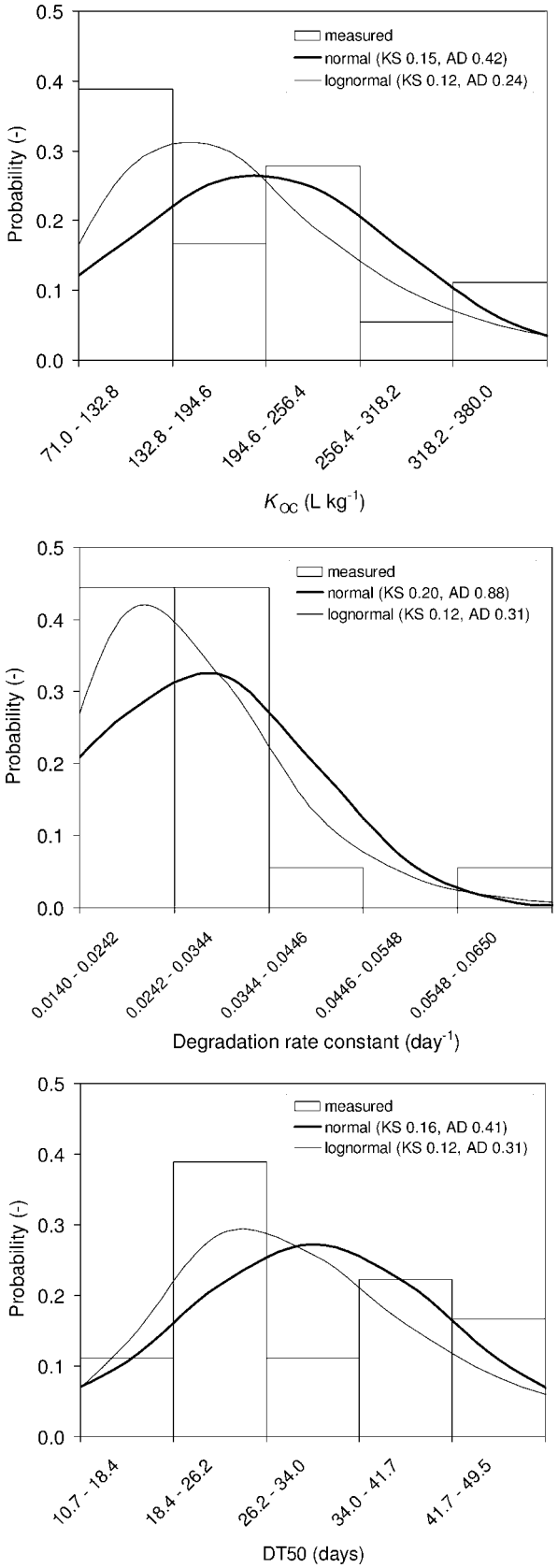

Fig. 1. Histograms of measured K_{oc} values (sorption coefficients normalized to organic carbon content), degradation rate constants, and 50% disappearance times (DT50 values), and fitted normal and log-normal distributions. The Anderson-Darling (AD) and Kolmogorov-Smirnov (KS) indices are statistical indices for the goodness of fit. Smaller values indicate a better fit.

cated at the 0.5th and 99.5th percentile (corresponding to K_{oc} values of 47.4 and 607.0 L/kg, respectively). These percentiles were chosen such that the measured values were included in the sampling interval. The log mean and log standard deviation of the distribution of DT50 values were 3.34 and 0.40, respectively. The 0.5th and 99.5th percentile cutoff values were 10.1 and 78.9 d, respectively. In total, 5,000 combinations of K_{oc} values and DT50 values were sampled with the Crystal Ball software using the Latin Hypercube Sampling method. This method divides the distribution into intervals of equal probability and then samples from each interval. The cumulative frequency curves of K_{oc} values and DT50 values sampled by modeler 2 were similar to those obtained by modelers 1 and 3 (Fig. 2).

Modeler 3 sampled 150 degradation rate constants and K_{oc} values from correlated log-normal distributions ($r = -0.39$) using a software package for sensitivity and uncertainty analysis (UNCSAM; National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands) [22]. The log-normal distribution was selected based on the Shapiro-Wilks test for normality of untransformed (test for normal distribution) and log-transformed (test for log-normal distribution) degradation rate constants and K_{oc} values. The parameters of the distributions were derived by fitting a normal distribution to log-transformed data. The mean and standard deviation of the distribution of the log-transformed degradation rate constants were -1.61 and 0.173, respectively. The mean and standard deviation of the distribution of log(K_{oc}) values were 2.23 and 0.215, respectively. Modeler 3 truncated the distributions at the log mean $\pm$ 2.58 $\times$ log standard deviation (i.e., 1st and 99th percentiles). The cumulative frequency curves of K_{oc} values and DT50 values sampled by modeler 3 were similar to those obtained by modelers 1 and 2 (Fig. 2). Figure 3 shows the influence of introducing correlation into the sampling. Combinations of small K_{oc} values and long DT50 values were sampled less frequently by modeler 3 than by modelers 1 and 2.

Modeler 4 investigated the fitting of a range of statistical distributions to K_{oc} values and degradation rate constants using graphical, quantile–quantile, and goodness-of-fit tests. Modeler 4 considered that none of the fits obtained was adequate from a statistical perspective. The modeler adopted an approach similar to those used in elicitation. This approach did not aim to find a distribution that matched the 18 measured data as closely as possible but, rather, used the more limited information that is commonly available from regulatory studies (mean or median K_{oc} values and DT50 values). The modeler assigned a triangular distribution with three parameters to the degradation rate constants (minimum, 0.008/d; maximum, 0.074/d; likeliest, 0.026/d) and K_{oc} values (minimum, 31.3 L/kg; maximum, 434.2 L/kg; likeliest, 172.0 L/kg). The likeliest value of the triangular distribution (i.e., the mode) was set to the median of 18 measurements, because the median often is assumed to be the most representative statistic for sorption and degradation data. The minimum of the distribution was selected so that the probability of sampling a value smaller than the measured minimum was 2.78%. Similarly, the maximum was set such that the probability of sampling a value larger than the measured maximum was 2.78%. Thus, the total probability of sampling a value outside the measured range was one divided by 18, with 18 being the number of measurements.

Modeler 4 selected this approach of setting the minimum and maximum values of the triangular distribution according

Table 2. Assumptions made by the six modelers to support the sampling of values for degradation and sorption parameters[a]

	Modeler 1	Modeler 2	Modeler 3	Modeler 4	Modeler 5	Modeler 6
Variables considered	K_{oc}, k	K_{oc}, DT50	K_{oc}, k	K_{oc}, k	K_{oc}, DT50	K_{oc}, DT50
Distributions assigned	Log normal	Log normal	Log normal	Triangular	K_{oc} log-normal, clay content triangular, DT50 from regression[b]	Normal
Selection criterion/goodness-of-fit statistics	Kolmogorov-Smirnov	Anderson-Darling	Shapiro-Wilk test for normality of log K_{oc} and log k	Statistical criteria, expert judgment	Anderson-Darling	Shapiro-Wilk test for normality of K_{oc} and DT50
Original paper consulted	No	No	No	No	Yes	No
Correlation	No	No	Yes	No	Not applicable[b]	No
Truncation	No	Yes	Yes	Not applicable	Yes (K_{oc})	No
Sample size	1,000	5,000	150	5,000	1,000	400
Software used	MS Excel[c]	Crystal Ball[d]	UNCSAM[e]	@RISK[f]	Crystal Ball[d]	MS Excel[c]
Sampling method	Random	Latin Hypercube	Latin Hypercube	Latin Hypercube	Latin Hypercube	Random

[a] K_{oc} = sorption coefficients normalized to organic carbon contents; k = degradation rate constant; DT50 = 50% disappearance time.
[b] A linear regression was used to calculate DT50 values from K_{oc} values and clay content.
[c] Microsoft, Redmond, WA, USA.
[d] Decisioneering, Denver, CO, USA.
[e] National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.
[f] Palisade, Newfield, NY, USA.

Fig. 2. Cumulative frequency distributions of K_{oc} values (sorption coefficients normalized to organic carbon content; top) and 50% disappearance times (DT50 values; bottom) sampled by the six modelers (solid lines) and frequency distribution of the measured data (open circles [20]).

to the inverse of the number of data points available because it offers the advantage of being adapted to the size of datasets. If a small number of data points are available, then the probability of obtaining values outside the range defined by the minimum and maximum observed is considered to be large. Conversely, for large datasets, the probability of seeing a value outside the minimum–maximum values observed is considered to be small. A total of 5,000 values of each parameter were sampled using the software @RISK (Palisade, Newfield, NY, USA). The number of runs was selected on the basis of literature information reported for previous results acquired with PELMO [23], given that a detailed study of convergence, such as that reported by Ballio and Guadagnini (http://www.agu.org/pubs/crossref/2004/2003WR002876.shtml), was considered to be outside the scope of the present exercise. The resulting distributions of sampled values were shifted toward shorter DT50 values and larger K_{oc} values compared with the measurements and with the values sampled by modelers 1, 2, and 3 (Fig. 2). This results from the fact that the likeliest value was set to the median measured value. In triangular distributions, the median value splits the total area of the triangle into two equal parts. The likeliest value (or mode) is the tip

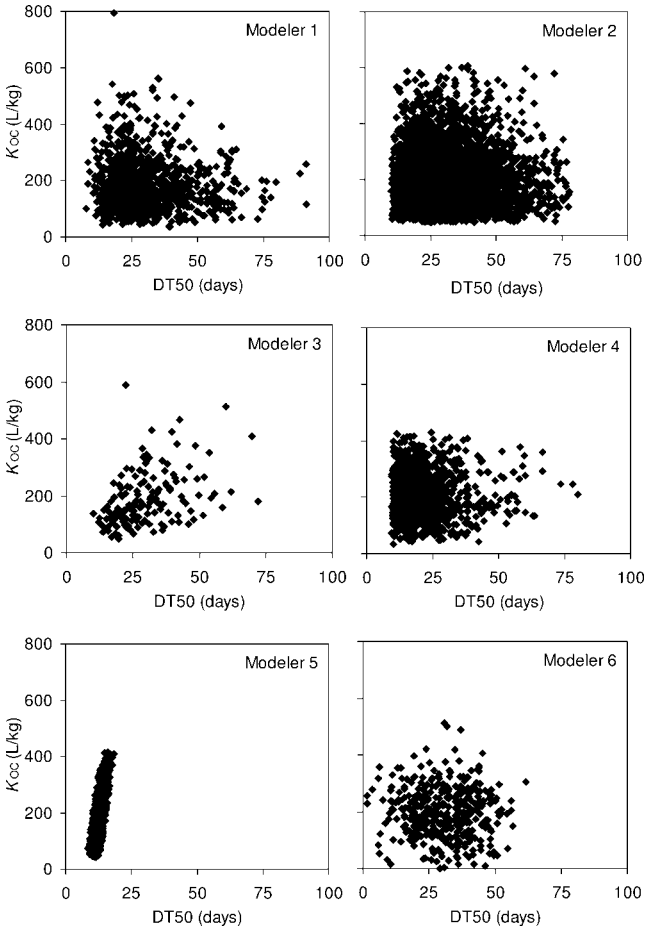

Fig. 3. Scatter plots of K_{oc} values (sorption coefficients normalized to organic carbon content) versus 50% disappearance times (DT50 values) sampled by the six modelers.

of the triangle. The median is larger than the likeliest value in distributions that are skewed to the left, is identical to the likeliest value in symmetrical distributions, and is smaller than the likeliest value in distributions that are skewed to the right. The likeliest value can be calculated theoretically from the minimum, maximum, and median of the distribution. Using the median of the measured K_{oc} values and the minimum and maximum values chosen by modeler 4, the likeliest value was calculated to be 93 L/kg, whereas the elicitation approach adopted by modeler 4 resulted in a K_{oc} of 172 L/kg. This caused the shift toward larger K_{oc} values. A significant discrepancy was found between the distribution of measured degradation rate constants and the triangular distribution specified by modeler 4, and a larger proportion of larger degradation rate constants (corresponding to shorter DT50 values in Figs. 2 and 3) was sampled by modeler 4 than by modelers 1, 2, 3, and 6.

In contrast to the other modelers and based on previous work with clay minerals [24,25], modeler 5 analyzed these data in the original publication [20] and proposed a relationship between the DT50 values for metamitron, the K_{oc} values, and the clay content of the soils tested:

$$\mathrm{DT50} = 3.677 + 0.897\mathrm{Clay} + 0.0147K_{oc}$$

where the multiple correlation coefficient (R) is 0.6326. A F test ($\alpha = 0.05$, one-sided; df (regression) = 2, df (residual) = 15, $n = 18$) indicated a statistically significant dependence of the DT50 values on clay content and K_{oc} values at $p < 0.025$ with this model. However, a Student's t test on the partial regression coefficients indicated that only the partial regression coefficient for clay was significantly different from zero ($\alpha = 0.05$, one-sided; $df = 15$, $n = 18$). The two independent variables (clay content and K_{oc}) were not highly correlated with each other ($r^2 = 0.0574$); thus, the interpretation of the two above-mentioned partial regression coefficients was considered to be justified [26]. An analysis of residuals indicated two distinct groups of data. Regression analyses were carried out for each of these two groups, and it was concluded that the regression line using all data gave the best approximation of DT50 values to extrapolate to small clay contents.

Modeler 5 assumed that the clay content of the Borstel soil for which the leaching assessment was undertaken was uncertain (i.e., experimental and natural variability were considered by allowing the clay content in the Borstel soil to vary by ±3% [i.e., ~42% relative variation]). The modeler sampled 1,000 values for the clay content from a triangular distribution with a likeliest value of 7%, and a minimum value of 4%, and a maximum value of 10% using Crystal Ball. Next, 1,000 K_{oc} values were sampled from a truncated log-normal distribution (mean, 191.8 L/kg; standard deviation, 101.3 L/kg; minimum, 38.3 L/kg; maximum, 418.5 L/kg). The DT50 was then calculated for each run from the sampled clay content and K_{oc} value using the regression equation given above.

The combinations of DT50 values and K_{oc} values generated by modeler 5 were much less scattered compared with those provided by the other modelers (Figs. 2 and 3). The range of DT50 values (8.6–18.5 d) was smaller than the range of the values sampled by the other modelers and also smaller than the range of measured DT50 values. It should, however, be noted that the clay contents of the soils included in the original study were between 15 and 41% [20]. The clay content of the Borstel soil (4–10%) is smaller and outside this range. The DT50 values calculated for the Borstel soil by modeler 5 are, therefore, extrapolations and cannot be compared directly with the measured values.

The reason for the relatively small range of sampled values may be that the unexplained error term in the regression was ignored. This will exaggerate the strength of the relationship between DT50, clay content, and K_{oc} value and reduce the randomness in the sampled DT50 values. An example of how functional relationships, such as the equation used by modeler 5, can be utilized in probabilistic Monte Carlo leaching assessments for pesticides is given by Lindahl et al. [27]. The authors accounted for the error terms in the regression equations used to estimate model parameters as random, normally distributed variables.

Modeler 6 assumed that sorption and degradation parameters of metamitron are independent based on correlation analysis, and no correlation was included in the sampling. Modeler 6 found that DT50 values and K_{oc} values were normally distributed based on the Shapiro-Wilks test (significance level, $p < 0.05$). He sampled from normal distributions with a mean DT50 of 30.2 d (standard deviation, 11.1 d) and a mean K_{oc} value of 190 L/kg (standard deviation, 91 L/kg). The distributions were not truncated. For each variable set for a given simulation, 400 values were generated by random sampling using a spreadsheet-based Monte Carlo approach (Microsoft Excel 2002; Microsoft). The modeler derived the sample size from the following equation [28]:

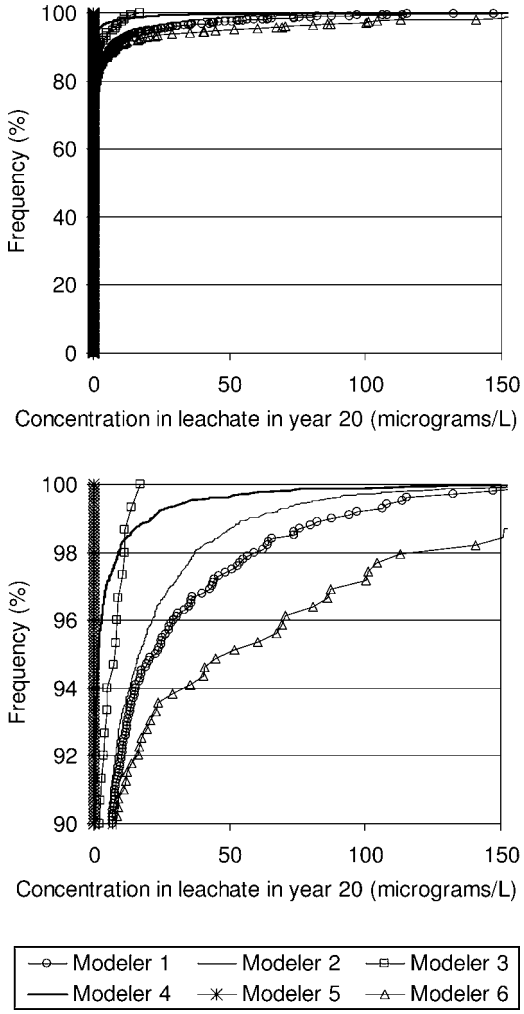

Fig. 4. Cumulative frequency distributions of simulated maximum annual average concentrations in leachate at 1-m depth generated from input data sampled by the modelers (top, full distributions; bottom, only 90th to 100th percentile shown).

$$n = 4pq/L^2$$

where n is the sample size, p is the event probability (here, the probability of chemical leaching), q equals $100 - p$, and L is the accepted error (100 − confidence interval). The number of samples calculated from this equation is largest when $p = q = 50$. For a confidence interval of 95%, the sample size was calculated to be 400 iterations. One of the sampled DT50 values and 10 of the K_{oc} values were negative. All 11 combinations in which either the DT50 or the K_{oc} value was negative were excluded from the analysis, and only 389 combinations of DT50 values and K_{oc} values were used in the leaching modeling. Figure 1 shows that the normal distributions selected by modeler 6 fit these data less well compared to the log-normal distributions chosen by modelers 1, 2, and 3.

Probabilistic leaching assessment

Figure 4 shows cumulative frequency curves for maximum annual average concentrations of metamitron in leachate at a 1-m depth simulated with PELMO using the six sets of sorption and degradation input parameters. Selected statistics, percentiles, and probability of exceeding the European regulatory threshold concentration of 0.1 μg/L are given in Table 3.

Simulated concentrations in leachate and the probability of exceeding 0.1 μg/L differed considerably between the modelers, and discrepancies between the various cumulative distribution functions were largest at the upper percentiles (Fig. 4 and Table 3). The effects of assumptions made in Monte Carlo modeling also were stronger at the upper percentiles in a study by Dubus et al. [19] (http://www.silsoe.cranfield.ac.uk/ecochemistry/publications/papers/pl0548.pdf). Simulated concentrations based on data provided by modeler 5 were much smaller than those for the remaining five sets of input data and did not exceed 0.001 μg/L for any of the 1,000 model runs. Modeler 5 sampled clay contents of the Borstel soil and K_{oc} values for metamitron and calculated DT50 values for each run from these two variables using a regression equation derived from the experimental data. Because the error term from the regression was ignored, the resulting range of DT50 values (8.6–18.5 d) was much smaller than that for the remaining sets of input parameters derived from sampled distributions (Figs. 2 and 3).

Concentrations at the upper percentiles calculated using these data provided by modeler 4 were smaller than those for modelers 1, 2, and 6 and also were smaller than those for modeler 3 except for the upper 2% of values (Fig. 4 and Table 3). This results from the fact that the distributions of parameters sampled by modeler 4 were shifted toward shorter DT50 values and larger K_{oc} values (Fig. 2).

Modeler 3 was the only participant in the ring test who included a correlation between the degradation and sorption parameters. As a result, combinations of small K_{oc} values and long DT50 values were sampled less frequently by modeler 3 than by modelers 1, 2, and 6 (Fig. 3). The model simulated

Table 3. Percentiles and statistics of the cumulative frequency distributions for simulated maximum annual average concentrations in leachate at 1-m depth and probabilities of exceeding 0.1 μg/L generated from input data sampled by the modelers

	Modeler 1	Modeler 2	Modeler 3	Modeler 4	Modeler 5	Modeler 6
50th Percentile	0.004	0.004	0.003	<0.001	<0.001	0.002
80th Percentile	1.2	0.88	0.39	0.009	<0.001	0.990
90th Percentile	7.0	5.6	2.0	0.25	<0.001	8.4
95th Percentile	22.8	16.5	7.8	1.7	<0.001	48.9
99th Percentile	87.8	57.2	12.7	21.7	<0.001	214.2
Exceedance probability (%)[a]	35.7	33.2	27.0	12.8	0.0	29.4
Mean	4.2	3.0	0.92	0.83	<0.001	8.6
Standard deviation	16.3	12.2	2.7	6.6	0.004	36.0
Coefficient of variation (%)	390	408	297	803	1080	421

[a] Probability that the concentration in leachate exceeds 0.1 μg/L.

greater potential for the pesticide to leach to depth for these extreme combinations. Concentrations at the upper percentiles of the cumulative frequency curve thus were larger when the parameters were sampled from uncorrelated distributions (Fig. 4 and Table 3). The 50th percentile concentration in leachate and the probability of exceeding 0.1 μg/L were similar with or without correlation. The correlation between K_{oc} and the degradation rate coefficient was only significant at $p = 0.12$ and, therefore, might be justifiably ignored. However, even weak correlations that are not significant at the confidence levels normally considered in statistics may have important effects on leaching model predictions [19] (http://www.silsoe.cranfield.ac.uk/ecochemistry/publications/papers/pl0548.pdf), and it is not clear that such correlations should be ignored. This illustrates one important subjective choice in probabilistic risk assessments for leaching.

The 95th and 99th percentile concentration in leachate calculated from these data provided by modeler 6 were much larger than those calculated for the remaining five sets of input parameters. This results from the parameters being sampled from untruncated normal distributions that generated a number of very small K_{oc} values (Fig. 1). The upper percentiles of the output distribution were strongly influenced by truncation of the input distributions.

Modelers 1 and 2 sampled K_{oc} values and DT50 values from almost identical log-normal distributions. This resulted in very similar combinations of sampled K_{oc} and DT50 values. However, modeler 1 sampled from untruncated distributions, whereas modeler 2 used the 0.5th and 99.5th percentiles as cutoff values. A larger number of small K_{oc} values in combination with long DT50 values thus were sampled by modeler 1, giving somewhat larger simulated concentrations at the upper percentiles (Figs. 3 and 4).

Modeler 3 sampled 150 values from the distributions. The distributions of the measured K_{oc} values and DT50 values were matched as well by the 150 sampled values as they were by the much larger number of values sampled by modeler 2 (5,000), as shown in Figure 1. However, simulated concentrations in leachate are determined by a combination of both K_{oc} values and DT50 values. Therefore, whether the number of combinations may have influenced the outcome of the present study was investigated. Dubus and Janssen [23] demonstrated that the predicted risk for pesticide leaching can differ considerably between simulations using different sets of parameter values sampled repeatedly from the same distribution with the Latin Hypercube Sampling method, even when 5,000 model runs were undertaken. The variability decreased with increasing sample sizes. In the present study, modeler 2 sampled 5,000 parameter combinations using the same sampling method as modeler 3. The probability of exceeding 0.1 μg/L and the 95th percentile concentration in leachate were calculated from 2, 3, 4, . . . , 5,000 consecutive model runs to evaluate the influence of sample size on the model outcome. The exceedance probability changed considerably up to approximately 50 model runs (Fig. 5). The changes in the exceedance probability were smaller from this point onward. The 95th percentile concentration was almost constant from approximately 450 model runs.

It is difficult to determine an adequate sample size objectively. Modeler 6 selected the sample size using an equation based on statistical considerations [28]. This equation can give an initial estimate. However, the sample size that is required for a robust assessment also depends on the number of combined parameters, the modeling scenario, the model output of interest, and the sampling method. Methodologies for investigating aspects of convergence in modeling have been reported (http://www.agu.org/pubs/crossref/2004/2003WR002876.shtml). Modelers 1 and 6 used conventional random sampling, whereas modelers 2, 3, 4, and 5 used Latin Hypercube Sampling. The latter method divides the distribution into intervals of equal probability and then samples from each interval. Latin Hypercube Sampling is more precise than conventional random sampling for producing random samples, because the full range of the distribution is sampled in a more consistent manner. Thus, fewer trials are required with Latin Hypercube Sampling to achieve the same accuracy. In practice, the number of model runs that can be undertaken often is limited by computing times, and the Latin Hypercube Sampling method can be advantageous in these cases.

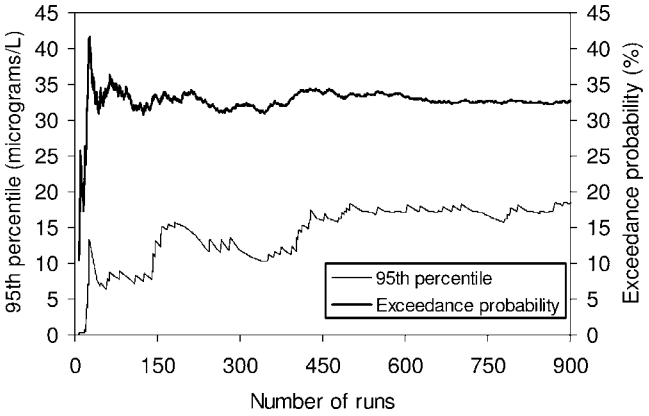

Fig. 5. The 95th percentile concentration in leachate and the probability of exceeding 0.1 μg/L calculated from an increasing number of consecutive model runs (input data generated by modeler 2).

CONCLUSION

The present study investigated the influence of user subjectivity in Monte Carlo sampling on the results of probabilistic pesticide exposure modeling. Six modelers made subjective decisions regarding the type of statistical distribution assigned to degradation and sorption data for an example pesticide, the upper and lower limits of the distribution within which samples were taken (truncation), correlation between parameters, the tool and method used for sampling (random vs Latin Hypercube Sampling), and the number of samples generated. All modelers had already undertaken Monte Carlo modeling in the past; therefore, the present results are believed to reflect real-life applications of the technique. Also, it is believed that the present results are applicable to a range of pesticides and scenarios in addition to the examples tested here.

User subjectivity can significantly influence the results of probabilistic exposure assessments. Therefore, detailed documentation and justification of the approach used is important to ensure the transparency and reproducibility of such assessments. User subjectivity can be reduced by standardizing protocols for Monte Carlo sampling. However, in the context of pesticide exposure modeling, no guidance agreed to by all stakeholders is currently available. This results, in part, from the fact that generalized, objective criteria for decision making are difficult to identify. For example, the number of measurements available for uncertain or variable pesticide parameters, such as K_{oc} values and DT50 values, usually is small. Statistical

tests often are not adequate to evaluate the goodness of fit of different types of distributions to small samples [29]. Therefore, the decision regarding the type of distribution and the parameterization of the distribution assigned to these data sometimes must be based on visual assessment and expert judgment, which may lead to significant differences in the results of probabilistic modeling exercises.

Uncertainties in probabilistic risk assessments because of user subjectivity and other factors are expected to introduce uncertainties, particularly at the extreme ends of the curves. This was confirmed by the present study, in which the effects of subjective choices made by the different modelers were most evident for high leaching losses with a low probability of occurrence. Differences in assigning correlations between parameters and truncation of the input distributions contributed to the discrepancies. Both truncation and correlation can be useful to restrict sampling to realistic parameter combinations, and they should be introduced when appropriate. The uncertainty in regulatory endpoints derived from probabilistic modeling is significantly smaller for lower percentiles (e.g., the 80th rather than the 90th percentile or larger). It should, however, be noted that rare occurrences of large concentrations can be relevant for regulators and other stakeholders. The percentile on which the regulatory decision is based depends on the situation at hand, and agreed-on levels of risk do not exist.

Deterministic exposure assessments for pesticides often are based on a single combination of model parameters, but the likelihood of this combination occurring under realistic conditions is not evaluated in such assessments. The level of protection afforded by deterministic assessments is unknown. In contrast, probabilistic modeling considers a range of possible values for uncertain parameters and allows better quantification of the magnitude and likelihood of the risk involved. Therefore, probabilistic modeling often is considered to be of greater accuracy than deterministic assessments. However, the present study demonstrates that the subjective assumptions and choices made in Monte Carlo modeling introduce uncertainty into probabilistic risk assessments, and the results need to be interpreted with care.

Acknowledgement—This project funded, in part, by the United Kingdom Department for Environment, Food, and Rural Affairs within project PL0548; the Italian national project Pesticide Fate Models Use: Regional Scenario Definition and Uncertainty Analysis; and the European-funded Framework 6 AquaTerra project (Integrated Modeling of the River–Sediment–Soil–Groundwater System: Advanced Tools for the Management of Catchment Areas and River Basins in the Context of Global Change; project GOCE-CT-2004-505428).

REFERENCES

1. Klein M. 1995. *PELMO: Pesticide Leaching Model, User Manual,* Ver 2.01. Fraunhofer-Institut für Umweltchemie und Ökotoxikogie, Schmallenberg, Germany.
2. Jene B. 1998. *PELMO 3.00 Manual Extension.* Staatliche Lehr- und Forschungsanstalt für Landwirtschaft, Weinbau und Gartenbau, Neustadt an der Weinstrasse, Germany.
3. Carsel RF. 1984. *PRZM-3, a Model for Predicting Pesticide and Nitrogen Fate in the Crop Root and Unsaturated Soil Zones: User's Manual for Release 3.0.* U.S. Environmental Protection Agency, Athens, GA.
4. Tiktak A, van den Berg F, Boesten JJTI, van Kraalingen D, Leistra M, van der Linden AM. 2000. *Manual of FOCUS PEARL version 1.1.1.* RIVM report 711401 008, Alterra report 28. RIVM, Bilthoven, The Netherlands.
5. Dubus IG, Brown CD, Beulke S. 2003. Sources of uncertainty in pesticide fate modeling. *Sci Total Environ* 317:53–72.
6. Rao PSC, Wagenet RJ. 1985. Spatial variability of pesticides in field soils: Methods for data analysis and consequences. *Weed Sci* 33:18–24.
7. Loague KM, Green RE. 1991. Statistical and graphical methods for evaluating solute transport models: Overview and application. *J Contam Hydrol* 7:51–73.
8. 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 (Am Soc Agric Eng)* 43:883–895.
9. Dubus IG, Brown CD. 2002. Sensitivity and first-step uncertainty analyses for the preferential flow model MACRO. *J Environ Qual* 31:227–240.
10. 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.
11. Miao Z, Trevisan M, Capri E, Padovani L, Del Re AAM. 2004. Uncertainty assessment of the model RICEWQ in Northern Italy. *J Environ Qual* 33:2217–2228.
12. Trevisan M, Vischetti C. 2005. Assessment of uncertainty associated with the extent of simulation processes from point to catchment: Application to 1D-pesticide leaching models. In Alvarez-Benedi J, Rafael Munoz-Carpena R, eds, *Soil-Water-Solute Process Characterization: An Integrated Approach.* CRC, Boca Raton, FL, USA, pp 673–692.
13. Loague K. 1991. The impact of land use on estimates of pesticide leaching potential: Assessments and uncertainties. *J Contam Hydrol* 8:157–175.
14. Fontaine DD, Havens PL, Blau GE, Tillotson PM. 1992. The role of sensitivity analysis in groundwater risk modeling for pesticides. *Weed Technol* 6:716–724.
15. Freissinet C, Erlich M, Vauclin M. 1998. A fuzzy logic-based approach to assess imprecisions of soil water contamination modeling. *Soil Tillage Research* 47:11–17.
16. Cryer SA, Applequist GE. 2003. Direct treatment of uncertainty: II—Applications in pesticide runoff, leaching, and spray drift exposure modeling. *Environ Eng Sci* 20:169–181.
17. Aldenberg T, Jaworska JS, Traas TP. 2002. Normal species sensitivity distributions and probabilistic ecological risk assessment. In Posthuma L, Suter GW II, Traas TP, eds, *Species Sensitivity Distributions in Ecotoxicology.* Lewis, Boca Raton, FL, USA, pp 49–102.
18. Brown CD, Dubus IG, Fogg P, Spirlet M, Gustin C. 2004. Exposure to sulfosulfuron in agricultural drainage ditches: Field monitoring and scenario-based modeling. *Pest Manag Sci* 60: 765–776.
19. Dubus IG, Brown CD, Beulke S, Turner NL. 2002. Uncertainty and probabilistic approaches to pesticide fate modeling. Cranfield University report to DEFRA for project PL0548. Study number JA3755E. Cranfield University, Silsoe, UK.
20. Allen R, Walker A. 1987. The influence of soil properties on the rates of degradation of metamitron, metazachlor and metribuzin. *Pestic Sci* 18:95–111.
21. Forum for the Co-ordination of Pesticide Fate Models and Their Use. 2000. FOCUS groundwater scenarios in the EU review of active substances—The report of the work of the Groundwater Scenarios Workgroup of FOCUS (Forum for the Co-Ordination of Pesticide Fate Models and Their Use). Sanco/321/2000 Rev. 2. Brussels, Belgium.
22. Janssen PHM, Heuberger PSC, Sanders R. 1992. UNCSAM 1.1: A software package for sensitivity and uncertainty analysis. Technical report 959101004. National Institute of Public Health and Environmental Protection, Bilthoven, The Netherlands.
23. Dubus IG, Janssen PHM. 2003. Issues of replicability in Monte Carlo modeling: A case study with a pesticide leaching model. *Environ Toxicol Chem* 22:3081–3087.
24. Campbell GD, Galicia HF, Schindler PW. 1987. Binding of cadmium by montmorillonite-humic acid mixtures: Miscible displacement experiments. *Aust J Soil Res* 25:391–403.
25. Galicia HF, Campbell GD, Schindler PW. 1989. Distribution of Cd(II) between (K,H)-montmorillonite and an aqueous solution of 0.10 M KNO_3 at pH 4.5 and 298 K—Comparison between potentiometric and miscible displacement experiments. *Geoderma* 189–202.

26. Zar JH. 1984. *Biostatistical Analysis,* 2nd ed. Prentice-Hall, NJ, USA, pp 333–345.
27. Lindahl A, Kreuger J, Stenström J, Gärdenäs A, Alavi G, Roulier S, Jarvis N. 2005. Stochastic modeling of diffuse pesticide losses from a small agricultural catchment. *J Environ Qual* 34:1174–1185.
28. Snedecor GW, Cochran EG. 1989. *Statistical Methods,* 8th ed. Iowa State University Press, Ames, IA, USA.
29. U.S. Environmental Protection Agency. 2000. Guidance for data quality assessment. EPA QA/G-9, QA00 update. EPA/600/R-96/084. Washington, DC.