Pest Manag Sci **58**:745–758 (online: 2002)
DOI: 10.1002/ps.526

Review

Calibration of pesticide leaching models: critical review and guidance for reporting

Igor G Dubus,* Sabine Beulke and Colin D Brown
Cranfield Centre for EcoChemistry, Cranfield University, Silsoe, Beds MK45 4DT, UK

Abstract: Calibration of pesticide leaching models may be undertaken to evaluate the ability of models to simulate experimental data, to assist in their parameterisation where values for input parameters are difficult to determine experimentally, to determine values for specific model inputs (eg sorption and degradation parameters) and to allow extrapolations to be carried out. Although calibration of leaching models is a critical phase in the assessment of pesticide exposure, lack of guidance means that calibration procedures default to the modeller. This may result in different calibration and extrapolation results for different individuals depending on the procedures used, and thus may influence decisions regarding the placement of crop-protection products on the market. A number of issues are discussed in this paper including data requirements and assessment of data quality, the selection of a model and parameters for performing calibration, the use of automated calibration techniques as opposed to more traditional trial-and-error approaches, difficulties in the comparison of simulated and measured data, differences in calibration procedures, and the assessment of parameter values derived by calibration. Guidelines for the reporting of calibration activities within the scope of pesticide registration are proposed.

Keywords: pesticide leaching; validation; evaluation; extrapolation; inverse modelling; guidelines

1 INTRODUCTION

Environmental exposure to crop protection products is traditionally assessed using a range of tools, including laboratory, lysimeter and field experiments, and the use of computer simulation models. Although the use of computer models in pesticide registration is an attractive option in terms of temporal, financial and manpower resources when compared with experimentation,[1] modelling is not sustainable on its own and experimental work is necessary. There is an intimate, although complex, relationship between models and experimental data. Extrapolation using predictive models may act as a substitute for field studies, since experimental investigation of the fate of crop-protection products for multiple locations and climatic conditions is impractical. However, experimental data are essential for model development, for evaluating the accuracy of models in the description of field behaviour, and thus for assessing the confidence that should be placed in model predictions. Calibration of fate models against experimental data is hence often at the heart of exposure assessment for crop protection products, especially at higher tiers.

Despite the complexity of pesticide leaching models in use and the large number of model input parameters that could be varied, the required activities for calibration are often given little consideration.[2] The calibration process is left to the discretion of the modeller and thus an *ad hoc* approach is adopted. There have been numerous calls for the development of guidelines in relation to modelling[1,3,4] to decrease the uncertainty and the large user-subjectivity associated with the use of pesticide leaching models.[4,5] Codes of 'Good Modelling Practice' have been proposed by Görlitz[6] and by Estes and Coody.[7] Good Modelling Practices were defined as 'the development, maintenance, distribution and use of computer simulation models whereby the integrity of the model, its various improvements and utilisation is assured'.[7] These documents provide a general framework for ensuring the quality, consistency and integrity of the models,[3] but do not provide guidelines on either the model parameterisation *per se* or on calibration.

The development of detailed modelling guidelines that are broadly applicable is a difficult task given the heterogeneity of modelling situations. Resseler *et al*[8] have issued recommendations for performing modelling studies for registration purposes, but these are mainly relevant to the German registration context. CAMASE, an EU-funded workgroup, has issued

* Correspondence to: Igor G Dubus, Cranfield Centre for EcoChemistry, Cranfield University, Silsoe, Beds MK45 4DT, UK
E-mail: i.dubus@cranfield.ac.uk
Contract/grant sponsor: Pesticides Safety Directorate; Department for Environment, Food and Rural Affairs; contract/grant number: PL0539
(*Received 20 November 2001; revised version received 19 February 2002; accepted 2 April 2002*)

general guidelines for modelling which cover the evaluation of models, sensitivity and uncertainty analyses and calibration.[9] The guidance for calibration was intended to be applicable to a large range of environmental models and this resulted in the derivation of general concepts rather than specific guidelines. More recently, the FIFRA Environmental Model Validation Task Force has issued a report containing guidance information for calibrating leaching and run-off models.[10] The report proposes some basic principles for calibration and identifies those parameters to be varied within the validation work undertaken by the Task Force. Although these two guidance documents for calibrating pesticide leaching models emphasise the need for a high quality report of calibration activities for improved transparency and reproducibility, detailed information that should be made available in calibration reports is not listed. Given the importance of written communication in the pesticide review process, the development of guidelines for reporting calibration activities appears desirable. The development of guidance for reporting is also expected to be useful in that indirect guidance for the performance of calibration can be suggested.

The present paper presents a critical review of the use of calibration and calibration procedures in pesticide fate modelling, and proposes guidelines for reporting calibration activities within the context of pesticide registration.

2 THE USE OF CALIBRATION IN MODELLING THE ENVIRONMENTAL FATE OF PESTICIDES

Calibration of pesticide leaching models may be undertaken for a range of purposes which broadly fall under four categories: (1) model parameterisation, (2) 'validation' of models or of the use of models, (3) extrapolation and (4) targeted parameter estimation.

2.1 Model parameterisation

Most pesticide leaching models were initially developed as research tools to describe the fate of compounds in heavily instrumented field or laboratory experiments. For this reason, there has been little emphasis on the use of parameters which can readily be derived from easily measured data or on the development of procedures to support the parameterisation of a model for cases where few data are available. This has restricted the extensive use of detailed mechanistic models.[11] In some instances, the derivation of adequate values for input parameters relies on the fitting of a relationship to experimental data. Examples include the derivation of DT_{50} values from laboratory degradation data or the derivation of parameters of the van Genuchten or Brooks and Corey equations from water-release data. Where such an independent assessment is not possible, parameter values may be attributed by calibration of the whole model ('indirect fitting')[12] or by 'expert judgement' where the experience and knowledge of the modeller prevail. Deterministic pesticide leaching models require a detailed set of theoretical parameters, due to the highly complex and variable character of the natural soil–plant–atmosphere conditions which are simulated.[13] Processes affecting the fate of pesticides in soil and water are numerous and difficult to characterise in terms of effective parameters. Some of the parameters integrated into the models cannot easily be measured or determined. Some authors have thus questioned the ability of pesticide leaching models to predict the fate of organic chemicals in the environment with acceptable accuracy, and argue that a calibration against measured data is always necessary to simulate the leaching of solutes.[13] The requirement for calibration appears particularly important for preferential flow models.[14,15]

2.2 Model testing or model 'validation'

The testing of a model against experimental data is an essential activity that contributes to estimating the confidence that should be assigned to the predictions of the model. Such evaluations have been carried out for pesticide leaching models used for pesticide registration in Europe and the USA.[4,14,16–18] The testing of the capacity of a model to describe or predict reality has often been referred to as 'model validation' or 'model verification'[19–21] even though it is demonstrated that complex environmental models cannot be proven or validated, but only tested and invalidated.[22,23] The terminology used is misleading with regard to the confidence that should be assigned to the models, and wording such as 'model testing' or 'model evaluation' is more appropriate.[22] Procedures for evaluating models have ranged from blind simulations where no calibration is carried out[24] to approaches where calibration is at the heart of the testing exercise.[3] Although some authors evaluated a number of pesticide leaching models using predictive simulations only,[5,17] the combination of blind and calibrated simulations in model evaluation has been the most common approach in recent years.[3,14,25] Testing based on blind simulations will assess the accuracy of models where a potential use without calibration is expected. Blind simulations will provide an assessment of the model as well as the associated parameterisation, whereas controlled calibrated simulation can be considered a truer test of the inherent capability of the model to represent field data. Loague[26] suggested an evaluation approach in which a solute transport model is first calibrated against field data from a specific period by adjusting input parameters until an acceptable fit is achieved and then run for a different time using the calibrated parameter set. The model is deemed validated if an 'acceptable' fit is found between the model predictions and the experimental data for the second period.[27] However, a successful calibration of a model against experimental data could imply either that the model structure and the parameter values are both realistic, or that they are both unrealistic but compensate for one another.[28] Some

authors consider that the only appropriate way of evaluating the accuracy and performance of a model is to attempt to predict the measured data with values for all parameters obtained independently (blind simulation).[29] However, the parameterisation of complex pesticide leaching models often requires the use of expert opinion or pedotransfer functions to select values for some of the numerous input parameters required.[30] A blind simulation will therefore not only test the model, but other aspects of the parameterisation as well, such as the data that were used to derive input values, the expert judgement of the modeller or the quality of the pedotransfer functions.[31,32]

2.3 Extrapolation

Field and lysimeter experiments represent a major financial commitment for agrochemical companies wishing to register a new compound, and modelling is commonly used to maximise the return on these studies. A possible approach is the use of experimental data to calibrate a leaching model and then use of the calibrated set of input parameters to make predictions for different environmental conditions (extrapolation). Extrapolations to radically different scenarios (eg between contrasting soil types or different climatic regions) are questionable given the uncertainty in the modelling, and extrapolations are thus most often limited to small deviations from the calibrated set. A common form of extrapolation is the calibration of a model against field data for a number of years and then the running of the model for longer time series for the same site.[33] The approach is considered of most interest when unusual weather conditions have been experienced during a field study.[34] However, it has been demonstrated that the use of a single set of input parameters for different agricultural seasons might not lead to a good description of results over the entire experimental period,[15,26,35,36] especially under circumstances where droughts are experienced.[13] Other examples of limited extrapolation are the use of an application rate different from that used in the calibration[11,33] and the simulation of leaching deeper than the profile depth used during calibration.[37]

Two opposite views coexist with regard to the extrapolation of results from a calibrated set to different conditions. For some, successfully calibrating a model demonstrates its ability to simulate a specific set of conditions and allows one to extrapolate to other points in space and time.[38] For others, the calibration of complex deterministic models tends to be specific to the conditions at the site for which experimental data were collected and no extrapolation should be carried out. Russell *et al*[39] considered that no extrapolation should be carried out without prior calibration, while Vanclooster *et al*[4] judged that calibration should be avoided wherever possible.

2.4 Targeted parameter estimation

Degradation and sorption parameters are environmental fate variables of particular importance in the registration of pesticides. Half-lives and sorption distribution coefficients are typically derived from controlled experiments in the laboratory, but there is continuing debate as to whether these are suitable for describing the field behaviour of compounds.[40] A possible supplement to laboratory determinations is to use data measured in field or lysimeter experiments to estimate sorption and degradation parameters through calibration of a pesticide leaching model.[41–43] The approach consists in calibrating a pesticide leaching model against field data such as pesticide residue profiles in soil, concentrations in drainflow or concentrations in lysimeter leachate, thereby back-estimating sorption and degradation parameters. Although the calibration could be carried out using a traditional trial-and-error approach, it is often performed automatically using parameter estimation packages such as PEST,[44] UCODE[45] or SUFI.[46] The general approach of estimating values for input parameters through model calibration is commonly referred to as 'inverse modelling'.[47]

3 ISSUES ASSOCIATED WITH THE CALIBRATION OF PESTICIDE LEACHING MODELS

3.1 Data requirements

Most of the data used for calibration in registration modelling are collected in field and lysimeter experiments. The amount of information collected in the field depends on the purpose of the study and differs significantly between field experiments carried out for research and those performed for regulatory purposes. Information should not be collected in the field only because it has been collected in the past[26] and Diekkrüger *et al*[48] suggested that more effort is put into the improvement of field measurement techniques rather than into the development of new models. Data which are useful for a modeller when simulating the results of a field study are site-specific meteorological data (including those necessary to calculate potential evapotranspiration using the Penman–Monteith equation: air temperatures, wind speed, humidity, sunshine hours or solar radiation); a detailed soil profile description, including soil structure; basic soil properties (such as organic carbon content, particle-size distribution, soil pH where the fate of ionisable compounds is simulated, bulk density); water retention properties of the soil; an assessment of soil variability at the field scale; the actual application rate of the compound and the proportion reaching the soil; an estimation of crop development; sorption and degradation parameters specific to the experimental soil (ideally at different depths); complete mass balances of water and solutes, including a non-reactive tracer (where possible); residue profiles; measurements of fluxes. Although these determinations represent additional costs, their absence would contribute to uncertainty in the modelling. Bromide is often used in field experiments carried out for research purposes as an inert (ie non-

degraded, non-sorbed) tracer and may provide information on soil hydrology and the extent of solute dispersion.[10] However, concerns about the suitability of bromide profiles to assess the transport components in models have recently been reiterated.[32]

Water and solute fluxes may be highly dynamic and change rapidly with time. Flux measurements (eg concentrations in drainflow or percolation) should thus be made at an adequate temporal resolution. A fine resolution is particularly required when preferential flow processes are an important pathway of transport or when volatilisation is to be estimated from the measurement of pesticide concentrations close to the surface in the hours and days following application.[30]

The nature, quantity and quality of data have a particular importance in calibration[49] as this will partly determine whether the calibration problem is 'ill-posed' or 'well-posed'.[50] Inverse modelling has been widely used in soil water physics to estimate soil hydraulic properties, such as the parameters of the van Genuchten–Mualem model,[51,52] from transient outflow experiments. If water outflow data are used on their own, the calibration problem is ill-posed and non-uniqueness issues are encountered. The inclusion of additional data such as water content or water tension will stabilise the inverse problem and allow a robust estimation of hydraulic parameters, provided the data are of quality.[53] The identification of the data requirements for an effective and robust calibration of pesticide leaching models (a 'well-posed' inverse problem) should be considered a research priority. Aspects of data quality are further discussed below.

3.2 Selection of a model for performing the calibration

The selection of a leaching model which is potentially capable of simulating the experimental data is essential if a calibration is to be carried out.[54] Although detailed information on individual models is usually widely available, little guidance is available to support model selection on the basis of objective criteria. Guidance such as that generated by Pennel *et al*[16] would be useful. These authors provided guidelines on the selection of a specific leaching model (CMLS, PRZM, LEACHP, MOUSE or GLEAMS) on the basis of the simulation of an experimental dataset using these five models. Del Re and Trevisan[55] identified a number of criteria for selecting models, but these were generic and cannot be used to select a specific model. The lack of guidance on model selection means that the choice of a model for parameterisation and calibration usually falls to the modeller undertaking the work. Differences between models used for pesticide registration in Europe have lessened in the last few years,[56] but they still present their own specificities and it is expected that this will lead to differences in predictions. The use of an inappropriate model will lead to a poor simulation of the data[57,58] and to the derivation of unrealistic values for input parameters where a calibration is carried out.[36] The role of model accuracy in limiting the end use of calibrated parameters should not be overlooked.

The choice of a model may be based on a number of decision criteria, including the objectives of the modelling and the availability of the data necessary to parameterise the model. For purposes of screening or general management guidance, the use of simpler models which are less data-intensive is justified.[59] For calibration purposes within the EU registration process, it is proposed that the main criterion for the choice of a specific model is the knowledge of the main processes affecting the fate of pesticides in the field context. A set of decision rules to choose one model from the four which are mainly used for pesticide registration in Europe is proposed in Table 1. A detailed description of the capabilities of the different models can be found elsewhere.[60] The decision criteria presented in Table 1 may not lead to the selection of a single model, and, in these instances, a decision should be made as to the most significant processes affecting the fate of crop-protection products. Preferential flow and pesticide volatilisation are both processes likely to dominate model selection

Table 1. Set of decision criteria to assist in the selection of a specific model for describing field data and performing calibrations. Only those leaching models selected by the FOCUS groundwater scenarios working group are considered. Brackets indicate that the use of the model is possible although the parameterisation is not straightforward

Decision criteria	*Model(s) suggested*
Accounting for pesticide losses by volatilisation	PEARL, PELMO, PRZM
Evidence or strong suspicion of a significant influence of preferential flow on water hydrology or pesticide loss	MACRO
Simulation of complex degradation schemes	PELMO, PEARL
Simulation of the fate of compounds susceptible to ionisation	PEARL, PELMO, (PRZM), (MACRO)
Simulation of the interaction between the unsaturated zone and the upper groundwater	PEARL
Need for an accurate description of soil hydrology	MACRO, PEARL
Simulation of lysimeter experiments	PEARL, MACRO, PELMO[a], PRZM[a]
Increase in sorption with time	PEARL, PRZM, PELMO, (MACRO)

[a] PELMO does not integrate a bottom boundary condition specific to the simulation of lysimeter flow, but the model has been considered capable of describing lysimeter datasets for coarse-textured soils.[17] Given the similarities between PELMO and PRZM, it is anticipated that this conclusion can be extended to PRZM.

where they have a significant impact on pesticide fate. A model which accounts for all processes affecting the fate of pesticides in the field is currently not available. The modeller therefore has to make concessions and select the least imperfect of those available. The rationale supporting the choice of a particular model should be carefully documented.

3.3 Critical assessment of the experimental data

A primary requirement for a successful calibration is that the experimental data which are used to calibrate against are of good quality. If this is not the case, then calibration should not be considered in the first place. The adequacy of the experimental data to be used for calibration should not be taken for granted, since sources of error and uncertainties in experiments investigating the fate of pesticides are potentially numerous.[61] Typical sources of uncertainty may include: the intrinsic variability in the field, the performance and adequacy of the sampling and measuring equipment, and the uncertainty associated with analytical determinations (limits of detection, definitive identification of analytes). Hence, although experimental data are traditionally considered to be certain, they can be largely uncertain and variable in reality,[57] and should be considered as such in the calibration.[10,36] Although critical assessments of the experimental data have only rarely been reported in the literature, such an assessment should be considered as a prerequisite to calibration and adequately reported. Particular attention should be paid to aspects of uncertainty, the quality of replication (where appropriate) and the presence of outliers in the dataset. Attempts to understand the reasons for large variability in replicated data should be made, and measures taken to address the variability in the data should be reported. Pennell *et al*[16] observed that the variability in replicated bromide and pesticide concentrations was large, and subsequently used the least variable depth to solute centre of mass to undertake model calibration. Within the scope of model evaluation, calibration carried out with a poor-quality experimental dataset could lead to the rejection of a good model or to the acceptance of a poor one.[12] In a context of parameter estimation and extrapolation, this could result in the derivation of parameter values with limited physical meaning and specific to the situation considered.

3.4 Choice of input parameters to be varied during the calibration

The rationale behind the selection of specific model input parameters to be varied during a calibration is rarely reported in the literature. Input parameters that need to be varied in the calibration are those that are both uncertain and have a strong influence on model output. The selection of parameters should therefore be based upon a combination of information on model sensitivity and parameter uncertainty.

The degree of influence of input parameters on model predictions can be assessed through a sensitivity analysis. Sensitivity analyses vary in complexity and include one-at-a-time sensitivity analyses, analyses based on random sampling, response surface methodology and Fourier amplitude sensitivity tests.[62] In the simplest and most common of these methods (one-at-a-time approach), each selected input parameter is varied and the impact of this variation on model output is scrutinised. Recently, sensitivity analyses have been carried out for the four leaching models used for pesticide registration in Europe, using four different scenarios and one-at-a-time and Monte Carlo approaches.[63,64] Information on the sensitivity of these leaching models has also been reported by Fontaine *et al*,[65] Jarvis,[66] Smith *et al*,[67] Boesten,[68] and Boesten and van der Linden.[69] Results of sensitivity analyses tend to depend on the initial scenario considered[70] and the sensitivity of the model also varies with the output considered (eg pesticide losses by leaching, drainflow or run-off). Consequently, although these studies give a general idea about the most influential parameters for a particular model and for particular scenarios, it is recommended that a limited sensitivity analysis be carried out if environmental conditions or pesticide properties in the modelling differ from those for which sensitivity information is available.

Some modellers consider that calibration should be restricted to those parameters which are non-measurable and those parameters for which site-specific measurements are not available.[11] It is the opinion of the authors that parameters for which site-specific estimations are available should nevertheless be allowed to vary in the calibration, albeit to a limited extent because their values are still uncertain (due to spatial variability and uncertainty arising from the laboratory and analytical procedures) and because they may not be representative of field behaviour.[40] An example of such procedure was presented by Klein *et al.*[71] Although field capacity was measured independently, it was allowed to vary in their calibration on the basis that field capacity had only operational significance. Parameter values determined from empirical relationships such as pedotransfer functions (eg for the determination of the water retention curve or the hydraulic conductivity at saturation) should be considered uncertain and may need to be included in the calibration if it is found that they have a significant influence on model predictions. Potential evapotranspiration (PET) data can be determined using a variety of equations which will lead to different estimates and should therefore be considered to be uncertain. PET data are expected to have a major impact on calculated water balances.[64]

There is general consensus on the need to use sensitivity and uncertainty information for selecting those parameters to be included in a calibration, but further research is required to identify the optimum number of parameters to be varied to allow a robust calibration of pesticide leaching models.

3.5 Trial-and-error *versus* automated calibration

Calibration of pesticide leaching models is traditionally performed in a non-automated way ('manual' or 'trial-and-error' calibration). This consists in manually modifying values for a small number of input parameters selected by the model user, running the model and examining the output files to see whether the modification led to a better description of the experimental data. This iterative procedure is repeated until the modeller is satisfied with the improvement in the fit between model predictions and experimental data. Manual calibration offers advantages where data are sparse and of poor quality, and where expert judgement is required to assess the reasonableness of parameter estimates. However, manual calibration also suffers from a number of shortcomings, including the subjectivity in a visual assessment of the fit between measured and predicted data,[72] the subjectivity in making the decision to end the calibration,[42] the difficulty in dealing with the calibration of more than two parameters at a time,[2] the lack of statistical information on the calibrated parameters,[47] the lack of explicit assessment of the confidence that should be assigned to the calibration,[73] and the tedious and time-consuming aspects of this process.[74,75] Furthermore, when a mismatch between data and model is obtained, it is difficult to know if this originates from model deficiencies or from an incomplete adjustment of the parameters.[2]

Software packages enabling the automation of the calibration process for complex models were developed in the early 1990s and are now widely used, especially in the fields of groundwater flow modelling[47] and soil water physics.[53] The principle consists in the minimisation of an objective function (usually the weighted sum of squared residuals between measured data and model predictions) through the modification of selected input parameters in an iterative process. Modification of the model input is based on a variety of non-linear estimation algorithms, such as the steepest descent, Gauss–Newton, Gauss–Marquardt–Levenberg and simplex procedures, which aim at keeping the number of iterations to a minimum. Prior information on the parameters, including limits to their variation can be integrated into the calibration. Stand-alone packages such as PEST[44] or UCODE[45] can be linked to virtually any DOS model (including the leaching models used for pesticide registration in Europe) without the need for modification of the model code. The packages will take control of the entire calibration process by running the model, examining the discrepancy between model output and experimental data, and adjusting selected input parameters. These steps are repeated until an optimised fit between model predictions and experimental data is achieved and statistical information on the quality of the calibration is generated. Examples of the application of automated calibration procedures to pesticide leaching models have been reported.[41–43] Guidelines for calibrating models using automated techniques have been provided by Hill.[76]

The choice of a particular mode of calibration (manual *versus* automatic) should be left to the modeller, but it should be thoroughly justified. It is the opinion of the authors that pesticide fate modellers should be encouraged to use automatic techniques for calibration, as this can help to establish the confidence to be assigned to calibrated values. It is essential, however, that the modeller remains active in the calibration through critical evaluation of all stages of the process.

3.6 Difficulties in the comparison of model output and experimental data

Investigations of the fate of a crop protection product in the field can involve a wide range of measurements. Data which are traditionally used in model calibration within regulatory modelling are (1) pesticide residues in soil for different times and depths (field leaching and dissipation studies), (2) concentration of pesticides at a given depth in water extracted by suction samplers (field leaching study), (3) drainflow and concentration of pesticides in drainflow (field drainflow study) and (4) water flow at the bottom of lysimeters, concentrations of pesticides in the leachate, total loss by leaching and final soil residue profile (lysimeter study). Additional data may comprise the distribution of soil moisture within the profile, measurements of water tension and height of any water table. Ideally, the observations should comprise fluxes (ie water flows and pesticide concentrations) as well as mass balances for water and solutes. Flux measurements enable the identification of preferential flow phenomena and the comparison of model output to patterns, peak magnitude, values at each individual sampling point or values integrated over a time period. Flux measurements for water permit an independent assessment of the hydrological component of a model.[77]

A direct comparison between the measured data and outputs from leaching models is rarely possible and, in most instances, model predictions will need post-processing before the model can be calibrated. A number of solutions to the difficulties encountered when comparing model output and pesticide residues, suction sampler, lysimeter or drainflow data are proposed below.

3.6.1 Pesticide residues in the soil profile

Pesticide residue data consist of the amount of the compound at different depths in the profile. The parameterisation of the profile in a modelling exercise requires the definition of layers, but those may differ from the layers used in the field sampling (pesticide residues are typically sampled in 5- to 10-cm increments in the topsoil). Where a match between modelling and sampling depth cannot be obtained, weighted averages of model outputs for each model-

ling layer will be required to enable a comparison between model predictions and experimental values.

3.6.2 Suction samplers

Although suction samplers are widely used to assess the leaching potential of pesticides in the field, they provide a challenge to the modeller who wishes to simulate the field data collected. Suction samplers extract water from a soil volume which is related to the suction applied and the characteristics of the soil.[78] In contrast, pesticide leaching models produce an output value for a layer comprised between two different depths (the value reported is usually the average for the layer) and a direct comparison between suction sampler data and model output is thus not valid. Spatial averaging of the model output across different layers is necessary, but there is a difficulty related to its extent since the radius of the soil volume of extraction may lie between 0.1 and 0.5 m.[79]

3.6.3 Lysimeters

Lysimeters are hydrologically isolated soil cores instrumented to allow the collection of the water flowing at their base. Samples of leachate are collected on discrete dates, often with irregular time intervals between sampling points. Any pesticide concentrations measured are integrations over time, with the integration time varying from one sample to the next. Most pesticide leaching models produce output on a regular time-step (hour, day, month or year) and are not designed to output integrated concentrations over time. A direct comparison between concentrations measured in lysimeter leachate and the model output for the sampling date is thus an inadequate procedure. Model outputs for flow and pesticide leaching must be integrated over time between the last and the actual sampling date by accumulation and calculation of a flow-weighted average pesticide concentration, respectively.[80]

3.6.4 Field drainflow studies

Field drainflow studies monitor the flow of water and concentrations of pesticides at the outlet of a field drainage system. Water flow is usually monitored continuously using an automatic flow meter, whereas water samples are generally collected at irregular intervals for pesticide determination and quantification. Since a pesticide leaching model produces output with a regular time-step, the comparison with total loadings estimated in the drainflow study is difficult. The modeller needs to make assumptions on the pattern of pesticide concentrations between sampling occasions. The modeller may assume stable concentrations between the two sampling times (at the concentration for the first sample) or a linear interpolation of concentrations between successive samples.[27]

3.7 Visual *versus* numerical assessment of fit

The calibration of models against experimental data is based on an iterative procedure where input parameters are varied at each iteration. Reasons for stopping the calibration may include the achievement of a fit between measured and simulated data which is considered 'acceptable', or inability to improve the fit any further. The goodness-of-fit (or lack-of-fit) of the model predictions to the measured data may be assessed graphically or by using a range of indices. Graphical displays are typically used when a trial-and-error calibration is carried out. These most often plot (1) changes in a variable as a function of time or depth and (2) the measured data against the simulated data. Although they are useful for showing trends, types of error and distribution patterns,[72] the level of adequacy between simulated and measured data to be considered 'acceptable' is user-dependent, and this limits the use of such displays. Also, graphical displays may not be adequate for examining the discrepancy between model and simulated data or for revealing problems with models as demonstrated by Kirchner *et al*[28] using two simple linear models.

A number of numerical indices have been used to try to reduce the subjectivity introduced by the modeller into the evaluation of model performance. These include the total error (TE), the maximum error (ME), the root mean square error (RMSE), the scaled root mean square error (SRMSE), the coefficient of determination (CD), the model efficiency (ME or EF), the Nash–Sutcliffe coefficient (CNS), the average difference (AVDIF), the coefficient of shape (CS), the cumulative value test (CVT), the coefficient of residual mass (CRM), linear regressions and the *t*- and *F*-tests. The reader is referred to Loague and Green[72] and to Janssen and Heuberger[2] for detailed mathematical expressions of these indices. The automatic calibration of models using dedicated packages usually relies on the minimisation of an objective function defined as the weighted sum of squared residuals between observed and measured data. Although statistical indices have been increasingly used in the comparison between measured and simulated data, standards and even the usefulness of these goodness-of-fit indices have not yet been established for the various applications in which they might be used.[26] Several of these statistics are sensitive to a few large errors, especially in small datasets.[72] Also, goodness-of-fit statistics do not take into account temporal offsets of the model predictions against experimental data. A timing difference in the prediction of onset of drainflow of a few hours over a period such as a whole winter is of little consequence for the interpretation of results, but may have a major effect on goodness-of-fit statistics.[20] Furthermore, the choice of levels of fit deemed acceptable defaults to the modeller and the goodness-of-fit is therefore subjective. Typical tests are probably not sufficiently strict to convince model sceptics about the accuracy and usefulness of models, and there would be a need for the establishment of agreed performance criteria[30] that invalid models are unlikely to pass.[28] Both graphical and numerical methods have limitations when considered individu-

ally[72] and the combination of the two sets of techniques should be preferred.[81,82]

3.8 Sequential procedures in calibration

The general consensus is that the most appropriate procedure for calibrating models is first to calibrate the hydrology of the model to provide a reasonable representation of water movement at the experimental site, and then to calibrate the solute transport component of the model.[10,20,38] It is generally considered that parameters calibrated against soil hydrology should be left unchanged during the calibration against pesticide data.[81] Although these general procedures are desirable, it may not always be possible to get a good fit to the hydrology or to derive a unique set of calibrated values. Whilst some authors claim that having a good simulation of water fluxes is necessary to predict both pesticide fluxes and concentrations accurately, simulating a good fit to the pesticide data with an inadequate description of the hydrology is possible.[83] Data for a non-interactive tracer such as bromide are frequently used as an intermediate step between calibration of hydraulic and pesticide routines. However, in a number of evaluation exercises where bromide was used, a simultaneous good model fit to the water, bromide and pesticide data was not possible.[25,84] Given these limitations and the uncertainty associated with both the model input parameters and the hydrological data, the modification of parameter values which were calibrated against the hydrology during a calibration of the pesticide section of the model can be justified provided that the hydrology-calibrated parameters are only varied within the bounds of their uncertainty and that the fit to the hydrology still meets the acceptability criteria of the modeller. Such conditional calibrations can be automated using packages such as PEST when used in its regularisation mode.[44] In specific instances (eg parameter estimation), a simultaneous calibration of the water and pesticide components of leaching models may be more appropriate, as a sequential calibration might lead to the derivation of lumped parameters. Parameter lumping is treated in more detail in the next section.

3.9 Potential pitfalls in the calibration of pesticide leaching models

Pesticide leaching models are large, non-linear, complex simulation systems and may hence suffer from non-uniqueness with regard to the set of calibrated parameters.[85] Non-uniqueness occurs when different combinations of parameters or parameter values provide an equally good fit to the data, and commonly results from large correlation between input parameters in the model and/or when the data are insufficient in terms of quantity and quality with respect to the number of parameters to be identified through model calibration. Pesticide fate models are likely to be subject to non-uniqueness issues because of their non-linear character and the inherent compensation of a number parameters with regard to the prediction of pesticide loss (eg sorption and degradation parameters with respect to total leaching). Non-uniqueness issues in the calibration of pesticide leaching models remain largely unnoticed when a manual calibration is carried out and are best revealed using automated calibration techniques.[47] Calibration uniqueness when deriving sorption and degradation parameters may be assessed by using different sets of starting values[73,86] or response surface analysis.[87,88] Additional research should be carried out to estimate the extent of non-uniqueness in the calibration of pesticide leaching models. The use of methodologies incorporating a framework for dealing with non-uniqueness in parameter estimation and subsequent extrapolation, such as that proposed by Beven and Binley,[89] deserves investigation.

Another pitfall of calibration is the 'lumping' of parameters. Here, lumping refers to the attribution to a parameter of a value that does not reflect its theoretical meaning. This is held to occur during the calibration process, rather than being inherent in the model. Lumping may originate from the facts that (1) the model may not be intrinsically capable of simulating the experimental data (eg by not including a description of key processes affecting the fate of pesticides), (2) the data may be of poor quality and uncertain, (3) other model input parameters may have been attributed inadequate values or (4) multiple sets of parameter values may satisfy the conditions to be a solution in ill-posed calibration problems. Lumping thus reflects inaccuracies, uncertainties and limitations associated with experimental data, modelling and calibration. Lumped parameters can usually only be obtained by calibration and have lost their physical, chemical or biological definition. Hence, lumped values will only be valid for the specific set of conditions for which they were obtained and will be of little value for deriving information regarding the specific processes controlling transport and fate[29] and for extrapolation purposes.

The degree of influence of the modeller on calibration results is also a significant issue in calibration. Since calibration procedures are left to the discretion of the modeller, differences are expected with respect to the selection of parameters to be calibrated, the variation applied to their values, the setting up of automatic calibration packages, where appropriate, and the assessment of the goodness-of-fit between measured and simulated data. The user-subjectivity in the parameterisation and calibration of pesticide leaching models is established[5,90] and may prevent model evaluation in some instances.[31]

There are numerous examples of calibrated values substantially differing from the values initially expected for the scenario considered (ie those used in the initial parameterisation of the model). Carsel *et al*[91] calibrated the PRZM model against field data for leaching of aldicarb, and both the decay constant in the lower zone and the linear sorption distribution

coefficient had to be increased by a factor of *ca* 2 to reach an adequate description of the data. Francaviglia *et al*[36] reported the need for the use of unrealistic values of bulk density and field capacity to calibrate PELMO against a lysimeter dataset. Mills and Simmons[92] had to increase laboratory sorption values in the top 5 cm of the soil by a factor of 10 and consider a linear increase of dispersion with depth to improve their fit to the data. Similarly, sorption coefficients for aldicarb derived from a calibration were outside the literature range.[33] Villholth *et al*[93] found that the sorption distribution coefficient derived by calibration against experimental data was about an order of magnitude smaller than that derived in the laboratory. Thorsen *et al.* [25] could only improve the simulation of the leaching of a pesticide in lysimeters by violating the physical description of the soil column.

3.10 Assessment of the parameter values derived from calibration

Owing to the non-uniqueness of calibrations and the potential for parameter lumping, changes in parameter values resulting from calibration need to be carefully assessed. Two approaches are typically used. First, calibrated parameters can be assessed against values used in the initial input file. Initial values usually reflect the best estimate of an adequate value the modeller can make on the basis of laboratory or field experiments, review of the literature or expert judgement. It is particularly critical to assess whether a calibrated value falls within the range of uncertainty expected for a particular parameter. Although calibrated pesticide properties derived by Carsel *et al*[91] were varied within a factor of *ca* 2 of the initial estimates for these parameters, they were within the range of uncertainty as estimated by a literature review.

The second approach to assess parameter values derived by model calibration is to use the calibrated values to generate model output which can then be compared to data different from those used in the calibration. Thorsen *et al*[25] consider that parameter values derived against an experiment under controlled conditions where less complex environmental conditions prevail should be tested against a dataset acquired under more complex conditions. The cross-validation step of the calibrated parameters against a dataset different from that used in the calibration is considered important,[2] especially if the calibration is used to derive values for sorption and degradation which are to be considered within pesticide registration.[43] Such testing exercises may include, in the case of a field drainflow study, (1) the calibration of the model hydrology against soil moisture contents in the profile and a verification against drainflow volumes, (2) the calibration of the model against pesticide residues and a verification against pesticide concentrations in drainflow, although the use of soil residue data for assessing pesticide leaching has been questioned,[77,94] and (3) the calibration of the model against data for one year and a verification against a subsequent year. Data collected in independent experiments have also been used for cross-validation. Gottesbüren *et al*[43] estimated sorption and degradation parameter values by calibrating PEARL and PESTRAS against lysimeter data and tested the optimised values against data for pesticide residues from a field experiment. However, the additional experimental data available could be directly integrated into the calibration instead of being used for separate evaluation. Increasing the amount of data available for calibration is expected to decrease the ill-posed nature of the calibration problem and hence non-uniqueness issues in the calibrated parameter sets.[53] A possible refinement to the calibration of pesticide leaching models and the evaluation of calibrated parameter sets could therefore include (1) an initial parameter estimation based on a calibration against a given experimental dataset, (2) a simulation using calibrated parameters with a comparison to data different from those used in the initial calibration, and (3) a model calibration integrating all available experimental data using calibrated values derived in step (1) as starting values for parameters to be estimated and strict limits on the variation of these parameters.

A discrepancy between measured and simulated data in the cross-validation run may not solely be attributed to inadequate calibrated parameter values, since the lack-of-fit might also be due to model deficiencies (the cross-validation is sometimes used as a model testing method) or the attribution of inadequate values to parameters other than the calibrated ones. Parameters should not be allowed to be varied outside their 'reasonable range' during the calibration[10,81] and setting a parameter to a specific value merely to achieve a good fit to the measured data should be avoided.[71] In some instances, substantial effort put into a calibration does not significantly improve the fit to the data. A poor match may suggest an inadequacy in the conceptual model, an error in the numerical solution, a poor set of parameter values, a poor set of experimental values or some combination of these. It may not be possible to distinguish between these different sources of error. Discrepancies between expected and calibrated values should be discussed and assumptions on the likely cause of such discrepancies proposed. The uncertainty left in the model parameters after calibration should be acknowledged and adequately accounted for in subsequent model applications.[2]

4 GUIDELINES FOR THE REPORTING OF CALIBRATION ACTIVITIES

Given the diversity in modelling situations and the importance of written communication when regulators assess modelling studies, it appears that the most appropriate way to improve quality in the modelling (including model parameterisation and calibration)

Table 2. Guidelines on the reporting of calibration activities carried out for pesticide registration. Reporting of calibration activities is expected to provide answers to the following questions

Introduction
What was (were) the specific aim(s) of the calibration?
Has there been any previous modelling activity for either the experimental site or the compound of interest?
Critical assessment of the experimental data used for calibration
What data were measured in the experimental study?
How many replicates were there?
What was the quality of the replication?
Were any unusual conditions experienced during the experimental period (weather conditions; flooding or freezing conditions)?
Were there difficulties with regard to operational (eg failure of the monitoring equipment) or analytical (eg analytical replication) procedures?
What were the limits of detection and quantification?
Were difficulties encountered in the identification or quantification of compounds?
Were there missing data for a period? Were there outliers?
What were the main uncertainties related to the data? What overall confidence should be assigned to the experimental data?
Detailed justification of the choice of a specific pesticide leaching model for the calibration
Which model (version, release date) was used?
What was the rationale behind the choice of the particular model used?
Is the model a *priori* suitable for describing the experimental data?
Were there processes important for describing the data not explicitly accounted for in the model? if so, were these processes accounted for in the modelling?
Have there been previous studies conducted with this model with the same compound? With the same soil? How did the model perform?
Detailed description of the initial parameterisation of the model and of the selection of parameters to be calibrated
How were values for the input parameters chosen for the initial parameterisation? Which values were determined by independent experiments? Which were determined by expert judgement or educated guess?
Where did the main uncertainties in the parameterisation lie?
Was information on the sensitivity of the model available? Was the sensitivity information transferable to the current modelling exercise?
If either no information was available on the sensitivity of the model or the information was not transferable, was a small-scale sensitivity analysis conducted?
If a small-scale sensitivity analysis was not conducted, how were the parameters to be varied in the calibration selected?
Detailed description of the calibration procedures used
Which experimental data were used in the calibration?
In the case where replicates were available, were the data for only one replicate considered in the calibration? Alternatively, how was the information from the different replicates combined? How was conflicting replication handled (eg differences in flow volumes for a replicated lysimeter experiment)?
How were concentrations below the limit of quantification handled?
How were outliers or missing samples (where applicable) handled?
What were the assumptions made for the concentrations between two sampling dates (drainflow studies)?
Which model output(s) was (were) used in the calibration?
Could the model output be directly compared to the experiment data (note that a post-processing is required in most cases)? If not, how was model output or the experimental data post-processed? Provide a numerical description of the post-processing performed.
Was the calibration done manually (trial-and-error calibration) or was it performed automatically?
In the case of a manual calibration
How was the goodness-of-fit between model output and experimental data assessed? Visually through graphical displays? Numerically using statistical indices? Using both types of assessment?
What was the main target of the calibration? Was it peak values, low values, average values, timing of peaks, first detection, detailed pattern, general trend?
Was the calibration performed sequentially (eg calibration of the hydrological part of the model then calibration of the pesticide section)?
Were parameters calibrated sequentially for one particular set of output (ie one parameter after the other)?
How many runs were carried out to achieve the end results?
What criterion was used to stop the calibration?
In the case of an automatic calibration
Which package was used? Which version of the package?
How was the objective function defined?
Were (some) parameters transformed?
Which weights were assigned to the experimental observations?
Was any relationship specified between parameters?
Were all the input parameters calibrated simultaneously?

Table 2. Continued

How many iterations and runs were necessary to achieve convergence?
Were consistent calibration results obtained when different starting values were specified?
What was the correlation between parameters during the calibration (as reported by the calibration package)?
Did the residuals show a particular pattern or were they randomly distributed?
Was the visual examination of the fit to the observations satisfactory when using the final set of calibrated parameters?
Assessment of calibration results
By how much did the parameters have to be changed to get a good fit to the data? Are the calibrated values plausible? Reasonable? Do they fit with what is known about the variability and uncertainty of these parameters?
Was a satisfactory (visually and statistically) fit to the data obtained? If not, what could it be attributed to (Inadequate choice of parameters to be varied? Inadequate values for parameters not included in the calibration? Inadequate calibration procedures? Inability of the model to describe the data?)?
Cross-validation of the set of calibrated parameters against other model output or against another field dataset
How much uncertainty has been left in the parameters after calibration?
Does the calibrated parameter set give satisfactory results when considering an output other than that used for the calibration?
Is there a good fit between model predictions and experimental data when the calibrated set of parameters is used to describe another dataset?
Conclusions
How much confidence should be assigned to the final values attributed to the parameters? Where do the uncertainties lie?
Can the results of the calibration be used for the intended purpose defined in the introduction?
Tables and figures
The following tables and figures are useful to assess the confidence that should be assigned to the calibration results:
A comparative table with initial (from the initial model parameterisation) and final (calibrated) values for the parameters included in the calibration,
A table comparing values of statistical indices (at least the sum of squared residuals) before and after the calibration (where appropriate),
A figure showing a comparison between the experimental data and the model predictions for the initial and calibrated runs (charts against time or depth for the variable used in the calibration),
A figure showing a comparison between the experimental data and the model predictions for the initial and calibrated runs (charts against time or depth for the variable(s) measured in the field, but not used in the calibration),
A figure plotting measured *versus* simulated data with a line of perfect agreement (or 1:1 line) for the variable used in the calibration and for other variables measured in the field, but not used in the calibration.

and to decrease the associated uncertainty is to issue guidelines on the reporting of the modelling. Such guidance provides flexibility to modellers, as opposed to guidelines on the modelling itself.[28] Recommendations for parameterising and calibrating pesticide leaching models have been issued and have highlighted the need for quality reporting.[8–10] However, the aspects that should be included in reports have not been explicitly set out. Guidance on the reporting of calibration activities with pesticide leaching models is proposed in Table 2. The guidelines are intended to be non-country-specific and are generic in nature. It is hoped that the guidelines will raise awareness among modellers of the issues associated with the calibration of pesticide leaching models. Their use is expected to improve calibration activities as a whole and help modellers and regulators to assess the confidence that should be attributed to predictions based on calibrated parameters.

5 CONCLUSIONS

Complex deterministic models are being used in pesticide registration in Europe to assess the potential for a pesticide to impact on the environment. Within this context, calibration may be used to derive input parameters which are difficult to obtain from independent measurements, to back-derive targeted parameter values, to validate the use of a model or to establish the basis for subsequent extrapolations. The need for calibration is controversial within the pesticide modelling community. Some individuals consider that calibration is a pre-requisite to a reliable simulation of pesticide fate, whereas others argue that a calibrated set of parameters is only valid for the conditions at hand and should not be used for other scenarios.

The calibration of leaching models is clearly one of the most arduous tasks a pesticide fate modeller is faced with. The success of a calibration is primarily limited by the nature, amount and quality of the available data, the appropriateness of the model used, the effectiveness of the applied calibration technique, the time available, computer power, expertise and financial resources.[2,83] Within the scope of the calibration of pesticide leaching models, the main factors which may lead to inappropriate calibrations include the lack of substantial data or their poor quality, the use of a model which is not capable of describing the experimental dataset, the inadequate selection of parameters to be varied and inadequate calibration procedures. Given these uncertainties, the inherent

nature of pesticide leaching models (non-linearity, large correlation between parameters) and the differences between calibration approaches adopted by individuals, it is clear that interpreting the strength of a calibration and resulting model output is no simple matter. The default assumption should be that calibration results are uncertain, if not demonstrated otherwise.

Given the importance of calibration activities within pesticide registration and their potential limitations, the regulator should be provided with sufficient information to allow an assessment of the confidence to be assigned to results from calibration or from extrapolation based on calibrated parameters. Use of the guidelines proposed in the present paper would help regulators in their assessment and indirectly improve calibration practice.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the funding of this work by the UK Department for Environment, Food and Rural Affairs on behalf of the Pesticides Safety Directorate (project PL0539). The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the study sponsors. The authors are grateful to the reviewers for constructive criticism on the manuscript.

REFERENCES

1 Gustafson DI, Use of computer models to assess exposure to agricultural chemical via drinking water. *Sci Total Environ* **171**:35–42 (1995).
2 Janssen PHM and Heuberger PSC, Calibration of process-oriented models. *Ecol Model* **83**:55–66 (1995).
3 FOCUS, *Leaching models and EU registration*, EC document reference 4952/VI/95, 123 pp (1995).
4 Vanclooster M, Boesten JJTI, Trevisan M, Brown CD, Capri E, Eklo OM, Gottesbüren B, Gouy V and van der Linden AMA, A European test of pesticide leaching models: methodology and major recommendations. *Agr Water Manag* **44**:1–19 (2000).
5 Brown CD, Baer U, Günther P, Trevisan M and Walker A, Ring test with the models LEACHP, PRZM-2 and VARLEACH: variability between model users in prediction of pesticide leaching using a standard data set. *Pestic Sci* **47**:249–258 (1996).
6 Görlitz G, *Rules for the correct performance and evaluation of model calculations for simulation of the environmental behaviour of pesticides*, Document prepared by the Biologische Bundesanstalt für Land- und Forstwirtschaft, the Fraunhofer Institut für Umweltchemie and Ökotoxikologie, the Arbeitsgruppe 'Simulationsmodelle' in Industrieverband Agrar and the Umweltbundesamt, 9 pp (1993).
7 Estes TL and Coody PN, Towards the development of good modelling practices in chemical fate modeling, in *Final report of the FIFRA environmental model validation task forcé - Appendix*, ed by Jones RL and Russell MH, www.femvtf.com (1994).
8 Resseler H, Schäfer H, Görlitz G, Hermann M, Hosang J, Klowskowski R, Marx R, Sarafin R, Stein B and Winkler R, Recommendations for conducting simulation calculations for the registration procedure. *Nachrichtenbl Deut Pflanzenschutzd* **49**:305–309 (1997).
9 CAMASE, Guidelines for modelling, http://www.bib.wau.nl/camase/modguide.html (1995).
10 Jones RL and Russell MH (eds), *Final report of the FIFRA environmental model validation task force*, www.femvtf.com, 767 pp (2000).
11 Bergström L, Model predictions and field measurements of chlorsulfuron leaching under non-steady-state flow conditions. *Pestic Sci* **48**:37–45 (1996).
12 Addiscott T, Smith J and Bradbury N, Critical evaluation of models and their parameters. *J Environ Qual* **24**:803–807 (1995).
13 Hanson JD, Rojas KW and Shaffer MJ, Calibrating the Root Zone Water Quality Model. *Agron J* **91**:171–177 (1999).
14 Bergström LF and Jarvis NJ, Evaluation and comparison of pesticide leaching models for registration purposes. *J Environ Sci Health A* **29**:1061–1072 (1994).
15 Beulke S, Brown CD, Dubus IG and Harris G, Evaluation of uncalibrated preferential flow models against data for isoproturon movement to drains through a heavy clay soil. *Pest Manag Sci* **57**:537–547 (2001).
16 Pennell KD, Hornsby AG, Jessup RE and Rao PSC, Evaluation of five simulation models for predicting aldicarb and bromide behavior under field conditions. *Water Resour Res* **26**:2679–2693 (1990).
17 Klein M, Müller M, Dust M, Görlitz G, Gottesbüren B, Hassink J, Klowskowski R, Kubiak R, Resseler H, Schäfer H, Stein B and Vereecken H, Validation of the pesticide leaching model PELMO using lysimeter studies performed for registration. *Chemosphere* **35**:2563–2587 (1997).
18 Walker A, Calvet R, Del Re AAM, Pestemer W and Hollis JM, Evaluation and improvement of mathematical models of pesticide mobility in soils and assessment of their potential to predict contamination of water systems. *Mitteilungen aus der Biologischen Bundesanstalt für Land- und Forstwirtschaft* **307**:1–115 (1995).
19 Brooke D and Matthiessen P, Development and validation of a modified fugacity model of pesticide leaching from farmland. *Pestic Sci* **31**:349–361 (1991).
20 Armstrong AC, Portwood AM, Leeds-Harrison PB, Harris GL and Catt JA, The validation of pesticide leaching models. *Pestic Sci* **48**:47–55 (1996).
21 Watanabe H and Takagi K, A simulation model for predicting pesticide concentrations in paddy water and surface soil. II. Model validation and application. *Environ Technol* **21**:1393–1404 (2000).
22 Konikow LF and Bredehoeft JD, Groundwater models cannot be validated. *Adv Water Resour* **15**:75–83 (1992).
23 Oreskes N, Shrader-Frechette K and Belitz K, Verification, validation and confirmation of numerical models in the earth sciences. *Science (Washington)* **263**:641–646 (1994).
24 Beulke S, Brown CD and Jarvis NJ, MACRO: a preferential flow model to simulate pesticide leaching and movement to drains, in *Modelling of environmental chemical exposure and risk*, ed by Linders JBHJ, Kluwer Academic Publishers, The Netherlands, pp 117–132 (2001).
25 Thorsen M, Jørgensen PR, Felding G, Jacobsen OH, Spliid NH and Refsgaard JC, Evaluation of a stepwise procedure for comparative validation of pesticide leaching models. *J Environ Qual* **27**:1183–1193 (1998).
26 Loague K, Simulation of organic chemical movement in Hawaii soils with PRZM: 3. Calibration. *Pac Sci* **46**:353–373 (1992).
27 Kumar A and Kanwar RS, Incorporating preferential flow and herbicide fate and transport into the DRAINAGE model. *Transac ASAE* **40**:977–985 (1997).
28 Kirchner JW, Hooper RP, Kendall C, Neal C and Leavesley G, Testing and validating environmental models. *Sci Total Environ* **183**:33–47 (1996).
29 Brusseau ML, Non-ideal transport of reactive solutes in heterogeneous porous media: 3. Model testing and data analysis using calibration *versus* prediction. *J Hydrol* **209**:147–165 (1998).
30 Jarvis NJ, Using preferential flow models for management purposes, in *Modelling of transport processes in soils*, ed by Feyen J

and Wiyo K, Wageningen Pers, Wageningen, The Netherlands, pp 521–535 (1999).
31 Tiktak A, Application of pesticide leaching models to the Vredepeel dataset. II. Pesticide fate. *Agr Water Manag* **44**:119–134 (2000).
32 Vanclooster M and Boesten JJTI, Application of pesticide simulation models to the Vredepeel dataset. I. Water, solute and heat transport. *Agr Water Manag* **44**:105–117 (2000).
33 Lorber MN and Offutt CK, A method for the assessment of groundwater contamination potential using a Pesticide Root Zone Model (PRZM) for the unsaturated zone, in *Evaluation of pesticides in groundwater*, ed by Garner WY, Honeycutt RC and Nigg HN, ACS Symposium Series 315, American Chemical Society, Washington, DC, pp 342–365 (1986).
34 Eckhardt DAV and Wagenet RJ, Estimation of the potential for atrazine transport in a silt loam soil, in *Herbicide metabolites in surface water and groundwater*, ed by Meyer MT and Thurman EM, ACS Symposium Series 630, American Chemical Society, Washington, DC, pp 101–116 (1996).
35 Brown CD, Beulke S and Dubus IG, Progress in simulating macropore flow through soil, in *Proc Brighton Crop Protect Conf—Pests and Diseases*, BCPC, Farnham, Surrey, UK, pp 85–90 (2000).
36 Francaviglia R, Capri E, Klein M, Hosang J, Aden K, Trevisan M and Errera G, Comparing and evaluating pesticide leaching models: results for the Tor Mancina dataset (Italy). *Agr Water Manage* **44**:135–151 (2000).
37 Loague K, Miyahira RN, Green RE, Oki DS, Giambelluca TW and Schneider RC. Chemical leaching near the Waiawa shaft, Oahu, Hawaii. 2. Modelling results. *Ground Water* **33**:124–138 (1995).
38 Durborow TE, Barnes NL, Cohen SZ, Horst GL and Smith AE, Calibration and validation of runoff and leaching models for turf pesticides and comparison with monitoring results, in *Fate and management of turfgrass chemicals*, ed by Clark JM and Kenna MP, ACS Symposium Series 743, American Chemical Society, Washington, DC, pp 195–227 (2000).
39 Russell MH, Layton RJ and Tillotson PM, The use of pesticide leaching models in a regulatory setting: an industrial perspective. *J Environ Sci Health A* **29**:1105–1116 (1994).
40 Beulke S, Dubus IG, Brown CD and Gottesbüren B, Simulation of pesticide persistence in the field on the basis of laboratory data—a review. *J Environ Qual* **29**:1371–1379 (2000).
41 Gottesbüren B, Estimation of pesticide parameters from lysimeter studies using an inverse modelling approach, in *Book of Abstracts*, 9th IUPAC Intern Congress, London, 2–7 August, Abstract No 6B-015 (1998).
42 Dubus IG, Beulke S and Brown CD, *A briefing document on the application of inverse modelling techniques to pesticide leaching models*, SSLRC Research Report to MAFF No PL0528, Cranfield University, Silsoe, UK (2000).
43 Gottesbüren B, Platz K, van de Veen JR, Dubus IG and Beulke S, Estimation of pesticide parameters from static and dynamic experiments by two independent modelling groups, in *Pesticide behaviour in soils and water*, ed by Walker A, Bending GD, Bromilow RH, Brown CD, Carter AD, Johnson AC and MacKay N, BCPC Symp Proc No 78, Brighton, UK, 13–15 November 2001, BCPC, Farnham, Surrey, UK, pp 89–94 (2001).
44 Doherty J, *PEST—model-independent parameter estimation*, Watermark Computing, Corinda, Australia (2000).
45 Poeter EP and Hill MC, *Documentation of UCODE: a computer code for universal inverse modelling*, US Geological Survey, Water Resources Investigations Report 98-4080 (1998).
46 Abbaspour K, van Genuchten MTh, Schulin R and Schläppi E, A sequential uncertainty domain inverse procedure for estimating subsurface flow and transport parameters. *Water Resour Res* **33**:1879–1892 (1997).
47 Poeter EP and Hill MC, Inverse models: a necessary next step in groundwater modeling. *Ground Water* **35**:250–260 (1997).
48 Diekkrüger B, Söndgerath D, Kersebaum KC and McVoy CW, Validity of agroecosystem models—a comparison of results of different models applied to the same data set. *Ecol Model* **81**:3–29 (1995).
49 Yapo PO, Gupta HV and Sorooshian S, Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *J Hydrol* **181**:23–48 (1996).
50 Carrera J and Neuman SP, Estimation of aquifer parameters under transient and steady state conditions. 2. Uniqueness, stability and solution algorithms. *Water Resour Res* **22**:211–227 (1986).
51 Mualem Y, A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour Res* **12**:513–522 (1976).
52 van Genuchten MT, A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* **44**:892–898 (1980).
53 Hopmans JW and Šimunek J, Review of inverse estimation of soil hydraulic properties, in *Characterization and measurement of the hydraulic properties of unsaturated porous mdia*, ed by van Genuchten MT, Leij FJ and Wu L, University of California, California, pp 643–659 (1999).
54 Larocque M and Banton O, Using field data and simulation modelling to determine nitrogen cycle parameters. *Soil Sci Soc Am J* **60**:1840–1845 (1996).
55 Del Re AAM and Trevisan M, Selection criteria of xenobiotic leaching models in soil. *Eur J Agron* **4**:465–472 (1995).
56 Travis KZ, Pesticide leaching models—past, present and future, in *Proc Brighton Crop Protect Conf—Pests and Diseases*, BCPC, Farnham, Surrey, UK, pp 69–76 (2000).
57 Klein M, Evaluation and comparison of pesticide leaching models for registration purposes. Results of simulations performed with the pesticide leaching model. *J Environ Sci Health* **A29**:1197–1209 (1994).
58 Mills MS and Simmons ND, Assessing the groundwater contamination potential of agricultural chemicals: a flexible approach to mobility and degradation studies. *Pestic Sci* **54**:418–434 (1998).
59 Di HJ and Aylmore LAG, Modeling the probabilities of groundwater contamination of pesticides. *Soil Sci Soc Am J* **61**:17–23 (1997).
60 FOCUS, *FOCUS groundwater scenarios in the EU plant protection review process*, EC document reference Sanco/321/2000 rev 2, 202 pp (2000).
61 Dubus IG, Brown CD and Beulke S, Sources of uncertainty in pesticide fate modelling, in *Pesticide behaviour in soils and water*, ed by Walker A, Bending GD, Bromilow RH, Brown CD, Carter AD, Johnson AC and MacKay N, BCPC Symp Proc No 78, Brighton, UK, 13–15 November 2001, BCPC, Farnham, Surrey, UK, pp 239–244 (2001).
62 Helton JC, Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal. *Reliab Eng Syst Safe* **42**:327–367 (1993).
63 Dubus IG, Brown CD and Beulke S, *Sensitivity analyses for leaching models used for pesticide registration in Europe*, SSLRC Research Report to MAFF No PL0532, Cranfield University, Silsoe, UK (2000).
64 Dubus IG and Brown CD, Sensitivity and first-step uncertainty analyses for the preferential flow model MACRO. *J Environ Qual* **31**:227–240 (2002).
65 Fontaine DD, Havens PL, Blau GE and Tillotson PM, The role of sensitivity analysis in groundwater risk modelling for pesticides. *Weed Technol* **6**:716–724 (1992).
66 Jarvis NJ, *MACRO—a model of water movement and solute transport in macroporous soils*, Monograph, Reports and Dissertations No 9, Department of Soil Science, Swedish University of Agricultural Sciences, Uppsala, Sweden (1991).
67 Smith WN, Prasher SO and Barrington SF, Evaluation of PRZM and LEACHMP on intact soil columns. *Trans ASAE* **34**:2413–2420 (1991).
68 Boesten JJTI, Sensitivity analysis of a mathematical model for

pesticide leaching to groundwater. *Pestic Sci* **31**:375–388 (1991).
69 Boesten JJTI and van der Linden AMA, Modeling the influence of the sorption and transformation on pesticide leaching and persistence. *J Environ Qual* **20**:425–435 (1991).
70 Ferreira VA, Weesies GA, Yoder DC, Foster GR and Renard KG, The site and condition specific nature of sensitivity analysis. *J Soil Water Cons* **50**:493–497 (1995).
71 Klein M, Hosang J, Schäfer H, Erzgräber B and Resseler H, Comparing and evaluating pesticide leaching models—Results of simulations with PELMO. *Agr Water Manag* **44**:263–281 (2000).
72 Loague K and Green RE, Statistical and graphical methods for evaluating solute transport models: overview and application. *J Contam Hydrol* 7:51–73 (1991).
73 Madsen H, Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J Hydrol* **235**:276–288 (2000).
74 Yapo PO, Gupta HV and Sorooshian S, Multi-objective global optimization for hydrologic models. *J Hydrol* **204**:83–97 (1998).
75 Vanclooster M, Ducheyne S, Dust M and Verrecken H, Evaluation of pesticide dynamics of the WAVE model. *Agr Water Manag* **44**:371–388 (2000).
76 Hill MC, *Methods and guidelines for effective model calibration*, US Geological Survey, Water-Resources Investigations Report 98-4005, 90 pp (1998).
77 Jarvis NJ, Bergström LF and Brown CD, Pesticide leaching models and their use for management purposes, in *Environmental behaviour of agrochemicals*, ed by Roberts TR and Kearney PC, Wiley, New York, pp 185–220 (1995).
78 Litaor I, Review of soil solution samplers. *Water Resour Res* **24**:727–733 (1988).
79 Grossmann J and Udluft P, The extraction of soil water by the suction-cup method: a review. *J Soil Sci* **42**:83–93 (1991).
80 Shirmohammadi A and Knisel WG, Evaluation of the GLEAMS model for pesticide leaching in Sweden. *J Environ Sci Health A* **29**:1167–1182 (1994).
81 Borah MJ and Kalita PK, Development and evaluation of a macropore flow component for LEACHM. *Trans ASAE* **42**:65–78 (1999).
82 Gaunt JL, Riley J, Stein A and Penning de Vries FWT, Requirements for effective modelling strategies. *Agr Syst* **54**:153–168 (1997).
83 Gottesbüren B, Aden K, Bärlund I, Brown C, Dust M, Görlitz G, Jarvis N, Rekolainen S and Schäfer H, Comparison of pesticide leaching models: results using the Weiherbach dataset. *Agr Water Manag* **44**:153–181 (2000).
84 Jarvis NJ, Brown CD and Granitza E, Sources of error in model predictions of pesticide leaching: a case study using the MACRO model. *Agr Water Manag* **44**:247–262 (2000).
85 Spear RC, Large simulation models: calibration, uniqueness and goodness of fit. *Environ Modell Softw* **12**:219–228 (1997).
86 Isabel D and Villeneuve JP, Importance of the convergence criterion in the automatic calibration of hydrological models. *Water Resour Res* **22**:1367–1370 (1986).
87 Toorman AF and Wierenga PJ, Parameter estimation of hydraulic properties from one-step outflow data. *Water Resour Res* **28**:3021–3028 (1992).
88 Šimunek J, Wendroth O and van Genuchten MT, Parameter estimation analysis of the evaporation method for determining soil hydraulic properties. *Soil Sci Soc Am J* **62**:894–905 (1998).
89 Beven K and Binley A, The future of distributed models—model calibration and uncertainty prediction. *Hydrol Process* **6**:279–298 (1992).
90 Boesten JJTI, Modeller subjectivity in estimating pesticide parameters for leaching models using the same laboratory dataset. *Agr Water Manag* **44**:389–409 (2000).
91 Carsel RF, Mulkey LA, Lorber MN and Baskin LB, The Pesticide Root Zone Model (PRZM): a procedure for evaluating pesticide leaching threats to groundwater. *Ecol Model* **30**:49–69 (1985).
92 Mills MS and Simmons ND, Assessing the groundwater contamination potential of agricultural chemicals: a flexible approach to mobility and degradation studies. *Pestic Sci* **54**:418–434 (1998).
93 Villholth KG, Jarvis NJ, Jacobsen OH and De Jonge H, Field investigations and modeling of particle-facilitated pesticide transport in macroporous soil. *J Environ Qual* **29**:1298–1309 (2000).
94 Malone RW, Shipitalo MJ, Douglass LW, Owens LB, Nelsen TC, Warner RC and Byers ME, Assessing herbicide movement using soil samples *versus* percolate samples. *Trans ASAE* **43**:343–348 (2000).