Research Article

Received: 27 January 2009 Revised: 6 May 2009 Accepted: 11 June 2009 Published online in Wiley Interscience:

(www.interscience.wiley.com) DOI 10.1002/ps.1825

A refined lack-of-fit statistic to calibrate pesticide fate models for responsive systems

Bernard T Nolan,[a,b]* Igor G Dubus[a,c] and Nicolas Surdyk[a]

Abstract

BACKGROUND: Calibration by inverse modelling was performed with the MACRO transport and fate model using long-term (>10 years) drainflow and isoproturon (IPU) data from western France. Two lack-of-fit (LOF) indices were used to control the inverse modelling: sum of squares (SS) and an alternative statistic called the vertical-horizontal distance integrator (VHDI), which is designed to account for offsets in observed and predicted arrival times of peak IPU concentration. With these data, SS was artificially inflated because it is limited to comparison of predicted and observed IPU concentrations that are concurrent in time. The LOFs were used along with the index of agreement (*d*) and the correlation coefficient (*r*) to ascertain the fit of the calibrated models.

RESULTS: Predicted arrival times of peak IPU concentration differed somewhat from observed times. All four indices indicated better model fit for the second of two validation periods when inverse modelling was controlled by VHDI rather than SS (SS = 26.4, $d = 0.660$, $r = 0.606$ and VHDI = 1.25). The VHDI statistic was markedly lower compared with the uncalibrated model (38.0) and SS calibration results (24.5). The final maximum predicted IPU concentration (44.5 $\mu g\ L^{-1}$) for the calibration period was very similar to the observed value (44 $\mu g\ L^{-1}$).

CONCLUSION: VHDI is seen as an effective alternative to SS for calibration and validation of pesticide fate models applied to responsive systems. VHDI provided a more realistic assessment of model performance for the transient flows and short-lived concentrations observed here, and also effectively substituted for the objective function in inverse modelling.

Keywords: pesticides; isoproturon; drains; macropores; MACRO model; inverse modelling

1 INTRODUCTION

Pesticide transport and fate models are routinely used by regulatory agencies to assess the effects of agricultural contaminants on groundwater quality.[1–3] The models are essential tools because they permit efficient screening of large numbers of pesticides, soil types, weather regimes and other factors in a reasonable timeframe and at relatively low cost.[4] Such models have become increasingly sophisticated in the past two decades, in attempts to improve prediction accuracy under field conditions. Models such as MACRO[5,6] are capable of simulating complex processes such as preferential flow in the unsaturated zone. The MACRO model has been extensively used by researchers and in European registration to predict pesticide transport through a variety of soil types to groundwater and to drains.[7–10] Preferential flow commonly occurs in field soils and is widely recognised as contributing to the rapid and significant transport of agricultural contaminants to depth, including pesticides.[11] In particular, macropore flow bypasses the soil matrix, dramatically reducing the residence time of water and solutes in the vadose zone. Well-structured or clayey soils are susceptible to rapid transport of even strongly sorbing chemicals.[12] As models become more complex, however, calibration becomes more difficult. Reliable methods are lacking for estimating macropore parameters, which are difficult to measure directly.[13] In some cases, macropore parameters have no direct physical meaning and are lumped to indicate preferential flow effects.[14]

Autocalibration by inverse modelling is a potential means of estimating model parameters that are otherwise difficult to obtain. Software routines such as PEST[15] and UCODE[16] provide distinct advantages over manual, trial-and-error approaches. Model fit is commonly evaluated with an objective function based on the sum of squared deviations between predicted and observed values.[15] Parameter values are adjusted automatically until the dependent variables (e.g. water flows, pesticide concentration in leachate) match observed values to the extent possible. Inverse modelling is advantageous in that parameters are adjusted simultaneously, and insensitive and/or highly correlated parameters can readily be identified.[17] A potential disadvantage of automatic calibration based on typical objective functions is that they do not account for temporal offsets between predicted values and observed data. A timing difference of a few hours or days does not imply bad fit when the modelling period encompasses months or years,[18] but can have a major effect on sum-of-squares measures.[19] In

* *Correspondence to: Bernard T Nolan, US Geological Survey, 413 National Center, Reston, Virginia 20192, USA. E-mail: btnolan@usgs.gov*

a *BRGM, 3 Avenue C Guillemin, BP 36009, 45060 Orléans Cedex 2, France*

b *Le STUDIUM, 3D avenue de la Recherche Scientifique, 45071 Orléans Cedex 2, France*

c *FOOTWAYS, 10 Avenue Buffon, 45071 Orléans Cedex 2, France*

practice it is not uncommon to obtain predicted and observed 'peak' concentrations that are similar in magnitude but that occur on different dates. In such cases, inverse modelling may reject the parameter combination because the arrival times do not coincide exactly, even though the peak concentrations may match.[14]

The present study featured a long-term, high-quality dataset from an agricultural experimental station in western France. Drainflow volumes and herbicide concentrations in tile drains were measured for more than 10 years at the site. The numerous observations provided a rare opportunity to estimate macropore flow parameters with a high degree of precision, and partitioning the long-term dataset permitted validation of calibrated models. The goal of the present modelling exercise was to improve on a previous blind calibration of the MACRO model[20] by finding an optimal combination of parameters. To assist in calibration and to extend the usefulness of the observed data, the authors developed an alternative lack-of-fit (LOF) statistic that considers the case where the predicted and observed peak concentrations arrive at different times. It was anticipated that use of the alternative statistic in conjunction with inverse modelling would improve model calibration for responsive systems with well-defined peaks. The alternative LOF was compared with sum-of-squares and two goodness-of-fit (GOF) measures, the correlation coefficient and the index of agreement. Previous researchers had found that the correlation coefficient was not consistently related to model accuracy,[21] and the index of agreement has been described as an improvement over the coefficient of determination (R^2).[22] However, these GOFs have not been simultaneously evaluated in the context of transient and highly responsive systems. Therefore, the specific objectives were to compare traditional and alternative LOF measures by inverse modelling for adjustment of macropore flow parameters and to evaluate model fit by these and selected GOF measures for calibration and validation datasets.

2 METHODS

The experimental site is located near La Jaillière, France, and is operated by Arvalis – Institut du Végétal. The site comprises 5 ha of land consisting of six drained and two undrained field plots on which mainly corn and winter wheat are grown. The area has an oceanic climate typical of western Europe and received an average of 725 mm of rain per year over the period 1982–1996. Slope at the site generally varies between 0 and 3%. The present research focused on a single drained plot (T4) about 1 ha in size.

2.1 Drainage and soil samples

Artificial drains constructed of 50 mm diameter PVC were installed by Arvalis in field plot T4 at spacings of about 10 m and at depths of 80–100 cm. Drainflows were routed to a collection chamber for periodic flow volume measurement and herbicide sampling by an autosampler. Drainage from the plot was sampled between 1 September 1993 and 28 July 2003 for analysis of isoproturon [3-(4-isopropylphenyl)-1,1-dimethylurea; IPU], a residual urea herbicide typically used on cereals for control of annual grasses and broadleaf weeds. The IPU samples were stored at −18 °C and analysed by liquid chromatography–mass spectrometry at the GIRPA laboratory in Angers, France (limit of quantitation 0.05 μg L^{-1}).

Soils at the La Jaillière site are medium loamy over clay and are classified as Stagnic Luvisols.[23] A soil pit was excavated in May 2005 at the edge of plot T4 to procure samples for laboratory analysis and to characterise soil layers, texture, size of peds and their arrangement, degree of mottling and likely water flow pathways. The field observations and laboratory data were used for initial parameterisation of MACRO, the pesticide transport model used in this study. Soil samples were analysed for sand, silt and clay percentages, organic matter (OM) content and bulk density at INRA in Arras, France (Table 1). The MACRO sorption coefficient, ZKD, was measured for IPU in batch experiments conducted for each soil horizon (Table 1)[20] and then adjusted slightly on the basis of initial, limited model calibration. The adjusted ZKD values (Table 2) were the starting values for calibration by inverse modelling and are referred to as the 'uncalibrated model' in Section 3. The ZKD values correspond to an organic carbon sorption coefficient (K_{oc}) of 125 cm^3 g^{-1} and reflect the organic carbon content of each soil layer. To simulate degradation in MACRO, an IPU half-life (DT_{50}) of 20 days was specified on the basis of field values in the FOOTPRINT Pesticide Properties Database[24] and the results of the initial, limited calibration. The corresponding degradation rates (DEG) in Table 2 are expressed on a per-day basis and reflect the soil water content at the micropore–macropore boundary and the temperature at which the degradation rates were measured. Degradation rates for each layer were adjusted for changes in organic carbon, bulk density and ZKD, as described in a prior study.[25]

Table 1. Measured soil properties used to parameterise the MACRO model

Horizon depth (cm)	Texture (%) Sand	Silt	Clay	Organic matter (%)	Bulk density (g cm^{-3})	Sorption coefficient ZKD (cm^3 g^{-1})
0–30	34.6	44.6	20.8	2.19	1.53	1.43
30–48	32.8	41.3	25.9	0.79	1.58	0.44
48–65	15.5	35.3	49.2	0.46	1.61	0.25
65–110	21.5	35.8	42.7	0.37	1.63	0.22

2.2 MACRO model parameterisation

Version 4.3 of MACRO, a one-dimensional non-steady-state model of water flow and solute transport,[26] was used to simulate drain flows and IPU concentration in drainage at field plot T4. The model considers two flow domains consisting of micropores and macropores so as to simulate preferential flow and solute transport in a variety of soil types. Unsaturated water flow in the micropores is simulated by Richards' equation, and water retention parameters are calculated using the Brooks and Corey equation.[27] Unsaturated hydraulic conductivity in micropores is described by Mualem's model.[28] For the macropores, gravity flow of water is assumed, and hydraulic conductivity is simulated using a kinematic wave approach. Solute transport is simulated by the convection–dispersion equation in the micropores and as mass flow only in the macropores. The division between micropores and macropores is characterised by the soil water content (XMPOR), soil water tension (CTEN) and hydraulic conductivity (K_b). At water contents greater than XMPOR, the hydraulic conductivity of the macropores increases very rapidly as saturation is approached.

Rainfall was measured at the La Jaillière field site, and irrigation amounts were obtained from a detailed schedule of operations for incorporation into the MACRO weather input file. Irrigation inputs of 20–30 mm each were distributed over a 1 h period at midday. Minimum and maximum temperatures and Penman estimates of

Table 2. MACRO model parameters varied during inverse modelling controlled by the internal objective function (sum of squares; SS) or externally by the vertical-horizontal distance integrator (VHDI). Initial inverse modelling varied all parameters shown, but the objective function failed to stabilise after 12 iterations

Parameter	Name	Starting (uncalib.) value	Lower bound	Upper bound	Calibration method SS[a]	Calibration method VHDI[a]
Pesticide properties						
Degradation rate (day^{-1}) layer 1	DEG1	0.0433	0.0177	0.885	0.0433	0.0433
Degradation rate (day^{-1}) layer 2	DEG2	0.0359	0.0159	0.375	0.0359	0.0359
Degradation rate (day^{-1}) layer 3	DEG3	0.0312	0.0153	0.231	0.0312	0.0312
Degradation rate (day^{-1}) layer 4	DEG4	0.0274	0.0139	0.184	0.0274	0.0274
Sorption coefficient ($cm^3\ g^{-1}$) layer 1	ZKD1	1.55	0.0620	3.72	1.55	1.55
Sorption coefficient ($cm^3\ g^{-1}$) layer 2	ZKD2	0.568	0.0227	1.36	0.568	0.568
Sorption coefficient ($cm^3\ g^{-1}$) layer 3	ZKD3	0.335	0.0134	0.804	0.335	0.335
Sorption coefficient ($cm^3\ g^{-1}$) layer 4	ZKD4	0.265	0.0106	0.636	0.265	0.265
Macropore properties						
Boundary soil water content (% vol.) layer 1	XMPOR1	37.38	16.20	42.87	38.20	35.55
Boundary soil water content (% vol.) layer 2	XMPOR2	39.29	17.06	41.30	38.65	36.60
Boundary soil water content (% vol.) layer 3	XMPOR3	39.81	27.32	40.26	40.24	37.18
Boundary soil water content (% vol.) layer 4	XMPOR4	38.80	24.75	39.54	39.54	38.86
Boundary soil water tension (cm) layer 1	CTEN1	18.00	10.00	50.00	28.94	18.03
Boundary soil water tension (cm) layer 2	CTEN2	20.00	10.00	50.00	15.21	19.43
Boundary soil water tension (cm) layer 3	CTEN3	40.00	10.00	50.00	10.00	40.16
Boundary soil water tension (cm) layer 4	CTEN4	35.00	10.00	50.00	25.31	35.00

[a] Degradation rate and sorption coefficient were fixed after the initial inverse modelling attempt.

PET were obtained from the weather station at Beaucouzé, France (latitude = 47° 28′ 0″ N, longitude = 0° 37′ 60″ W), which is 32 km from La Jaillière. Crop growth parameters in MACRO (foliar indices, crop height, root distribution and depth, foliar senescence) were specified on the basis of recommendations by the FOCUS (FOrum for Coordination of pesticide fate models and their USe) workgroup.[29] Dates of IPU application and of planting and harvest of corn, winter wheat and cover crops were obtained from the La Jaillière schedule of operations. IPU was applied by spraying a commercial product at a rate of 200 L ha^{-1}, which resulted in input concentrations of 2500–6250 g m^{-3} of active ingredient to the field plot. These application rates are representative of those used in France.

For initial parameterisation of MACRO, the authors estimated XMPOR and CTEN for each soil layer using van Genuchten water retention curves derived from pedotransfer functions associated with the HYPRES soil properties database, which contains water retention data from 2894 soil horizons in Europe.[30] The HYPRES pedotransfer functions used as input for the observed bulk density values and percentages of clay, silt and organic matter in Table 1. Starting values of CTEN shown in Table 2 were selected by expert judgement based on prior experience with MACRO applied to clayey soils in north-western Europe. The value of XMPOR equivalent to CTEN was derived from the water retention curves, and K_b was estimated as described in a prior study.[7]

2.3 Lack of fit and calibration by inverse modelling

The parameters DEG, ZKD, XMPOR and CTEN are among the most sensitive parameters in MACRO, according to prior modelling conducted for fine loamy soils in north-western Europe.[31] Model calibration by inverse modelling was initially attempted for all four parameters using PEST parameter estimation software.[15] However, the objective function had not stabilised after 12 optimisation iterations and 404 model calls, indicating that the calibration problem may have been ill posed. DEG and ZKD were therefore excluded from inverse modelling, and, instead, response surface analysis was used to evaluate these two parameters. Running MACRO for exhaustive combinations of two parameters and analysing the GOF response surface (Φ^{-1}) provides an effective means of identifying whether a unique best-fit combination exists and shows the best-fit values of the two parameters.[4] The GOF statistic is the inverse of the LOF statistic (Φ). Additionally, response surface analysis effectively identifies potential problems such as correlation between two parameters.[17] A total of 224 MACRO simulations was conducted for DT_{50} values of 1, 2, 3, 4 and 5–50 days (in increments of 5 days) and for K_{oc} values of 5, 10, 15, 20 and 25–300 $cm^3\ g^{-1}$ (in increments of 25 $cm^3\ g^{-1}$). The resulting Φ^{-1} values were plotted in three dimensions using SURFER v.8.03.[32]

The authors focused on adjustment of XMPOR and CTEN by inverse modelling for two cases: 1 – using the SS-based objective function in PEST; 2 – controlling PEST with an alternative LOF statistic described below. In case 1, a typical way of conducting inverse modelling, PEST computed correlations between parameters and a composite sensitivity for each. Relative composite sensitivities were then computed as PEST's composite sensitivity multiplied by the magnitude of the parameter and the square root of N (Hill M, US Geological Survey, private communication, January 2009). The parameter sensitivities and model residuals are used in inverse modelling to determine the magnitude and direction of parameter adjustments required to reduce the objective function.[17]

To compare model predictions with the field observations by inverse modelling, a Perl program was written to sum the daily predictions and compute average IPU concentrations for sampling periods. Although MACRO provides daily and weekly estimates of

drainflow and pesticide concentration, the data were collected at irregular time intervals ranging from 1 to 18 days. The PEST control file invoked a batch file to run MACRO, open model output files and process the output using the Perl program. The Perl output was specified as the model output in the PEST control file. Thus, the model predictions read by the PEST instruction file for case 1 were the aggregate flows and average IPU concentrations in the Perl output. The Perl program also calculated the index of agreement (d) and correlation coefficient (r) between observed and predicted values, and an alternative LOF statistic for each of three time periods that were subsets of the overall period of IPU measurement: period 1 (28 December 1993–30 October 1995), used for calibration by inverse modelling, and periods 2 and 3 (18 October 1999–10 September 2001 and 17 September 2001–28 July 2003 respectively) which were reserved for model validation. These latter periods occur late in the overall period of record and were therefore considered a good test of the extrapolation capabilities of the calibrated model. Whereas r can indicate positive or negative relations between two variables, d varies between 0 and 1 and has a meaning similar to R^2 (note that $R^2 = r^2$). The index of agreement is the SS of observed and predicted values divided by the potential error and subtracted from 1. The potential error is the 'sum of the squared absolute values of the distances from P_i to $\bar{O}$ to O_i', where P_i and O_i are predictions and observations, and $\bar{O}$ is the mean of the observations.[22]

An alternative LOF statistic, called the 'vertical-horizontal distance integrator' (VHDI), was developed for use in case 2 modelling. Case 2 is proposed as an alternative method of model calibration for responsive systems and is designed to handle small time offsets such as those observed at La Jaillière. The VHDI statistic locates the maximum predicted IPU concentration regardless of when it occurs within a time period and compares it with the maximum observed concentration in the same period. Because it considers only the peak concentrations, it is intended for responsive systems that show significant and rapid response to water and chemical inputs at the land surface, and short-lived chemical concentrations. The VHDI is defined as

$$\mathrm{VHDI} = \sqrt{\left(C_{\max\ \mathrm{pred}} - C_{\max\ \mathrm{obs}}\right)^2 + \left(t_{\max\ \mathrm{pred}} - t_{\max\ \mathrm{obs}}\right)^2} \quad (1)$$

where $C_{\max\ \mathrm{pred}}$ is the maximum predicted IPU concentration ($\mu g\ L^{-1}$), $C_{\max\ \mathrm{obs}}$ is the maximum observed IPU concentration ($\mu g\ L^{-1}$), $t_{\max\ \mathrm{pred}}$ is the sampling period during which the maximum predicted IPU concentration occurs (given as the day on which the sampling period begins, with days numbered consecutively from the beginning of the experiment), and $t_{\max\ \mathrm{obs}}$ is the sampling period during which the maximum observed IPU concentration occurs. VHDI comprises vertical and horizontal distance components: $C_{\max\ \mathrm{pred}} - C_{\max\ \mathrm{obs}}$ gives the vertical distance, that is, the difference in maximum IPU concentrations; and $t_{\max\ \mathrm{pred}} - t_{\max\ \mathrm{obs}}$ represents the horizontal distance, that is, the difference in peak arrival times. VHDI essentially is a Pythagorean measure of distance between the observed and predicted maximum IPU concentrations (Fig. 1). As the vertical and/or horizontal distances increase, VHDI and model LOF increase. Although not shown here, the user can supply weights to reduce the influence of either the vertical or the horizontal distance. This might be appropriate if, for example, data have high uncertainty because of sampling and/or measurement errors.

In case 2 modelling, VHDI was calculated outside PEST using the Perl program and was substituted in the following manner for the model prediction read by PEST. A value of zero was specified as a single observed data point in the PEST control file for comparison with VHDI in the Perl output. PEST therefore attempted to drive VHDI to zero through modification of MACRO input parameters within specified bounds by inverse modelling. The PEST objective function in this case was simply the squared value of VHDI, and correlations for parameters were not computed because there were fewer observations than parameters being estimated. However, the approach does yield parameter sensitivities. The reader may consult the PEST user manual[15] for more details on parameter optimisation and uncertainty analysis by inverse modelling.

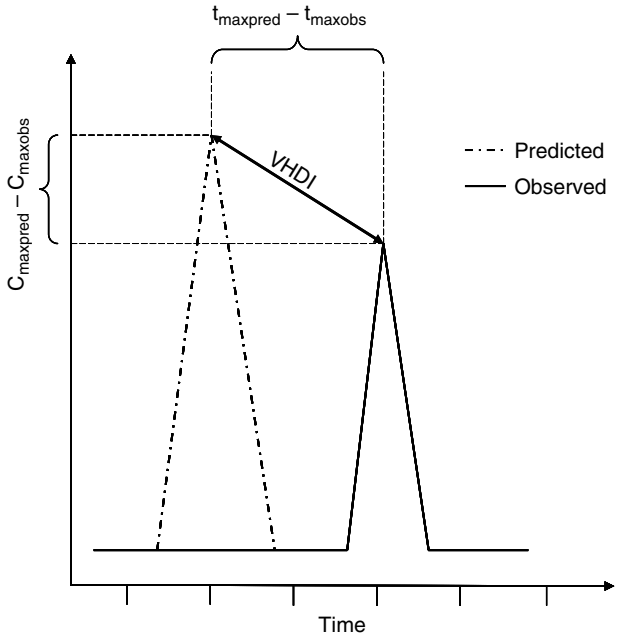

Figure 1. Diagram showing the vertical-horizontal distance integrator (VHDI) and its two contributing distances. $C_{\max\ \mathrm{pred}} - C_{\max\ \mathrm{obs}}$ is the difference in predicted and observed maximum IPU concentrations, and $t_{\max\ \mathrm{pred}} - t_{\max\ \mathrm{obs}}$ is the difference in peak arrival times.

3 RESULTS AND DISCUSSION

3.1 Observed conditions and the uncalibrated MACRO model

Observed drainflow and IPU concentration indicated rapid transport of water and solute to 1 m depth at the La Jaillière site, which is consistent with macropore flow (Figs 2a and b). The peak observed IPU concentration generally occurred within 6–15 days of pesticide application, based on the last date of the sampling period. The fast arrival times agree with prior studies conducted in clay soils. In undisturbed lysimeter studies in the UK, maximum IPU concentration ($17.2\ \mu g\ L^{-1}$) occurred within 20–40 days after treatment for two of seven clay soil replicates.[33] Fast arrival of IPU occurred more frequently in the clay loam (clay content 24.9–29.3%) than in a sandy loam (7.83–13.4% clay), suggesting macropore flow. Similarly, the maximum observed concentration of sulfosulfuron ($2.3\ \mu g\ L^{-1}$) occurred within 10 days after treatment in another UK study involving clay soils (27.6–37.8% clay).[8] The fast arrival times underscore the importance of macropore flow to drains in structured soils with high clay content. The clay contents of the horizons at plot T4 (20.8–49.2%) encompass the amounts cited above for clay soils.

The uncalibrated MACRO model used parameter values shown in Table 2. Both the timing and magnitude of predicted drainflows

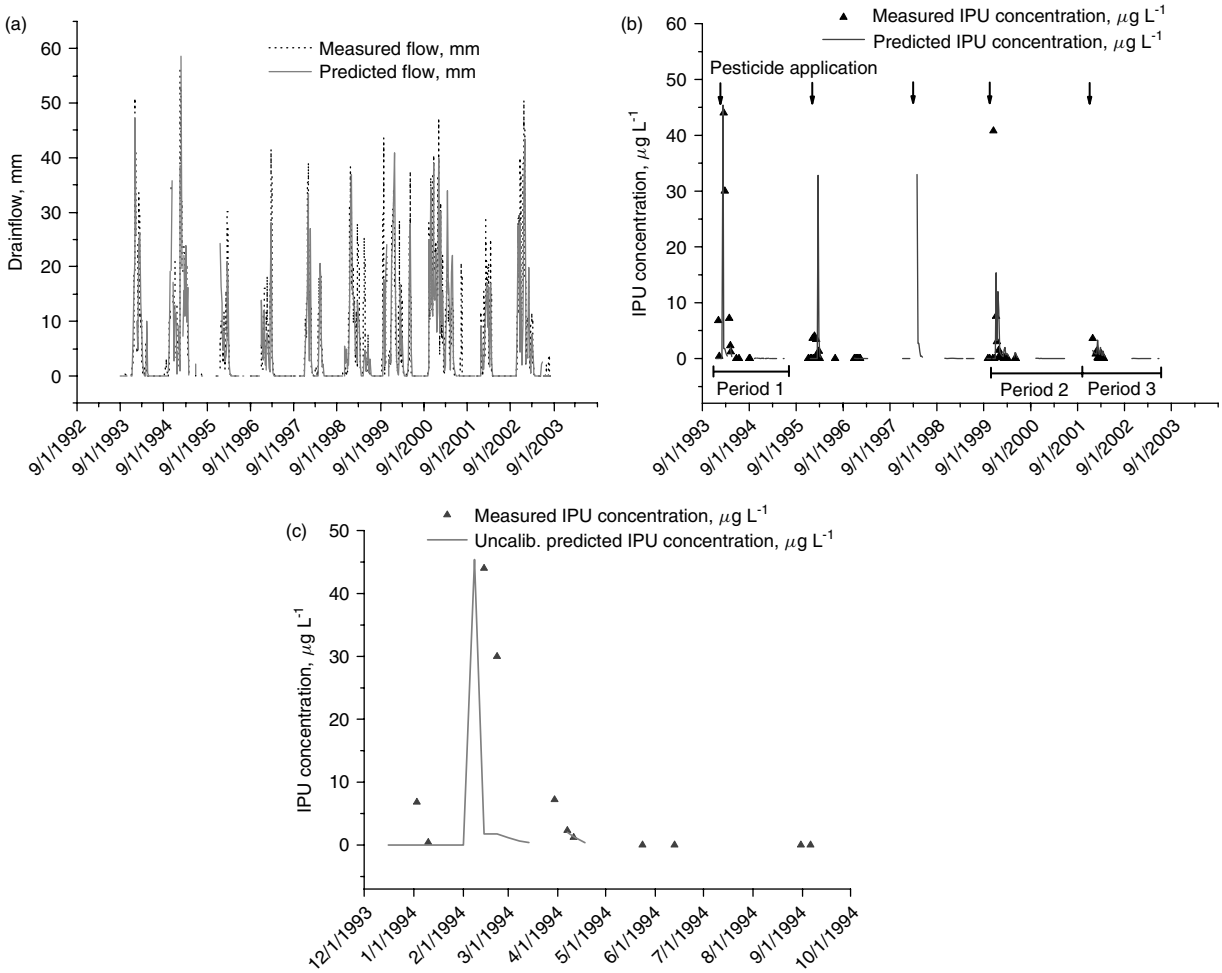

Figure 2. Observed versus predicted (a) drainflows and (b) isoproturon (IPU) concentrations obtained without calibration for the total period of record, and (c) observed and predicted IPU concentrations for period 1.

matched the observed values reasonably well throughout the >10 year period of modelling ($d = 0.942$, $r = 0.899$) (Fig. 2a). The sharp peaks indicated that the system was highly responsive even after aggregating drain flows for the irregular sampling periods. System responsiveness was preserved because the sampling periods were short (1–18 days) compared with the overall period of record. The uncalibrated model also provided fairly good agreement between predicted and observed IPU concentrations, especially for periods 1 and 3 (Fig. 2b). The arrival time of IPU was generally well simulated, in that predicted IPU peaks occurred within 1–4 weeks of the observed peaks. However, the 1 week early arrival of the predicted IPU peak in period 1 (Fig. 2c) resulted in a large SS (2630) and comparatively low values of d and r (0.479 and 0.468 respectively). Additionally, predicted IPU concentrations were substantially less than the corresponding observed values of 6.8 μg L^{-1} on 3 January 1994, 7.2 μg L^{-1} on 30 March 1994, 3.6 μg L^{-1} on 8 January 1996 and 40.8 μg L^{-1} on 15 November 1999 (these dates indicate the start of the sampling periods associated with these concentrations).

3.2 Response surface analysis

The authors used a combination of response surface analysis and inverse modelling in an attempt to improve the timing and amount of predicted peak IPU concentrations for the calibration and validation periods. Period 1 was focused upon for calibration, and periods 2 and 3 were reserved for validation. For starting (uncalibrated) values of parameters, the maximum predicted IPU concentration in period 1 (45.4 μg L^{-1}) arrived 1 week before the observed peak. During the response surface analysis, DT_{50} and K_{oc} were varied in the hope of finding an optimal combination of parameters that might delay the arrival of the predicted peak. However, the various combinations of DT_{50} (1–50 days) and K_{oc} (5–300 cm^3 g^{-1}) only changed the magnitude of the predicted peak and not its arrival time (data not shown). A secondary objective of the response surface analysis was to evaluate SS and the alternative statistic VHDI for the various combinations of DT_{50} and K_{oc}. The inverses of these measures are shown in Fig. 3 such that, the higher the value of the response surface, the better the fit. SS^{-1} continued to increase as K_{oc} and DT_{50} decreased, with no clear optimal combination of parameter values (Fig. 3a). A DT_{50} ridge at values greater than ≈30 days suggests non-uniqueness that might have led to an ill-posed calibration problem when inverse modelling was initially attempted for simultaneous optimisation of both pesticide and macropore parameters. The maximum value of SS^{-1} occurred at $DT_{50} = 5$ days and $K_{oc} = 5$ cm^3 g^{-1}, which contradicts the known behaviour of IPU in both field and laboratory studies. This suggests problems with model structure, the data and/or SS as a measure of model

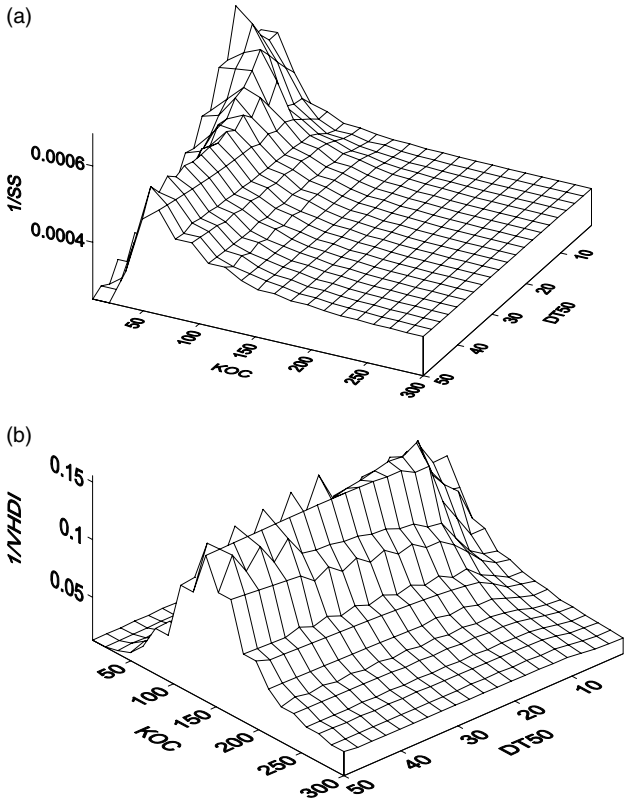

Figure 3. Goodness-of-fit (GOF) response surfaces for various combinations of half-life (DT_{50}, days) and organic carbon sorption coefficient (K_{oc}, $cm^3\ g^{-1}$) for (a) sum of squares (SS) and (b) the vertical-horizontal distance integrator (VHDI). The GOFs are the inverses of these lack-of-fit statistics.

fit with these data. DT_{50} values for IPU obtained under field and laboratory conditions were 19 days for extractable parent and 22–48 days for extractable residue in a prior study.[34] K_{oc} values of 132–174 $cm^3\ g^{-1}$ were observed for IPU in laboratory studies of European soils, one of which was from a drainage ditch in western France with a high clay content (50%).[35,36] The lack of a clear and realistic optimal combination of DT_{50} and K_{oc} in the SS^{-1} response surface underscores limitations with traditional LOF statistics based on one-to-one comparisons in time.

In contrast to the above, the surface plot of $VHDI^{-1}$ clearly indicated improved fit at $K_{oc} = 125\ cm^3\ g^{-1}$ and suggested improved fit at $DT_{50} \approx 20$ days (Fig. 3b). Both values are reasonable and are comparable with those cited in the literature. For example, the K_{oc} value is similar to that shown for IPU in an online pesticide properties database (122 $cm^3\ g^{-1}$).[24] However, the GOF surface, in spite of showing a maximum, indicates potential non-uniqueness in the form of a ridge along the DT_{50} axis.

The values of DEG (20 days) and ZKD (125 $cm^3\ g^{-1}$) based on VHDI response surface analysis and literature values were considered appropriate, and were fixed in subsequent inverse modelling. SS and VHDI were used in case 1 and case 2 modelling for adjustment of XMPOR and CTEN only. It was anticipated that changes in these parameters would alter predicted flow in macropores, potentially delaying the arrival of the maximum predicted IPU concentration in period 1. A delay of only one sampling period would bring the predicted and observed arrival times into agreement.

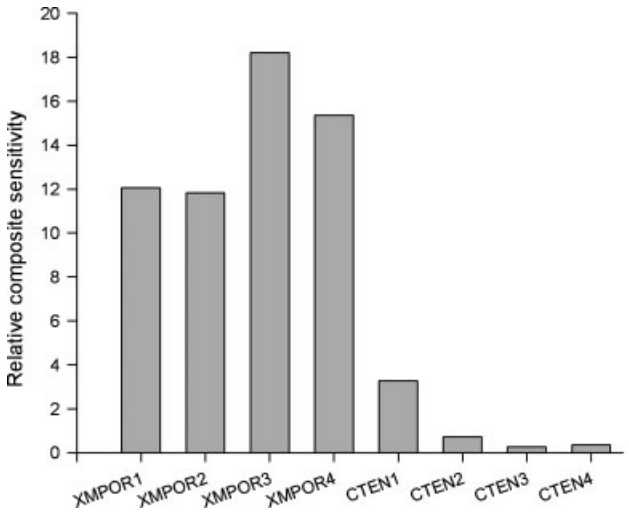

Figure 4. Relative composite sensitivities for MACRO parameters resulting from parameter adjustment by inverse modelling controlled by sum of squares. Definitions of model parameters are provided in Table 2.

3.3 MACRO model calibration and validation

Case 1 modelling controlled by PEST's internal objective function (SS) resulted in six optimisation iterations and 83 model calls and produced the macropore parameter values shown in Table 2. In this case, both relative composite sensitivities and correlation coefficients for parameters were obtained. The relative composite sensitivity measures the composite change in model output resulting from a fractional change in the value of a parameter.[15] The relative composite sensitivities of XMPOR1–4 and CTEN1 were >1, which indicates that these were more precisely estimated than the remaining parameters (Fig. 4).[37] None of the sensitivities was <1% of the maximum sensitivity (18.2 for XMPOR3), suggesting that the precision of the parameter estimates was adequate. Correlations among XMPOR and CTEN were all below 0.95, which indicates that problems with non-uniqueness were not encountered with these parameters (Table 3). The maximum correlation between any two parameters (0.791) occurred with CTEN4 and XMPOR4. Parameter correlations below 0.95 indicate that the parameter values can be individually estimated.[37] Extreme correlation would have required collection of more data to uniquely define the parameters, or setting one of the parameters to a fixed value such that only the ratio of parameters was being estimated.[17]

In case 1 modelling, predicted peak IPU concentration in period 1 was reduced from 45.4 to 44.4 µg L^{-1} compared with the uncalibrated result, which agreed more closely with the maximum observed concentration of 44 µg L^{-1}. However, the predicted arrival time was still 1 week early (Fig. 5a). The result of the time offset was that SS was artificially inflated for the sampling interval starting on 14 February 1994 (which includes the observed peak) because the corresponding predicted IPU concentration was only 1.8 µg L^{-1}. PEST attempted to reduce period 1 SS by increasing the nominal predicted IPU concentrations that were concurrent with the high observed values, which pulled up the predicted IPU curve in mid–late February and also in early April (Fig. 5a). The April predictions (1.4–5.8 µg L^{-1}) overestimated observed IPU concentration. This is in contrast to the uncalibrated modelling results, in which early April predictions (1.3–2.0 µg L^{-1}) were very similar to observed IPU values (1.2–2.3 µg L^{-1}) (Fig. 2c). Additionally, maximum predicted IPU concentration in late January 2002 (8.6 µg L^{-1}) (period 3 used for validation) is

Table 3. Correlation matrix for boundary soil water content (XMPOR) and boundary soil water tension (CTEN) corresponding to soil layers 1 to 4 and derived from inverse modelling based on sum of squares

	XMPOR1	XMPOR2	XMPOR3	XMPOR4	CTEN1	CTEN2	CTEN3	CTEN4
XMPOR1	1.000	−0.459	0.188	−0.072	−0.315	−0.102	−0.056	−0.114
XMPOR2	−0.459	1.000	−0.387	0.108	−0.116	−0.160	0.021	0.023
XMPOR3	0.188	−0.387	1.000	−0.578	−0.487	−0.304	0.388	−0.260
XMPOR4	−0.072	0.108	−0.578	1.000	0.484	0.402	0.102	0.791
CTEN1	−0.315	−0.116	−0.487	0.484	1.000	0.412	0.014	0.281
CTEN2	−0.102	−0.160	−0.304	0.402	0.412	1.000	0.050	0.409
CTEN3	−0.056	0.021	0.388	0.102	0.014	0.050	1.000	0.194
CTEN4	−0.114	0.023	−0.260	0.791	0.281	0.409	0.194	1.000

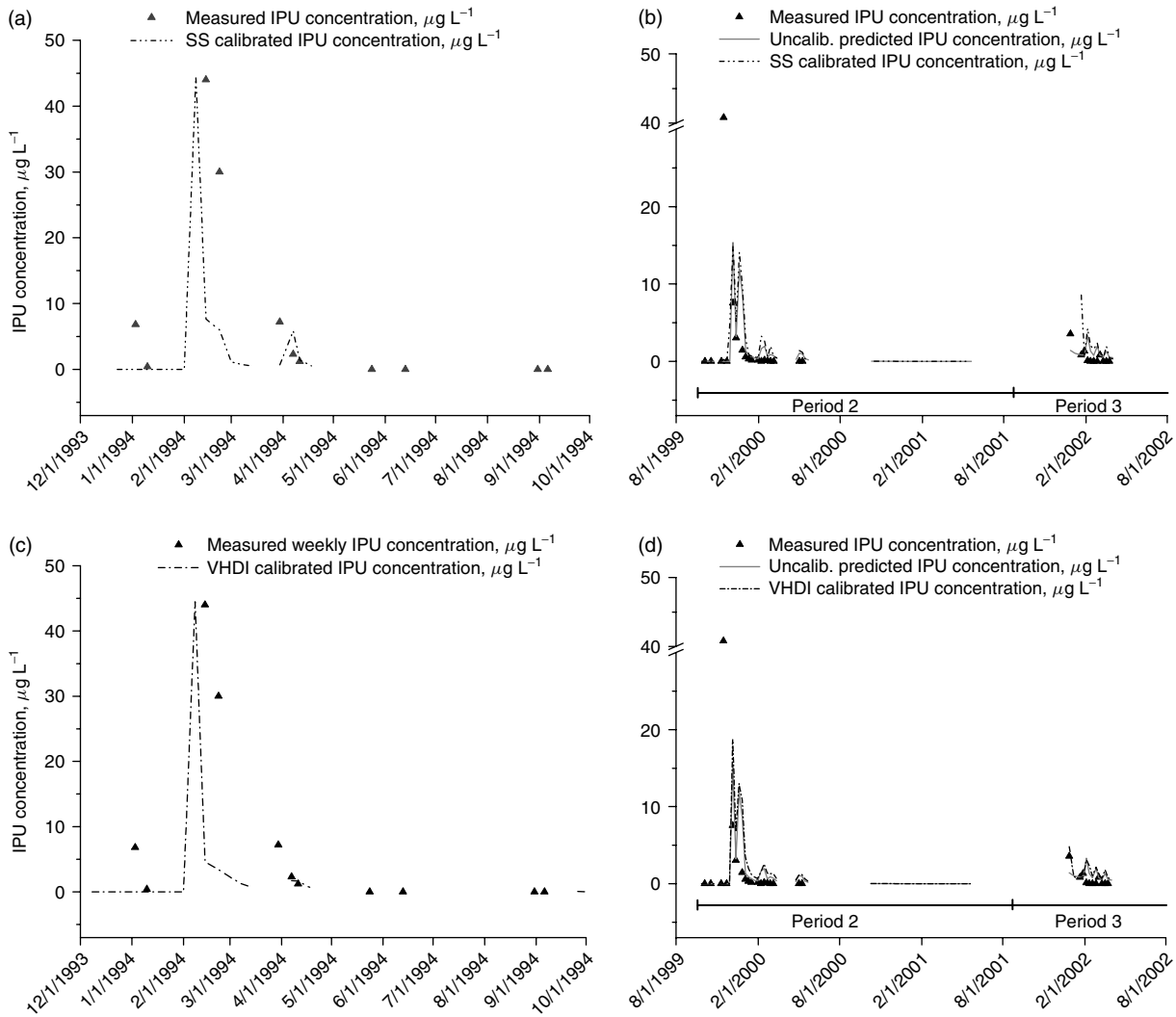

Figure 5. Observed versus predicted IPU concentrations for (a) calibration and (b) validation periods after adjustment of macropore parameters by inverse modelling controlled by sum of squares (SS); (c, d) the same, but inverse modelling controlled by the vertical-horizontal distance indicator (VHDI).

an order of magnitude greater than the corresponding observed value (0.8 μg L^{-1}) (Fig. 5b), and differs substantially from the earlier prediction by the uncalibrated model (0.9 μg L^{-1}).

The correlation between predicted and observed IPU concentrations was somewhat improved by case 1 modelling for calibration period 1 ($r = 0.757$) compared with the uncalibrated model ($r = 0.468$) (Table 4). The index of agreement indicated only marginal improvement for period 1 ($d = 0.520$ for case 1 modelling and $d = 0.479$ for the uncalibrated model). The index of agreement has an SS term in the numerator and is sensitive to extreme values.[22] Therefore, it is subject to the same limitations as the SS LOF statistic when observed and predicted peaks arrive at different times. As SS in the numerator increases, d decreases. The correlation coefficient is also susceptible to extreme values[38] and

Table 4. Model-fit statistics for calibration and validation periods before and after adjustment of macropore parameters by inverse modelling controlled by sum of squares and by the vertical-horizontal distance integrator. Both the sum of squares and index of agreement use the squared differences of observed and predicted values that are concurrent in time

	Calibration	Validation	
Model fit statistic	Period 1 (28 December 1993–30 October 1995)	Period 2 (18 October 1999–10 September 2001)	Period 3 (17 September 2001–28 July 2003)
Uncalibrated			
Sum of squares	2630	1790	23.8
Index of agreement	0.479	0.247	0.316
Correlation coefficient	0.468	0.038	−0.047
Vertical-horizontal difference integrator	6.16	33.0	38.0
Case 1 – calibration by sum of squares			
Sum of squares	2000	1800	94.6
Index of agreement	0.520	0.226	0.170
Correlation coefficient	0.757	0.026	0.172
Vertical-horizontal difference integrator	6.01	33.3	24.5
Case 2 – calibration by vertical-horizontal distance integrator			
Sum of squares	2310	1920	26.4
Index of agreement	0.511	0.233	0.660
Correlation coefficient	0.884	0.023	0.606
Vertical-horizontal difference integrator	6.02	30.5	1.25

should be interpreted with caution. However, the squared differences in r are based on the means of the observed and predicted concentrations and may be more resistant to the effects of the timing offsets. The index of agreement and r were low for period 3 (0.170 and 0.172 respectively), which had substantially lower observed IPU concentrations than period 1, relative to predicted IPU. Case 1 modelling yielded the lowest calibration SS (2000), which was expected because PEST attempts to minimise an SS-based objective function. However, SS increased slightly for periods 2 and 3 compared with the uncalibrated model (Table 4).

Case 2 modelling controlled by VHDI resulted in three PEST iterations and 101 model calls and produced the parameter values shown in Table 2. The maximum predicted IPU concentration for period 1 (45.4 μg L^{-1}) was reduced to 44.5 μg L^{-1}. However, the arrival time of the predicted IPU peak remained unchanged (Fig. 5c). Nonetheless, predicted IPU concentrations in early April (1.7–1.8 μg L^{-1}) were bracketed by the observed values, and predicted concentrations in validation periods 2 and 3 fitted the observed data fairly well (Fig. 5d). Model fit for period 1 improved compared with the uncalibrated version according to all four indices. For example, the correlation coefficient increased to 0.884 and the value of VHDI decreased from 6.16 to 6.02 (Table 4). The index of agreement (0.511) saw less improvement than r discussed above. The improved model fit was particularly dramatic for validation period 3, which saw a reduction in VHDI from 38.0 to 1.25, an increase in r from −0.047 to 0.606 and an increase in d from 0.316 to 0.660. The peak predicted IPU concentration for period 3 (4.8 μg L^{-1}) agrees more closely with the maximum observed value of 3.6 μg L^{-1} (Fig. 5d), which was improved over both the uncalibrated model and case 1 modelling results. VHDI also indicated some improvement in model fit for validation period 2 compared with case 1 modelling (Table 4). The maximum predicted IPU concentration increased from 15.4 (uncalibrated version) to 18.7 μg L^{-1} for period 2, which is somewhat closer to the observed peak value of 40.8 μg L^{-1} (Fig. 5d).

Figures 6a to d are provided to enable more direct comparison of cases 1 and 2 for the four model fit indices (SS, d, r and VHDI). Period 1 SS decreased from 2630 to 2000 in the following order: uncalibrated model > case 2 modelling > case 1 (Fig. 6a). As mentioned above, this is reasonable because case 1 modelling was controlled by the internal objective function in PEST. Correlations between predicted and observed IPU concentrations increased for both periods 1 and 3 in the following order: uncalibrated model < case 1 < case 2 (Fig. 6c). Thus, the correlations were maximised for periods 1 ($r = 0.884$) and 3 ($r = 0.606$) when the inverse modelling was controlled by VHDI. SS-based inverse modelling improved correlations ($r = 0.757$ and 0.172 for periods 1 and 3 respectively) compared with the uncalibrated model, but not to the same degree.

All four indices indicate that model fit was better for the second of two validation periods when the inverse modelling was controlled by VHDI rather than SS (SS = 26.4, $d = 0.660$, $r = 0.606$, and VHDI = 1.25). VHDI decreased for period 3 in the following order: uncalibrated model > case 1 > case 2 (Fig. 6d). As discussed above, reduction in VHDI in case 2 modelling was particularly dramatic for period 3 and was corroborated by visual inspection of Fig. 5d. Model fit in period 3 was also improved, compared with the uncalibrated version, by SS-controlled inverse modelling (VHDI = 24.5), but not to the same degree (Fig. 6d).

Neither method of calibration was able to shift the timing of the predicted peak IPU concentration in period 1. This lack of movement may have resulted from aggregation of IPU concentrations over the irregular sampling periods. For calibration purposes, the daily predicted values were processed by the Perl program to obtain average values for comparison with the composite samples. Thus, the predicted daily values are somewhat buffered from the inverse modelling step. Plots of daily values are useful in order more readily to discern the effects of macropore parameter adjustment on predicted drainflows and IPU concentrations, and are discussed here in the context

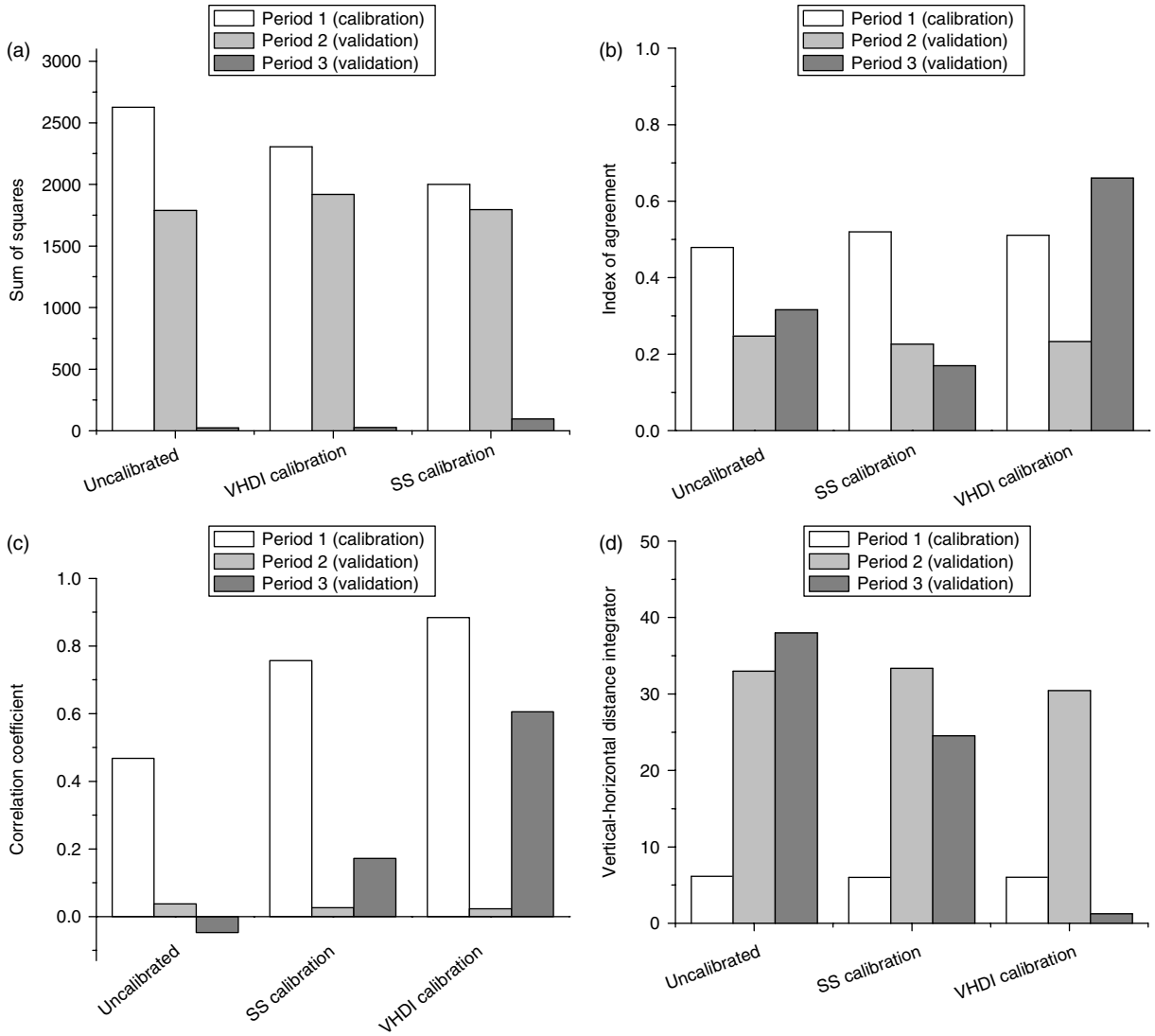

Figure 6. Model fit indices for the uncalibrated MACRO model and for inverse modelling controlled by sum of squares (SS) and by the vertical-horizontal distance integrator (VHDI), for calibration and validation periods. The four model fit indices are (a) SS, (b) index of agreement, (c) correlation coefficient and (d) VHDI. Both SS and the index of agreement use the squared differences of observed and predicted values that are concurrent in time.

of calibration using VHDI. For an individual drainage event, drainflow and IPU concentration at early times were slightly reduced after parameter adjustment, while those at later times increased somewhat. For example, MACRO predicted almost twofold increases in drainflow (Fig. 7a) and over threefold increases in IPU concentration (Fig. 7b) on the recession part of the curves during 13–20 February 1994 as a result of the adjusted macropore parameters. Relative composite sensitivities computed by PEST during case 2 modelling suggest that most of the effect resulted from changes in XMPOR1, XMPOR2 and XMPOR3 (data not shown). Estimates of all three parameters were less than the starting values (Table 2). These effects were confirmed by performing additional MACRO simulations in which XMPOR1, XMPOR2 and XMPOR3 were decreased one at a time while all other parameters were held constant. Parameter XMPOR1 (boundary water content in the first soil layer) influenced drainflows and IPU concentrations at early times for a particular drainage event, whereas XMPOR2 and XMPOR3 (layers two and three) influenced later times. In general, decreasing XMPOR resulted in greater generation of macropore flow by the model. This contributed to an overall increase in predicted drainflows and IPU fluxes attributable to macropores during the >10 year simulation period. Cumulative predicted drainflow in macropores increased from 3910 to 4090 mm after parameter adjustment by VHDI, and cumulative predicted IPU flow from macropores to drains increased from 7.1 to 8.2 mg m^{-2}. These changes were sufficient to have increased predicted peak IPU concentration in both validation periods, which benefited case 2 model fit for these periods. The effect was more subtle than in case 1 modelling, which saw overprediction of IPU for both calibration and validation periods.

4 CONCLUSIONS

The following conclusions were drawn from this research:

1. The uncalibrated MACRO model fitted the observed drainflow and IPU concentration data reasonably well, especially considering that the simulation period exceeded 10 years.

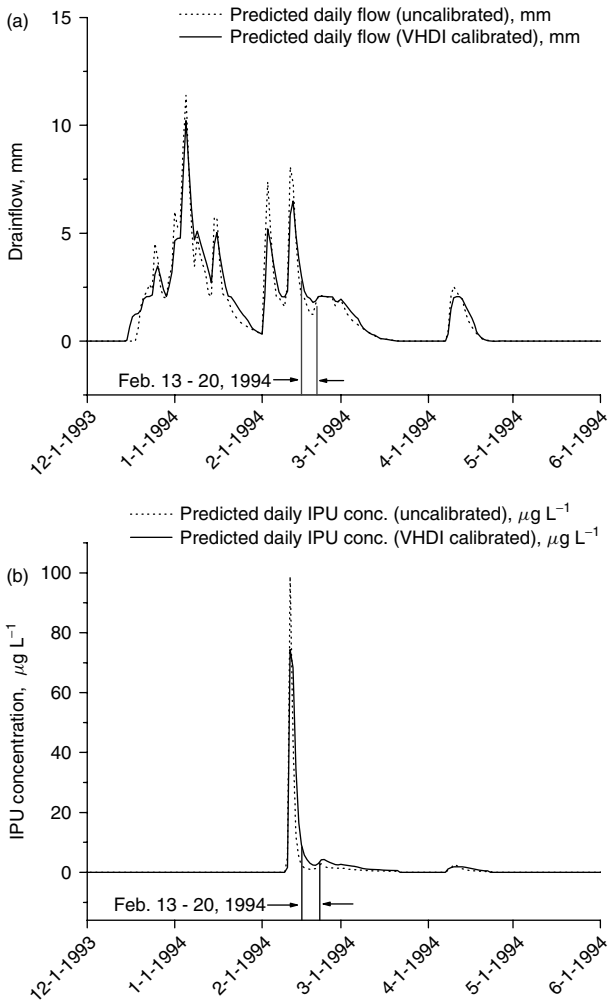

Figure 7. Observed and predicted daily flows (a) and daily IPU concentrations (b) for period 1 before and after parameter adjustment by inverse modelling controlled by the vertical-horizontal distance indicator (VHDI).

Isoproturon arrival times were simulated more accurately than peak concentrations.

2. VHDI used in conjunction with inverse modelling software provided an effective, alternative means of calibrating the MACRO model for a system characterised by transient drainflows and short-lived pesticide concentrations. This approach outperformed SS-based inverse modelling for the second of two validation periods according to all four model fit indices (SS = 26.4, $d = 0.660$, $r = 0.606$ and VHDI =1.25). SS did not account for offsets in arrival times of observed and predicted peak solute concentrations, which is potentially limiting in responsive systems. Similarly, d was susceptible to the time offsets because it includes SS in the numerator, which may have limited its effectiveness somewhat with these data.
3. Macropore parameters generally were well estimated by these data, according to parameter correlations and sensitivities from inverse modelling controlled by SS. All of the parameter correlations were well below 0.95 (absolute value basis), a threshold above which parameter non-uniqueness is a problem. XMPOR1–4 and CTEN1 had relative composite sensitivities greater than 1, indicating that they were more precisely estimated than the remaining CTEN parameters. Inverse modelling controlled by VHDI did not yield parameter correlations. However, the $VHDI^{-1}$ response surface provided useful information on potential parameter non-uniqueness and allowed identification of an optimal K_{oc}–DT_{50} combination. Thus, response surface analysis simplified the inverse modelling and was a useful complement to VHDI.
4. Neither method of inverse modelling (VHDI or SS controlled) improved the predicted IPU arrival time for period 1, which was early by one sampling interval. This is probably because the predicted daily flows had to be aggregated to match the irregular sampling periods in order to compute the inverse modelling objective functions.

ACKNOWLEDGEMENTS

This work was undertaken within the scope of a research project funded by the French pesticide registration authorities to evaluate the potential use of preferential flow models to support the development of national drainage scenarios. The field drainage data at La Jaillère were collected by Arvalis-Institut du Végétal. Dr Nolan benefited from a visiting scientist contract with the regional research agency 'Le STUDIUM' from the Centre region in France. The funding of the work by the French registration authorities, the granting of access to the field data by Arvalis-Institut du Végétal and the financial support provided by STUDIUM are gratefully acknowledged.

REFERENCES

1 FOCUS, Leaching models and EU registration. EC document reference 4952/VI/95 (1995).
2 FOCUS, Soil persistence models and EU registration. The final report of the work of the Soil Modelling WG of FOCUS (1997).
3 USEPA, FIRST: A screening model to estimate pesticide concentrations in drinking water. US Environmental Protection Agency, Washington, DC (2008).
4 Dubus IG, Beulke S, Brown CD, Gottesburen B and Dieses A, Inverse modelling for estimating sorption and degradation parameters for pesticides. *Pest Manag Sci* **60**:859–874 (2004).
5 Jarvis NJ, MACRO, a model of water movement and solute transport in macroporous soils. Swedish University of Agricultural Sciences, Department of Soil Sciences (1991).
6 Larsbo M and Jarvis NJ, MACRO 5.0. A model of water flow and solute transport in macroporous soil. Technical description. Swedish University of Agricultural Sciences Emergo 20-03 : 6 Report (2003).
7 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).
8 Brown CD, Dubus IG, Fogg P, Spirlet M and Gustin C, Exposure to sulfosulfuron in agricultural drainage ditches: field monitoring and scenario-based modelling. *Pest Manag Sci* **60**:765–776 (2004).
9 Holman IP, Dubus IG, Hollis JM and Brown CD, Using a linked soil model emulator and unsaturated zone leaching model to account for preferential flow when assessing the spatially distributed risk of pesticide leaching to groundwater in England and Wales. *Sci Total Environment* **318**:73–88 (2004).
10 Larsson MH and Jarvis NJ, Evaluation of a dual-porosity model to predict field-scale solute transport in a macroporous clay soil. *J Hydrol* **215**:153–171 (1999).
11 Jarvis NJ, Review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality. *Eur J Soil Sci* **58**:523–546 (2007).
12 Flury M, Leuenberger J, Studer B and Flühler H, Transport of anions and herbicides in a loamy and a sandy field soil. *Water Resour Res* **31**:823–835 (1995).
13 Scorza RP, Jr, Jarvis NJ, Boesten JJTI, van der Zee SEATM and Roulier S, Testing MACRO (version 5.1) for pesticide leaching in a Dutch clay soil. *Pest Manag Sci* **63**:1011–1025 (2007).

14 Beulke S, Renaud F and Brown CD, Development of guidance on parameter estimation for the preferential flow model MACRO 4.2. Cranfield Centre for EcoChemistry Final Report, DEFRA Project PL0538, Contract No. JA3749E (2002).
15 Doherty J, *PEST-Model Independent Parameter Estimation User Manual*, 5th edition. Watermark Numerical Computing, Brisbane, Australia (2004).
16 Poeter EP, Hill MC, Banta ER, Mehl S and Christensen S, UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation. US Geological Survey Techniques and Methods 6-A11 (2005).
17 Poeter EP and Hill MC, Inverse models: a necessary next step in ground-water modeling. *Ground Water* **35**:250–260 (1997).
18 Armstrong A, Aden K, Amraoui N, Diekkrüger B, Jarvis N, Mouvet C, *et al*, Comparison of the performance of pesticide-leaching models on a cracking clay soil: results using the Brimstone Farm dataset. *Agric Water Manag* **44**:85–104 (2000).
19 Dubus IG, Beulke S and Brown CD, Calibration of pesticide leaching models: critical review and guidance for reporting. *Pestic Manag Sci* **58**:745–758 (2002).
20 Carluer N, Dubus IG, Gouy V, Gril JJ and Surdyk N, Contribution à l'amélioration de l'évaluation du risque a priori pour les eaux superficielles par les phytosanitaires: soutien à la construction de scénarios nationaux pour l'évaluation des préparations. Cemagref – BRGM – Arvalis Institut du Végétal Final Report (2007).
21 Willmott CJ, Some comments on the evaluation of model performance. *Bull Am Meteorol Soc* **63**:1309–1313 (1982).
22 Legates DR and McCabe GJ, Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model evaluation. *Water Resour Res* **35**:233–241 (1999).
23 Beard G, *Report on Soil Profile Characteristics and Topsoil Structure at the Arvalis Experimental Site, La Jaillière, France.* [Online]. Available: http://www.landlook.fsnet.co.uk/default.htm [September 2006].
24 *FOOTPRINT Pesticide Properties Database.* [Online]. Available: http://www.herts.ac.uk/aeru/footprint/en/ [July 2007].
25 Jarvis NJ, Hollis JM, Nicholls PH, Mayr T and Evans SP, MACRO-DB: a decision-support tool for assessing pesticide fate and mobility in soils. *Environmental Modelling and Software* **12**:251–265 (1997).
26 Jarvis NJ, *The MACRO Model (Version 4.3) – Technical Description.* [Online]. Available: http://www-mv.slu.se/bgf/Macrohtm/macro43b/ TechMacro43.pdf [November 2006].
27 Brooks RH and Corey AJ, Hydraulic properties of porous media. Colorado State University Hydrology Paper 3 (1964).
28 Mualem Y, A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resour Res* **12**:513–522 (1976).
29 FOCUS, Surface Water Scenarios in the EU Evaluation Process under 91/414/EEC. EC Document Reference SANCO/4802/2001-rev. 2 (2001).
30 Wösten JHM, Lilly A, Nemes A and Bas CL, Development and use of a database of hydraulic properties of European soils. *Geoderma* **90**:169–185 (1999).
31 Dubus IG and Brown CD, Sensitivity and first-step uncertainty analyses for the preferential flow model MACRO. *J Environ Qual* **31**:227–240 (2002).
32 *Golden Software I.* [Online]. Available: http://www.goldensoftware.com/ [July 2007].
33 Beulke S, Brown CD, Dubus IG, Fryer CJ and Walker A, Evaluation of probabilistic modelling approaches against data on leaching of isoproturon through undisturbed lysimeters. *Ecol Model* **179**:131–144 (2004).
34 Perrin-Ganier C, Breuzin C, Portal J-M and Schiavon M, Availability and persistence of isoproturon under field and laboratory conditions. *Ecotoxicol Environ Safety* **35**:226–230 (1996).
35 Nemeth-Konda L, Füleky G, Morovjan G and Csokan P, Sorption behaviour of acetochlor, atrazine, carbendazim, diazinon, imidacloprid and isoproturon on Hungarian agricultural soil. *Chemosphere* **48**:545–552 (2002).
36 Patria L, Griseau M, Le Brun O, Merlet N and Dore M, Approach of degradation ways of two selected substituted ureas, isoproturon and chlortoluron in the water/sediments system of a swamp's drainage ditches. *J Eur d'Hydrol* **26**:101–120 (1994).
37 Hill MC and Tiedeman CR, *Effective Groundwater Model Calibration with Analysis of Data, Sensitivities, Predictions, and Uncertainty*. John Wiley & Sons, Inc., Hoboken, NJ (2007).
38 Helsel DR and Hirsch RM, *Statistical Methods in Water Resources*. Elsevier, New York, NY (1992).

