Uncertainty assessment in watershed-scale water quality modeling and management:

1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach

Yi Zheng and Arturo A. Keller

Received 14 July 2006; revised 8 May 2007; accepted 18 May 2007; published 4 August 2007.

1 Watershed-scale water quality models involve substantial uncertainty in model output because of sparse water quality observations and other sources of uncertainty. Assessing the uncertainty is very important for those who use the models to support management decision making. Systematic uncertainty analysis for these models has rarely been done and remains a major challenge. This study aimed (1) to develop a framework to characterize all important sources of uncertainty and their interactions in management-oriented watershed modeling, (2) to apply the generalized likelihood uncertainty estimation (GLUE) approach for quantifying simulation uncertainty for complex watershed models, and (3) to investigate the influence of subjective choices (especially the likelihood measure) in a GLUE analysis, as well as the availability of observational data, on the outcome of the uncertainty analysis. A two-stage framework was first established as the basis for uncertainty assessment and probabilistic decision-making. A watershed model (watershed analysis risk management framework (WARMF)) was implemented using data from the Santa Clara River Watershed in southern California. A typical catchment was constructed on which a series of experiments was conducted. The results show that GLUE can be implemented with affordable computational cost, yielding insights into the model behavior. However, in complex watershed water quality modeling, the uncertainty results highly depend on the subjective choices made by the modeler as well as the availability of observational data. The importance of considering management concerns in the uncertainty estimation was also demonstrated. Overall, this study establishes guidance for uncertainty assessment in management-oriented watershed modeling. The study results have suggested future efforts we could make in a GLUE-based uncertainty analysis, which has led to the development of a new method, as will be introduced in a companion paper. Eventually, the study should assist in the development of a new generation of watershed water quality models.


1. Introduction

Managing nonpoint sources of pollution at the watershed scale is a major global issue. Increasingly, watershed-scale water quality models, such as HSPF [e.g., Bicknell et al., 2001], SWAT [e.g., Neitsch et al., 2001] and watershed analysis risk management framework (WARMF) [Chen et al., 1996, 1999, 2004; Keller et al., 2004], are being used to assist in managing watershed nonpoint sources of pollution. Although these physically based models include considerable hydrologic, biogeochemical, spatial and temporal complexity, their simulations are subject to significant uncertainty. Nevertheless, the output of these complex models is valuable for many reasons: (1) the temporal pattern of water quality parameters, such as long-term trends and critical timing of concentration spikes, can be adequately simulated and can lead to better management decisions and practices; (2) the spatial distribution of critical areas can be discerned to develop better monitoring programs and target management actions; and (3) comparison between different scenarios can generate valuable information for management, even though there is uncertainty about specific scenario values.

Systematic uncertainty assessment would be very useful for those who implement these models and/or depend on the output of these models to make decisions. For example, while these models have been applied in total maximum daily load (TMDL) calculations to support decision making, the unaddressed estimation of uncertainty often leads to indefensible decisions with regards to the margin of safety for the TMDL. However, a process for...
conducting an uncertainty analysis for these models has not been adequately addressed by previous studies. In addition, the role of the management objectives is largely overlooked in most uncertainty assessments. These critical issues motivated this study.

[5] There have been several studies [e.g., Beck, 1987; Craig et al., 2001; Kennedy and O’Hagan, 2001] classifying sources of uncertainty in the modeling process from different perspectives. However, for watershed water quality modeling which involves significant temporal and spatial variability and requires a large amount of input data, the results of these studies cannot be directly applied. Also, the consideration of management concerns in the modeling process was not discussed in these studies. In section 2 of this paper, we first introduce a framework for assessing the sources of uncertainty involved in a management-oriented watershed modeling process. This framework constitutes the basis for implementing appropriate techniques to conduct an uncertainty analysis of a watershed model, as well as for evaluating probabilistic management decisions.

[5] A number of general methods for uncertainty analysis are available, including Taylor expansion-based methods [e.g., Najj et al., 1998], Rosenblueth’s method [e.g., Rosenblueth, 1975], stochastic response surface (SRS) methods [e.g., Cryer and Applequist, 2003a, 2003b], Karhunen-Loeve moment equation (KLME) approach [e.g., Chen et al., 2005, 2006] and Monte Carlo (MC) based methods [e.g., Helton, 1993]. Although computationally intensive, MC-based methods are considered to be the most applicable for complex watershed models. The feasibility of other methods significantly decreases because of the nonlinearity, discontinuity, analytical intractability and complicated parameter interactions within the watershed models. One of the few studies that applied MC-based uncertainty analysis for a watershed water quality model (SWAT) was done by Sohrabi et al. [2003]. However, only model parameter uncertainty was addressed in the study, which is a common limitation of most uncertainty studies. For watershed modeling where uncertainty from other sources is substantial, an analysis of global uncertainty is necessary.

[5] Different strategies have been developed, mostly with regards to hydrology, to deal with global uncertainty within an MC framework. The first category of strategies includes calibration-based methods such as Bayesian recursive estimation (BaRE) [Thiemann et al., 2001]. Such methods model a lumped error term explicitly and search for an optimal parameter set, using MC simulation. Multiobjective approaches such as multiobjective complex evolution (MOCOM-UA) [Yapo et al., 1998] and multiobjective calibration iterative procedure (MCIP) [Demarty et al., 2005], are in the second category. These methods pursue a small set of Pareto optimal parameter sets, instead of one single optimum, through calibration/optimization. Error terms are implicitly treated. The third category, represented by the generalized likelihood uncertainty estimation (GLUE) method [Beven and Binley, 1992], replaces the calibration/optimization ideology with acceptance of the existence of many behavioral (i.e., acceptable with regard to a given criterion) parameter sets. Error terms are also implicitly considered in GLUE. The GLUE approach has been applied to a variety of models up to a medium level of complexity [e.g., Freer et al., 1996; Franks and Beven, 1997; Schulz et al., 1999; Zak and Beven, 1999; Page et al., 2003; Romanovicz and Beven, 2003; Mertens et al., 2004]. A brief discussion about the GLUE approach is provided in section 3.

[7] This paper is a detailed discussion on the application of GLUE to complex watershed models within the management context. There are several reasons for selecting GLUE. First, GLUE accounts for all sources of uncertainty, either explicitly or implicitly. Second, GLUE is conceptually simple, and requires no restricted error assumptions if a goodness-of-fit measure is used as its likelihood function. Third, GLUE is less vulnerable to model discontinuity since no optimum searching is necessary. Finally, compared to multiobjective approaches, GLUE’s uncertainty bounds are more likely to reflect the real magnitude of uncertainty, since GLUE identifies many nonoptimal simulations that yet conform to the observations to certain degree. We have seen only one study [Muleta and Nicklow, 2005] using GLUE for a watershed water quality model (SWAT). Twelve uncertain parameters were considered in the GLUE analysis for streamflow simulation, and eight for sediment simulation. Nevertheless, the study is a straightforward application of GLUE. The dependence of uncertainty results on subjective choices (e.g., likelihood measures, behavioral criteria, etc.) made by modelers was not examined, and the implication of such dependence for water quality management was not discussed.

[8] For this study, WARMF was used as an example of a complex watershed model. Pesticide (diazinon) concentration was the output variable studied. On the basis of the results of a previous sensitivity analysis (A. A. Keller and Y. Zheng, Uncertainty and variability in the source terms of a TMDL calculation: 1. Point sources, submitted to Journal of the American Water Resources Association, 2006, hereinafter referred to as Keller and Zheng, submitted manuscript, 2006a), 39 catchment, river reach or general parameters were taken into account in the uncertainty analysis. A series of GLUE experiments were carried out on a typical catchment of the Santa Clara River in California, as a case study.

[9] It should be made clear that the GLUE approach does not impose any restrictions on likelihood measures or the conditions for conducting the analysis. However, there are some goodness-of-fit measures that have been traditionally used as likelihood measures, and the modeler must make some decisions on certain conditions for the analysis, such as the number of runs, the criteria for deciding simulations that are behavioral, etc. In this manuscript the goal was to implement GLUE following the traditional likelihood measures used in several previous studies, and to evaluate the effect of various conditions on the uncertainty analysis. Thus this study is not an evaluation of the GLUE approach in general; we evaluate the traditional approach used in specific GLUE analysis, which for the most part have focused on hydrologic modeling rather than water quality modeling.

[10] The main objectives of this study were to (1) develop a framework to characterize the sources of uncertainty and their interactions in management-oriented watershed modeling; (2) apply the GLUE approach for quantifying simulation uncertainty for complex watershed models; (3) investigate the influence of subjective choices (especially the likelihood measure) in a specific GLUE analysis, as well
2. Framework for Characterizing Uncertainty

A suitable framework for determining the sources of uncertainty and their interaction is necessary in assessing global uncertainty, as well as for reducing its magnitude. Figure 1 illustrates the framework that we propose for management-oriented watershed modeling. The entire process is divided into a linking/calibration stage and a planning/management stage. Here “linking” refers to connecting cause and effect through explicit processes. In the first stage, stochastic model simulations are generated assimilating the uncertainty from process parameter values, input (in our framework, “input” specifically refers to dynamic input data, such as rainfall, pollutant loads, etc., and geographic information, such as elevation, land use, etc.) and model structure, and then constrained by observational data (involving observational uncertainty) to derive a linkage with quantifiable uncertainty. Linkage can be achieved with an optimal parameter set, using explicitly assumed error term(s), as in calibration-based methods [e.g., Doherty, 2003], or multiple behavioral parameter sets linked with respective probability values, as in the GLUE approach.

In the second stage, the established linkage is used to generate stochastic predictions which lead to probabilistic management judgment. Two new types of uncertainty are involved in this stage. The first one is change in linkage (hereafter referred as “linkage uncertainty”) which means that the established linkage between driving forces and concerned water quality parameters could actually vary in the future, or because of a management action. Such variability may be due to internal changes of the watershed system not captured by the watershed model (e.g., surface roughness could be substantially altered by habitat restoration). It can also be caused by external changes (e.g., different loading scenarios may result in different reaction kinetics for a pollutant if redox conditions are modified). Another uncertainty is future input uncertainty, such as uncertainty associated with future weather or future loads. Considering the uncertainty in the planning stage is crucial for evaluating probabilistic management options, but has been overlooked both in the literature and in management practice. Note that the “environmental objectives” in Figure 1 may also have uncertainty (e.g., the water quality objective may be derived on the basis of uncertain data [Reckhow et al., 2005]). Nevertheless, since environmental objectives are usually preset for a specific management problem, their uncertainty is not considered in our framework.

With regards to the framework, the output variable of interest (e.g., measured water quality) in the linking stage can be decomposed as follows:

\[
Z\left(X^*, \theta^*, t\right) = y(X, \theta, t) + \delta(X, \theta, t) + \epsilon\left(X^*, \theta^*, t\right)
\]

(1)

where \(X\) and \(\theta\) represent the input matrix and parameter vector required by the watershed model \(y(\cdot)\); \(X^*\) and \(\theta^*\) represent true inputs and true parameters, both of which are unknown system attributes in reality; \(t\) is the time index; and \(Z(\cdot, t), y(\cdot, t), \delta(\cdot, t)\) and \(\epsilon(\cdot, t)\) stand for observation, model response, model structure error and observational error at time \(t\), respectively. Differentiating \(X^*\) and \(\theta^*\) from \(X\) and \(\theta\) implies that observational error \(\epsilon\) is independent of model simulation \(y\) and structure error \(\delta\). Additionally, the true watershed response can be expressed as

\[
\hat{y}(X^*, \theta^*, t) = Z(X^*, \theta^*, t) - \epsilon\left(X^*, \theta^*, t\right)
\]

(2)
In the planning stage, what is of real importance for management is the future true watershed response, not the potential observation. Thus this stage can be modeled as

\[ y_f^* (x_f^*, \theta_f^*, t) = y(x_f, \theta_f, t) + \delta_f (x_f, \theta_f, t) \]  

Here the subscript “f” (future) is the index for the planning stage. The equation assumes that the same watershed model \( y(\cdot) \) is used in both stages. With respects to equations (1) and (3), parameter uncertainty and input uncertainty are associated with \( \theta \) and \( X \), respectively; model structure uncertainty and observational uncertainty are correspondingly represented by \( \delta \) and \( \varepsilon \); future input uncertainty is reflected by the change from \( X \) to \( X_f \) (or \( X^* \) to \( \theta_f^* \)); and linkage uncertainty is reflected by the change from \( \theta \) to \( \theta_f \) (or \( \theta^* \) to \( \theta_f^* \)), as well as from \( \delta \) to \( \delta_f \). Note that this paper only addresses the uncertainty in the first stage. The linkage uncertainty and future input uncertainty will be studied separately.

In semidistributed watershed models like SWAT and WARMF, most of the model parameters are conceptual representations of abstract watershed characteristics (e.g., catchment-wide hydraulic conductivities), and therefore cannot be directly measured in the field. Even if a parameter is potentially measurable at the watershed scale, insufficient data, measurement inaccuracy or intrinsic randomness (e.g., temporal variability) could result in uncertainty as well. In general, parameter uncertainty is treated explicitly (e.g., probabilistic distributions assigned in MC simulation), because it significantly contributes to simulation inaccuracy and is relatively easy to analyze probabilistically.

In many cases, a water quality measurement would be a point measurement in time and in space, while prediction and management are normally targeted at a coarser resolution (e.g., daily and for a whole reach, respectively). Hence observational error \( \varepsilon \) generally consists of temporal variability and commensurability error (which depend on the watershed system’s properties \( X^* \) and \( \theta^* \)), as well as traditional measurement errors.

In watershed water quality modeling, there are two types of inputs (\( X \)). The first one involves the forcing functions including meteorological data (e.g., daily precipitation, temperature, etc.) and pollutant loads from various sources (e.g., wastewater treatment plant, land application, atmospheric deposition, etc.). The other type of input is the geographic information including topography, land use and river network data. Directly considering the uncertainty in \( X \) is difficult since \( X \) contains both dynamic and spatially distributed variables with largely unknown covariance matrices. Making strong assumptions on \( X \)’s covariance matrices without solid physical knowledge could only complicate the estimation of uncertainty. The strategy that we applied was to convert the uncertainty of \( X \) to parameter uncertainty and model structure uncertainty, as discussed below.

In semidistributed watershed models, weather data from discrete sites are used to represent conditions of continuous areas. Generally, spatial variability is the main cause of uncertainty for the meteorological component of \( X \). In WARMF, each meteorological station is assigned a set of multipliers including precipitation weighting coefficients and temperature lapse rates. These multipliers are specific model parameters with which uncertainty in meteorological inputs is partially converted to parameter uncertainty. This “multiplier approach” can be easily extended to other watershed models. Note that such conversion is linear, and not complete. The unexplained uncertainty is then assimilated into model structure uncertainty \( \delta \) which needs to be addressed with other strategies.

For the pollutant load component of \( X \), data resolution, both temporal and spatial, is the major cause of uncertainty (Keller and Zheng, submitted manuscript, 2006a). The “multiplier approach” can be implemented as well. For each pollutant source, one or several “load adjustment coefficients” can be created externally, and treated as elements of \( \theta \). For example, we may create two multipliers for pesticide application rates in two different periods (in each period, the multiplier is time invariant). Again, “load adjustment coefficients” can only explain a certain portion of input uncertainty, and the remaining portion is captured in \( \delta \).

In general, watershed models adopt certain algorithms to process geographic input into spatially distributed parameters (hereafter referred to as geographic parameters) including topographic parameters for the catchments and geometric parameters for river segments. For instance, SWAT has an internal module for processing topographic data such as digital elevation model (DEM), while WARMF uses external tools (e.g., BASINS or ArcHydro GIS) to do the job. The geographic parameters are numerous, and therefore it is practically infeasible to simulate them in a probabilistic way. In practice, modelers use the best geographic data available, and treat these distributed parameters as deterministic. The unaccounted uncertainty due to geographic data quality, data resolution and processing algorithms is again assimilated in \( \delta \).

Consequently, \( \delta \) by our definition represents not only the imperfection of model structure, but also some of the imperfection of the inputs. This is not just an expedient solution. Rather, it is a solution that makes physical sense, because the imperfection of inputs is largely due to the incompatibility between available data and model requirements in terms of scale, resolution and conceptual consistency. The uncertainty analysis in the linking stage should directly deal with the three components in equation (1): \( \theta \) (including the multipliers but excluding the geographic parameters), \( \varepsilon \) and \( \delta \).

3. GLUE Approach for Watershed Water Quality Modeling

The GLUE approach is based on the concept of equifinality [Beven and Freer, 2001; Beven, 2006] of model structure and/or parameter sets in providing acceptable (i.e., behavioral) fits to observational data. Like other uncertainty analysis methods, GLUE deals with parameter uncertainty explicitly. Basically, \( \varepsilon \) and \( \delta \) are implicitly considered in GLUE. Although GLUE allows evaluation of multiple model structures, in practice at most a few structures can be considered concurrently, so only a small proportion of \( \delta \)
can be treated directly. For now we consider only one model structure in the analysis.

[24] In GLUE, a number of candidate parameter sets are first identified, usually through MC simulation. Next, Bayes theorem is applied to update the prior probability of each parameter set. All parameter sets meeting the predefined behavioral criterion (i.e., threshold) are then retained as behavioral parameter sets. Finally, posterior probability weighted simulations generated by the behavioral parameter sets can be used to calculate percentile on the basis of uncertainty limits at each time step, and dynamic uncertainty bounds can be consequently constructed. The use of Bayesian updating is the core part of GLUE approach. Per our notation, Bayes theorem can be expressed as

\[ p(y(\theta)|Z) = p(\theta|Z) = l(\theta|Z)p_0(\theta) \]  

where \( p(y(\theta)|Z) \) and \( p(\theta|Z) \) are posterior probability distributions, \( p_0(\theta) \) is prior probability distribution, and \( l(\theta|Z) \) represents the likelihood measure which in theory is the conditional probability distribution \( p(Z|\theta) \). Bolded \( y \) and \( Z \) are the time series representations of model simulation and observation (same for \( Y, \delta \) and \( \varepsilon \) in this paper), respectively. Determining \( l(\theta|Z) \) is the critical part of this theorem. GLUE allows the flexibility to use different forms of likelihood measures for different problems. In GLUE applications where the error terms are not explicitly modeled, \( l(\theta|Z) \) is in a form of goodness-of-fit measure which actually measures the degree (not probability) to which \( y(\theta) \) approximates observation \( Z \).

[25] The reasons for selecting GLUE for the uncertainty analysis were mentioned in section 1. The point to reinforce is that GLUE does not necessarily require explicit error assumptions and/or an optimization procedure, both of which are extremely difficult to consider for complex watershed models. However, there are several things to be aware of. First, GLUE’s uncertainty results are conditioned on some subjective choices (hereafter referred to as the “GLUE conditions”) made by the modeler including the likelihood measure, the number of runs for the MC simulation and the behavioral criteria. In the field of hydrologic modeling, some studies [Christensen, 2003; Montanari, 2005] examined how GLUE conditions may affect the uncertainty estimation. However, the use of subjective conditions in the context of watershed water quality management has not been studied. Second, GLUE’s uncertainty limits are percentiles of model simulations, not direct estimates of the probability of observing a particular observation. As a result, one should not expect that a specific proportion of observations will lie within the GLUE uncertainty bounds [Beven and Freer, 2001]. Nevertheless, in practice modelers do need certain criteria to check the adequacy of uncertainty results, if comparing observations against uncertainty bounds itself is not intuitive enough.

[26] Watershed water quality models add entire new hierarchical levels (i.e., simulation of sediment, temperature and chemicals) to consider, with regard to conventional hydrologic simulation. In water quality simulations, \( \delta \) and \( \varepsilon \) are in general much more substantial and complicated, and observational data for constraining the simulation is often very limited. In this special context, it is of great interest to examine whether a GLUE analysis using traditional goodness-of-fit measures as its fuzzy likelihood functions can handle those error terms effectively. In addition, incorporating management considerations in judging the adequacy of uncertainty results is an important concern in a management application.

4. Data and Methods

4.1. Watershed Model

[27] In management practice, it is rare that more than one complex watershed model is set up in parallel, because of limited resources. More commonly, a modeler would use one appropriate watershed model for the specific watershed problem unless the modeling results eventually invalidate the model, in which case either improving the model or using another model should be considered. In this study, the watershed analysis risk management framework (WARMF) was selected as an example of a complex watershed model. It simulates a variety of watershed processes, including catchment and stream hydrology, temperature fluctuations, sediment movement, plant growth, nutrient cycles, pesticide fate and transport, oxygen demand, etc. The typical timescale of simulation is a daily time step, although recently the model was implemented at an hourly time step. The submodels embedded are adapted from many well established algorithms. The computing engine of WARMF was taken from ILWAS [Gherini et al., 1985]. Algorithms for sediment erosion and pollutant transport from farm lands and other land uses were adapted from ANSWERS [Beasley et al., 1980] and USLE [Meyer and Wischmeier, 1969]. The pollutant accumulation and wash-off from urban areas was adapted from SWMM [Rossman, 2005]. Significant modifications were also made on sediment dynamics, point source simulation, lake simulation, incorporation of additional conventional and nonconventional contaminants, among other changes [Systech Engineering, 2001]. WARMF has been integrated into the USEPA’s BASINS framework as an additional tool for watershed analysis. A detailed description of the equations considered in WARMF is included in the technical guide (http://www.epa.gov/athens/wwqtec/html/warmf.html).

[28] The watershed delineation can be carried out using any delineation algorithm (e.g., within BASINS or ArcHydro) and is imported as a shapefile into WARMF. The watershed is divided into multiple catchments, each of which contains a river or stream segment, or a lake. Up to five soil layers can be specified for each catchment, with different thicknesses and hydrogeochemical properties. A catchment can contain several land uses, specified as a percent contribution, with the resulting loading values averaged on the basis of their land use weights. Parameters for surface processes are land use dependent, accounting for in-catchment heterogeneity, but soil parameters are uniform within each layer in a given catchment. Catchment size is a function of the delineation, which can be refined down to a fine grid if the analysis and data warrant it. Other model parameters include watershed-wide coefficients and river parameters.

4.2. Watershed and Pollutant

[29] Evaluating the watershed model is the most time-consuming part of an MC analysis. A watershed is usually divided into tens or hundreds of catchments to account for
Table 1. Estimated Monthly Application Rates of Diazinon (g/ha)

<table>
<thead>
<tr>
<th>Month</th>
<th>Residential Area (21.9%)</th>
<th>Cropland and Pasture (24.2%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>1.621</td>
<td>0.811</td>
</tr>
<tr>
<td>February</td>
<td>1.536</td>
<td>0.768</td>
</tr>
<tr>
<td>March</td>
<td>1.561</td>
<td>0.781</td>
</tr>
<tr>
<td>April</td>
<td>2.228</td>
<td>1.114</td>
</tr>
<tr>
<td>May</td>
<td>2.619</td>
<td>1.31</td>
</tr>
<tr>
<td>June</td>
<td>2.568</td>
<td>1.284</td>
</tr>
<tr>
<td>July</td>
<td>2.526</td>
<td>1.263</td>
</tr>
<tr>
<td>August</td>
<td>2.695</td>
<td>1.348</td>
</tr>
<tr>
<td>September</td>
<td>2.429</td>
<td>1.215</td>
</tr>
<tr>
<td>October</td>
<td>1.901</td>
<td>0.951</td>
</tr>
<tr>
<td>November</td>
<td>2.242</td>
<td>1.121</td>
</tr>
<tr>
<td>December</td>
<td>2.067</td>
<td>1.034</td>
</tr>
</tbody>
</table>

*Percentage of land use in the catchment.

spatial heterogeneity, which significantly increases the computational cost. Since this study is focused on a theoretical analysis, a single catchment extracted from a large watershed was used for conducting the numerical experiments. The advantages of using a single catchment were discussed by Zheng and Keller [2006]. The catchment was constructed using data (e.g., geographic information, hydrologic data, etc.) from the Santa Clara River Watershed [Zheng and Keller, 2006], a semiarid region typical of coastal southern California. More details about this watershed were presented by Keller et al. [2004]. [30] The output variable of concern is in-stream diazinon concentration at the outlet of the catchment. Diazinon is an organophosphate pesticide which has the potential for wide impact on human and ecological receptors. Its water solubility $C_{sol}^{w}$ (40–60 mg/L at 20°C) is significant, several orders of magnitude above the levels associated with toxicity, leading to significant mobility in runoff from irrigation or rainfall. It also has very low volatility (saturated vapor pressure $P_{sat}$ = $9.7 \times 10^{-3}$ Pa at 20°C) and medium partitioning coefficient (log octanol-water partitioning coefficient $\log K_{ow} = 3.3$), and thus adsorbs significantly to colloids, sediment and soil organic matter (Extension Toxicology Network, Pesticide information profiles, diazinon, 2005, Oregon State University, Corvallis, Oregon, http://extoxnet.orst.edu/pips/diazinon.htm). Diazinon was once widely used in southern California, but is now being partially phased out by USEPA. In this study, we only considered the scenario where diazinon pollution is mainly driven by rainfall runoff, not by irrigation. Two freshwater quality targets were considered: diazinon’s freshwater criterion maximum concentration (CMC) as a criterion for acute effects and the freshwater criterion continuous concentration (CCC) as a criterion for chronic effects. Per the California Department of Fish and Game [Siepmann and Findlayson, 2000], the CMC and CCC for diazinon are 80 ng/L and 50 ng/L, respectively.

4.3. Inputs, Parameters, and Synthetic Observations

[31] Model simulation was performed for a 3-year period from 1 October 1996 to 30 September 1998, using weather data from the National Climate Data Center station Piru 2 ESE (34°24’N, 118°45’W) located within the Santa Clara River Watershed. The application rates of diazinon on different land uses, ranging from 0.8 g/ha-month to 2.7 g/ha-month (see Table 1), were estimated from data for the Newport Bay area (Orange County, CA) [California Department of Pesticide Regulation, 1998; Harrison et al., 2005]. [32] WARMF has more than one hundred catchment, river reach and general parameters (excluding geographic parameters) involved in simulating diazinon fate and transport, but the number of parameters to include in the uncertainty analysis can be significantly reduced. Thirty-nine catchment, river reach or general parameters (see Table A1 in Appendix A) were taken into account in the uncertainty analysis. These parameters include the sensitive parameters for diazinon concentrations identified in a previous sensitivity analysis [Zheng and Keller, 2006], as well as some potentially important parameters based on the critical processes identified in the study [Zheng and Keller, 2006], modeling theory and our experience. Some parameters may not be important in this specific case study. Such selection is protective (i.e., we include more parameters than in the original sensitivity analysis). Among these parameters, there are two load adjustment coefficients: one is for the diazinon application rate in cropland and another in residential areas. These two parameters partially account for the uncertainty associated with pollutant loads. A uniform prior distribution was assumed for all uncertain parameters, which reflects a common situation where no good knowledge about parameter distributions is available to the modeler.

[33] For pesticides like diazinon, long-term high-frequency monitoring data are generally unavailable. Synthetic daily observations of diazinon concentration were therefore used in this study on the basis of a reduced numbers of observations in similar watersheds. The mean values of all the uncertain parameters (for certain parameters, the mean values were slightly modified to meet the constraints described in Table A1 in Appendix A) are first arbitrarily selected as the true parameter values $\theta^*$.

By running the watershed model with $\theta^*$, we generated $y(X, \theta^*)$ as a synthetic true model response. The temporal pattern of the synthetic $y(X, \theta^*)$ (see Figure 2a) reflects the assumption that the diazinon pollution is mainly driven by rainfall runoff, not by irrigation water. Our preliminary analysis also showed that a slight deviation from $\theta^*$ would not lead to an unreasonable $y(X, \theta^*)$. Next, data corruption was carried out by subtracting a synthetic error term $E$ from $y(X, \theta^*)$ to produce observation $Z$ (i.e., $Z = y(X, \theta^*) - E$). In this synthesis, we assumed that, on jth day, the ratio $E_i/y_i$, denoted as $w_j$, is uniformly distributed in an interval $[a_j, b_j]$ without autocorrelation. Therefore $Z_j$ can be generated as $y_j \cdot (1 - w_j)$. In the base flow period associated with low concentration, $E$ is small and unbiased with regard to $y(X, \theta^*)$. In this case, $w_j$ was assumed to be centered at 0 and have a small range. In the high-concentration period, generally coincident with rainfall events, $y(X, \theta^*)$ can be significantly higher than $Z$, and the discrepancy increases with the magnitude of concentration. In this case, $w_j$ was assumed to be centered at a positive number and have a relatively large range which increases with $y_j$. To account for time lag that may occur during storm events, we also assumed that there is a backward shift for simulated concentration peaks with respect to observed peaks. With all the above assumptions, we can generate synthetic observations randomly. It is worth emphasizing that the
process described above is only for generating the synthetic
data sets for our experiments. It is NOT a formal process for
modeling the errors, and is entirely independent of any
methods that could be used to explicitly model the errors.

Figure 2 shows the synthetic data sets. For better
illustration, the data sets were plotted only from simulation
days 420 to 600. Precipitation is shown in the second y
axis, for reference. Note that the data sets in Figure 2 represent
one possible outcome among many. Nevertheless, the data
sets are applicable for our experiments, since they represent
a typical situation consistent with our modeling experience
and knowledge of diazinon fate and transport, as well as
field observations [e.g., Santa Ana Regional Water Quality
Control Board, 2003]. Previous studies have shown that
WARMF can in general reproduce the critical timing of
diazinon concentration accurately. On the other hand, while
providing reasonable simulation results for low-flow peri-
ods, the model is likely to notably overestimate the con-
centration during rainfall events.

4.4. Description of the Experiments

[35] In this study, four experiments were carried out: (1) five likelihood measures (Table 2), denoted as ME, IRV(1), IRV(0.5), IMAE and ILRV, were compared; (2) four different numbers of MC runs (R = 4000, 6000, 8000 and 10,000) were tested; (3) five different behavioral thresholds (k = 10%, 20%, 40%, 60% and 80%) were examined, where
k = 20% means that when all the sampled models are sorted in
descending likelihood values, the models in the bottom 20%
are defined as nonbehavioral ones; and (4) the performance
of GLUE in data-scarce situations was investigated. Two
arbitrary semimonthly sampling plans (i.e., 72 samples in the
3-year period) were considered: water samples collected on
4th and 19th of a month in the first plan, or on 7th and 22nd in
the second plan. Arbitrary sampling programs are common
for regulatory processes. We used R = 6000 in experiments 1,
3 and 4, k = 20% in experiments 1, 2 and 4, and the most
appropriate likelihood measure identified in experiment 1 for
experiments 2, 3 and 4. For higher efficiency of MC simu-
lation, Latin Hypercube Sampling (LHS) [Iman and Helton,
1985] was adopted.

[36] In this study, 95% and 5% uncertainty limits (or bounds) were used to define the 90% uncertainty interval
(or, band). The percentage of observations above the upper
bound (P_u) and the percentage of observations below the
lower bound (P_l) were examined in each GLUE analysis. As
one should not expect that a specific proportion of observa-
tions will lie within the GLUE uncertainty bounds [Beven
and Freer, 2001], an additional criterion is needed to
evaluate uncertainty results achieved using GLUE. Here
we introduced management as an additional dimension in
the uncertainty analysis, since supporting management
decisions is the ultimate goal of using complex watershed
models. The management variable nonattainment frequency

Table 2. Likelihood Measures Compared

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Formula^b</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>Model efficiency or coefficient of determination</td>
<td>1 - \frac{\sigma_r^2}{\sigma_o^2}, \sigma_r^2 &lt; \sigma_o^2</td>
</tr>
<tr>
<td>IRV(N)</td>
<td>Inverse of residual variance to power of N (N = 1 or 0.5)</td>
<td>(\sigma_r^2)^{-N}</td>
</tr>
<tr>
<td>IMAE</td>
<td>Inversed mean absolute error</td>
<td>\frac{1}{n} \sum_{i=1}^{n}</td>
</tr>
<tr>
<td>ILRV</td>
<td>Inverse of log-transformed residual variance</td>
<td>(\log(\sigma_r^2))^{-1}</td>
</tr>
</tbody>
</table>

^bME and IMAE were derived from the objective functions commonly used in hydrologic calibration [Gupta et al., 1998]; IRV(N) was frequently used in GLUE applications [e.g., Beven and Binley, 1992]; and ILRV is a variant of IRV(N) that we proposed for this study.

Notations: r_i, residual between model simulation and observation at time i; \sigma_r^2, variance of residuals; \sigma_o^2, variance of observations; n, number of time steps; N, shaping factor.
(NAF) was defined. NAF is the frequency (or percentage time) with which the concentration time series exceeds a given target. A management variable is a metric that is directly linked with a specific management concern(s). NAF can be evaluated for observations and model simulations, as well as for GLUE uncertainty bounds. Eventually, our criterion for evaluating the uncertainty results is whether the uncertainty limits estimated for NAF make sense for the management decision.

The experiments were not designed to identify the ideal GLUE conditions, but to reveal the influence of GLUE conditions on uncertainty results. The likelihood measures tested are all goodness-of-fit measures, since in watershed-scale water quality modeling, information about the errors (especially \( \delta \)) is rarely enough for developing a statistically rigorous likelihood measure. In addition, GLUE’s behavioral criteria may take forms other than the percentage threshold (i.e., \( k\% \)) discussed in this paper. Page et al. [2003] based the behavioral separation on the error between \( y(\theta_i) \) and \( Z \). If the error is within a predefined range of acceptability, \( \theta_i \) is identified as a behavioral parameter set. Nevertheless, we considered the simple criterion of \( k\% \), just because (1) our main purpose is to illustrate the influence of behavioral criteria on uncertainty results, and (2) the “error approach” is mathematically connected with the “percentage approach”; an error range could be eventually projected into a percentage threshold.

5. Results and Discussion

The values of \( P_u \) and \( P_l \) in different experiments are graphically presented in Figure 3. As Figure 3a displays, the performances of the five likelihood measures are considerably different, and in this case ILRV results in \( P_u \) and \( P_l \) closest to 5%. Figure 4 presents an example of a GLUE uncertainty band confined by 95% and 5% uncertainty.
bounds. The solid circles are the observations outside of the uncertainty band, and the open circles are the observations inside of the band. As shown by the width of the uncertainty band, the simulation uncertainty is most substantial at high concentrations (note that the y axis is in logarithmic scale for better illustration), which is consistent with the a priori knowledge. Uncertainty bands generated with other GLUE conditions exhibit a similar pattern, but differ in band width. It was also found that the difference in band width results mainly from a shift up or down of the 95% uncertainty bound. Observations lying outside of the bands could be due to their extraordinary magnitude and/or inconsistent timing. Further investigation revealed that inconsistent timing is the major cause of the outliers in this case (see the solid circles in Figure 4). The findings imply that, with the selected likelihood measures, δ and ε (or, the grand error E) could be adequately handled by the likelihood measure in magnitude, but not in timing.

[39] Figure 3b shows that, with the current model settings, the number of MC runs (R) has no significant impact on GLUE results for R ≥ 4000. R = 6000 was used in experiments 1, 3 and 4 since this number insures a good balance between numerical efficiency and stability of GLUE results. A 6000-run MC simulation for this case study takes around 20 hours on a high-performance personal computer. The computing time needed for the post analysis of the MC simulations is almost negligible.

[40] The effect of the behavioral threshold k is demonstrated in Figure 3c. k = 20% (i.e., 4800 behavioral simulations retained) led to P_u and P_l closest to 5%. Considerably larger P_u and slightly smaller P_l were observed for k larger than 20%. On the other hand, P_u calculated with k = 10% is extremely small which indicates that the 95% bound is overprotective for the observations. Figure 5 compares the mean temporal pattern of three groups of model simulations: (1) 600 nonbehavioral simulations with k = 10%; (2) 600 simulations that are behavioral with k = 10% but nonbehavioral with k = 20%; and (3) 4800 behavioral simulations at k = 20%. It appears that k = 10% is not large enough to exclude some ‘bad’ simulations which significantly overestimate the concentrations, especially for dry periods. Including such simulations in constructing uncertainty bounds inevitably leads to unreasonable results. These “bad” simulations could be due to unrealistic parameter value combinations. A separate study would be necessary to identify the reasons for these behaviors. The main point here is that the behavioral threshold could significantly affect uncertainty estimation, and should be very carefully determined.

[41] Figure 3d compares the different sampling plans. Both semimonthly plans resulted in overprotective 95% uncertainty bounds for similar reasons as in the k = 10% case. The results indicate that the semimonthly observations in the assumed plans are not frequent enough to adequately constrain model simulations so that all the influential ‘bad’ simulations can be effectively eliminated.

[42] More importantly, we evaluated the uncertainty results from a management point of view. Figure 6 displays the estimated uncertainty limits for the management variable NAF for both criterion maximum concentration (CMC) (first column) and criterion continuous concentration (CCC) (second column) in different numerical experiments. The observed NAF based on Z is 18.9% for the CMC and 22.7% for the CCC, for the case study (dashed lines). For better comparison, the y axes in Figures 6c and 6d are shown for up to 40%, and the numbers (with arrows) in Figures 6c and 6d indicate the upper limits for NAF with k = 10%. Figure 6 reinforces the relevance of the GLUE conditions discussed in section 3, and displays the impact that subjective choices in the GLUE approach may have on decision making. In some cases, such as when ME is used as the likelihood measure (Figures 6a and 6d), the 90% uncertainty intervals fail to bracket the observed NAF value. In other cases like k = 10% (Figures 6c and 6d) or semimonthly sampling (Figures 6e and 6f), The uncertainty intervals are too wide to be useful for decision making. Additionally, the uncertainty intervals are considerably biased toward the lower end in most of these cases. All these inadequacies could result in highly indefensible management decisions. For instance, overestimating NAF may lead to an overprotective pollutant load reduction in pursuit of attainment of the water quality objectives.

[43] Figure 6 also reveals the importance of incorporating management considerations in the uncertainty analysis. In GLUE, the uncertainty bounds for simulations using the watershed model (e.g., the 95% and 5% bounds in Figure 4) are not conditioned on the management concerns. However, the uncertainty with respect to management variables such as NAF does depend on specific management concerns (e.g., CMC or CCC as the water quality target). First of all, the differences between various likelihood measures with regards to the NAF interval, especially for the 95%
uncertainty limit, is more significant for CCC than for CMC (e.g., Figure 6a versus Figure 6b). This is because the CCC (50 ng/L) is closer to the lower concentrations of the 95% uncertainty bound (see Figure 4 as an example), and therefore the upper limit of NAF for CCC is more sensitive to shifts of the upper uncertainty bound. Second, the uncertainty limits for NAF are more significantly biased for CCC than for CMC. Third, ILRV, the most appropriate likelihood measure for CMC, resulted in a much wider uncertainty interval for CCC (Figures 6a and 6b) than for CMC. The possible causes of the differences between CMC and CCC are as follows. (1) The MC simulation did not generate enough simulations that properly characterize the dynamics of low-concentration periods, especially for concentrations around the CCC. This problem may stem from inappropriate modeling settings (e.g., parameter ranges, boundary conditions, etc.), inadequate model structure, or the combination of these two. (2) All of the selected likelihood functions are goodness-of-fit metrics, which only measure the overall fit between simulations and observations. No specific considerations are given to the concentrations around the CCC which are of major interest for management. Thus the calculated likelihood values may not correctly reflect the usefulness of the simulations in characterizing exceedance for CCC. The first problem could be solved by changing model settings or improving model structure (if feasible), but the second problem is rooted in the use of goodness-of-fit likelihood definition for the GLUE analysis, therefore may require a modification in watershed water quality modeling.

[44] The minimal $R$ for a GLUE analysis is dependent both on management concerns and the level of confidence. We also tested $R = 200, 1000$ and $2000$ with ILRV and $k = 20\%$. For CMC, even $R = 200$ led to similar uncertainty limits for NAF as those calculated using $R \geq 4000$ (the difference is less than 1% in absolute value). However, for CCC, even $R = 1000$ introduced significant differences to the 5% limit for the NAF (increased from around 63% to 71%). Hence in management practices, a higher $R$ would be recommended when using GLUE or similar methods.

[45] On the other hand, we would argue that $R = 10000$ is already sufficient for this level of complexity. Our major reasoning is as follows: watershed water quality simulation involves much more significant $d$ and $e$. In terms of traditional goodness of fit, there could be relatively few high-likelihood simulations. For instance, with ILRV, the standardized likelihood values for the 6000 simulations only range from $1.25 \times 10^{-4}$ to $2.04 \times 10^{-4}$ (the prior probability is $1/6000 = 1.67 \times 10^{-4}$). Also, given a large $R$ and the uniform prior distribution (which is generally assumed), the prior probability of all simulations is small ($1.00 \times 10^{-4}$ in the case of $R = 10000$). Thus the few high-likelihood simulations would not make significant difference to uncertainty results. As an experiment, we randomly selected 20 behavioral simulations in the case of $R = 6000$, and arbitrarily increased their likelihood values by a factor...
of 10 (likelihood values of other behavioral simulations changed accordingly). This modification did not have significant effect on the uncertainty results such as the GLUE uncertainty bounds and the uncertainty limits for NAF. Nevertheless, if modeling is more complicated, more model simulations may be necessary.

6. Conclusions

This study developed a conceptual framework which characterizes all uncertainty sources and their interactions in management-oriented watershed modeling. The GLUE approach was selected for estimating uncertainty in linking watershed pollutant loads with observed water quality, using WARMF as an example of a complex watershed model. The results show that applying GLUE to complex watershed models can be computationally feasible. For the catchment used in this study, even 2000 model runs were sufficient to provide stable uncertainty estimates. In real applications with more complicated watersheds, computing time can be reduced by considering a smaller number of uncertain parameters (with a more case-specific sensitivity analysis), refining parameter ranges, and/or using a workstation instead of a personal computer. One could also use distributed computing facilities based on grid technology to improve the computing efficiency.

GLUE deals with structure error $\delta$ and observational error $\varepsilon$ implicitly. Whether these error terms can be appropriately characterized depends on the type of likelihood measures the modeler uses in a GLUE analysis.

---

Table A1. Uncertain Model Parameters

<table>
<thead>
<tr>
<th>ID</th>
<th>Units</th>
<th>Parameter Description</th>
<th>Range and Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>PWF</td>
<td></td>
<td>Precipitation weighting factor (adjusting coefficient for local variations in precipitation)</td>
<td>0.8 – 1.2</td>
</tr>
<tr>
<td>TL</td>
<td>deg C</td>
<td>Temperature lapse (amount of temperature to subtract due to regional variation in temperature)</td>
<td>–5 – –5</td>
</tr>
<tr>
<td>IMP</td>
<td></td>
<td>Fraction of impervious area for a land use</td>
<td>Residential (0.2 – 0.7); Commercial and industrial (0.3 – 0.8)</td>
</tr>
<tr>
<td>SSF</td>
<td>days</td>
<td>Frequency of street sweeping (time period between street sweeping passes)</td>
<td>1 – 60</td>
</tr>
<tr>
<td>SSE</td>
<td>%</td>
<td>Street sweeping efficiency (portion of deposited material removed in one street sweeping pass)</td>
<td>60 – 100</td>
</tr>
<tr>
<td>KC</td>
<td>1/d</td>
<td>Rate of diazinon decay within catchment</td>
<td>0.1 – 0.2</td>
</tr>
<tr>
<td>THCKj</td>
<td>cm</td>
<td>Thickness of $j$th soil layer ($j = 1, 4 and 5$)</td>
<td>$j = 1 (4-27), j = 4 (12-19); j = 5 (59-727)$</td>
</tr>
<tr>
<td>SMj</td>
<td></td>
<td>Saturation moisture (maximum volume fraction of water) in $j$th soil layer ($j = 1, 4 and 5$)</td>
<td>0.2 – 0.6</td>
</tr>
<tr>
<td>IMj</td>
<td></td>
<td>Initial moisture (initial volume fraction of water) in $j$th soil layer ($j = 1, 4 and 5$)</td>
<td>IMj is a fraction $\beta$ of SMj; $\beta$ is assumed to uniformly distribute between 0.1 and 0.9</td>
</tr>
<tr>
<td>FLDj</td>
<td>cm/d</td>
<td>Field capacity in $j$th soil layer ($j = 1, 4 and 5$)</td>
<td>FLDj is a fraction $\gamma$ of SMj; $\gamma$ is assumed to uniformly distribute between 0.1 and 0.9</td>
</tr>
<tr>
<td>HHCj</td>
<td>cm/d</td>
<td>Horizontal hydraulic conductivity (in saturation) of $j$th soil layer ($j = 1, 4 and 5$)</td>
<td>1 – 1000</td>
</tr>
<tr>
<td>ADSj</td>
<td>L/kg</td>
<td>Adsorption isotherm for diazinon in $j$th soil layer</td>
<td>5 – 15</td>
</tr>
<tr>
<td>IDR</td>
<td>m</td>
<td>Initial depth of stream water</td>
<td>0.1 – 0.3</td>
</tr>
<tr>
<td>NR</td>
<td></td>
<td>Surface roughness (Manning’s $n$) of river channel</td>
<td>0.02 – 0.045</td>
</tr>
<tr>
<td>DVM</td>
<td></td>
<td>Detachment velocity multiplier (ratio between flow velocity and bottom sediment detachment)</td>
<td>$0.5 \times 10^{-6} – 1.5 \times 10^{-6}$</td>
</tr>
<tr>
<td>DVE</td>
<td></td>
<td>Detachment velocity exponent (exponent of velocity in determining bottom sediment detachment)</td>
<td>1 – 2</td>
</tr>
<tr>
<td>IDS</td>
<td>m</td>
<td>Initial depth of bed sediment</td>
<td>0 – 2</td>
</tr>
<tr>
<td>VF</td>
<td></td>
<td>Vegetation factor (empirical coefficient proportional to amount of sediment release due to lack of vegetation)</td>
<td>$0 – 1 \times 10^{-4}$</td>
</tr>
<tr>
<td>BDD</td>
<td>m$^3$/d</td>
<td>Bed diffusion rate</td>
<td>$0 – 2 \times 10^{-4}$</td>
</tr>
<tr>
<td>ICPW</td>
<td>mg/L</td>
<td>Initial concentration of diazinon in stream water</td>
<td>$0 – 5 \times 10^{-4}$</td>
</tr>
<tr>
<td>ADW</td>
<td>L/kg</td>
<td>Adsorption isotherm for diazinon in stream water</td>
<td>5 – 20</td>
</tr>
<tr>
<td>ICPB</td>
<td>mg/g</td>
<td>Initial concentration of diazinon in river bed sediment</td>
<td>$0 – 5 \times 10^{-4}$</td>
</tr>
<tr>
<td>ADB</td>
<td>L/kg</td>
<td>Adsorption Isotherm for diazinon in river bed</td>
<td>5 – 20</td>
</tr>
<tr>
<td>KW</td>
<td>l/day</td>
<td>Rate of diazinon decay in stream water</td>
<td>0.1 – 0.18</td>
</tr>
<tr>
<td>KB</td>
<td>l/day</td>
<td>Rate of diazinon decay in river bed sediment</td>
<td>0.01 – 0.03</td>
</tr>
<tr>
<td>LAC</td>
<td>l/day</td>
<td>Load adjustment coefficients for a land use</td>
<td>Residential (0.8 – 1.2); Cropland (0.8 – 1.2);</td>
</tr>
<tr>
<td>Total number of uncertain parameters</td>
<td></td>
<td></td>
<td>39</td>
</tr>
</tbody>
</table>
complicated watershed water quality modeling, goodness-of-fit measures are the most convenient choices. However, to ensure reasonable uncertainty results, the modeler must make those subjective choices (especially the likelihood measure) carefully, which requires appropriate evaluation criteria (in our case, management adequacy) and significant trial-and-error work. For the case study, ILRV proved to be the most appropriate likelihood measure among the five candidates, and \( k = 20\% \) appeared to be the best choice for behavioral separation. Availability of observational data also has significant influence on the uncertainty results. As demonstrated, the hypothetical semimonthly sampling plans led to unreasonable uncertainty bounds or limits for the dynamic simulation and the nonattainment frequency. The problem may result from the goodness-of-fit likelihood measures used in the specific GLUE analysis.

The results also indicated the relevance of incorporating management concerns into the uncertainty analysis. As shown, the GLUE analysis generated better results for the water quality target CMC than for the target CCC. We conclude that in order to achieve good uncertainty results for specific management concerns, adequate model structure, appropriate model settings and suitable likelihood measure are all necessary.

Furthermore, from the perspective of planning, it is true watershed response in future (\( Y_f \)), not future observation, that is relevant. Because of the existence of \( \varepsilon \) and its potential variability in time, behavioral simulations with respect to observation \( Z \) in the linking stage are not necessarily behavioral in the planning stage with respect to \( Y_f \). This issue is yet to be addressed within GLUE’s framework. As Beven [2006, p. 27] stated, “in forming prediction limits in this way, there is an implicit assumption (as in previous applications of GLUE) that the errors in prediction will be ‘similar’ (in all their complexity) to those in the evaluation period."

Overall, this paper provides a detailed discussion on the application of GLUE to complex watershed models within the management context. The results suggest future efforts which could be made to better suit a GLUE-based uncertainty analysis to watershed water quality modeling, including (1) defining likelihood functions in a more rigorous way instead of using goodness-of-fit measures, (2) directly factoring management concern(s) into the likelihood measure, (3) developing more specific strategies for behavioral separation, (4) treating the error terms, especially \( \varepsilon \), in a more explicit way, and (5) targeting the true watershed response \( Y \) instead of observation \( Z \). In the companion paper, based on GLUE, we developed MOCAU which reflects these suggested efforts.

Appendix A

Table A1 presents the uncertain parameters analyzed in this study. To be general, fairly broad ranges were set for these parameters. For many of the parameters, WARMF’s default values were used as the mean value, and a percentage deviation was assigned to each of these parameters, usually larger than ±25%. Other resources [Beasley and Huggins, 1981; Rose et al., 1991; Noordwijk et al., 2004] were also consulted to set the ranges. The physicochemical properties of Diazinon were obtained from the U.S. Department of Agriculture’s Agricultural Research Service (http://www.ars.usda.gov) database.

Acknowledgments. The authors acknowledge funding from the UC Tenure Substances Research and Teaching Program, the California State Water Resources Control Board under agreement 01-161-140, as well as partial support from the U.S. Geological Survey USDI Program under agreement 00HQGR0089. The opinions expressed in this paper are those of the authors and do not necessarily reflect those of the funding agencies.

References


California Department of Pesticide Regulation (1998), Annual pesticide use data, user documentation, Sacramento.


Page, T., K. J. Beven, J. Freer, and A. Jenkins (2003), Investigating the uncertainty in predicting responses to atmospheric deposition using the model of acidification of groundwater in catchments (MAGIC) within a generalised likelihood uncertainty estimation (GLUE) framework, Water Air Soil Pollut., 142, 71–94.


Santa Ana Regional Water Quality Control Board (2003), Diazinon and chlorpyrifos TMDL: Upper Newport Bay and San Diego Creek, report, Riverside, Calif.


———. Keller and Y. Zheng, Bren School of Environmental Science and Management, University of California, 3420 Bressi Hall, San Diego, CA 92037-5131, USA.