Jul 1st, 12:00 AM

Dealing with unidentifiable sources of uncertainty within environmental models

Ann Van Griensven
T. Meixner

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

https://scholarsarchive.byu.edu/iemssconference/2004/all/42

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Dealing with unidentifiable sources of uncertainty within environmental models

A. van Griensven\textsuperscript{a,b} and T. Meixner\textsuperscript{b}

\textsuperscript{a}BIOMATH, Ghent University, Belgium (ann.vangriensven@biomath.ugent.be)
\textsuperscript{b}University of California Riverside, California, USA (tmeixner@mail.ucr.edu)

Abstract: Sources of Uncertainty Global Assessment using Split Samples (SUNGLASSES) is a method for assessing model global uncertainty to aid in the development of integrated models. The method is complementary to the commonly investigated input and parameter uncertainty, as it accounts for errors that may arise due to unknown or unassessable sources of uncertainty, such as model hypothesis errors, simplifications, scaling effects or the lack of the observation period to represent long-term variability and fluctuations in the system. Such sources are typically dominant for most environmental models and they undermine the reliability of environmental models.

The SUNGLASSES algorithm directly estimates the overall predictive uncertainty without identifying or quantifying the underlying sources of uncertainties. The method uses the split sample approach to generate an estimate of model output uncertainty by selecting a threshold below which model simulations are determined to be acceptable. Where this methodology differs from other methods that use a threshold, is that the threshold is determined by evaluating the confidence bounds on model outputs during an evaluation time period of data that was not used to initially calibrate the model and generate parameter estimates. Where parameter uncertainty is often assessed using some goodness-of-fit criterion such as the mean squared errors, SUNGLASSES focuses on a criterion that evaluates the correctness of the model output values to be used directly in decision making, such as total mass balance assessments or violations of standards as imposed by legislation. The described method is applied to the integrated water quality modelling tool, SWAT2003, applied to Honey Creek, a tributary of the Sandusky catchment in Ohio. Water flow and sediment loads are analysed. The incorporation of the split sample approach in the methodology produces a reasonable error bound that captures most of the observations during both the initial calibration period and during the evaluation period.

Keywords: Uncertainty; Catchment; Water quality; Modelling

1. INTRODUCTION

Applications of environmental models are important tools for decision making. But, the model results can be highly uncertain and may therefore adversely impact decisions. Therefore, it is important to know the reliability of model results. In this context, it is important to distinguish between confidence intervals for model results and reliability. Confidence intervals are the results of an uncertainty analysis, while reliability depends on the completeness of the uncertainty analysis that should ideally cover all sources of uncertainty, i.e., physical input uncertainty, parameter input uncertainty, model structure and code hypothesis uncertainty. Therefore, it is important to assign reliability levels on the model predictions of interest (i.e. the model outputs being used for decision making) even when the sources of uncertainty are not completely known or understood.

2. METHODS

2.1 Introduction

Two methods for assessing uncertainty are presented: ParaSol and SUNGLASSES. ParaSol is an optimisation and statistical method for the assessment of parameter uncertainty. In the literature, many methods for parameter uncertainty exist. Compared to these methods, ParaSol can be classified as being global, efficient and being able to deal with multiple objectives. These requirements for an uncertainty method are typical for environmental models of many types.

On top of ParaSol, SUNGLASSES uses all parameter sets and simulations that were generated by ParaSol. SUNGLASSES aims at detecting
additional sources of uncertainty by using an evaluation period in addition to the calibration period.

2.2 ParaSol

ParaSol – Parameter Solutions - (van Griensven and Meixner, 2004a) operates by a parameter search method for model parameter optimisation – a modified version of the SCE-UA method (Duan et al., 1992) – followed by a statistical method that uses the model runs that were performed during the optimisation to provide parameter uncertainty bounds and the corresponding uncertainty bounds on the model outputs.

The Shuffled complex evolution algorithm

This algorithm conducts a global minimisation of a single function for up to 16 parameters (Duan et al., 1992). In a first step (zero-loop), SCE-UA selects an initial ‘population’ by random sampling throughout the feasible parameters space for p parameters to be optimised (delineated by given parameter ranges). The population is portioned into several “complexes” that consist of 2p+1 points. Each complex evolves independently using the simplex algorithm. The complexes are periodically shuffled to form new complexes in order to share the gained information. SCE-UA has been widely used in watershed model calibration and other areas of hydrology such as soil erosion, subsurface hydrology, remote sensing and land surface modelling (Duan, 2003). It has been found to be robust, effective and efficient (Duan, 2003).

Objective functions

Sum of the squares of the residuals (SSQ): This objective function is similar to the Mean Square Error function (MSE) and aims at estimating the matching of a simulated series to a measured time series.

\[
SSQ = \sum_{n=1}^{N} [TF(x_{n, \text{measured}}) - TF(x_{n, \text{simulated}})]^2
\]  

(1)

with n the number of pairs of measured (\(x_{\text{measured}}\)) and simulated (\(x_{\text{simulated}}\)) variables and TF a user defined transformation function.

The sum of the squares of the difference of the measured and simulated values after ranking (SSQR): The SSQR method aims at the fitting of the frequency distributions of the observed and the simulated series. As opposed to the SSQ method, the time of occurrence of a given value of the variable is not accounted for in the SSQR method (van Griensven and Bauwens, 2003).

After independent ranking of the measured and the simulated values, new pairs are formed and the SSQR is calculated as

\[
SSQR = \sum_{r=1}^{N} [x_{r, \text{rank, measured}} - x_{r, \text{rank, simulated}}]^2
\]  

(2)

where \(r\) represents the rank.

Multi-objective optimisation

Several SSQ’s or SSQR’s can be combined to a Global Optimisation Criterion (GOC) using (van Griensven and Meixner, 2004):

\[
GOC = \frac{M}{m=1} \frac{SSQ_m \ast N_m}{SSQ_{m, \text{min}}}
\]  

(3)

The probability of a given parameter solution being the best one is related to the GOC according to (van Griensven and Meixner, 2004):

\[
p(\theta | Y_{\text{obs}}) \propto \exp[-GOC]
\]  

(4)

Thus the sum of the squares of the residuals get weights that are equal to the number of observations divided by the minimum. This equation allows also for uncertainty analysis as described below.

Parameter change options

In this optimisation and uncertainty algorithm parameters affecting hydrology or pollution can be changed either in a lumped way (over the entire catchment), or in a distributed way (for selected subbasins or HRU’s). They can be modified by replacement, by addition of an absolute change or by a multiplication of a relative change. A parameter is never allowed to go beyond predefined parameter ranges. A relative change allows for a lumped calibration of distributed parameters while they keep their relative physical meaning (soil conductivity of sand will be higher than soil conductivity of clay). This last method of relative change is the method utilised here.

Uncertainty analysis method

The uncertainty analysis divides the simulations that have been performed by the SCE-UA optimisation into ‘good’ simulations and ‘not good’ simulations. The simulations gathered by SCE-UA are very valuable as the algorithm samples over the entire parameter space with a focus of solutions near the optimum/optima. There are two separation techniques, both are based on a threshold value for the objective function (or global optimisation criterion) to select the ‘good’ simulations by considering all the simulations that give an objective function below this threshold.
The threshold value can be defined by $\chi^2$-statistics where the selected simulations correspond to the confidence region (CR) or Bayesian statistics that are able to identify the high probability density region (HPD) for the parameters or the model outputs (figure 1).

$\chi^2$-method

For a single objective calibration for the SSQ, the SCE-UA will find a parameter set $\Theta^*$ consisting of the $p$ free parameters ($\theta^*_1, \theta^*_2, \ldots, \theta^*_p$), that corresponds to $SSQ_{\text{min}}$, the minimum of the sum the square SSQ. According to $\chi^2$ statistics, we can define a threshold “c” for “good” parameter sets using equation:

$$c = SSQ_{\text{min}}^* (1 + \frac{\chi^2_{p,0.95}}{n - p}) \quad (5)$$

whereby $n$ is the number of observations and $p$ the number of free parameters. The $\chi^2_{p,0.95}$ gets a higher value for more free parameters $p$.

For multi-objective calibration, the selections are made using the GOC of equation (3). A threshold for the GOC is the calculated by:

$$c = GOC_{\text{min}}^* (1 + \frac{\chi^2_{p,0.95}}{NTOT - p}) \quad (6)$$

with NTOT the total number of observations for all the objective functions considered in the GOC.

All parameter sets that give simulations with a GOC below the value “c” will be selected as “good” parameter sets.

This option is described briefly since it is not chosen for the case study discussed in this paper.

In accordance to the Bayesian theorem, the probability $p(\theta|Y_{\text{obs}})$ of a parameter set $\theta$ is proportional to the GOC (equation 4) upon the assumption that the initial parameter distribution is equal to the uniform distribution. After normalizing the probabilities (to ensure that the integral over the entire parameter space is equal to 1) a cumulative distributions can be made and hence a 95% confidence regions can be defined. As the parameters sets were not sampled randomly but were more densely sampled near the optimum during SCE-UA optimisation, it is necessary to avoid having the densely sampled regions dominate the results. This problem is prevented by determining a weight for each parameter set $\theta_i$ by the following calculations (van Griensven and Meixner, 2004).

2.3 SUNGLASSES

To develop a stronger evaluation of the model prediction power, the Sources of Uncertainty Global Assessment using Split SampleS (SUNGLASSES) was designed to assess predictive uncertainty that is not captured by the parameter uncertainty estimated by Parasol. The SUNGLASSES method accounts for strong increases in errors when simulations are done outside the calibration period by using a split-sample strategy whereby the validation period is used to set uncertainty ranges.

These uncertainty ranges depend on the GOC, used during a calibration period representing the objective functions, and an evaluation criterion (to be used in decision making) used during an evaluation period. The GOC is used to assess the degree of error in the process dynamics, while the evaluation criterion defines a threshold on the GOC. This threshold should be as small as possible, but the uncertainty ranges on the criteria should include the “true” value for both the calibration and the validation period, e.g. when mass balance is used as criteria, these “true” values are a model bias equal to zero. Thus, the threshold is increased till the uncertainty ranges on the mass balance bias includes zero. SUNGLASSES operates by ranking the GOCs (Figure 2). Statistical methods can be used to define a threshold considering parameter uncertainty. In this case, Parasol was used to define such a threshold. However, when we look at the predictions, it is possible that unbiased simulations are not within the Parasol uncertainty range, which means that there are some more unknown uncertainties acting on the model outputs (Figure 3). Thus, a new, higher threshold is needed in order to have

![Figure 1: Confidence region for the $\chi^2$-statistics and the Bayesian statistics for the 2 parameters Smax and k of a simple 2-parameter model.](image)

*Bayesian method (Box and Tiao, 1974)*
unbiased simulations included in the uncertainty bounds (figure 2 and 3).

3. CASE STUDY

The methods are applied to a river basin model of the Honey creek, a tributary of the Sandusky river, Ohio (Figure 4) using the modelling tool “SWAT”.

3.1 SWAT

The Soil and Water Assessment Tool (SWAT) [Arnold et al., 1998] is a semi-distributed and semi-conceptual program that calculates water, nutrient and pesticide transport at the catchment scale on a daily time step. It represents hydrology by interception, evapo-transpiration, surface runoff (SCS curve number method [USDA Soil conservation Service, 1972]), soil percolation, lateral flow and groundwater flow and river routing (variable storage coefficient method [Williams, 1969]) processes. Other processes include nutrient, erosion, crop and pesticide, in-stream water quality processes. The catchment is divided into sub-basins, river reaches and Hydrological Response Units (HRU’s). While the sub-basins can be delineated and located spatially, the further sub-division into HRU’s is performed in a statistical way by considering a certain percentage of sub-basin area, without any specified location in the sub-basin. The methods ParaSol and SUNGLASSES are programmed by the authors within the SWAT2003 version.

Figure 2: Selection of good parameter sets using a threshold imposed by ParaSol or by SUNGLASSES

Figure 3: Confidence regions for the sediment loads calculations according to ParaSol and SUNGLASSES

Figure 4: Location of the Honey creek within the Sandusky basin.

3.2 Model description

A simple SWAT2003 model for Honey creek that covers 338 km$^2$ and consists of 1 subbasin (5 HRU’s), 1 river reach and 1 point source near the mouth of the creek (van Griensven et al., 2004). Daily data for water flow and sediment concentrations were used for calibration and evaluation of the model. Table 1 lists the 10 most important parameters for water flow and sediments concentrations, according to the results of a sensitivity analysis (van Griensven et al., 2004).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMFX</td>
<td>Maximum melt rate for snow during (mm/°C/day)</td>
</tr>
<tr>
<td>ALPHA_BF</td>
<td>Baseflow alpha factor (days)</td>
</tr>
<tr>
<td>ch_k2</td>
<td>Channel conductivity (mm/hr)</td>
</tr>
<tr>
<td>USLE-P</td>
<td>USLE equation support practice (P) factor.</td>
</tr>
<tr>
<td>CN2</td>
<td>SCS runoff curve number for moisture condition II.</td>
</tr>
<tr>
<td>sol_awc</td>
<td>Available water capacity of the soil layer (mm/mm soil).</td>
</tr>
<tr>
<td>surlag</td>
<td>Surface runoff lag coefficient</td>
</tr>
<tr>
<td>SFTMP</td>
<td>Snowfall temperature (°C)</td>
</tr>
<tr>
<td>SMTMP</td>
<td>Snow melt base temperature (°C)</td>
</tr>
<tr>
<td>Sol_z</td>
<td>Soil depth</td>
</tr>
</tbody>
</table>
3.3 Objective functions

SWAT was applied to the Honey Creek catchment to estimate sediment export from the catchment. Therefore, the joint calibration included the SSQ and SSQR for streamflow and SSQR for sediment loads, with a Box-Cox transformation to reduce the heteroscedastic nature of the residuals. The results allow an investigation of the joint uncertainty when both flow and water quality variables are used for model calibration as should be common practice for water quality models.

3.4 Evaluation criterion

Based on the assumption that the model purpose was to assess global fluxes of sediments load at the outlet of the creek, the evaluation criteria was described by the model biases on the mass flux that were calculated as:

\[
BIAS = \left( \frac{\sum_{n=1}^{N} SIM_n - \sum_{n=1}^{N} OBS_n}{\sum_{n=1}^{N} OBS_n} \right) \times 100. \tag{8}
\]

for \( N \) the number of pairs (simulation, observation), \( SIM_n \) the simulation at day \( n \) and \( OBS_n \) the observation of day \( n \). The bias was calculated for the water flow and the sediment loads in the calibration and validation period.

4 RESULTS AND DISCUSSION

The confidence region for the sediment load calculations for ParaSol using the option \( \chi^2 \)-statistics with \( \chi^2_{0.975} \) (Figure 5a) is much narrower and captures fewer observations than the confidence region for SUNGLASSES (Figure 5b). This result suggests that a traditional parameter uncertainty only covers a small share of the total uncertainty for cases where enough observations exist. Similarly, the confidence regions for the bias on the outputs of interest, i.e. the total loads, is much larger under SUNGLASSES than under ParaSol (Figure 3).

The result that SUNGLASSES has a much larger uncertainty bound than the ParaSol method indicates that other causes of uncertainty are involved including: the inappropriateness of the data set to identify the important processes, model structural errors, and model discretisation errors. The latter sources are likely true of most distributed environmental models as they share many of the attributes of distributed water quality models (processes scaled up from point scale to landscape scale, multiple criteria to meet, and inadequate data availability to properly parameterise these models). Erosion processes require thus an even higher physically based analysis of the system in order to define proper processes and scaling. Erosion processes are as well demanding for the underlying hydrological processes, where a proper representation of the small scale processes is needed rather than a just some good curve fitting.

The result that model structural error is a critical problem in water quality models is not unexpected as others have shown that structural changes in models can dramatically improve simulation results when focused on predicting floods [Boyle et al., 2001], predicting the effects of land use change on streamflow and salinity [Kuczera and Mroczkowski, 1998; Mroczkowski et al., 1997], or in finding flaws in models of stream chemical composition [Meixner et al., 2002]. Given this past experience of success in altering model structure and improving prediction results it is not surprising that model structural uncertainty is the
major source of predictive uncertainty when using water quality models. The result is thus reassuring since it indicates that what is needed for these models are a better way to represent the processes in them.

SUNGLASSES somehow operates not only as a validation procedure for the model structure but also as a validation of the parameter uncertainty procedure of the model (in this case the application of ParaSol on the Honey creek model). This validation is related to model structure since a good model structure should require less data to capture all dynamics and to average out the errors than is the case for a poor model structure. In general, if all underlying assumptions of the parameter uncertainty method are correct and if the dataset is adequate to translate the variability of the system into a model, SUNGLASSES should not lead to larger uncertainty bounds for the model outputs.

5 CONCLUSIONS

The ParaSol results show an important drawback in traditional statistical uncertainty methods: these do account for the number of observations, but do not consider additional sources of uncertainty that are in general not known and not quantifiable, such as model hypothesis errors, simplifications, scaling effects or the lack of the observation period to represent, in the model, the long-term variability and fluctuations of the real world. These uncertainties lead to wrong assessments of indicators (like global mass balances) that might be used in decision making. Therefore, SUNGLASSES is proposed for assessing total uncertainty to aid in the development of integrated models. It reveals problems of bias in the model outputs to be used for decision making by evaluating predictions outside the calibration period. SUNGLASSES leads to more selections of parameter combinations and much wider uncertainty ranges. SUNGLASSES enables thus to assess predictive uncertainty and helps decision makers understand how uncertain their models are so that they can put the proper level of trust in computational models of the environment as they move forward to make decisions. The results here indicate that the main concern should be about the uncertainty associated with model structural error and less so on model parametric uncertainty.

6 ACKNOWLEDGEMENTS

Support for this work was provided by the National Science Foundation (EAR-0094312 ). The authors are grateful to Sabine Grunwald and Tom Bishop of the University of Florida for sharing the Sandusky catchment model and R. Srinivasan and J. Arnold for their support of this research.

REFERENCES


Williams, J.R. Flood routing with variable travel time or variable storage coefficients. Trans. ASAE 12(1), 100-103.