

The Evaluation of Uncertainty Propagation into River Water Quality Predictions to Guide Future Monitoring Campaigns.

Vandenbergh V.*, Bauwens W.** and Vanrolleghem P.A.*

* Ghent University, Department of Applied Mathematics, Biometrics and Process Control, BIOMATH Coupure Links 653, B-9000 Ghent, Belgium (Email: veronique.vandenbergh@biomath.Ugent.be)

**Free University of Brussels, Laboratory of Hydrology and Hydraulic Engineering, Pleinlaan 2, B-1050 Brussels, Belgium

Abstract: To evaluate the future state of river water in view of actual loading or different management options, water quality models are a useful tool. However, the uncertainty on the model predictions is sometimes too high to draw proper conclusions. It is of high importance to modellers to minimise the uncertainty of the model predictions. Therefore different research is needed according to the origin of the uncertainty. If the uncertainty stems from input data uncertainty or from parameter uncertainty, more reliable results can be obtained by performing specific measurement campaigns. To guide these measurement campaigns, an uncertainty analysis can give important information.

In this article an overview of different techniques that give valuable information for the reduction of input and parameter uncertainty is given. The practical case study is the river Dender in Flanders, Belgium.

First a global sensitivity analysis shows the importance of the different uncertainty sources. Here it is seen that the parameters influence the model results more than the input data. Further an analysis in time and space of the uncertainty bands is performed to find differences in uncertainty between certain periods or places. More measurements are needed during periods or on places with high uncertainty. This research also shows that finding a link between periods with high uncertainty and specific circumstances (climatological, eco-regional, etc...) can help in gathering data for the calibration of submodels (eg. diffuse pollution vs. point pollution). The methods can be used for every variable under study and for all kind of rivers but the conclusions made for the practical case study are only applicable for the Dender.

Keywords: monitoring, optimal experimental design, river water quality modelling, uncertainty analysis

1. INTRODUCTION

In the field of environmental modelling and assessment, uncertainty analysis (UA) is a necessary tool to provide, next to the simulation results, also a quantitative expression of the reliability of those results. Next to the expression of uncertainty bounds on the results, uncertainty studies have mainly been used to provide insight in the parameter uncertainty. However, uncertainty analysis can also be a means to prioritise uncertainties and focus research efforts on the most problematic points of a model. As such, it helps to prepare future measurement campaigns and to guide policy decisions.

In this study, the use of an UA as an evaluation tool is assumed to be applied on an already calibrated model that can simulate measured data well but with an unacceptably high uncertainty. We only consider parameter and input uncertainty that can be minimised by gathering additional data. Model uncertainty and mathematical uncertainty

are not taken into consideration. The aim of this research is to show how UA can be used to guide future monitoring campaigns to make model results more reliable by minimising the parameter and input data uncertainty of the model.

The practical case study is the river Dender in Flanders, Belgium.

2. CASE STUDY: THE DENDER BASIN

The Dender river, a tributary of the river Scheldt in Belgium, drains an area of 1384 km². The main channel is partly canalised and contains 14 sluices. The river is heavily polluted by domestic, industrial and agricultural pollution.

A water quantity and quality model for the river Dender for 1994 was implemented in ESWAT. ESWAT is an extension of SWAT (van Griensven and Bauwens, 2000), the Soil and Water Assessment Tool developed by the USDA (Arnold et al., 1998). ESWAT was developed to allow for an integral modelling of the water quantity and quality processes in river basins.

3. METHODS

To reduce the overall uncertainty on the model results for a certain variable the following steps are proposed.

1. Identify which sources contribute mainly to the overall uncertainty on the model results
2. Estimate or calculate the uncertainty related to those main contributors
3. Propagate the uncertainty through the model
4. Analyse the model results to set up a future monitoring campaign
5. Perform the measurements
6. Recalibrate the model with new inputs
7. Repeat step 3 till 6 until satisfying results are obtained

For every step of this process different techniques exist that can be chosen among according to the experience of the modeller. In the practical example the methods we used will be described.

Step 1: Identification of the main uncertainty contributors, uncertainty characterisation.

This step is mainly carried out via a global or local sensitivity analysis. Because it is assumed that an already calibrated model is available, a local sensitivity analysis will certainly identify the most important parameters and data of the model. Indeed, local analysis is done around an a priori assumed value of the parameter. For a local sensitivity analysis the following methods exist: finite difference method, (b) the direct differential method, (c) the Green's function method, (d) the polynomial approximation method and (e) automatic differentiation.

For a detailed review of existing sensitivity techniques reference is made to the reviews of Turanyi (1990) and Rabitz et al. (1983)

Step 2: Estimation or calculation of uncertainty

Parameter uncertainty can be calculated using the covariance matrix obtained during the local sensitivity analysis or the calibration process. (Beck, 1987)

If no direct calculations are possible, e.g. for the uncertainty on the inputs, it is best to estimate the uncertainty for this. One can divide the parameters and data in uncertainty classes (accurately known, very poorly known and an intermediate class) and assign a percentage uncertainty to them. A similar approach was adopted by Reichert and Vanrolleghem, 2001.

Step 3: Propagate the uncertainty through the model

For this step Monte Carlo methods can be used, in which the input data or parameters are sampled between the uncertainty bounds that are detected in the previous step. Another option is to apply linear error propagation. The advantage of the latter is computational efficiency. However, if model non-linearities are significant within the uncertainty range, the results will be inaccurate. Monte Carlo simulation is a simple technique but requires a large number of model runs, which is computationally very demanding. Less runs with the same results as 'ad random sampling' are needed with 'the Latin Hypercube sampling' (McKay et al., 1988).

Step 4: Analyse the model results to set up a future measurement campaign

Two different approaches can be used according to the aim for which the additional measurements are collected. If it is the aim to reduce parameter uncertainty an automated optimal experimental design method that is explained in Vandenberghe et al (2002) can be used. It is based on maximisation of the determinant of the Fisher Information Matrix, which corresponds to the minimisation of the variance of the parameters. This method requires a lot of simulation runs but is totally automated and as such requires no additional information or knowledge from the modeller.

However, when only focussing on the input data uncertainty that leads to output uncertainty expert-knowledge is required. It is then the aim to find a link between periods of high/low uncertainty and external circumstances (rain, discharge points, seasons, solar radiation,...) This information is then used to make decisions about, place, period, frequency,... of future measurements.

Step 5: Perform the measurements

At this stage it is essential to ensure a good quality control on the measurements to minimise measurement errors. Important is also to carefully add information concerning hour, place and depth of the sample.

Step 6: Recalibrate the model with new inputs

An important issue here is that the calibration method has to be able to find the optimum. First, a choice is made between manual and automated methods. The former depends totally on the

experience of the modeller. Automated methods can differ in search method: global search methods scan the whole parameter space and are as such able to find the global optimum, but do not provide uncertainty measures. Local search methods start on a certain point in parameter space and end when they find an optimum. However, there is no assurance that this is the global optimum, so it is best to start in the neighbourhood of the optimum for those methods. With these methods covariance matrices for the optimum parameters are often calculated.

Step 7: Repeat step 3 till 6 until satisfying results are obtained

The stop criterion for this trial and error method is dictated by an ‘a priori’ desired reliability of the model results. In practice however, personnel, time and equipment matters will be the limiting factor and will indicate when this process stops.

4. RESULTS AND DISCUSSION

The seven steps are now demonstrated on a case study: simulations of the water quality of the river Dender, Flanders, Belgium for 1994. The evaluation of the uncertainty on model results is performed for Nitrate in the river water.

Step 1: Identification of the main uncertainty contributors.

We evaluate the sensitivity of the model on the following result: the time that NO₃ is higher than 3 mg/l at Denderbelle, near the mouth of the river in 1994. A sensitivity analysis for all input data and parameters in the ESWAT model is too complex for the program we use: UNCSAM (Janssen et al, 1992). This program cannot handle more than 50 parameters at the time. So we split the problem in different parts: 1) sensitivity to model parameters 2) sensitivity to point pollution input and 3) sensitivity to diffuse pollution input. Each sub problem gives a ranking of the parameters by using the Standardised Regression Coefficient (SRC) (1)

$$SRC_i = \frac{\Delta y / S_y}{\Delta x_i / S_{x_i}} \quad (1) \quad \text{with} \quad \Delta y / \Delta x_i = \text{change}$$

in output due to a change in an input factor and S_y, S_{x_i} the standard deviation of respectively the output and the input. The input standard deviation S_{x_i} is specified by the user.

The technique is explained in Vandenberghe et al. (2001). For each of the subproblems the parameters or data that contribute significantly to the output (5 % level) are then taken together in

one overall sensitivity analysis to compare the contribution of the different outputs. The column with the SRC as a result of that analysis is indicated in table 1 with “combined parameter-input”.

Table 1: Results of the sensitivity analysis for the model output “hours NO₃ >3mg/l” at Denderbelle, 1994. (pa16 = Amount of fertilisation on pasture in subbasin 16; fa4 = Amount of fertilisation on farming land in subbasin 4; gro pa = growth date of pasture; plfa = Plant date on farming land; co5 = Amount of fertilisation on corn in subbasin 5; co15 = Amount of fertilisation on corn in subbasin 15; pa12 = Amount of fertilisation on pasture in subbasin 12; co11 = Amount of fertilisation on corn in subbasin 11; ai5 = O₂ uptake per unit of NH₃ oxidation; rk5 = denitrification rate; rk2 = oxygen reeration rate; ai6 = O₂ uptake per unit of HNO₂ oxidation; bc2 = rate NO₂ to NO₃; rk3 = rate of loss of bod due to settling; ai4 = O₂ uptake per unit of algae respiration; Rs5 = organic phosphorous settling rate)

Diffuse pollution input	SRC	Point pollution input	SRC	parameter	SRC	Combined Parameter-input	SRC
Pa16	-0.30	BOD point 6	-0.61	Ai5	-0.7	Ai5	-0.51
Fa4	0.23	NO3 point 7	0.42	Rk5	-0.34	Ai6	-0.50
gro pa	-0.18	BOD point 5	-0.38	Rk2	0.32	Rk5	-0.40
plfa	0.17	BOD point 8	-0.24	Ai6	-0.21	Bc2	0.38
Co5	-0.17	NH3 point 1	0.23	Bc2	-0.2	Ai4	-0.31
Co15	-0.16	BOD point 3	-0.23	Rk3	0.17	Rk2	0.12
Pa12	0.16	BOD point 7	-0.22	Ai4	0.12	plfa	-0.08
Co11	0.15	BOD point 1	-0.14	Rs5	-0.09	BOD point 6	-0.07
		NO3 point 5	0.11		0.07	BOD point 1	-0.07
		BOD point 4	-0.09			Pa16	0.07
		NH3 point 2	0.09				
		BOD point 2	-0.08				
		NH3 point 3	0.06				

For the parameters, the sampling for the sensitivity analysis was based on own experience and

literature ranges. The ranges for the diffuse pollution inputs are given in table 2 and the way they are determined is explained in Vandenberghe et al. (2003). For the point pollution inputs we sampled uniform between halve and double the values, as we decided that those inputs belong to the uncertainty class 'poorly known', indeed, the loads coming from point pollution were only available as yearly averages.

Table 2. Uncertainty ranges for diffuse pollution input.

Input	Uncertainty
Plant date for the crops	+/- 1 month
Harvest date of the crops	+/- 1 month
Amount of fertiliser applied per subbasin and per crop (kg/ha)	+/-25%

The global sensitivity of the parameters and the inputs shows that some parameters, O₂ uptake per unit of NH₃ oxidation, O₂ uptake per unit of HNO₂ oxidation, denitrification rate, rate NO₂ to NO₃, O₂ uptake per unit of algae respiration and the reaeration rate are most influencing followed by the input data, plant date on farming land, Amount of fertilisation on pasture in subbasin 12 and bod loads from point 1 and 6. This could not be seen from the separate analyses of inputs and parameters. So the parameters can make the model give different results that are not much influenced by the input data. This again shows the importance of a well-calibrated model.

Step 2: Estimation or calculation of uncertainty

For both the point and diffuse pollution input the same uncertainties were taken as the sampling range used for the sensitivity analysis because we obtained no new information between the SA and the UA. For the uncertainty on the parameters a recalibration with the most influencing parameters so that uncertainty ranges can be calculated with the covariance matrix is best, but is not done here. Uncertainties of 50 % were assigned to each of the parameters.

Step 3: Propagation of the uncertainty through the model

Here again the uncertainties are split: parameter uncertainty, diffuse pollution uncertainty and point pollution uncertainty.

Then for each an uncertainty analysis was performed in which all of the uncertainty sources are varied at the same time to see the effects of the uncertainty on parameters and inputs. For this analysis we calculate the uncertainty bands (i.e. the 5% and 95% percentiles) for the results of the time series.

Figure 2 and 3 shows the time series of nitrate in the river water at Denderbelle, situated near the

mouth, with the 5% and 95% uncertainty bounds with resp. uncertainty on diffuse input and point pollution input. Figure 1 shows the uncertainty bounds for nitrate at the same location due to parameter uncertainty.

Step 4: Analyse the model results to set up a future measurement campaign

Figure 1 shows the propagation in time of the parameter uncertainty for Nitrate in the river at Denderbelle, 1994. Parameter uncertainty becomes at certain moments.

To cope with the parameter uncertainty optimal experimental design based on the Fisher Information Matrix should be done (as explained in the methods section) as this is the most objective method to find important measurement places to better estimate the parameters. This design of new experiments is not presented here as we focus here on the uncertainty analysis and what information can be revealed from it.

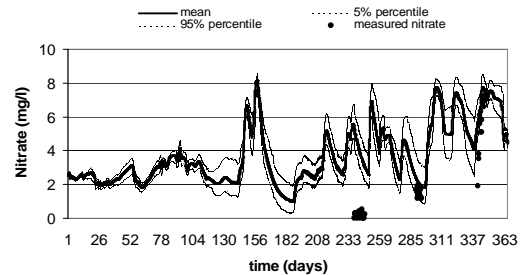


Figure 1. Simulation of nitrate with confidence intervals related to parameter uncertainty at Denderbelle, 1994.

Figure 2 and 3 give shows the simulations and their confidence intervals related to the uncertainty on the model inputs.

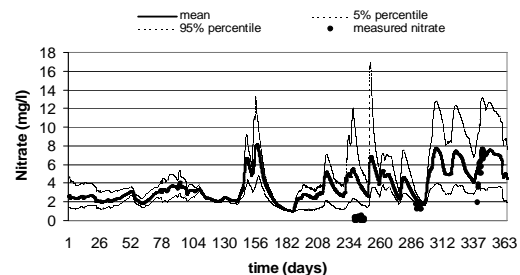
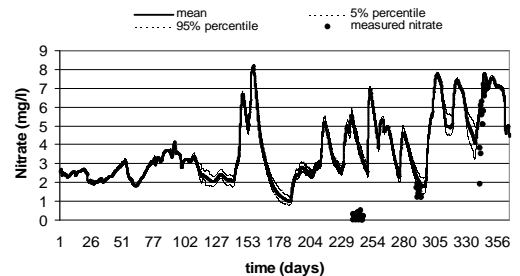


Figure 2 and 3. Simulation of nitrate with confidence intervals related to diffuse and point pollution input uncertainty at Denderbelle, 1994.

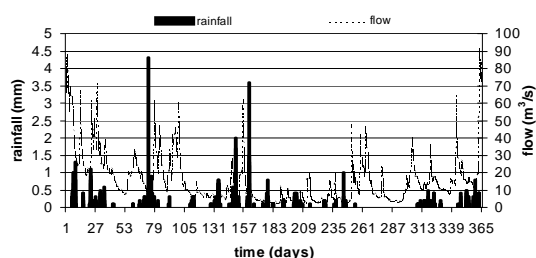


Figure 4. Rainfall and Flow in 1994 at Denderbelle.

Linking the obtained results in step 3 to the external circumstances, rain and flow (fig.4), we can see that diffuse pollution inputs are important during periods with high rainfall and high flows. During dry weather flows, the input uncertainty of the loads is also propagated. Hence this UA learns that we can obtain a better calibration for the diffuse pollution part of the model with data that are taken during wet periods with high flows, because the model output nitrate is more sensitive towards inputs of diffuse pollution in those periods. If one focusses on calibrating the in-stream behaviour and point pollution then measurements during dry periods are needed, as the model is in such conditions not sensitive towards input of diffuse pollution.

Further it is seen on fig. 1 that the 95 % bounds show much higher peaks than the mean concentrations time series. This means that some peak values of nitrate in the river water at Denderbelle may not be predicted properly due to an underestimation of the amount of fertiliser used. Those peaks (eg. day 156 and 260) are significantly higher than the levels of nitrate for basic water quality.

It is also of interest to know how the uncertainty is propagated from one place to the other. This analysis was done for the uncertainty propagation due to diffuse pollution inputs. The amount of time that NO_3 was higher than 3 mg/l was calculated. This was done for the time series of the mean, the 5 % - bound and the 95% - bound (Fig. 5). The uncertainty bounds become larger when approaching the mouth due to the summation of the uncertainties on all diffuse pollution inputs that enter the river. However, it is interesting to see that with the available quality of input data no conclusions can be drawn concerning the question whether the diffuse pollution causes more hours nitrate exceedance downstream than upstream. More accurate data are needed to draw good conclusions from the model results.

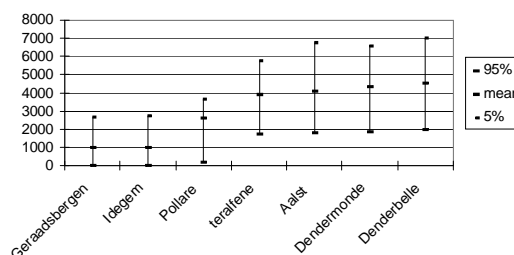


figure 5. Uncertainty propagation from upstream to the mouth of the Dender in 1994 related to diffuse pollution input uncertainty.

Step 5: Perform the measurements

Step 6: Recalibrate the model with new inputs

Step 7: Repeat step 3 till 6 until satisfying results are obtained

Those three steps are only relevant for future measurement campaigns. However, no additional measurements were done until now.

5. CONCLUSIONS & RECOMMENDATIONS

The results of uncertainty analysis were here evaluated to guide future monitoring campaigns. Diffuse and point pollution inputs are considered separately and give information of the model sensitivity to the inputs. Measurements during dry periods can be used to better calibrate the model for point source pollution because the inputs of diffuse pollution are not important then. On the other hand, periods with rainfall and high flows are needed for the calibration of the model with diffuse pollution because the model output nitrate is then very sensitive towards the inputs related to farmer's practices.

When comparing the influence of the uncertainty of the diffuse pollution inputs, the uncertainty bounds appeared to be too high to draw reliable conclusions from the model results. So, it showed the importance of accurate measurements and input data if the model results serve for decision support.

It is obvious from the comparison between the global sensitivity analysis for the subgroups and for all most influencing parameters together that the parameters are most important. This shows that it is best to start with a good calibration of your model and then focus on more accurate input data.

Too often a model is calibrated with only one comprehensive measurement campaign. This is mostly not the most efficient way. When e.g. only measurements during dry periods are made, the model cannot be well calibrated for the diffuse pollution part. So it is better to perform two separate smaller measurement campaigns with the

first one being 'exploring', while the second campaign is guided by previous analysis of the model results. The combination of the two monitoring campaigns can assure that at least some measurements are performed at 'the right moment', making the calibration process easier and more reliable.

It is necessary to combine all previous uncertainty analysis to evaluate the total uncertainty on the model results and to compare them with the measurements. In this way, model structure uncertainty can also be quantified (Willems and Berlamont, 2002).

In this research the second monitoring campaign is missing and could have shown the possibilities of the proposed succession of steps.

8. ACKNOWLEDGEMENT

The authors give special thanks to the financial support of the EU Harmoni-CA project (EVKI-CT-2002-20003) and the EU CD4WC project (EVK1-CT-2002-00118).

7. LITERATURE

- Arnold J.G., Williams J.R., Srivivasan R. and King K.W. (1996). *SWAT manual*. USDA, Agricultural Research Service and Blackland Research Center, Texas.
- Beck M.B. (1987). Water quality modeling: A review of the analysis of uncertainty. *Wat. Res. Res.*, **23**(5), 1393-1441.
- Demuyne C., Bauwens W., De Pauw N., Dobbelaere I. and Poelman E. (1997). Evaluation of pollution reduction scenarios in a river basin: application of long term water quality simulations. *Wat.Sci.Tech.*, **35**(9), 65-75.
- Janssen P.H.M., Heuberger, P.S.C. and Sanders S. (1992). *Manual Uncsam 1.1., a software package for sensitivity and uncertainty analysis*. Bilthoven, The Netherlands.
- McKay M.D. (1988). *Sensitivity and uncertainty analysis using a statistical sample of input values*. In: Uncertainty analysis, Y. Ronen, ed., CRC Press, Inc., Boca Raton, Florida, 145-186.
- Rabitz H., Kramer M. and Dacol D. (1983). Sensitivity analysis in chemical kinetics. *Ann. Rev. Phys. Chem.*, **34**, 419-461.
- Reichert P. and Vanrollegem P.A. (2001). Identifiability analysis of the River Water Quality Model No. 1 (RWQM1). *Wat. Sci. Tech.*, **43**(7), 329-338.
- Turanyi T. (1990), Sensitivity analysis of complex kinetic systems, tools and applications. *J. Math. Chem.*, **5**, 203-248.
- Vandenberghe V., van Griensven A., Bauwens W. (2002). Detection of the most optimal measuring points for water quality variables: Application to the river water quality model of the river Dender in ESWAT. *Wat.Sci.Tech.*, **46**(3),1-7.
- Vandenberghe V., van Griensven A., Bauwens W. and Vanrollegem P.A. (2003). Propagation of uncertainty in diffuse pollution into water quality predictions: Application to the river Dender in Flanders, Belgium. In: *Proceedings of the 7th International Specialised Conference on Diffuse Pollution and Basin Management*, 17-22 August 2003, Dublin, Ireland.
- Van Griensven A. and Bauwens W. (2000). Integral modelling of catchments. *Wat.Sci.Tech.*, **43**(7), 321-328.
- Willems P. and Berlamont J., (2002). Probabilistic emission and immission modelling: case-study of the combined sewer-WWTP-receiving water system at Dessel (Belgium). *Wat.Sci. Tech.*, **45**(3), 117-124.

Editors: Claudia Pahl-Wostl, Sonja Schmidt,
Andrea E. Rizzoli, Anthony J. Jakeman

University of Osnabrück, Germany



Complexity and Integrated

Resources Management

Transactions

of the 2nd Biennial

Meeting of the International

Environmental Modelling

and Software Society

Vol. 2

14-17 June 2004