Multi-Objective Calibration Of A River Water Quality Model For The Elbe River, Germany

Rode, M. and U. Suhr

UFZ Centre for Environmental Research Leipzig Halle, Department of Hydrological Modelling, Brückstraße 3a, D-39114 Magdeburg, Germany, michael.rode@ufz.de

Keywords: *River water quality modelling; automatic model calibration; data information content*

EXTENDED ABSTRACT

There is a growing need to use river water quality models for the development of river basin management plans. Substantial uncertainties exist in the identification of river water quality models which partially relates to parameter uncertainties and the information content of calibration data. The specific objectives of this study were to i) assess the parameter identification] of a river water quality model using the Parameter Estimation Program (PEST), and ii) evaluate the dependencies between available calibration data and model prediction using a cross-validation procedure.

The investigation was conducted based on five extensive flow time related longitudinal surveys with 14 sampling locations along a 536 km free-flowing reach of the German part of the Elbe River. The five surveys represent seasonal development of algal biomass and nutrient concentration under a wide range of boundary conditions like low flow conditions, high global radiation and nutrient limiting algal growth conditions. The nonlinear parameter estimator PEST was used for the multi-objective calibration of the deterministic river water quality model QSIM. Calibration runs were conducted using five different data sets and all their possible combinations. Cross-validation of the model was carried out using the remaining 4, 3, 2 and 1 data sets not included in the calibration process.

To identify the 'optimal' parameters for the River Elbe, the river water quality model was calibrated using all five longitudinal surveys. The results show good agreement between measured and simulated values for all variables (Chl a, NO_3, O_2, dissolved phosphorus (DP), dissolved Si). Considering the z-distribution of normalized residuals for all five output variables, we found a good approximation of the normal distribution (Figure 1). This indicates that the multi-objective model calibration leads to a well-defined model for the Elbe River. Hence, we have confidence that the PEST multi-objective calibration procedure is able to identify an optima for all eight parameters used for model calibration.

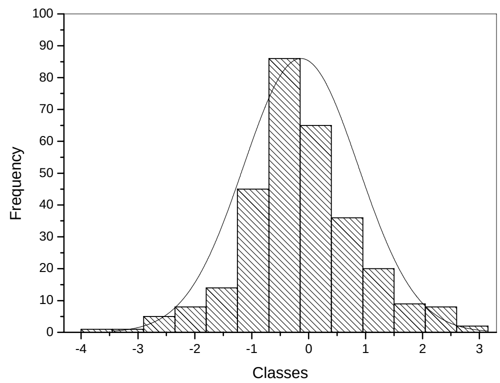

Figure 1. Z-distribution of normalized residuals of five water quality variables using all five calibration data sets

The Elbe case study also showed that calibration with a single survey data set leads to substantial errors if these parameters are applied to deviating boundary conditions. These uncertainties can be decreased with an increased calibration database. The overall performance of the validation improves only substantially when increasing the number of calibration data sets from one to two data sets.

It can be concluded that the PEST optimization tool is very efficient using a comparably small number of optimization runs for model calibration. In the presented case study at least two longitudinal data sets of differing boundary conditions should be used for calibration of the primary production and nutrient modules of the QSIM model. The results of this study will help model users e.g. at environmental agencies to evaluate the uncertainties in river water quality models and to define appropriate data collections and monitoring schemes to achieve a sufficient accuracy of model predictions.

1. INTRODUCTION

There is a growing need to use water quality models for the development of river basin management plans. Most of the widely used river water quality models (like QUAL2, WASP5, Mike 21, AQUASIM) use physically-based process descriptions of the hydrodynamics and more or less conceptual descriptions of matter transformation and primary production in river system. Due to the spatial and temporal variability, measurement errors, the simplicity of the model description or simply the lack of data, often parameter values will not be exactly known. Therefore, in most cases a model calibration will be necessary. Moreover uncertain model identification often is caused by small amounts and information content of calibration data.

The time-consuming nature of manual trial- and error model calibration has led to the development of the more complex inverse modelling techniques for parameter estimation, where parameters are optimized while minimizing a suitable objective function that expresses the discrepancy between the output of a dynamic model and the measurement. Most commonly used automated calibration procedures are the shuffle complex evolution method (Duan et al., 1992) and the Levenberg-Marquardt method (Olsthoorn, 1995; Doherty and Johnston, 2003). Omlin et al. (2001) used a systematic approach for model identification combining an analysis of the sensitivity of model results to single parameters with an analysis of the approximate linear dependency of sensitivity functions of parameter subsets.

In the case of water quality models there is an important advantage compared to hydrological models that multiple criteria like different nutrient and biological variables can be used for model calibration. In the past only very few attempts were made to use multi-objective criteria for river water quality model identification (van Griensven & Bauwens, 2003, McIntyre et al. 2003). Omlin et al. (2001) used a multi-objective parameter estimation process for a sensitivity analysis of a water quality model of Lake Zürich. He also gave estimates of the uncertainties of model predictions. The same procedure was used with the River Water Quality Model Number 1 for a simple hypothetical case study (Reichert & Vanrolleghem, 2001). In the present study we used the parameter estimation software PEST which uses the Gauss-Marquardt-Levenberg method for the automatic model calibration (Doherty, 2000). When a continuous relationship exists between model parameters and model outputs, it is a very efficient method to identify the minimum in the objective function compared to other methods (Doherty and Johnson 2003).

The calibration process is closely related to the estimation of parameter uncertainties and hence the corresponding model predictions. As Beck (1991) pointed out a calibrated model incorporates acquired knowledge about the studied system. Therefore, calibration aims not only at finding parameter sets that will minimize a given objective function, which is itself not an easy task since the models are frequently non-linear. It also aims at reducing the uncertainties in the parameter values as well. Because of limited calibration data, it is often hard to identify a sufficiently certain parameter set of a river water quality model. This is due to the non-uniqueness of the optimized parameters. Non-uniqueness leads to more than one set of parameters, each yielding minimum values for the objective function determined by local minima (Vrugt et al., 2001). However, research into data requirements has led to the understanding that a larger amount of data or information will not necessarily improve the identification of the parameters (Kuczera, 1982). In most cases the simultaneous use of more than one signal can improve parameter identifiability (Gupta et al. 1998).

The specific objectives of this study were to i) assess the parameter identification of a river water quality model using the PEST methodology and to ii) evaluate the dependencies between available calibration data and model prediction using a cross-validation procedure. The results of this study will help model users e.g. at environmental agencies to evaluate the uncertainties in river water quality models and to define appropriate data collections and monitoring schemes to achieve a sufficient accuracy of model predictions.

2. MATERIAL AND METHODS

2.1 Case Study

The River Elbe is one of the largest rivers in Central Europe with a length of 1091.5 km. Most water quality data used for this study were acquired during five longitudinal surveys of phytoplankton development and nutrient concentrations in autumn 1996, late summer 1997 and 1998, in spring 1999 (Guhr et al., 2003) and July 2000 (Schöl et al.,2004). These surveys were carried out for the unregulated river reach from Schmilka at the German Czech border to Neu Darchau in the lower part of the river with a length of 536 km. The five surveys represent seasonal

Tabel 1. Start values of simulation runs, mean values at station Schmilka (km 3,9)

Input variable	Unit	Oct. 1996	Aug. 1997	Sept. 1998	May 1999	July 2000
Discharge	m**3/s	225	151	127	300	140
Chlorophyll a	µg/l	8.6	37.2	17	63	45
Share of diatoms	-	0.7	0.6	0.5	0.6	0.6
Ammonium	mgN/l	0.18	0.12	0.06	0.12	0.10
Nitrate	mgN/l	4.2	4.2	3.9	4.5	3.7
SRP	mgP/l	0.16	0.16	0.20	0.05	0.23
Dissolved phosphorus	mgP/l	0.17	0.16	0.22	0.06	0.25
Silicon	mgSi/l	5.4	2.9	3.0	2.4	3.2
Oxygen	mgO2/l	9.3	7.9	8.5	11.4	7.3
Suspended particulate matter	mg/l	20	20	25	20	18
Water temperature	°C	13.6	22.8	17.5	12.9	19.2
Global radiation, daily sum	J/cm**2	400	1500	1130	1700	1600

development of algal biomass and nutrient concentration under a wide range of boundary conditions like low flow conditions (1996), high global radiation and nutrient limiting algal growth conditions (1999). Upper boundary conditions (start values) relevant to primary production and nutrient concentration development for the five measurement surveys, i.e. discharge, temperature and light, are given in Table 1. Up to nine samples were taken for each cross section. In the river reach from Schmilka to Neu Darchau flow times varied according to the discharge between seven days for the survey in May 1999 and nine days for the survey in September 1998.

2.2 Model and calibration tool

For the Elbe River case study the widely used river water quality model QSIM, developed by the German Federal Institute of Hydrology, was applied (Kirchesch and Schöll, 1999; Schöl et al., 1999). The model has a modular structure with main modules concerning hydraulic, physical, chemical, and biotic processes. Driving forces of the model are discharge at the upper boundary and main tributaries as well as meteorological conditions including global radiation, air temperature, cloudiness, and wind velocity. Phytoplankton growth is simulated after Monod and Michaelis-Menten? kinetics. Hydraulic calculations are based on more than 3000 river cross sections. The duration of the longitudinal samplings corresponds well with the computed flow times for all five surveys.

For model calibration the automatic Parameter Estimation Program (PEST) was used, which implements the Gauss–Levenberg–Marquardt method (Doherty, 2000). The optimization process is a “hill-climbing” technique in which from a starting point the steepest gradient of the objective function in the parameter space is calculated. (Doherty, 2004).Derivatives of observations with respect to parameters are calculated using finite differences. The objective function used is the weighted sum of squares ϕ :

$$\phi = \sum_{i=1}^{m} (\omega_i r_i)^2 \qquad (1)$$

where ω_i is the weight attached to the i’th observation, *m* the number of observations, and r_i (the i’th residual) expresses the difference between the model outcome and the measured river water quality variable. Weights are inversely proportional to the standard deviations of the observed values. The values for ω_i are used as empirical weights with a goal to make the contributions of different model variables to ϕ similar in size and, therefore, give all measured variables a similar influence on the estimates of the parameters. Table 2 gives the values of ω_i used for multi-objective calibration. After optimizing the

Table 2. Scales used as empirical weights for different variables

Variable	Value	Coefficient of variation	Unit
Chl a	1	0,768	mg Chl a l^{-1}
O2	5	0,146	mg O l^{-1}
DP	500	0,615	mg P l^{-1}
DSi	10	0,633	mg Si l^{-1}
NO_3	10	0,119	mg N l^{-1}

parameters PEST calculates additional information on the 95% confidence limits for the adjustable parameters if the covariance matrix has been calculated. Furthermore based on the covariance matrix a correlation coefficient matrix is calculated and can be used to assess the dependencies between the parameters. As a by-product of the parameter estimation process, PEST calculates the composite scaled sensitivity of the parameter. The calculated uncertainty information of the parameters is determined on the same linearity assumption which was used to derive the equations for parameter improvement implemented in the optimization process. Various tests were made to tune the PEST optimization algorithm to the specific case study. Based on a sensitivity analysis in a first step most important kinetic parameters where included in the parameter estimation process. In a second step seven parameters were selected for the optimization process. Notice that due to dependencies between model parameters these parameters used for the automatic calibration process were not the seven most sensitive ones.

2.3 Evaluating data information content

The investigation of the water quality data information content was carried out for the parameter identification of the river water quality model. For this purpose calibration runs were conducted using data of five different data sets and all their possible combinations. This consists of a total of 30 data sets with different flow time related longitudinal data and water quality conditions. For each set, a single multi-objective optimization using the PEST program was conducted, where each optimization leads to up to about 200 model runs depending on the amount of calibration data. The 8 most sensitive kinetic model parameters were included in the model calibration. A common technique of validating a model is to use only a subset of all available data for calibration and test the model performance with the rest of the data. In this study the remaining 4, 3, 2 and 1 data sets not included in the calibration process were used for cross-validation of the model. In order to assess model performance precisely we used the index of agreement (Willmott, 1981) defined as:

$$d = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} \left(\left| P_i - \overline{O} \right| + \left| O_i - \overline{O} \right| \right)^2} \qquad (1)$$

where $\overline{O}$ is the mean of the observed value, O_i is the observed and P_i is the simulated value. The index of agreement varies between 0.0 and 1.0 with higher values indicating better agreement between the model and observations, similar to the coefficient of determination R^2. The index of agreement is also sensitive to extreme values, owning to the squared differences (Legates and McCabe, 1999). The objective criteria were used both for the calibration and validation of the model.

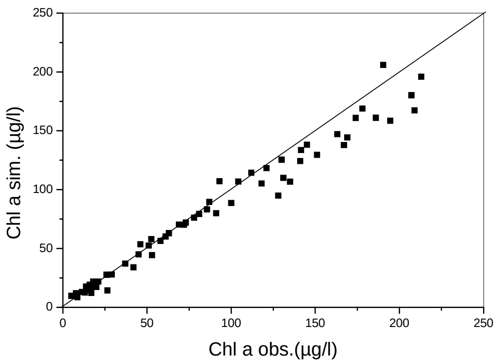

Figure 1. Observed and simulated Chl a concentrations based on five calibration data sets

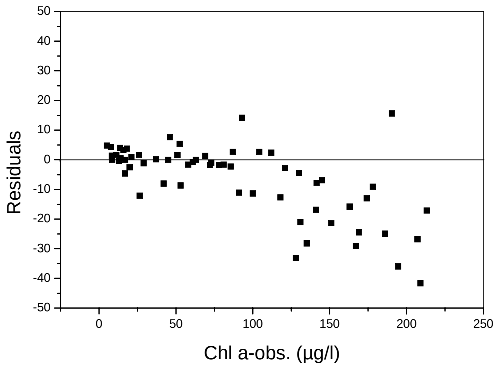

Figure 2. Residuals of Chl a concentrations based on five calibration data sets

3. RESULTS

To identify the ‘optimal’ parameters for the Elbe River, the river water quality model was calibrated using all five longitudinal surveys. The results show good agreement between measured and simulated values for all variables (Chl a, NO_3, O_2, dissolved phosphorus (DP), dissolved Si). The simulated and observed values of Chl a, which represents algal biomass, are shown as an example in Figure 1 and 2. The results indicate that the model is able to simulate algal growth for a wide range of boundary conditions with the same model parameters. In the case of very high Chl a concentration, model residuals tended to increase, contradicting the assumption of homogenous error variance. Considering the Z-distribution of normalized residuals for all five output variables we found a good approximation of the normal distribution (Figure 3). Z is defined as the difference between the value and the mean divided by the standard deviation of the residuals of a given variable. This indicates that the multi-objective model calibration leads to a well defined model for the Elbe River. Hence, we have confidence that the PEST multi-objective calibration procedure is able to identify a reasonable optimum for all eight parameters used for model calibration.

The empirical cumulative distribution function (cdf) for the Index of Agreement (IA) for different numbers of data sets was constructed. Figure 4 shows the results for the calibration and Figure 5 for the corresponding validation of the remaining data sets. Each cdf indicates a chance of obtaining a statistic of the magnitude if a data set of that number of longitudinal surveys is selected at random and used for calibration. The IA cdfs shift to the right as we increase the number of longitudinal surveys used for calibration. This finding is also true for the corresponding validation data sets. The shift to the right indicates improvement of model performance with increasing number of longitudinal data sets (compare Yapo et al. 1996). The calibration cdfs only shift significantly to the right when increasing the number of data sets from one to two, showing that only small improvements of model performance can be achieved when using more than two calibration data sets.

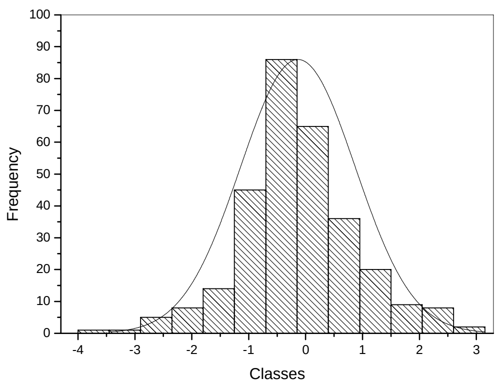

Figure 3. Z-distribution of normalized residuals for calibration run of all 5 data sets

In contrast to the calibration cdfs, the IA cdfs for validation steepen progressively with increasing number of calibration data sets and hence decreasing number of validation data sets. Increasing the steepness indicates reducing sensitivity of model performance to selection of data sets. The validation cdfs show a decrease of the range of IA values when increasing the number of calibration data sets and hence decreasing the number of corresponding validation data sets. The overall performance of the validation improves only substantially when increasing the number of calibration data sets from one to two data sets. This leads to the conclusion that at least two longitudinal data sets of differing boundary conditions should be used for model calibration and the benefit of the use of more than two data sets may be marginal. It should be noted that the findings are restricted to the specific conditions of the River Elbe with its long unregulated section, high algal growth rates and significant dependencies between algal biomass and nutrient concentrations. In the case of more complex boundary conditions, e.g. high inputs from sewage plants, and less significant relationships between water quality constituents more than two longitudinal surveys may be needed for reasonable model identification.

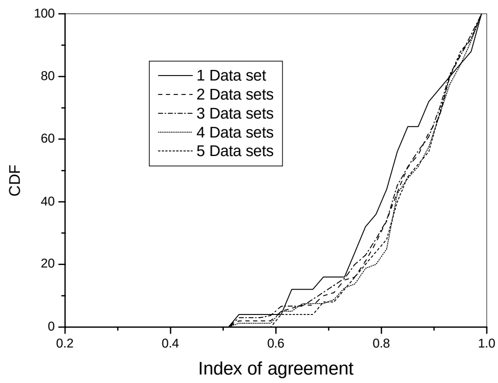

Figure 4. Empirical cumulative distribution functions of the Index of Agreement statistics for different numbers of calibration data sets

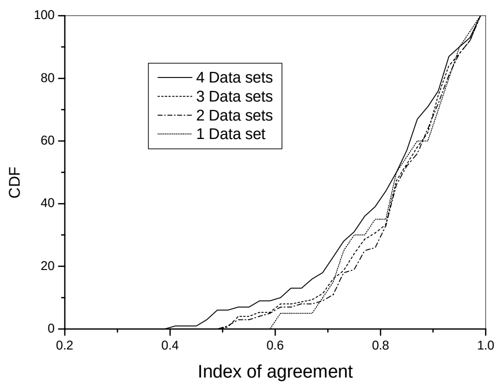

Figure 5. Empirical cumulative distribution functions of the Index of Agreement statistics for different numbers of validation data sets

4. CONCLUSION

The River Elbe case study on the calibration of the river water quality model QSIM shows that an automatic multi-objective calibration using the optimization tool PEST leads to reasonable model identification. The optimization tool is very efficient using a comparatively small number of optimization runs for model calibration. Furthermore PEST offers additional information for assessing uncertainties of calibrated model parameters. The investigation on data information content showed that calibration with single flow time related measuring surveys lead to substantial errors if these parameters are applied to deviating boundary conditions. These uncertainties can be decreased with an increased calibration database. In the case of the Elbe River, two or more data sets of flow time related longitudinal measuring surveys will be needed for a reasonable model identification of the primary production and associated nutrient model components of the river water quality model. These findings are restricted to cases were data sets of deviating boundary conditions are available. If data sets with comparable boundary conditions are used additional information for model parameter identification may be limited. The suggested methodology for model calibration including a cross validation procedure is especially suited for case studies with limited available data which is common for river water quality modelling investigations. Recently cross validation procedures have also been used for precipitation runoff modelling.

5. REFERENCES

Beck, M.B. (1991), Principles of modelling, *Water Sci. Technol.*, 24 (6), 1-8.

Doherty, J. (2000), Visual PEST - Model independent Parameter Estimation. Watermark Numerical Computing & Waterloo Hydrogeologic (http://www.flowpath.com/)

Doherty, J. and J.M. Johnston (2003), Methodologies for calibration and predictive analysis of a watershed model, *Journal of the American Water Resources Association* (4), 251-265.

Duan, Q., Gupta, V.K., Sorooshian, S. (1992), Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, 28(4), 1015-1031.

Guhr, H., Spott, D., Bormki, G., Baborowski, M., and Karrasch, B. (2003), The Effects of Nutrient Concentrations in the River Elbe. *Acta hydrochim.hydrobiol.* **31** (4-5), 282-296.

Gupta, H.V., S. Sorooshian, and P.O. Yapo (1998), Toward improved calibration of hydrological models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34, 751-763.

Kirchesch, V., Schöll, A. (1999), Das Gewässergütemodell Qsim – Ein Instrument zur Simulation und Prognose des Stoffhaushalts und der Planktondynamik von Fließgewässern. *Hydrologie und Wasser-bewirtschaftung*, 43 (6) 302-309.

Kuczera, G. (1982), On the relationship of the reliability of parameter estimation and hydrological time series data used in calibration. *Water Resources Research* 18, 146-154.

Legates, D.R., McCabe, G.J. (1999), Evaluating the use of “goodness-of-fit” measure in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35, 233-241

McIntyre, N.R., Wagener, T., Wheater, H.S., and Chapra,St. (2003), Risk-based modeling of surface water quality: a case study of the Charles River, Massachusetts. *Journal of Hydrology* **274** 225-247.

Olsthoorn, T.N. (1995), Effective parameter optimization for groundwater model calibration. *Ground Water* 33:42-48.

Omlin, M., Brun, R., Reichert, P. (2001), Biogeochemical model of lake Zürich: sensitivity, identifiability and uncertainty analysis. *Ecological Modelling*, **141**, 105-123.

Reichert, P. and Vanrolleghem, P. (2001), Identifiability and uncertainty analysis of the River Water Quality Model No. 1 (RWQM1) *Water Sci. Technol.* 43 (7) 329-338.

Schöl, A., Eidner, R., Böhme M., Kirchesch, V., Wörner, U., Müller, D. (2004), Bedeutung der Stillwasserzonen für die Nährstofflimitierung in der Elbe. Bundesanstalt für Gewässerkunde, Bericht 1406, 143 S.

Schöl, A., Kirchesch, V., Bergfeld, T., and Müller, D. (1999) Model-based analysis of oxygen budget and biological processes in the regulated rivers Moselle and Saar: modelling the influence of benthic filter feeders on phytoplankton. *Hydrobiologia*,410, 167-176.

Sincock, A.M., Wheater, H.S. and Whitehead, P.G. (2003), Calibration and sensitivity analysis of a river quality model under unsteady flow conditions. *Journal of Hydrology* **277** 214-229.

van Griensven A. and Bauwens W. (2003) Multi-objective autocalibration for semidistributed water quality models. *Water Resources Research*, 39 (12) 1348-1357.

Vrugt, J.A., Bouten, W., and Weerts, A. (2001), Information Content of Data for Identifying

Soil Hydraulic Parameters from Outflow Experiments. *Soil Sci. Soc.* **65** 19-27.

Wagener, T., McIntyre, N.R., Lees, M., Wheater, H.S., and Gupta, H.V. (2003), Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis. *Hydrological processes* **17** 455-476.

Wilmott, C.J. (1981), On the validation of models, *Phys.Geogr.*, 2, 184-194.

Yapo,P.O., Gupta,H.V., and Sorooshian,S. (1996), Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *Journal of Hydrology* **181** 23-48.