Predictions in Ungauged Basins: Promise and Progress (Proceedings of symposium S7 held during the Seventh IAHS Scientific Assembly at Foz do Iguaçu, Brazil, April 2005). IAHS Publ. 303, 2006.

Identification of a river water quality model and assessment of data information content

MICHAEL RODE & GUNTER WRIEDT

Department of Hydrological Modelling, UFZ Centre for Environmental Research Leipzig Halle, Brückstraße 3a, D-39114 Magdeburg, Germany

michael.rode@ufz.de

Abstract Substantial uncertainties exist in the identification of river water quality models, which partially depend on the information content of the calibration data. To evaluate the dependencies between available calibration data and model predictions investigations were conducted based on a 536 km free-flowing reach of the German part of the River Elbe. Five extensive flowtime related longitudinal surveys with 14 sampling locations were used. The multi-objective calibration of the deterministic river water quality model QSIM of the BfG (Germany) was carried out with the nonlinear parameter estimator PEST. The Elbe case study showed that calibration with less than two survey data sets leads to substantial errors if these parameters are applied to deviating boundary conditions. These uncertainties can be reduced with an increased calibration database. The results of this study will help model users to define appropriate data collections and monitoring schemes.

Key words calibration; data information content; QSIM; river water quality model

INTRODUCTION

An important consideration when modelling is the uncertainty in the model results due to inadequate calibration data for model identification as well as due to errors in sampled data and model structure. Uncertain model identification often is caused by small amounts and information content of calibration data. Sources of error include: (a) errors in the data used as initial and boundary conditions in the model; (b) errors in the data used to calibrate and validate the model results; and (c) errors due to the structure of the model which includes the equations, solutions and parameters used for the simulations.

Investigations regarding the information content of data have been conducted mainly to increase the information made available for parameter identification through the calibration process. It has been shown that the use of multiple objectives for single-output models, measuring the models performance during different response modes, can give more detailed information and allows the modeler to link model performance and model components (Boyle *et al*., 2001; Wagener *et al*., 2003). Time series data have been analysed to identify data points or time periods based on their information content. Vrugt *et al*. (2001) showed that the use of different localized subsets of time series data can improve parameter estimation if the range of measurements is sufficient. However, studies investigating the data information content for highly parameterized river water quality models based on complex multi-objective model identification are rare and model validation is often neglected.

The specific objectives of the study were to: (a) assess the identification of a river water quality model using the PEST methodology, and to (b) 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.

MATERIAL AND METHODS

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 by 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 summer 2000 (Schöl *et al*., 2004). These surveys were carried out for the free flowing 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 development of algal biomass and nutrient concentration under a wide range of boundary conditions like low flow conditions (1998), high global radiation and nutrient limiting algal growth conditions (1999). Upper boundary conditions (start values) relevant to primary production and nutrient concentration development, i.e. discharge, temperature and light, for the five measurement surveys are given in Table 1. Up to nine samples were taken for each cross section. On the river reach from Schmilka to Neu Darchau flow times varied according to the discharge between 7 d for the survey in May 1999 and 9 d for the survey in September 1998.

Table 1 Start values of simulation runs, mean values at station Schmilka (3.9 km).

Input variable	Unit	Oct. 1996	Aug. 1997	Sept. 1998	May 1999	July 2000
Discharge	$m^3 s^{-1}$	225	151	127	300	140
Chlorophyll *a*	$\mu g L^{-1}$	8.6	37.2	17	63	45
Share of diatoms	–	0.7	0.6	0.5	0.6	0.6
Ammonium	mg N L^{-1}	0.18	0.12	0.06	0.12	0.10
Nitrate	mg N L^{-1}	4.2	4.2	3.9	4.5	3.7
SRP	mg P L^{-1}	0.16	0.16	0.20	0.05	0.23
Dissolved phosphorus	mg P L^{-1}	0.17	0.16	0.22	0.06	0.25
Silicon	mg Si L^{-1}	5.4	2.9	3.0	2.4	3.2
Oxygen	mg O_2 L^{-1}	9.3	7.9	8.5	11.4	7.3
Suspended particulate matter	mg L^{-1}	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

For the River Elbe case study the widely used river water quality model QSIM (BFG model) was applied (Kirchesch & 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 of 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 and the model distinguishes between two different algal classes. Hydraulic calculations are based on more than 3000 gauged river cross sections. The duration of the longitudinal samplings corresponds well with the computed flow times for all five surveys.

Model calibration

For model calibration the automatic Parameter Estimation Program (PEST) was used, which implements the Gauss–Levenberg–Marquardt method (Doherty, 2004). 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. 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. Since stoichiometric and physical parameters are usually more accurately known or show smaller variations from one system to another than parameters of kinetic processes (Omlin *et al*., 2001) primarily kinetic parameters were used for model calibration. Based on the sensitivity analysis in a first step most sensitive kinetic parameters were included in the parameter estimation process. In a second step, seven parameters were selected for the optimization process: maximum growth rate of diatoms ($k_{gro,dia}$), minimum respiration rate of diatoms ($k_{min,res,dia}$), half-saturation coefficient of diatoms growth with respect to phosphorus ($K_{HPO4,dia}$), half-saturation coefficient of diatoms growth with respect to nitrogen ($K_{NO3,dia}$), minimum respiration rate of chlorophyceae ($k_{min,res,chlo}$), half-saturation coefficient of diatoms growth with respect to silicon ($K_{Si,dia}$) and half-saturation coefficient of chlorophyceae growth with respect to phosphorus ($K_{HPO4,chlo}$). These parameters were calibrated simultaneously according to the variables Chl *a*, NO_3, O_2, dissolved phosphorus (DP) and dissolved Si. Note that due to dependencies between model parameters the seven most sensitive parameters were not used for the automatic calibration process. All other kinetic parameters had only minor impact on the selected output variables.

In order to assess model performance more precisely it is useful to consider two or more objective criteria. The objective criteria used in this study were the Nash-Sutcliffe coefficient *E* and the index of agreement *d* (Willmott, 1981) defined as:

$$E = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2} \tag{1}$$

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

where $\overline{O}$ is the mean of the observed value, O_i is the observed and P_i is the simulated value. E measures the fraction of the variance of the observed water quality variables explained by the model in terms of the relative magnitude of the residual variance to the variance of the water quality variables; the optimal value is 1.0 and values should be larger than 0.0 to indicate minimal acceptable performance. 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, owing to the squared differences (Legates & McCabe, 1999). The objective criteria were used both for assessing the calibration and validation of the model.

Evaluation of data information content

The investigation of the data information content for the identification of the river water quality model calibration runs were conducted using data of 1, 2, 3, 4 and 5 data sets and all their possible combinations. This is a total of 30 data sets with differing number of flow time related longitudinal data sets and differing water quality conditions. For each set, a single multi objective optimization using the PEST programme was conducted, where each optimization leads on an average of about 180 model runs depending on the amount of calibration data. Within the optimization process all calibration runs were conducted with the same initial parameter values and upper and lower parameter bounds. With respect to the five selected output variables, for all calibration runs the same seven kinetic model parameters were adjusted. A common technique of validating a model is using only a subset of all available data for calibration and testing 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.

RESULTS

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). The simulated and observed values of Chl *a*, which represents algal biomass, are shown as an example in Fig. 1. The results indicate that the model is able to simulate algal growth for a wide range of boundary conditions with the identified model parameters. In the case of very high Chl *a* concentration model residuals tend 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 with a mean value of –0.14 (Figure 2). *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 River Elbe. Hence, we got confidence that the PEST multi objective calibration procedure is able to identify optima for all seven parameters used for model calibration. Due to the selected "hill-climbing" parameter estimation method previous knowledge about the parameter values and their lower and upper bounds support the calibration process.

Empirical cumulative distribution function (cdf) of the calibration runs for the *d* and the *E* statistics of different numbers of data sets was constructed (Fig. 3). Each cdf indicates the chance of obtaining a statistic of magnitude if a data set of that number of longitudinal surveys is selected at random and used for calibration. The *d* and *E* cdfs shift to the right as we increase the number of longitudinal survey data sets used

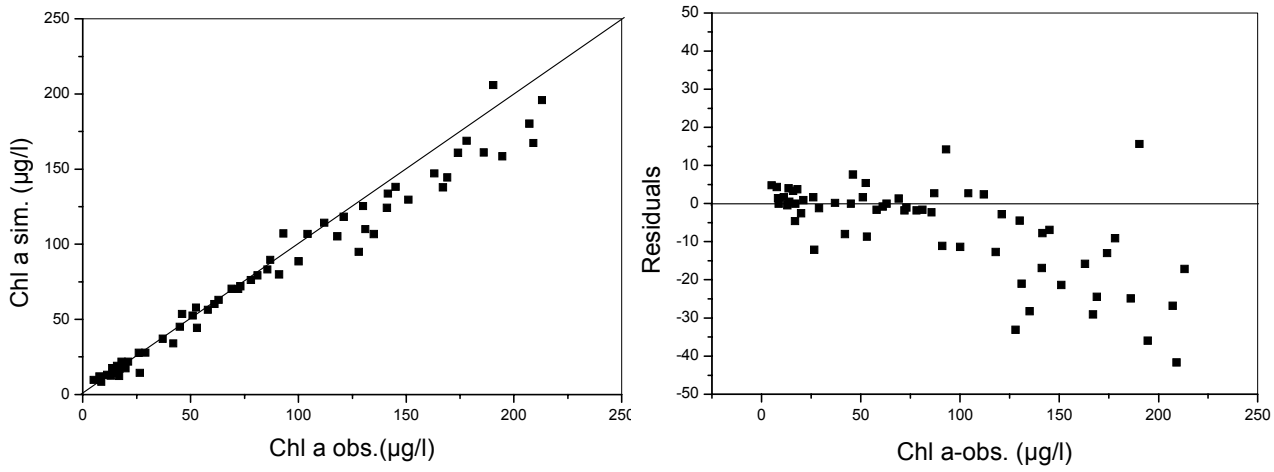

Fig. 1 Observed and simulated Chl *a* concentrations based on five calibration data sets.

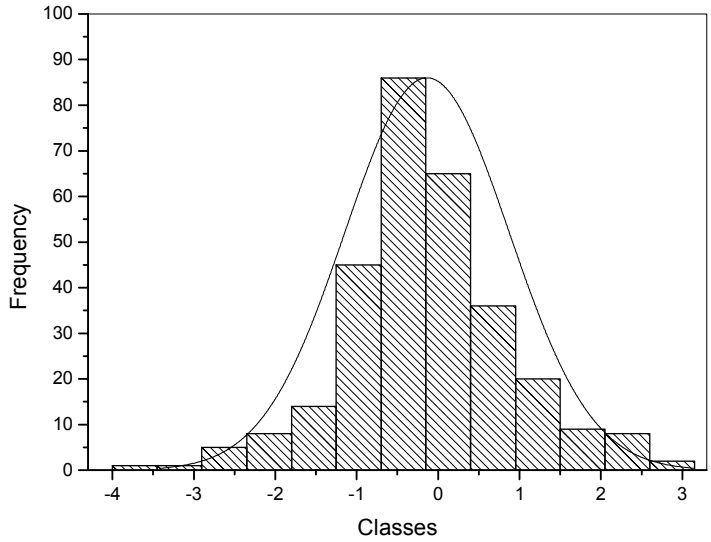

Fig. 2 *Z*-distribution of normalized residuals for the calibration run with five data sets.

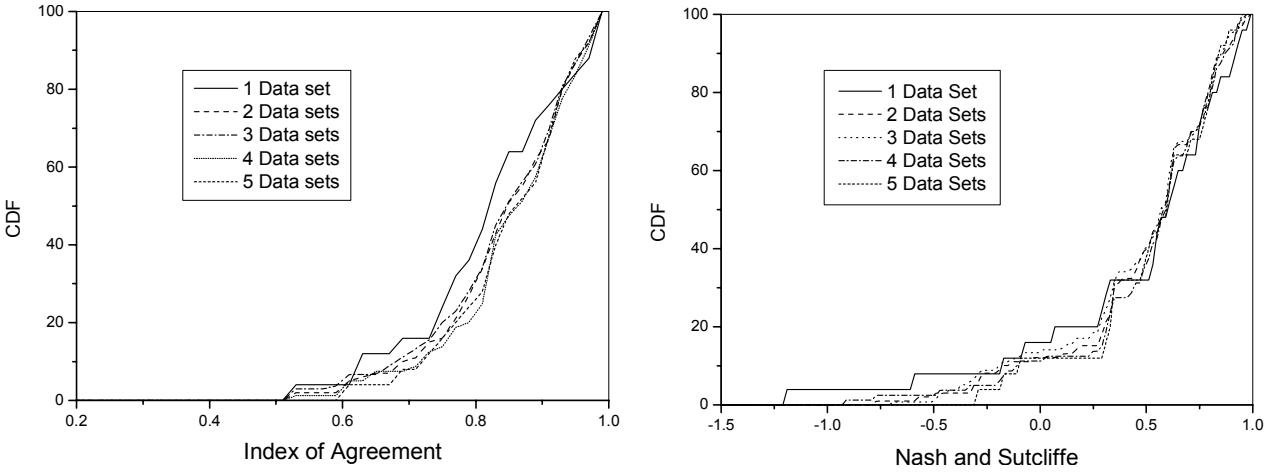

Fig. 3 Empirical cumulative distribution functions (cdf) of the *d* and *E* statistics for different numbers of data sets for the calibration runs.

for calibration. 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.

The *E* values show for all investigated combinations considerable lower levels to the *d* values. Increasing the number of calibration data sets leads to a shift of the cumulative distribution function to the right only for very low *E* values, reflecting the high sensitivity of this performance criterion. The *E* cdfs show a similar behaviour than the *d* objective criteria. The mean value of *E* is with 0.45 lowest using one data set for model calibration and the standard deviation shows the maximum decrease form 0.51 to 0.37 using two calibration sets instead of one (see Table 2). No substantial improvement could be achieved when more than two calibration runs were used.

Table 2 Mean and standard deviation of *E* and *d* objective criteria for calibration and corresponding validation data sets.

Number of data sets	Number of objective criteria	Nash and Sutcliffe		Index of Agreement	
		Mean	SD	Mean	SD
Calibration					
1	25	0.45	0.51	0.85	0.12
2	100	0.49	0.37	0.85	0.11
3	150	0.47	0.38	0.84	0.11
4	80	0.51	0.38	0.85	0.10
5	25	0.52	0.33	0.84	0.11
Validation					
4	100	0.29	0.65	0.80	0.14
3	150	0.44	0.41	0.83	0.12
2	100	0.47	0.39	0.84	0.11
1	20	0.50	0.30	0.83	0.11

SD, Standard deviation.

Furthermore, the means and standard deviations show that *d* is much less sensitive compared to the *E* objective criteria. To test whether the samples from the different numbers of data sets come from the same distribution we used the Kolmogorov-Smirnov two-sample test. The K-S statistic uses the maximum vertical differences between two empirical cumulative distribution functions as a test criterion. This absolute difference was always highest when the *d* and *E* cdfs for one data set were related to the cdfs of more than one data set compared to other possible combinations. Although these findings are supported by the visual inspection of the cdfs the K-S statistic, which was calculated for comparison of all data set results, does not allow distinguishing the cdfs at the significant level of 0.10.

The *d* cdfs for validation steepen progressively with increasing number of calibration data sets and hence decreasing number of validation data sets (Fig. 4). Increasing the steepness indicates reducing sensitivity of model performance to selection of data sets. The validation cdfs show a decrease of the range of *d* and *E* values when increasing calibration data sets and hence decreasing corresponding validation data sets. The overall performance of the validation improves only significantly when increasing the number of calibration data sets from one to two data sets. These findings are restricted to the specific conditions of the River Elbe during the growing season with its high algal biomass concentrations of up to 200 µg L^{-1} Chl *a* and hence significant dependencies between algal biomass and nutrient concentrations. In the case of more complex boundary conditions, e.g. high inputs from sewage plants, higher abundance of benthic filterers which could reduce the level of Chl *a* and less significant relationships between water quality constituents more than two longitudinal surveys may be needed for reasonable model identification.

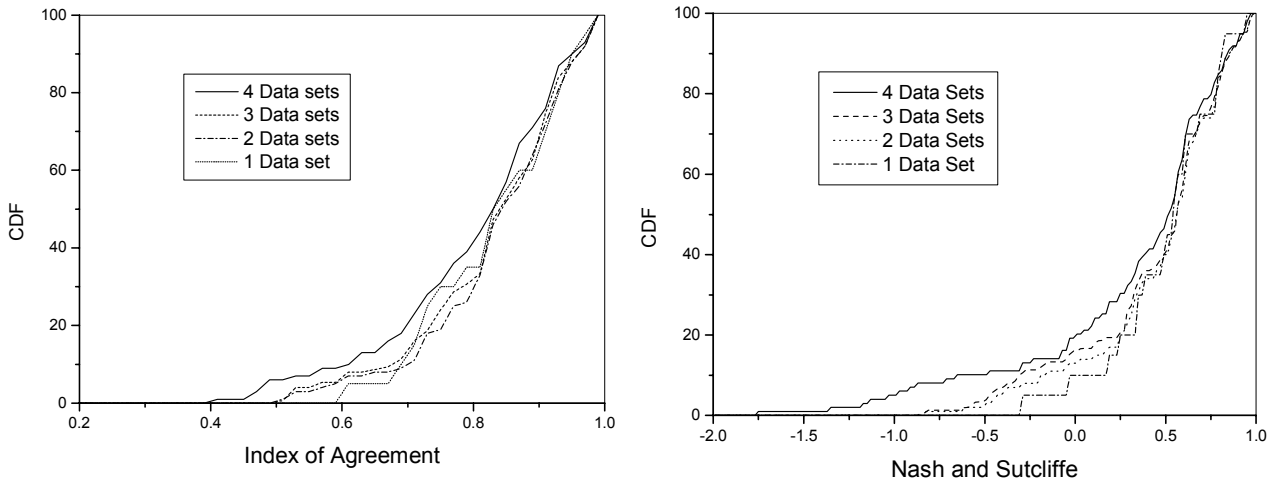

Fig. 4 Empirical cumulative distribution functions (cdf) of the *d* and *E* statistics for different numbers of data sets for the validation runs.

CONCLUSION

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. For the specific conditions of the River Elbe, at least two longitudinal data sets of different boundary conditions during the growing season should be used for model calibration and the benefit of the use of more than two data sets may be marginal. Since the River Elbe represents relatively strong and significant relationships between algal growth and nutrient concentrations these findings represent minimum requirements of data sets of flow-proportional longitudinal river water quality surveys. In most cases with less significant relationships between water constituents more than two data sets of flow proportional longitudinal surveys of the growing season will be needed for reasonable river water quality model identification. The introduced cross validation procedure is a good opportunity to identify data needs for model calibration in the case of limited high quality data, which is a common situation in river water quality studies.

REFERENCES

Boyle, D. P., Gupta, H.V. & Sorooshian, S. (2000) Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods. *Water Resour. Res.* **36,** 3663–3674.

Doherty, J. (2004) *PEST—Model independent Parameter Estimation—User Manual*, 5th edn. Watermark Numerical Computing (http://www.sspa.com/pest/).

Guhr, H., Spott, D., Bormki, G., Baborowski, M. & Karrasch, B. (2003) The effects of nutrient concentrations in the River Elbe. *Acta Hydrochim. Hydrobiol.* **31**(4–5), 282–296.

Legates, D. R. & McCabe, G. J. (1999) Evaluating the use of "goodness-of-fit" measure in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241.

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

Schöl, A., Eidner, R., Böhme M., Kirchesch, V., Wörner, U. & Müller, D. (2004) Bedeutung der Stillwasserzonen für die Nährstofflimitierung in der Elbe. *Federal Institute of Hydrology, Report 1406.*

Schöl, A., Kirchesch, V., Bergfeld, T. & 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.

Vrugt, J. A., Bouten, W. & 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. & Gupta, H.V. (2003) Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrol. Processes* **17**, 455-476.

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

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall–runoff models: sensitivity to calibration data. *J. Hydrol.* **181**, 23–48.