

On the use of Pareto optimization for multi-criteria calibration of hydrological models

HENRIK MADSEN¹ & SOON-THIAM KHU²

¹ DHI Water & Environment, Agern Allé 5, DK-2970 Hørsholm, Denmark
hem@dhi.dk

² Centre for Water Systems, School of Engineering, Computer Science and Mathematics,
University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK

Abstract The calibration of distributed and integrated hydrological models is formulated as a general multi-objective optimization problem that utilises multiple sources of information in a consistent way. This formulation allows a comprehensive evaluation of the trade-off between different calibration objectives and an elaborate comparison of different model structures. It is a natural solution and extension to the weighted nonlinear least squares regression problem, and it allows the modeller or decision-maker to choose (or weight) solutions at a later stage according to the specific model application being considered. The proposed framework is illustrated for calibration of two MIKE SHE models of different complexity for the same catchment.

Keywords calibration; MIKE SHE; multiple objectives; optimization problem; Pareto solutions

INTRODUCTION

In the calibration of distributed and integrated hydrological models it is generally recognized that multiple sources of information should be applied. Multi-site and multi-variable calibration and validation should be performed if distributed predictions are needed for different state variables (Refsgaard, 1997). It should also be noted that model parameters are usually better determined when new types of field data are used for the calibration rather than adding more data of the same variable (e.g. McLaughlin & Townley, 1996). This calls for formulation of the calibration problem using a general multi-objective framework. Such a framework has been widely applied for calibration of lumped, conceptual rainfall-runoff models (e.g. Gupta *et al.*, 1998; Madsen, 2000; Khu & Madsen, 2005) but has only recently been adopted in distributed hydrological modelling (Madsen, 2003).

In groundwater modelling, the parameter estimation problem has traditionally been solved using weighted or generalized nonlinear least squares regression which allows weighting of individual measurements (of the same or different state variables). The statistical theory of this approach requires the weighting to be proportional to the inverse of the covariance matrix of the “true” observation errors. This method, however, has severe theoretical and practical limitations. The observation error represents in a lumped way all the different error sources involved in the modelling. The model prediction depends, in general, on these errors in complex and nonlinear ways, hence violating the assumption of additive errors. Furthermore, it is in practice impossible to identify and quantify all the different error sources.

The multi-objective calibration framework presented in this paper offers a more general procedure for balancing different types of information in the model calibration. This approach is more pragmatic in the sense that it does not rely on statistical theory, although weighting according to observation errors can be included as part of the framework.

MULTI-OBJECTIVE CALIBRATION PROBLEM

In a multi-objective context model calibration can, in general, be performed on the basis of: (a) multi-variable measurements, such as groundwater levels, river flows, and concentration measurements; (b) multi-site measurements consisting of several measurement sites distributed within the modelling domain; and (c) multi-response modes, i.e. calibration criteria that measure various responses of the hydrological processes such as the general water balance, peak flows, and low flows. Mathematically, the calibration problem can be stated as:

$$\text{Min}\{F_1(\theta), F_2(\theta), \dots, F_m(\theta)\}, \quad \theta \in \Theta \quad (1)$$

where θ are the model parameters to be estimated, $F_i(\theta)$, $i = 1, 2, \dots, m$ are the different objective functions, and Θ is the feasible parameter space. The objective functions in equation (1) can be any statistic that measures the fit between observations and corresponding simulated values. When calculating these measures, weights can be assigned to individual measurements to account for measurement uncertainties and correlations between the measurements. Equation (1) can also include penalty terms such as prior information about the parameters (Mertens *et al.*, 2004).

The solution to the calibration problem will not be a single unique set of parameters but will consist of the Pareto set of solutions (non-dominated solutions), according to various trade-offs between the different calibration objectives. For a Pareto solution none of the objective functions can be further minimized without an increase of some of the other objective functions, thus forming the trade-off. Each of the Pareto solutions is, from a multi-objective point of view, equally good. This multi-objective equivalence of parameter sets should not be confused with the concept of parameter equifinality (Beven & Binley, 1992) which refers to the multitude of equally good parameter sets as measured according to a single likelihood or objective function.

An essential component of solving the multi-objective calibration problem in equation (1) is how to compare parameter sets when there are two or more calibration objectives. Available techniques can be grouped into: (a) aggregation approaches, and (b) Pareto domination approaches. In the aggregation approach individual weights are assigned to the objective functions (or transformations thereof) and aggregated into one measure that is used in the optimization (e.g. Madsen, 2000, 2003). By performing several individual optimization runs with different weight combinations, the entire Pareto surface can be explored. The Pareto domination approach does not rely on a single comparative measure but on whether one solution is dominated by another for the different calibration objectives considered. In the optimization, a Pareto rank is assigned to each parameter set according to the domination criterion (e.g. Goldberg, 1989).

Numerical optimization procedures that are especially suited for multi-objective optimization are the population based procedures such as genetic algorithms (GA) (e.g. Deb *et al.*, 2000), the shuffled complex evolution (SCE) algorithm (Duan *et al.*, 1992), and multi-objective extensions of SCE (Vrugt *et al.*, 2003). These procedures evolve a population of parameter sets rather than a single parameter set as in traditional gradient based search procedures.

Model calibration based on Pareto optimization is a powerful method that has several important advantages:

- It allows a comprehensive evaluation of the trade-offs between different calibration objectives and hence highlights possible model structural deficiencies. In the case of a perfect model and ideal measurements, the Pareto surface will collapse to one single point in objective function space with one best parameter set. Departures from this ideal solution may indicate the presence of model structural errors.
- It offers an elaborate framework for comparison of different models or model conceptualizations by considering several performance criteria in a consistent manner. For instance, one model structure may be better in simulating groundwater heads than river runoff as compared to another model structure. Such differences will not be highlighted if only one weight combination of groundwater and river runoff observations is considered.
- It allows choosing single Pareto-optimal solutions according to the specific model application being considered. The modeller or decision-maker avoids having to specify preferences to any of the calibration objectives at the calibration stage but can choose a preferred solution among the Pareto optimal solutions at a later stage. When several objective functions are included, a huge number of Pareto optimal solutions exist, and hence the calibration problem quickly becomes a decision-making problem. Khu & Madsen (2005) introduced a Pareto preference-ordering scheme that chooses preferred solutions according to a stronger dominance criterion than Pareto dominance. In this case, solutions that are also Pareto optimal in the different sub-space combinations of the objective functions are preferred.
- It provides a better discrimination between model structure and parameter sets and hence a better posed optimization problem by “unfolding” the equifinality problem.
- It provides a natural solution and extension to the weighting problem in weighted and generalized nonlinear least squares regression.

EXAMPLE APPLICATION

The multi-objective calibration framework was applied for calibration of a MIKE SHE model set-up of the Danish Karup catchment. The Karup catchment has an area of 440 km² and is drained by the Karup River and about 20 tributaries. The topography varies from about 20 to 100 m. The geology is relatively homogeneous with highly permeable sand and gravel deposits and small lenses of moraine clay. The aquifer is mainly unconfined and varies in thickness from about 10 m at the western and central

part to more than 90 m at the upstream eastern water divide. The depth of the unsaturated zone varies from 25 m at the eastern water divide to less than 1 m in the wetland areas along the river. The land use consists of agriculture (67%), forest (18%), heath (10%), and wetland areas (5%).

Two different models are considered: (a) a groundwater model with prescribed recharge, and (b) a fully integrated model that includes evapotranspiration processes and unsaturated flow using Richards' equation. The two models have the same parameterization of the saturated zone, the river and the drainage system. The geological conceptualization is taken from the Danish National Water Resources model (DK-model) which is interpreted in 10-m thick layers using a grid size of 1×1 km. For each grid element, a soil type is assigned with a given code and hydrogeological parameters. Six general soil units are defined, see Table 1. The hydraulic conductivities of these units are subject to calibration.

Table 1 Soil types used in the conceptualization of the saturated zone.

Soil code	Soil name	Description
1	Melt water sand	Quaternary and Post-Glacial sand and gravel
2	Clay	Glacial, Inter-Glacial and Post-Glacial clay and silt
3	Quartz sand	Miocene, medium to coarse grained sand and gravel
4	Mica sand	Miocene, fine to medium grained sand
5	Mica clay/silt	Pre-Quaternary clay and silt
6	Limestone	Limestone

The Karup River and the main tributaries are included in the model. A thin permeable layer is assumed between the river and the main aquifer. The leakage coefficient characterizing this layer is subject to calibration. The wetland areas are drained by ditches and drain pipes. Drainage is modelled conceptually using a linear reservoir description in each cell. The drainage level (relative to ground surface) and the time constant of the linear reservoir model are assumed to be homogeneous in the catchment and are subject to calibration.

For the fully integrated model two soil profiles are specified for the unsaturated zone model. A soil profile (*General*) used for the main part of the catchment and a soil profile (*Heath*) for the heath areas. The *General* soil profile comprises loamy sand to a depth of 100 cm and fine sand below. The *Heath* soil profile consists of fine sand to a depth of 55 cm and coarse sand below. For each of the four soil types, van Genuchten retention and conductivity curve parameters are subject to calibration.

In the calibration, groundwater level data from 17 wells sampled every two weeks and daily runoff data from the catchment outlet are used. Based on these data two objective functions are defined: (a) the average root mean squared error (RMSE) of the groundwater levels, and (b) the RMSE of the runoff at the catchment outlet. Data in the period 1 January 1971–31 December 1974 were used for calculation of the objective functions. To minimize the effect from the initial conditions, a 2-year warm-up period was applied in the simulations.

A preliminary sensitivity analysis was adopted to identify the most sensitive parameters to be included in the optimization. For the groundwater model, the hydraulic

conductivity of soil unit 1 (melt water sand) and unit 3 (quartz sand), the leakage coefficient, the drainage level, and the drainage time constant, were selected for calibration. A constant relation of 0.1 between vertical and horizontal hydraulic conductivity was assumed. Thus, five parameters were included in the calibration in total. Additional sensitive parameters for the integrated model were the saturated hydraulic conductivity and the inverse of the air-entry value in the van Genuchten retention curve of the two soil types in the *General* soil profile (loamy sand and fine sand). Thus, for the integrated model, nine parameters were included in the calibration. For the optimization a population based algorithm, the Population Simplex Evolution (PSE) method implemented in the AUTOCAL software (DHI, 2005) was applied. For both models a population size of 50 was used and in total 2000 model evaluations were performed.

The results of the Pareto optimization for the two models are shown in Fig. 1. For both models a trade-off between groundwater levels and runoff simulation is observed. The optimum solution with respect to groundwater levels provides a poor simulation of the runoff, and *vice versa*. For the integrated model the Pareto front has a sharp structure, whereas the trade-off for the groundwater model has a gentler slope. The sharp structure for the integrated model implies that by relaxing only slightly on the runoff performance, the groundwater head simulations can be significantly improved.

The estimated Pareto fronts allow an elaborate comparison of the performance of the two models. One would, in general, expect that the integrated model would perform better than the groundwater model due to the more advanced modelling of the recharge processes. However, the integrated model is not strictly better than the groundwater model. When focusing on the runoff simulation, the integrated model has a significantly better performance than the groundwater model, whereas for groundwater level simulation the groundwater model has better performance. This may indicate problems with the conceptualization of the evapotranspiration and unsaturated zone model or that other parameters not included in the optimization are important.

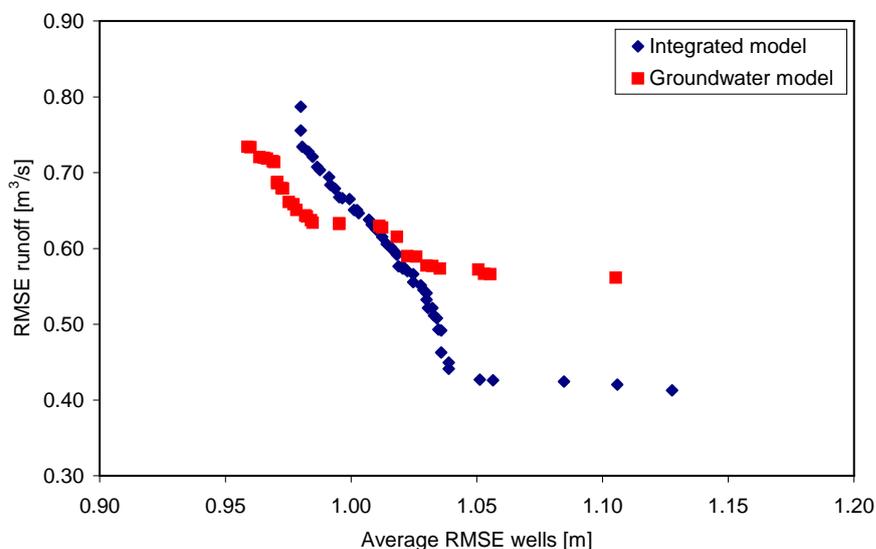


Fig. 1 Estimated Pareto fronts of the two models.

The modeller or decision maker can subsequently use the results of the Pareto optimization to choose preferred solutions. This allows emphasis to be put on certain modelling aspects (runoff or groundwater level simulations) or use of other information not directly included in the calibration. If the Pareto front is sharp, the break point on the front is often preferred as a good compromise solution (Madsen, 2003). However, in general, the final choice of preferred solutions should depend on the actual model application being considered. Preferred solutions may also be determined prior to the calibration by specifying appropriate weights to the different objective functions. However, due to differences in units and magnitudes of the different observations, and differences in variability of the objective functions in the feasible parameter space, appropriate weighting is very difficult to determine prior to the calibration. Estimation of the entire Pareto surface allows consistent weighting of objectives at a later stage.

CONCLUSIONS

Calibration of distributed and integrated hydrological models has been formulated as a general multi-objective optimization problem that allows consistent use of different types of information. Pareto optimization provides an elaborate framework for evaluating trade-offs between the different objectives and to compare different model structures. It is a natural solution and extension to the problem of weighting observations from different information sources in weighted nonlinear least squares regression that is traditionally applied in groundwater modelling.

An example application has been presented that illustrates the use of the proposed calibration framework. Two MIKE SHE models of different complexity for the same catchment have been calibrated using groundwater level and runoff data. For both models a trade-off between groundwater level and runoff simulations existed. The more complex integrated model is not strictly better than the simpler groundwater model, suggesting possible model structural errors related to the conceptualization of the recharge processes, or the existence of important parameters not included in the optimization. Based on the estimated Pareto front, the modeller or decision-maker can choose preferred solutions that tailor the model calibration to the specific objectives of the model application being considered.

REFERENCES

- Beven, K. J. & Binley, A. M. (1992) The future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Processes* **6**, 279–298.
- Deb, K., Agrawal, S., Pratap, A. & Meyarivan, T. (2000) A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: *Proc. of Parallel Problem Solving from Nature VI Conf.*, 849–858. Springer, Berlin, Germany.
- DHI (2005) AUTOCAL, Auto Calibration Tool, User Guide. DHI Software 2005, DHI Water & Environment, Denmark.
- Duan, Q., Sorooshian, S. & Gupta, V. (1992) Effective and efficient global optimization for conceptual rainfall–runoff models. *Water Resour. Res.* **28**(4), 1015–1031.
- Goldberg, D. E. (1989) *Genetic Algorithms in Search, Optimisation and Machine Learning*. Addison-Wesley, Reading, Massachusetts, USA.
- Gupta, H. V., Sorooshian, S. & Yapo, P. O. (1998) Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. *Water Resour. Res.* **34**(4), 751–763.

- Khu, S. T. & Madsen, H. (2005) Multi-objective calibration with Pareto preference ordering: an application to rainfall-runoff model calibration. *Water Resour. Res.* **41**, 10.1029/2004WR003041.
- Madsen, H. (2000) Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* **235**, 276–288.
- Madsen, H. (2003) Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour.* **26**, 205–216.
- McLaughlin, D. & Townley, L. R. (1996) A reassessment of the groundwater inverse problem. *Water Resour. Res.* **32**(5), 1131–1161.
- Mertens, J., Madsen, H., Feyen, L., Jaques, D. & Feyen, J. (2004) Including prior information in the estimation of effective soil parameters in unsaturated zone modelling. *J. Hydrol.* **294**, 251–269.
- Refsgaard, J. C. (1997) Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.* **198**, 69–97.
- Vrugt, J. A., Gupta, H. V., Bastidas, L., Bouten, W. & Sorooshian, S. (2003) Effective and efficient algorithm for multiobjective optimization of hydrologic models. *Water Resour. Res.* **39**(8), 10.1029/2002WR001746.