Automatic calibration of the MIKE SHE integrated hydrological modelling system

Henrik Madsen and Torsten Jacobsen
DHI Water & Environment

Abstract

In this paper, automatic calibration of an integrated hydrological modelling system, MIKE SHE, is considered. A consistent framework for automatic calibration is formulated which focuses on the different steps in the estimation process from model parameterisation and selection of calibration parameters, formulation of calibration criteria, and choice of optimisation algorithm. The calibration problem is formulated in a general multi-objective context in which different objective functions that measure individual process descriptions can be optimised simultaneously. A test example is presented that illustrates the use of the automatic calibration scheme. Different sensitivity tests are performed to demonstrate the trade-off between the calibration objectives with respect to the use of different calibration data and objective function measures.

Introduction

Traditionally, calibration of hydrological catchment models has been performed manually using a trail-and-error parameter adjustment procedure. The process of manual calibration, however, may be a very tedious and time consuming task, depending on the number of free model parameters and the degree of parameter interaction. Due to the subjectivity involved, it is difficult to explicitly assess the confidence of the model simulations. Consequently, a great deal of research has been directed to development of more efficient and more objective automatic calibration procedures.

In recent years, application of automatic calibration routines in hydrological modelling has advanced considerably. The routines, however, has evolved in various directions in different application areas. For parameter estimation in groundwater modelling, gradient-based local search techniques have mainly been applied (e.g. McLaughlin and Townley, 1996). In lumped, conceptual hydrological models, population-evolution-based global optimisation methods, such as the shuffled complex evolution algorithm (Duan et al., 1992), have shown to be more efficient. Application of automatic calibration in complex, integrated and distributed hydrological catchment models is an ongoing research area with very limited experience.

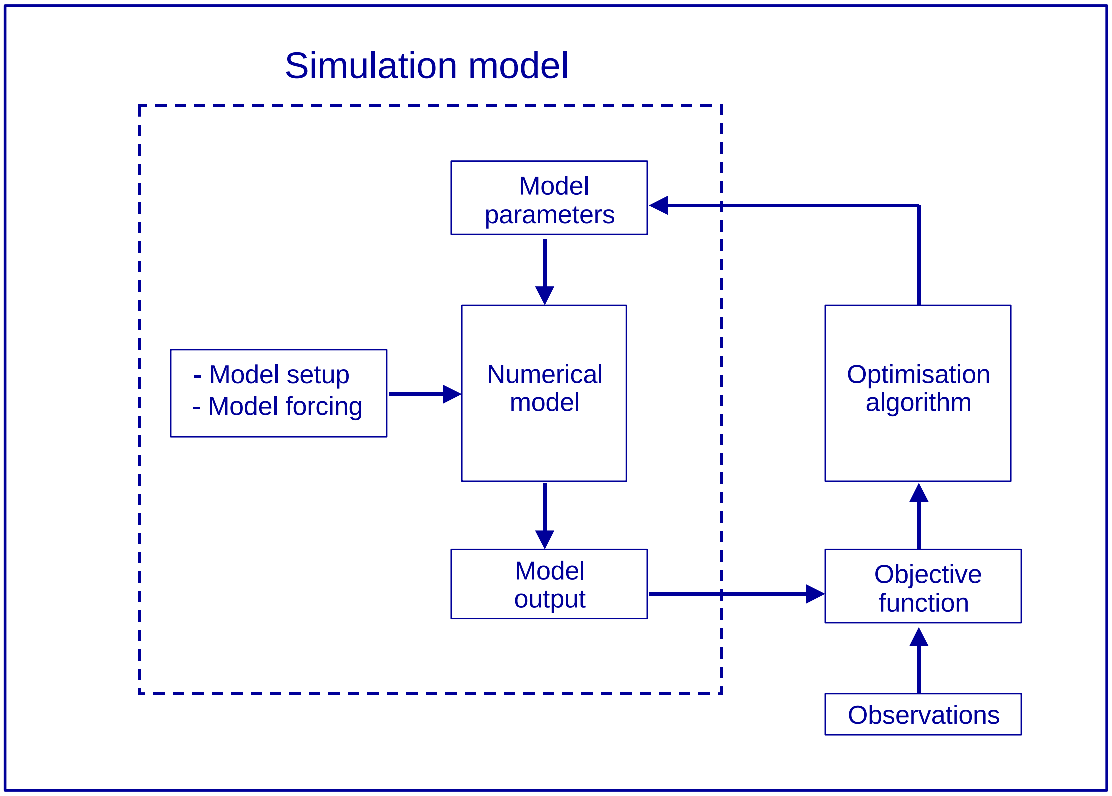

Figure 1 Outline of automatic calibration scheme, interfacing an optimisation algorithm with the simulation model.

In this paper, automatic calibration of the MIKE SHE integrated hydrological modelling system is considered. A framework for automatic calibration is presented which focuses on the different steps in the estimation process, including model parameterisation and choice of calibration parameters, specification of calibration criteria, and choice of optimisation algorithm. A test example is presented that illustrates the use of the proposed framework.

Formulation of calibration framework

The process of model calibration is illustrated in Fig. 1. In automatic calibration, parameters are adjusted automatically according to a specific search scheme for optimisation of certain calibration criteria (objective functions) that measure the goodness-of-fit of the simulation model. The process is repeated until a specified stopping criterion is satisfied, e.g. maximum number of model evaluations, convergence of the objective function, or convergence of the parameter set.

Formulation of a proper framework for automatic calibration involves the following key elements:

- Model parameterisation and choice of calibration parameters
- Specification of calibration criteria
- Choice of optimisation algorithm

Model parameterisation and choice of calibration parameters

A distributed hydrological modelling system such as MIKE SHE potentially involves a large number of model parameters to be specified by the user during the model setup. Some of these parameters may be assessed from field data, e.g. geological descriptions from well-logs, pumping test analysis, maps of soil profiles, soil analysis (texture, density, retention curves), and vegetation maps. Comprehensive field data, however, are seldomly available to fully support specification of all model parameters. In addition, some model parameters are of a more conceptual nature and cannot be directly assessed from field data.

In the model parameterisation, the available field data should be used to define the spatial patterns of the parameter values to describe the most significant variations. This is often done by defining a conceptual model with appropriate parameter classes of geological units, soil types, vegetation types etc. For each class, some parameters are then assessed directly from field data while other parameters may be subject to calibration. The challenge is to formulate a relatively simple model parameterisation in order to provide a well-posed calibration problem but at the same time keep it sufficiently complex in order to capture the spatial variability of key model parameters. The importance of a rigorous model parameterisation for calibration of distributed hydrological models was emphasised by Refsgaard (1997). This aspect becomes even more important when automatic procedures are applied for parameter estimation.

Sensitivity analysis can be conducted to investigate which parameters can be considered to be well determined (sensitive) and which are poorly determined (insensitive) with respect to the available observations. In Hill (1998) dimensionless, scaled sensitivities are used which measure the change in simulated values with respect to each of the parameters. Spear and Hornberger (1980) introduced a generalised sensitivity analysis procedure based on Monte Carlo sampling where a number of randomly generated parameter sets is evaluated and compared. Sensitivity analysis can be used in the initial model parameterisation process to decide which parameters are insensitive and can be set to fixed values. The results of such an analysis, however, should be carefully interpreted. The dimensionless, scaled sensitivities in Hill (1998) depends on the parameter values, and hence sensitivity statistics evaluated at some initial parameter values may be very different from the statistics obtained using other parameter sets. In addition, sensitivity statistics do not properly account for parameter correlations, implying that parameters that seem to be insensitive may have important correlations with other parameters that are essential for the model behaviour (Madsen, 2000b)

It should be noted that model parameterisation and model calibration is an iterative process. If the calibration results in poorly defined parameter values, one should reconsider the model parameterisation and define a simpler conceptual model that includes fewer calibration parameters. On the other hand, if the model is not able to sufficiently describe the spatial variability reflected in the observations, one should consider distributing key model parameters or including other process descriptions in the calibration.

Specification of calibration criteria

The automatic calibration scheme involves optimisation of numerical measures (objective functions) that compare observations of the state of the system with corresponding simulated values. For hydrological model calibration groundwater level data and river runoff or water level data are usually available. In some cases also observations of the water content in the unsaturated zone can be used. In this respect, it is important to note that for a proper evaluation of the validity of a distributed catchment model, distributed data rather than just catchment-integrated values such as river runoff are necessary for calibration. In addition, parameters are usually better determined (more sensitive) when new types of field data are used for calibration rather than adding more data of the same variable.

Now, denote by $F(\theta)$ an objective function that measures the goodness-of-fit of the simulated model with respect to the parameter set θ. The optimal parameter set θ_{opt} is found by solving the optimisation problem

$$\theta_{opt} = Min\{F(\theta)\} \qquad , \quad \theta \in \Theta \tag{1}$$

In this case the optimisation problem is constrained in the sense that θ is restricted to the feasible parameter space Θ. The parameter space is usually defined as a hypercube by specifying lower and upper limits on each parameter. These limits are chosen according to physical and mathematical constraints, information about physical characteristics of the system, and from modelling experiences. The feasible parameter space can also be defined as a hyperellipsoid by using prior knowledge about the correlation between the different parameters (Kuczera, 1997).

Automatic calibration can also be defined as an unconstrained optimisation problem. In this case, prior information about the parameters can be used by adding a penalty term in the objective function that measures the departure of the parameters from their prior estimates (e.g. Hill, 1998).

The most commonly used objective function adopted in automatic optimisation is the sum of squared errors between the observed and simulated model response. Calibration based on a single performance measure, however, is often inadequate to properly measure the simulation of all the important characteristics of the system that are reflected in the observations. Recently, automatic procedures have been developed that allow a simultaneous optimisation of a number of different calibration objectives (Gupta et al., 1998; Madsen, 2000a).

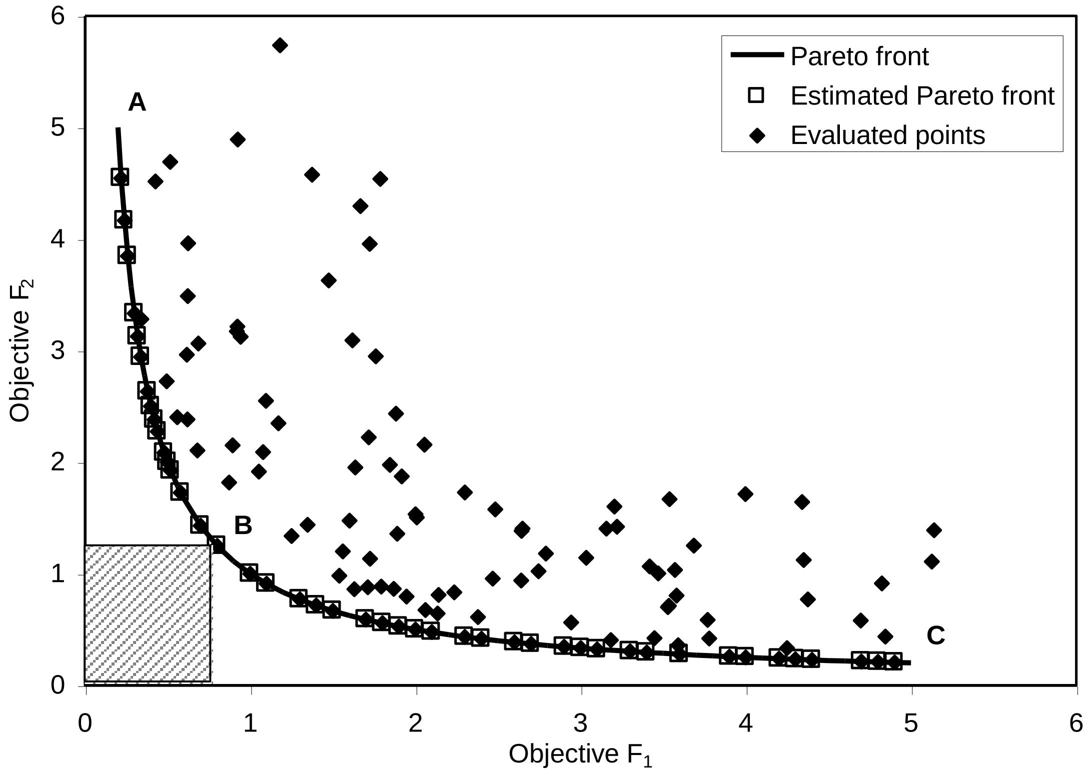

Figure 2 Definition of Pareto front for optimisation of two objectives.

In a multi-objective context, model calibration can, in general, be performed on the basis of:

- Multi-variable measurements, i.e. groundwater level, river runoff and other types of measurements.
- Multi-site measurements, i.e. several groundwater level and runoff measurement sites distributed within the cactchment.
- Multi-response modes, i.e. objective functions that measure various responses of the hydrological processes such as e.g. the general water balance, peak flows, and low flows.

When using multiple objectives, the calibration problem can be stated as follows:

$$Min\{F_1(\theta), F_2(\theta), ..., F_m(\theta)\} \quad , \quad \theta \in \Theta \tag{2}$$

where $F_i(\theta)$, i = 1, 2,.., m are the different objective functions. The solution of Eq. (2) will not, in general, be a single unique set of parameters but will consist of the so-called Pareto set of solutions (non-dominated solutions), according to various trade-offs between the different objectives. The definition of the Pareto front is illustrated in Fig. 2 in the simple case of two objectives F_1 and F_2. Points on the Pareto front have the characteristics that no other points have both a smaller value of F_1 and a smaller value of F_2 (illustrated for point

B in Fig. 2 where no points exist in the hatched rectangle). When moving along the Pareto front from A to C results in successively smaller values of F_2 at the expense of larger values of F_1.

When solving the multi-objective calibration problem, the problem is usually transformed into a single-objective optimisation problem by defining a scalar that aggregates the various objective functions. One such aggregate measure is the weighted average

$$F_{agg}(\theta) = \sum_{i=1}^{m} w_i F_i(\theta) \qquad , \quad \sum_{i=1}^{m} w_i = 1 \tag{3}$$

where w_i are weights assigned to the different objectives. For investigating the entire Pareto front, the aggregated measure can be adopted by performing several optimisation runs using different values of w_i.

In practical applications, the entire Pareto set may be computationally too expensive to calculate, and one is only interested in part of the Pareto optimal solutions. In this case, the user can specify the weights to reflect the relative priorities given to certain objectives, depending on the specific model application being considered. Furthermore, the weights should also reflect the measurement errors, i.e. smaller weights are given to measurements with larger errors. The selection of weights, however, is not straightforward, since the priority also depends on the value of F_i itself. For instance, if all w_i are equal, implicitly larger weights are given to objectives with larger F-values. A proper scaling of the objective functions in the aggregated measure can be defined as

$$F_{agg}(\theta) = \sum_{i=1}^{m} w_i g_i\left(F_i(\theta)\right) \qquad , \quad \sum_{i=1}^{m} w_i = 1 \tag{4}$$

where $g_i(.)$ are functions that transform the different objectives to a common scale. When using a population-based optimisation algorithm, as the one considered herein, an initial population within the feasible region is evaluated. From this initial population, the transformation functions can be evaluated. Madsen (2000a) used an Euclidian distance function in which all the objective functions are transformed to having about the same distance to the origin near the optimum. Van Griensven and Bauwens (2001) adopted a probability distribution function for F_i for transformation of the objective functions into a probability scale.

Choice of optimisation algorithm

Optimisation algorithms can, in general, be categorised as "local" and "global" search methods (Sorooshian and Gupta, 1995). Depending on the hill climbing strategy employed, local search algorithms may be further divided into "direct" and "gradient-based" methods. Direct search methods use only information on the objective function value, whereas gradient-based methods also use information about the gradient of the objective function. Local search methods are efficient for locating the optimum of a uni-modal function since in this case the hill climbing search will eventually reach the global optimum, irrespective of the starting point. In groundwater modelling, parameter estimation has mainly been based on local gradient-based search techniques (e.g. McLaughlin and Townley, 1996).

Numerical simulation models may have numerous local optima on the objective function surface, and in such cases local search methods are less effective because the estimated optimum will depend on the starting point of the search. For such multi-modal objective functions global search methods are more effective ("global" in the sense that these algorithms are especially designed for locating the global optimum and not being trapped in local optima). Popular global search methods are the so-called population-evolution-based search strategies such as the shuffled complex evolution (SCE) algorithm (Duan et al., 1992) and genetic algorithms (GA) (Wang, 1991).

For calibration of lumped, conceptual hydrological catchment models a large number of studies have been conducted that compare different automatic algorithms (e.g. Duan et al., 1992; Gan and Biftu, 1996; Cooper et al., 1997; Kuzcera, 1997; Franchini et al., 1998; Thyer et al., 1999). The main conclusion from these studies is that the global population-evolution based algorithms are more effective than multi-start local search procedures, which in turn perform better than pure local search methods. Global search procedures has recently been applied in steady state groundwater modelling (Heidari and Ranjithan, 1998; Solomatine et al., 1999), but to the authors' knowledge no attempt has yet been made to apply these techniques for calibration of integrated and distributed catchment models.

In this paper, the SCE algorithm is adopted for calibration of the MIKE SHE integrated modelling system. In the following is given a brief description of this algorithm. For a more detailed presentation the reader is referred to Duan et al. (1992).

The SCE algorithm includes the following steps:

(1) *Initialisation*. An initial sample of parameter sets θ_i are randomly generated from the feasible parameter space Θ. For each parameter set the objective function value $F_i = F(\theta_i)$ is calculated. The initial sample has the size $s = pm$ where p is the number of complexes and m is the number of points in each complex.

(2) *Partitioning into complexes*. The s points are ranked in order of increasing objective function value ($F_{(1)} \leq F_{(2)} \leq ... \leq F_{(s)}$). The s points are partitioned into p complexes, such that points corresponding to function values $\{F_{(1)}, F_{(p+1)}, .., F_{((s-1)p+1)}\}$ form the 1st complex, points corresponding to function values $\{F_{(2)}, F_{(p+2)}, .., F_{((s-1)p+2)}\}$ form the 2nd complex, etc.

(3) *Evolution*. A sub-complex of size q is formed from the complex by randomly choosing q points from the p points in the complex. A trapezoidal probability distribution is used for assigning the probability of a point to be included in the sub-complex (i.e. larger probability for points with smaller objective function value). The sub-complex is evolved (offspring generation) according to the simplex algorithm (Nelder and Mead, 1965). Offspring generation from the same sub-complex is performed α times, and β evolution steps are taken by each complex.

(4) *Complex shuffling*. The new sample of s points is shuffled, cf. (2).

Steps (2)-(4) are repeated until the stopping criteria are met.

The most important algorithmic parameter is the no. of complexes p. Sensitivity tests show that the dimensionality of the calibration problem (no. of calibration parameters) is the primary factor determining the proper choice of p (Duan et al., 1994). In general, the larger value of p is chosen the higher the probability of converging into the global optimum but at the expense of a larger number of model simulations (the number of model simulations is virtually proportional to p), and vice versa.

The SCE algorithm simultaneously evolves a number of potential solutions towards the region of the global optimum of the objective function. Thus, when optimising the aggregated objective function, the SCE algorithm is expected to provide a reasonable approximation of the Pareto front near the point that corresponds to the global optimum of the objective function. By performing several independent optimisation runs with different weights in Eq. (3), the entire Pareto front can then be explored (Madsen, 2000a).

Application example

Model setup

The automatic calibration scheme is applied for calibration of the MIKE SHE model setup of the Karup catchment used in Refsgaard (1997), see Fig 3. The Karup catchment has an area of 440 km^2 and is located in the western part of Denmark. The topography varies from about 20 m 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 mainly of agriculture (67%), forest (18%), heath (10%), and wetland areas (4%). The catchment is drained by the Karup River and about 20 tributaries.

A brief description of the model parameterisation is given in the following. See Refsgaard (1997) for more details.

<u>Hydrogeology and saturated zone</u>
The catchment is assumed to consist of one main unconfined aquifer covering the entire catchment with the same hydraulic parameters and five lenses with different parameters. The horizontal conductivity (K_h) of the main aquifer is subject to calibration.

<u>Soil and unsaturated zone</u>
Two soil profiles are used in the model. Soil profile "general" that comprises loamy sand to a depth of 100 cm and fine sand below is used for the main part of the catchment. Soil profile "heath" that comprises fine sand to a depth of 55 cm and coarse sand below is used for the heath areas. For each of the four soil types, the saturated hydraulic conductivity

(K_{sat}) and the exponent (n) in the formula for the hydraulic conductivity as a function of soil moisture (S_e), ($K = K_{sat}S_e^{\,n}$), are subject to calibration. In addition, the capillary pressure (pF) at field capacity which is assumed to be the same for all four soil types is chosen as a calibration parameter.

Drainage system

The wetland areas near the river are drained by ditches and tile drain pipes. The drainage levels and the drainage coefficients are assumed constant in the entire wetland areas. The drainage coefficient is subject to calibration.

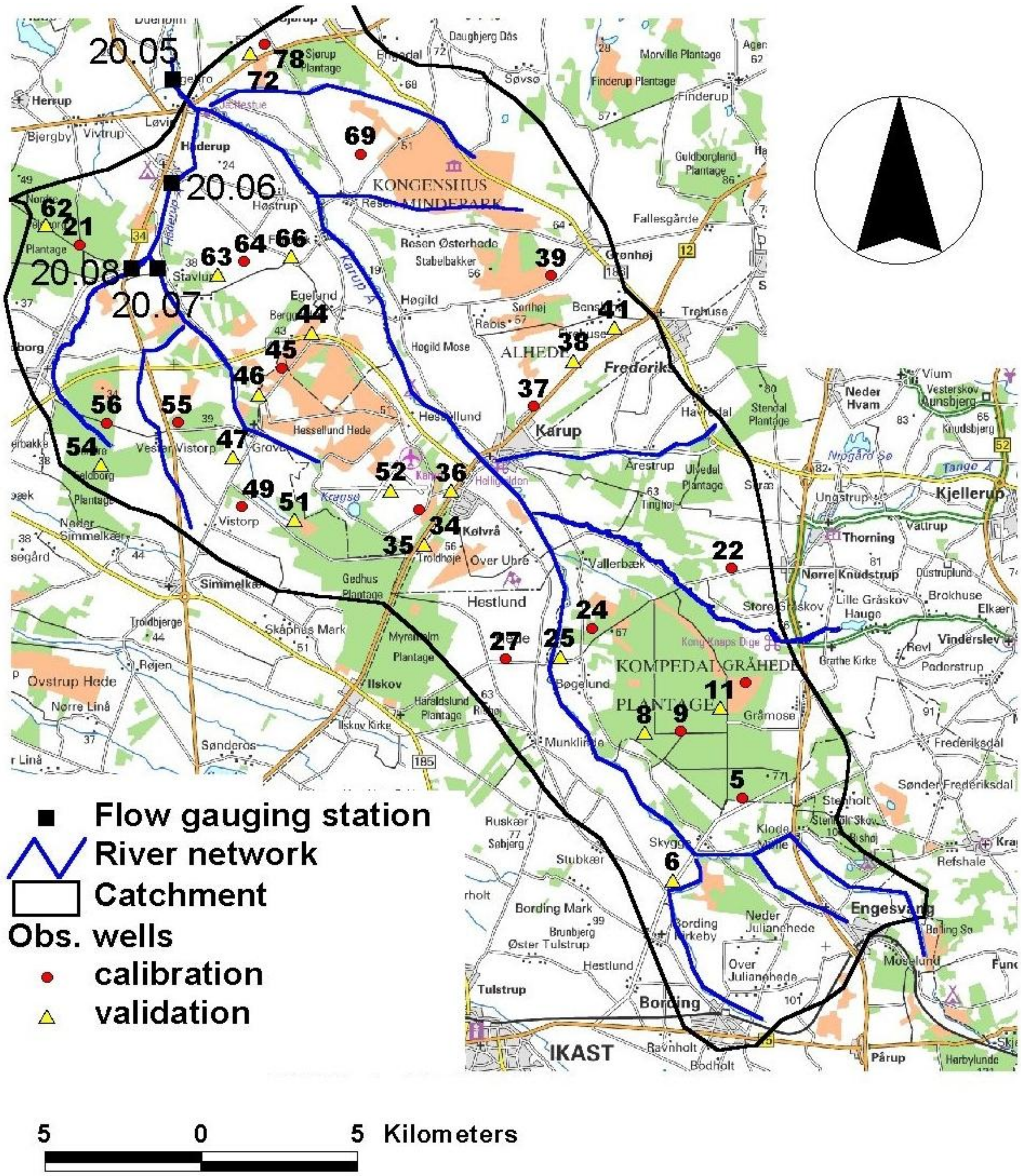

Figure 3 Karup catchment.

Stream-aquifer interaction
The main river and tributaries are included in the model (Fig 3). A leakage coefficient, which is assumed constant for the entire river system, is subject to calibration.

Vegetation
Four vegetation/cropping classes are assumed: (1) agricultural, (2) forest, (3) heath, and (4) wetland. In the present application, no vegetation parameters are subject to calibration.

The chosen calibration parameters and the parameter ranges used in the automatic calibration are summarised in Table 1 together with the parameter values given in Refsgaard (1997).

Table 1 Parameter ranges applied for the different parameters in the automatic calibration together with the balanced Pareto optimum solution and the parameter estimates found in Refsgaard (1997).

Parameter		*Parameter range*	*Balanced Pareto optimum*	*Refsgaard (1997)*
K_h main aquifer	[m/s]	10^{-5} - 10^{-3}	4.4 10^{-4}	3.5 10^{-4}
Drainage coefficient	[s^{-1}]	10^{-7} - 10^{-5}	4.4 10^{-7}	3.5 10^{-7}
K_{sat} (loamy sand)	[m/s]	10^{-6} - 10^{-4}	2.6 10^{-6}	1.0 10^{-5}
n (loamy sand)	[-]	10 - 25	24.0	18.8
K_{sat} (fine sand)	[m/s]	10^{-6} - 10^{-4}	5.8 10^{-6}	3.5 10^{-5}
n (fine sand)	[-]	5 - 15	8.7	10.0
K_{sat} (fine sand - heath)	[m/s]	10^{-6} - 10^{-4}	2.5 10^{-6}	3.5 10^{-5}
n (fine sand - heath)	[-]	5 - 15	10.5	10.0
K_{sat} (coarse sand)	[m/s]	10^{-5} - 10^{-3}	6.7 10^{-4}	2.0 10^{-4}
n (coarse sand)	[-]	5 - 10	8.9	6.0
Leakage coefficient	[s^{-1}]	10^{-8} - 10^{-5}	2.4 10^{-7}	3.0 10^{-7}
pF field capacity	[-]	1.5-2.5	1.68	2

Calibration data and definition of objective functions

The available measurements consist of groundwater level data sampled every two weeks from 35 locations in the catchment and daily discharge data from four stations in the river system, including the runoff at the catchment outlet (see Fig. 3). For calibration, groundwater level data from 17 wells as well as runoff data from the catchment outlet are used. Based on these data two objective functions are defined

$$F_h(\theta) = \frac{1}{M}\sum_{j=1}^{M}\sqrt{\frac{1}{n_i}\sum_{i=1}^{n_i}\left[h_{obs,i,j} - h_{i,j}(\theta)\right]^2} \tag{5}$$

$$F_q(\theta) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[q_{obs,i,j} - q_{i,j}(\theta)\right]^2} \quad (6)$$

Eq. (5) is the average root mean squared error (RMSE) of the groundwater levels (h) at the M = 17 locations, and Eq. (6) is the RMSE of the runoff (q) at the catchment outlet. The weighted average measure is given by

$$F_{agg}(\theta) = w_h F_h(\theta) + w_q F_q(\theta) \quad (7)$$

where w_h and w_q are the weights assigned to groundwater level and runoff data, respectively.

For calibration, data in the period 1 January 1971 - 31 December 1974 were used. To minimise the effect from the initial conditions for calculation of the objective functions, a 2-year warm-up period was applied in the simulations. For validation of the calibrated model, data in the period 1 January 1975 - 31 December 1977 were adopted.

Results

A number of tests were carried out in order to estimate the Pareto front and analyse the trade-offs between the two objectives defined above. For all tests the hypercube search space shown in Table 1 was used. The most important algorithmic parameter is the number of complexes p. A large value will provide a more robust algorithm in the sense that the chance of being trapped in a local optimum is small but this requires a large number of model evaluations, and hence implies an unacceptable computational burden. Preliminary sensitivity tests showed that p = 2 is a reasonable compromise between robustness and computing time. In this case the population of evaluated parameter sets converges to an acceptable limit within 400 model evaluations. Note that it is not the intention to let the optimisation algorithm evolve until *the* global optimum is found but rather to get a good estimate of the Pareto front in the vicinity of the optimum.

To explore the entire Pareto front, 5 different calibration runs (corresponding to a total of 2,000 model evaluations) were conducted; two calibration runs using only one of the objectives for estimation of the tails of the front, and 3 runs with different weights in (7) to estimate the intermediate part. The objective function values corresponding to the evaluated parameter sets are shown in Fig. 4 together with the estimated Pareto front which consists of 19 points. From a multi-objective point of view these 19 parameter sets are equally good but obviously for practical applications some points are more relevant than others. In this case the Pareto front has a very sharp structure, implying that calibration using only one of the variables will provide a very bad simulation of the other variable. However, a minor relaxation of the performance of one of the variables implies a significant improvement in the performance of the other variable. From the Pareto front, a balanced optimum solution can be defined that puts equal weight to the two objectives. Such a solution is shown in Fig. 4, and the corresponding parameter values are shown in Table 1.

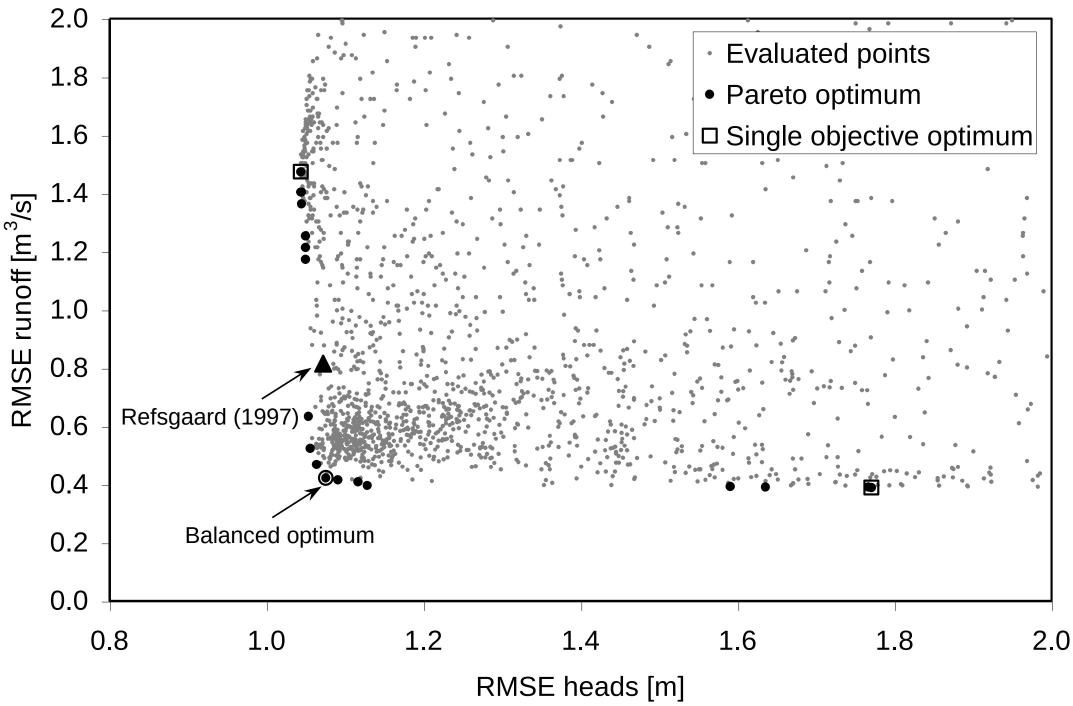

Figure 4 Estimated Pareto front. Marked points correspond to optimum values using only one objective, balanced Pareto optimum, and parameter set given by Refsgaard (1997).

The variation of the optimum parameter sets along the Pareto front is shown in Fig. 5. The parameter values are normalised with respect to the upper and lower limits given in Table 1 so that the feasible range of all parameters is between 0 and 1. A very large variability is observed in most of the parameter values when moving along the Pareto front. The hydraulic conductivity of the aquifer and the leakage coefficient are less variable, indicating that these parameters are important for the general model behaviour. For some of the parameters a general trend is present when moving along the Pareto front, i.e. these parameters have a significant trade-off when focusing on the different objectives. For example, the optimum values of the drainage coefficient and the capillary pressure at field capacity decrease when larger weights are given to the catchment runoff.

Performance statistics for groundwater levels and runoff data corresponding to the balanced Pareto optimum are shown in Tables 2-3. The statistics calculated for the groundwater levels include average values of bias and RMSE for both calibration data (17 observation wells) and validation data (18 observation wells). The statistics for the runoff data include bias, RMSE, and coefficient of determination R^2 (Nash and Sutcliffe, 1970) for the runoff at the catchment outlet (station 20.05) and three internal stations (20.06, 20.07, 20.08). Performance statistics are calculated for both the calibration and the validation period. Time series of simulated and observed runoff at station 20.05 for the calibration period and at station 20.06 for the validation period are shown in Figs. 6-7. Time series of simulated

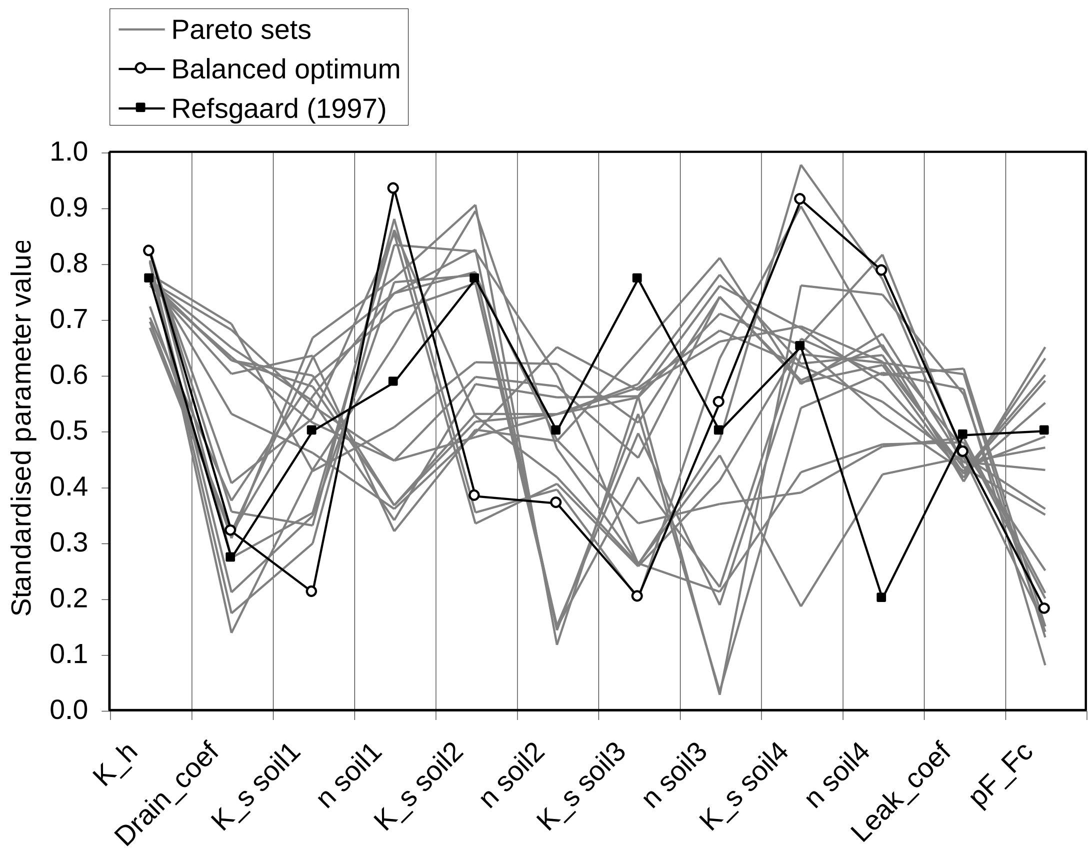

Figure 5 Standardised parameter values for Pareto optimum points, balanced optimum, and estimates given by Refsgaard (1997).

and observed groundwater levels for the calibration period at four selected observation wells are shown in Fig. 8.

Table 2 Performance statistics for groundwater level simulations.

		Calibration period		Validation period	
		Avg. bias [m]	Avg. RMSE [m]	Avg. bias [m]	Avg. RMSE [m]
Balanced Pareto optimum	Calibration points	0.00	1.08	0.08	1.06
	Validation points	0.27	1.02	0.36	1.07
Refsgaard (1997)	Calibration points	-0.07	1.07	-0.07	1.05
	Validation points	0.16	0.98	0.17	1.03

Table 3 Performance statistics for runoff simulations.

		Calibration period			Validation period		
	Station	Bias [%]	RMSE [m^3/s]	R^2	Bias [%]	RMSE [m^3/s]	R^2
Balanced	20.05	-0.5	0.42	0.88	-10.0	0.75	0.67
Pareto	20.06	0.2	0.13	0.62	1.2	0.15	0.74
optimum	20.07	-12.1	0.13	0.20	-6.2	0.10	0.67
	20.08	-15.6	0.06	-0.36	-11.4	0.06	0.24
Refsgaard	20.05	12.7	0.82	0.56	3.7	0.58	0.81
(1997)	20.06	14.3	0.20	0.15	17.6	0.24	0.36
	20.07	4.4	0.11	0.40	12.5	0.11	0.60
	20.08	1.4	0.05	0.06	8.8	0.05	0.40

Although bias is not included as an explicit calibration objective in the optimisation, the balanced Pareto optimum solution provides virtually unbiased catchment runoff and average groundwater levels. The individual observation wells may have significant biases but generally the groundwater level dynamics are simulated satisfactory (see Fig. 8). For the catchment runoff, the performance statistics become worse in the validation period with an increase of water balance error to 10 % and a decrease in R^2 from 0.88 to 0.67. The internal runoff stations, however, have generally better performance in the validation period. The simulation of groundwater levels has essentially the same performance in both the calibration and the validation period. The validation points are slightly more biased than the calibration points. Overall, the validation results show the same level of accuracy as the calibration results, confirming that the balanced Pareto optimum provides a hydrologically sound calibration.

For comparison, the parameter sets obtained by manual calibration in Refsgaard (1997) have been applied. The results obtained by using these parameters are shown in Tables 2-3 and Figs. 4-8. It should be noted that the results differ slightly from the calibration results presented in Refsgaard (1997). This is mainly due to the use of a somewhat different river network geometry and river flow solver in the present setup. For calibration of the model, Refsgaard (1997) used the runoff at the catchment outlet and groundwater level data from 7 observation wells (observation wells 8, 9, 11, 12, 21, 41, and 55, see Fig. 3). The same calibration and validation periods used in the present study were applied in Refsgaard (1997).

The parameter set obtained in Refsgaard (1997) is very close to a Pareto optimum (see Fig. 4) with more weight put on calibration of groundwater levels than on the catchment runoff. Compared to the balanced Pareto optimum point, the manual calibration provides worse performance on the catchment runoff in the calibration period with respect to both the water balance (bias) and overall shape of the hydrograph (RMSE and R^2) mainly due to an

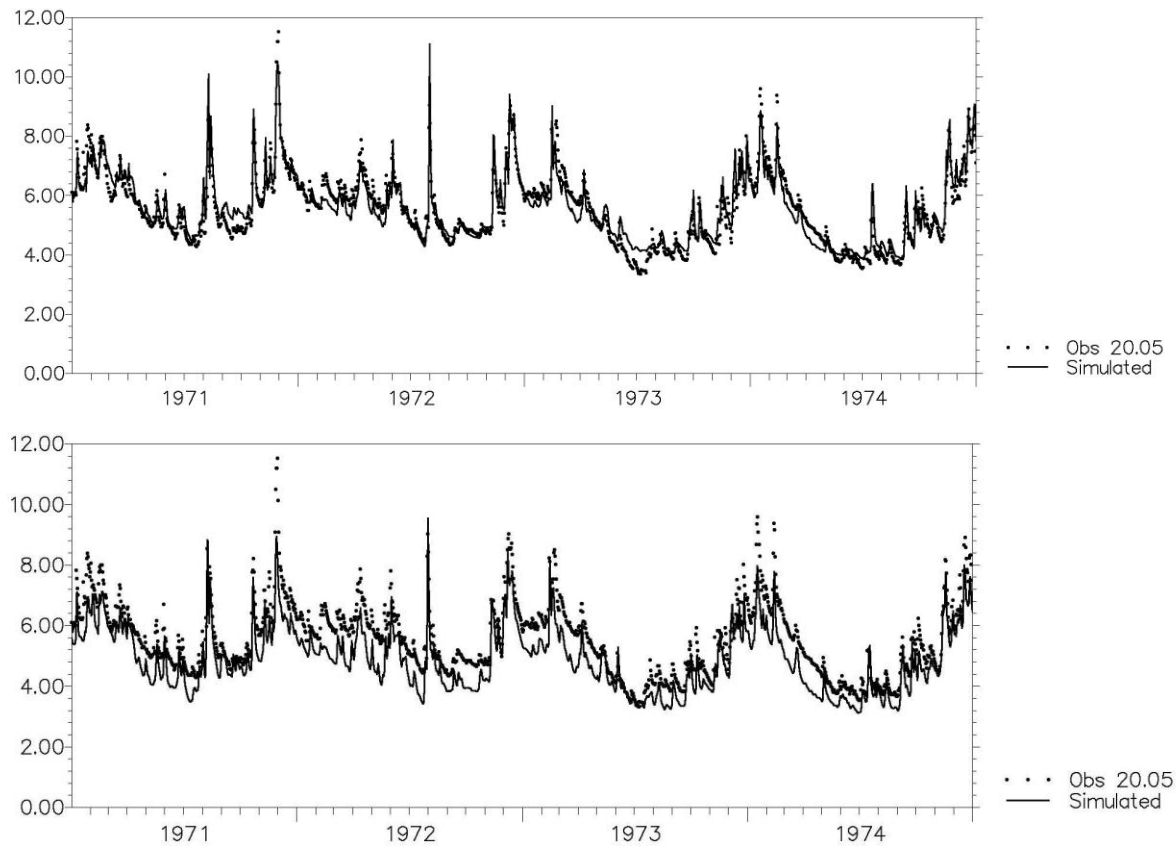

Figure 6 Simulated and observed runoff at the catchment outlet for the calibration period. Top: balanced Pareto optimum. Bottom: Refsgaard (1997).

underestimation of the baseflow (see Fig. 6). In the validation period, however, the catchment runoff from the manual calibration has less bias and higher R^2 than the balanced Pareto optimum solution. The internal runoff is generally better simulated with the balanced Pareto optimum solution than the manual calibration for both the calibration and the validation period. The manual calibration produces severe underestimation of the baseflow level (see Fig. 7).

With respect to simulation of groundwater levels in the calibration points, the manual calibration has virtually similar performance as the balanced Pareto optimum solution for both average bias and average RMSE, although there are some differences in the simulation results for the individual wells (see Fig. 8). For the validation points, the manual calibration has slightly smaller average values of bias and RMSE for both calibration and validation periods.

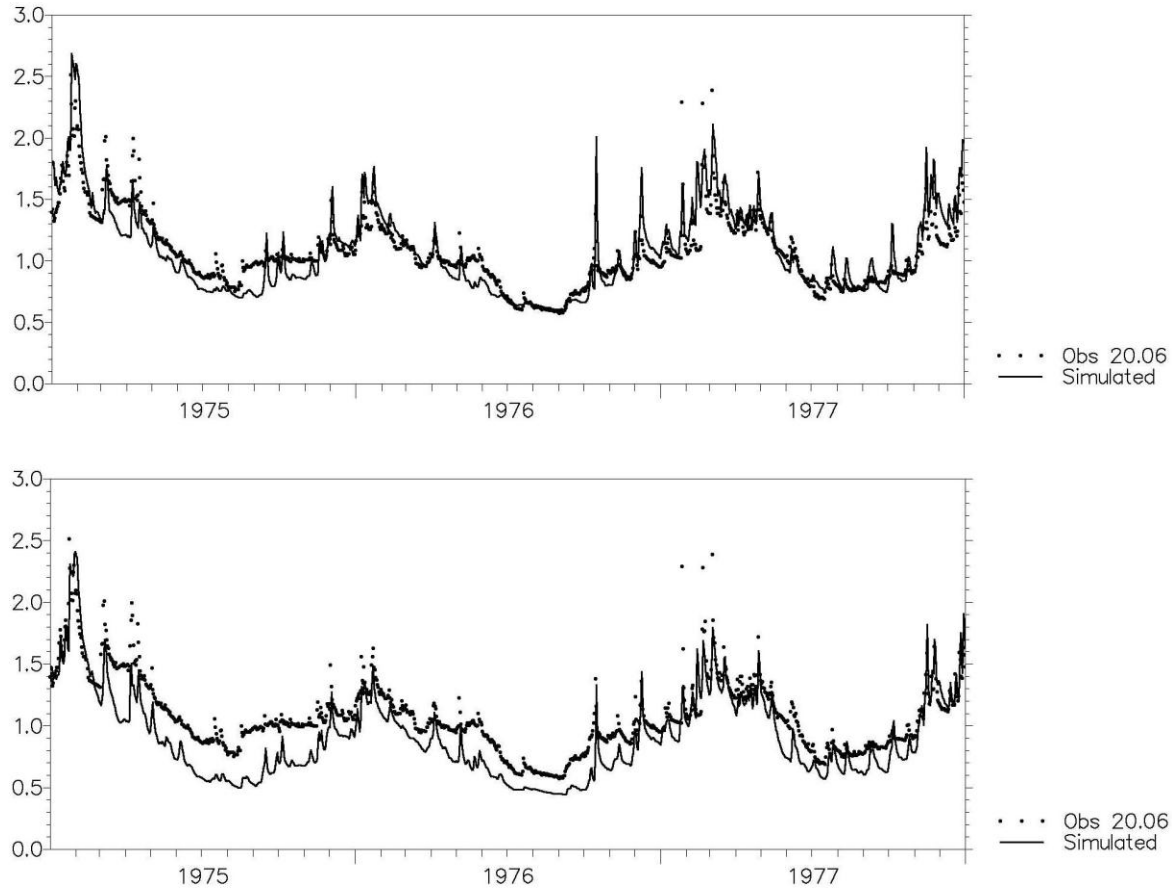

Figure 7 Simulated and observed runoff at station 20.06 for the validation period. Top: balanced Pareto optimum. Bottom: Refsgaard (1997).

Discussion and conclusions

A general framework for automatic calibration has been presented that includes three basic elements: (1) model parameterisation and choice of calibration parameters, (2) specification of calibration criteria, and (3) choice of optimisation algorithm. The importance of a rigorous model parameterisation for calibration of distributed hydrological models has been emphasised. The calibration problem has been formulated in a general multi-objective context that allows simultaneous optimisation of a set of objectives that involves different state variables, multi-site measurements, and multiple response modes. An automatic optimisation procedure based on the SCE algorithm has been introduced for solving the multi-objective calibration problem.

An application example has been presented that illustrates the use of the proposed calibration framework. An integrated MIKE SHE model setup of the Karup catchment was calibrated using both groundwater level and runoff data. A significant trade-off between the performance of the groundwater level simulations and the catchment runoff was observed,

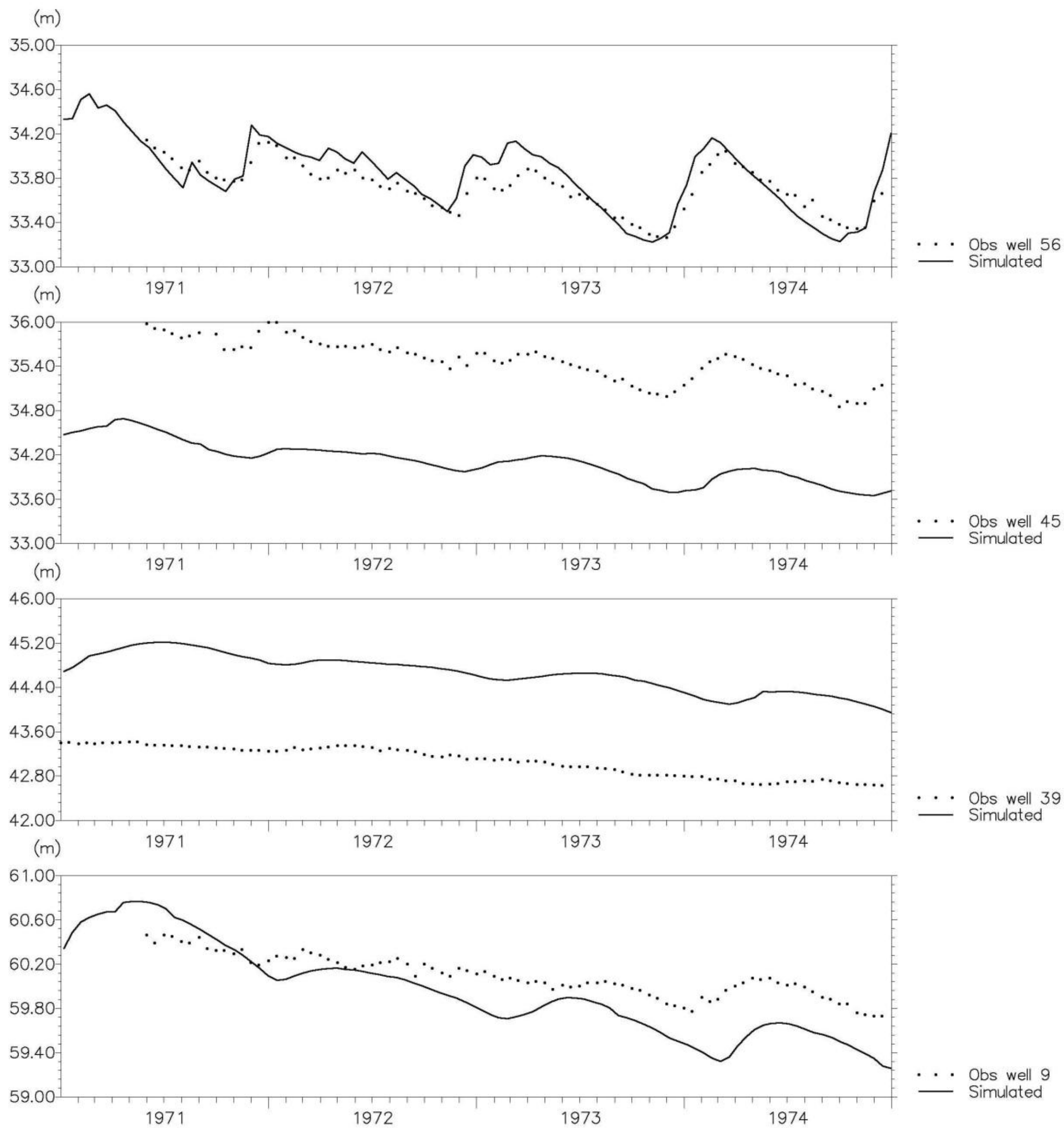

Figure 8a Simulated and observed groundwater levels at selected observation wells for the calibration period using the balanced Pareto optimum.

defining a Pareto front with a very sharp structure. From a multi-objective viewpoint the solutions on the Pareto front are equally good, and one can choose a single solution according to specific objectives for the model application being considered. A compromise balanced Pareto optimum solution corresponding to the break point on the Pareto front was selected in this case. The manual calibration obtained in Refsgaard (1997) was shown to be very close to a Pareto optimum with more weight put on groundwater level simulations than on catchment runoff.

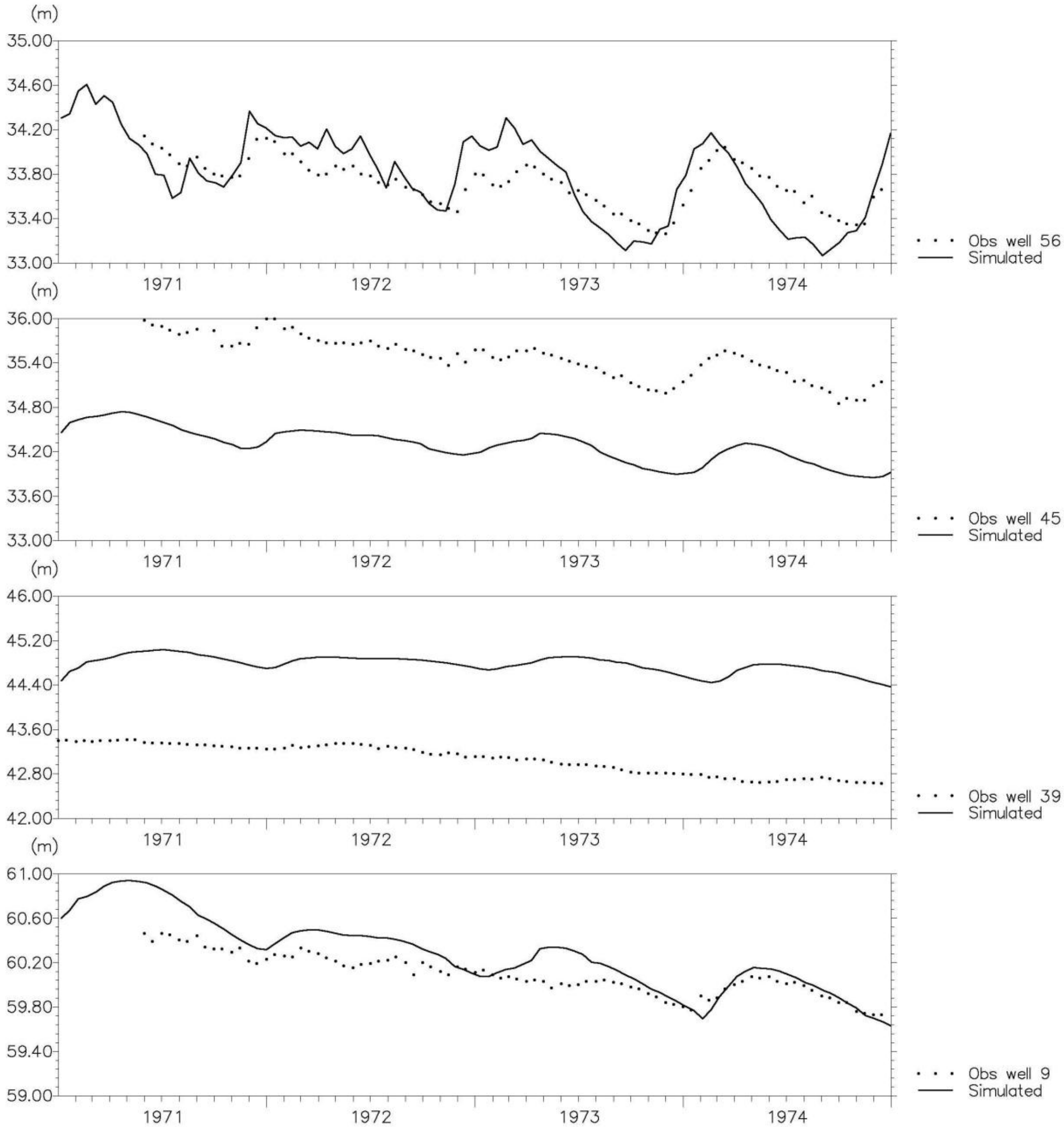

Figure 8b Simulated and observed groundwater levels at selected observation wells for the calibration period using the parameter estimates given in Refsgaard (1997).

The balanced Pareto optimum solution was validated on a different period than used for calibration as well as on data not used in the calibration. The validation results showed, in general, the same level of accuracy as the calibration results, confirming that the balanced Pareto optimum solution is hydrologically sound. Compared to the manual calibration, the balanced Pareto optimum solution provides generally better simulation of the runoff, whereas virtually similar performance is obtained for the groundwater level simulations.

The application example presents the first step in a model calibration. The proposed calibration framework is an iterative procedure where model parameterisation can be changed during the calibration process. In the present setup, one main aquifer was assumed, and the conductivity defining the hydraulic properties of this aquifer was seen to be the most important calibration parameter. The next step in the calibration process will then be to apply a more distributed setup of the saturated zone, e.g. by using zonation to improve the calibration. In addition, more complex descriptions of the drainage and the river-aquifer interaction could be considered by defining distributed drainage and leakage coefficients. Future calibration iterations should also analyse model simplifications, i.e. the unsaturated zone may be reparameterised to include fewer calibration parameters.

Acknowledgements

This work was in part funded by the Danish Technical Research Council under the Talent Project No. 9901671 "Data Assimilation in Hydrodynamic and Hydrological Modelling".

References

Cooper, V.A., Nguyen, V.T.V. and Nicell, J.A., 1997, Evaluation of global optimization methods for conceptual rainfall-runoff model calibration, *Wat. Sci. Tech.*, **36**(5), 53-60.

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

Duan Q., Sorooshian, S. and Gupta, V.K., 1994, Optimal use of the SCE-UA global optimization method for calibrating watershed models, *J. Hydrol*., **158**, 265-284.

Franchini, M., Galeati, G. and Berra, S., 1998, Global optimization techniques for the calibration of conceptual rainfall-runoff models, *Hydrol. Sci. J*., **43**(3), 443-458.

Gan, T.Y. and Biftu, G.F., 1996, Automatic calibration of conceptual rainfall-runoff models: Optimization algorithms, catchment conditions, and model structure, *Water Resour. Res.*, **32**(12), 3513-3524.

Gupta, H.V., Sorooshian, S. and Yapo, P.O., 1998, Toward improved calibration of hydrological models: Multiple and noncommensurable measures of information, *Water Resour. Res*., **34**(4), 751-763.

Heidari, M. and Ranjithan, S.R., 1998, A hybrid optimization approach to the estimation of distributed parameters in two-dimensional confined aquifers, *J. Amer. Water Resour. Ass.*, **34**(4), 909-920.

Hill, M.C., 1998, Methods and guidelines for effective model calibration, U.S. Geological Survey, Water Resources Investigations Report 98-4005.

Kuczera, G., 1997, Efficient subspace probabilistic parameter optimization for catchment models, *Water Resour. Res*., **33**(1), 177-185.

Madsen, H., 2000a, Automatic calibration of a conceptual rainfall-runoff model using multiple objectives, *J. Hydrol.*, **235**, 276-288.

Madsen, H., 2000b, Automatic Calibration and Uncertainty Assessment in Rainfall-Runoff Modelling, 2000 Joint Conference on Water Resources Engineering and Water Resources Planning & Management, Hyatt Regency Minneapolis,USA, July 30 – August 2, 2000.

McLaughlin, D. and Townley, L.R., 1996, A reassessment of the groundwater inverse problem, *Water Resour. Res.*, **32**(5), 1131-1161.

Nash, J.E. and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models. Part 1: A discussion of principles, *J. Hydrol.*, **10**, 282-290.

Nelder, J.A. and Mead, R., 1965, A simplex method for function minimization, *Comput. J*., **7**(4), 308-313.

Refsgaard, J.C., 1997, Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, **198**, 69-97.

Solomatine, D.P., Dibiki, Y.B. and Kukuric, N., 1999, Automatic calibration of groudwater models using global optimisation techniques, *Hydrol. Sci. J.*, **44**(6), 879-894.

Sorooshian, S. and Gupta, V.K., 1995, Model calibration, *Computer Models of Watershed Hydrology* (ed. Singh, V.P.), Water Resources Publications, Colorado, 23-68.

Spear, R.C. and Hornberger, G.M., 1980, Eutrophication in peel inlet. II: Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, **14**, 43-49.

Thyer, M., Kuczera, G. and Bates, B.C., 1999, Probabilistic optimisation for conceptual rainfall-runoff models: A comparison of the shuffled complex evolution and simulated annealing algorithms, *Water Resour. Res.*, **35**(3), 767-773.

Van Griensven, A. and Bauwens, W., 2001, Multi-objective calibration in integrated river basin modelling, Geophysical Research Abstracts, Vol. 3, European Geophysical Society, 26th General Assembly.

Wang, Q.J., 1991, The genetic algorithm and its application to calibrating conceptual rainfall-runoff models, *Water Resour. Res*., **27**(9), 2467-2471.