Journal of Hydrology 261 (2002) 48–59

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Comparison of different automated strategies for calibration of rainfall-runoff models

Henrik Madsen*, Geoffrey Wilson, Hans Christian Ammentorp

DHI Water and Environment, Agern Allé 11, DK-2970 Hørsholm, Denmark

Received 22 June 2001; revised 5 December 2001; accepted 11 December 2001

Abstract

Three different automated methods for calibration of rainfall-runoff models are presented and compared. The methods represent various calibration strategies that utilise multiple objectives and allow user intervention on different levels and different stages in the calibration process. The methods have been applied for calibration of a test catchment and compared on validation data with respect to overall performance measures in terms of water balance error and general hydrograph shape, and simulation of high and low flow events. The results illustrate the problem of non-uniqueness in model calibration since none of the methods are superior with respect to all performance measures considered. In general, the different methods put emphasis on different response modes of the hydrograph. Calibration based on the use of generic search routines in combination with user-specified calibration priorities is seen to compare favourably with an expert system that is designed for the specific model being considered and requires user intervention during the entire calibration process.

Keywords: Rainfall-runoff models; Calibration; Parameter estimation; Optimisation routine; Expert system

1. Introduction

Computer-based lumped, conceptual rainfall-runoff models have been widely applied in hydrological modelling since they were first introduced in the late 1960s and early 1970s. Well known examples of this type of model which are still used today are the Sacramento model (Burnash et al., 1973; Burnash, 1995), the HBV model (Bergström and Forsman, 1973; Bergström, 1995), and the NAM model (Nielsen and Hansen, 1973; Havnø et al., 1995). A lumped, conceptual rainfall-runoff model consists of a set of linked mathematical equations, describing in a simplified form the behaviour of the land phase of the hydrological cycle with parameters that represent average values for the entire catchment. While in some cases a typical range of likely parameter values can be given, it is not, in general, possible to determine the parameters from physiographic, climatic and soil physical characteristics of the catchment under consideration. Thus, the final parameter estimation must be performed by calibration against observed data.

Traditionally, calibration 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. Furthermore, because of the subjectivity involved, it

* Corresponding author. Tel.: +45-4516-9242; fax: +45-4516-9292.

E-mail address: hem@dhi.dk (H. Madsen).

PII: S0022-1694(01)00619-9

is difficult to explicitly assess the confidence of the model simulations. Due to this, a great deal of research has been directed to development of more effective and efficient automatic calibration procedures.

In recent years, several automatic global search algorithms have been developed that are especially designed for locating the global optimum on a response surface with numerous local optima which is often observed in rainfall-runoff models (Duan et al., 1992). Popular global search methods are the population-evolution-based optimisation algorithms such as, amongst others, genetic algorithms (Wang, 1991), shuffled complex evolution (SCE) (Duan et al., 1992), and simulated annealing (Sumner et al., 1997). A large number of studies have been conducted that compare different automatic algorithms for calibration of rainfall-runoff models (e.g. Duan et al., 1992; Gan and Biftu, 1996; Cooper et al., 1997; Kuczera, 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.

While much research has been directed to developing effective and efficient generic search routines, less effort has been made to tailor these methodologies against specific model applications. In this regard, an important element is the proper formulation of the objective function that is optimised numerically. Application of automatic calibration routines has mainly been based on a single objective measure of comparison, e.g. the sum of squared errors between observed and simulated runoff. A single measure, however, is often inadequate to properly take into account the simulation of all the important characteristics of the system that are reflected in the observations and which are implicitly used by the hydrologist to evaluate the goodness-of-fit of the calibrated model. Recently, automatic routines that use a multi-objective formulation of the calibration problem have been introduced in rainfall-runoff modelling (Lindström, 1997; Liong et al., 1996, 1998; Gupta et al., 1998; Yapo et al., 1998; Madsen, 2000a; Boyle et al., 2000).

Another group of automatic calibration methods comprises the knowledge-based expert systems. While the global optimisation routines are generic and require minimum user intervention, knowledge-based methodologies are model-specific (i.e. are designed for a specific rainfall-runoff model) and often require the user to intervene during the calibration process. The basic philosophy behind these methods is to automate the course of a manual calibration carried out by an experienced hydrologist. Harlin (1991) formulated a knowledge-based calibration scheme for the HBV model where the parameters are calibrated individually, focusing on different process descriptions. Zhang and Lindström (1997) formulated a different routine for the HBV model where the parameters are calibrated in two stages based on a division of the parameters in two subsets according to the model structure. Gupta et al. (1999) compared the SCE algorithm with a semi-automated expert system for calibration of the Sacramento model. They showed that the generic SCE algorithm compares favourably with the expert system, suggesting that state-of-art automatic calibration performs with a level of skill similar to that of an experienced hydrologist.

In this paper three different automated strategies are adopted for calibration of the NAM rainfall-runoff model. The methods were presented at a workshop arranged at the third DHI Software Conference (DHI, 1999). Before the conference, runoff and meteorological data sets from a test catchment were given to the participants for calibration of the NAM model. At the conference itself, validation data covering another period were made available for evaluation and intercomparison of the calibrated models. The three methods represent automated strategies that allow user intervention on different levels and different stages in the calibration process. These include:

1. A generic search routine where the user specifies a weighting of several calibration objectives that are aggregated into one measure which is then optimised automatically.
2. A method that uses different automatic search techniques in combination with different calibration objectives and requires user intervention at different stages in the calibration process.
3. A knowledge-based expert system where user intervention is required during the entire calibration

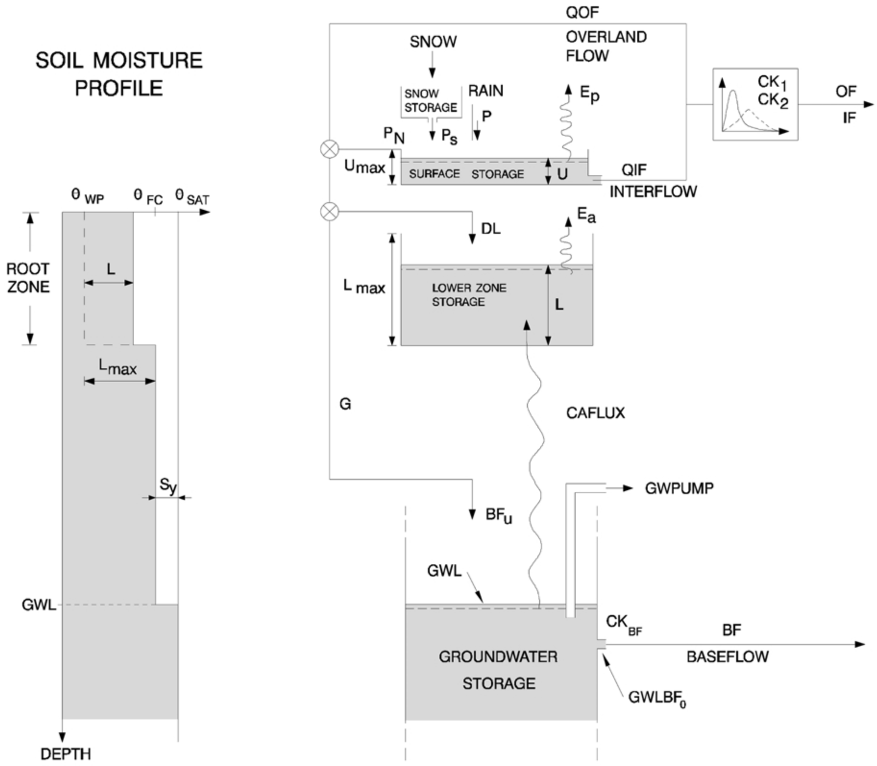

Fig. 1. NAM model structure.

process for subjective evaluation of different calibration criteria.

The paper is organised as follows. In Section 2, the experimental set-up is outlined, including a description of the NAM model, the test catchment, and the evaluation criteria employed. In Section 3, the three different calibration schemes are presented, and the results of the calibration exercise are summarised in Section 4. In Section 5 discussion of the results is given with focus on the formulation of calibration strategy and inclusion of specific model knowledge in the calibration process. Conclusions are given in Section 6.

2. Experimental set-up

The model used in this study is the NAM rainfall-runoff model that forms part of the rainfall-runoff module of the MIKE 11 river modelling system (Havnø et al., 1995; DHI, 2000). The NAM model represents the various components of the rainfall-runoff process by continuously accounting for the water content in four different and mutually interrelated storages where each storage represents different physical elements of the catchment. These storages are: (1) snow storage, (2) surface storage, (3) lower zone (root zone) storage, and (4) groundwater storage.

The meteorological input data to the model are precipitation, potential evapotranspiration and temperature (if snow modelling is included). On this basis, it produces, as its main results, catchment runoff and groundwater level values as well as information about other elements of the land phase of the hydrological cycle, such as the soil moisture content and the groundwater recharge. The resulting catchment runoff is split conceptually into overland flow, interflow and baseflow

Table 1
NAM model parameters

Parameter	Description
U_{MAX} (mm)	Maximum water content in the surface storage. This storage can be interpreted as including the water content in the interception storage, in surface depression storages, and in the uppermost few cm's of the soil.
L_{MAX} (mm)	Maximum water content in the lower zone storage. L_{MAX} can be interpreted as the maximum soil water content in the root zone available for the vegetative transpiration.
CQOF	Overland flow runoff coefficient. CQOF determines the distribution of excess rainfall into overland flow and infiltration ($0 \leq \mathrm{CQOF} \leq 1$).
TOF, TIF, TG	Threshold values for overland flow, interflow and recharge, respectively. Flow is only generated if the relative moisture content in the lower zone storage is larger than the threshold value.
CK_{IF} (h)	Time constant for interflow from the surface storage. CK_{IF} is the dominant routing parameter of the interflow because $CK_{IF} >> CK_{12}$.
CK_{12} (h)	Time constant for overland flow and interflow routing. Overland flow and interflow are routed through two linear reservoirs in series with time constants CK_{12}.
CK_{BF} (h)	Baseflow time constant. Baseflow from the groundwater storage is generated using a linear reservoir model with time constant CK_{BF}.

components representing, respectively, the quick, intermediate and slow response modes of the hydrograph. The structure of the model is shown in Fig. 1. In the present study, the basic NAM model is applied, including nine parameters to be determined by calibration. A brief description of these parameters is given in Table 1. Snow modelling is included in the simulation but calibration is not performed on the snow module parameters.

The NAM model was applied to the Danish Tryggevaelde catchment. This catchment has an area of 130 km^2, an average rainfall of 710 mm/year, and an average runoff of 240 mm/year. The catchment is dominated by clayey soils, implying a relatively flashy flow regime. For the calibration, a 5-year period (1 January 1984–31 December 1988) was used where daily data of precipitation, potential evapotranspiration, mean temperature, and catchment runoff are available.

For comparing the calibrated models, validation data covering the period 1 January 1989–31 December 1993 were used. In order to unify the presentation and intercomparison of the different models, several statistical measures were calculated, including:

1. Water balance error (difference between average simulated and average observed runoff).
2. Coefficient of determination (R^2) (Nash and Sutcliffe, 1970).
3. Peak flow statistics for four selected peak flow events, including peak error, bias and root mean square error (RMSE).
4. Low flow statistics for four selected low flow periods, including bias and RMSE.

3. Calibration methods

3.1. Multi-objective calibration using shuffled complex evolution

Madsen (2000a) presented an automatic calibration procedure for the NAM model based on optimisation of several objective functions simultaneously. In this method four basic objective functions are considered:

1. Overall volume error

$$F_1(\theta) = \left| \frac{1}{N} \sum_{i=1}^{N} [Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)] \right| \tag{1}$$

2. Overall RMSE

$$F_2(\theta) = \left[\frac{1}{N} \sum_{i=1}^{N} [Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)]^2 \right]^{1/2} \tag{2}$$

3. Average RMSE of peak flow events

$$F_3(\theta) = \frac{1}{M_\mathrm{p}} \sum_{j=1}^{M_\mathrm{p}} \left[\frac{1}{n_j} \sum_{i=1}^{n_j} [Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)]^2 \right]^{1/2} \tag{3}$$

4. Average RMSE of low flow events

$$F_4(\theta) = \frac{1}{M_l}\sum_{j=1}^{M_l}\left[\frac{1}{n_j}\sum_{i=1}^{n_j}[Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)]^2\right]^{1/2} \tag{4}$$

In Eqs. (1)–(4), $Q_{\mathrm{obs},i}$ is the observed discharge at time i, $Q_{\mathrm{sim},i}$ is the simulated discharge, N is the total number of time steps in the calibration period, M_p is the number of peak flow events, M_l is the number of low flow events, n_j is the number of time steps in peak/low flow event no. j, and θ is the set of model parameters to be calibrated. Peak flow events are defined as periods where the observed discharge is above a given threshold level. Similarly, low flow events are defined as periods where the observed discharge is below a given threshold level.

When using multiple objectives, the solution to the calibration problem will not, in general, be a single unique set of parameters but will consist of the set of Pareto optimal (non-dominated) solutions. Madsen (2000a) discusses the principles and implications of using multiple objectives. In the present application, the different objectives have been transformed into a single aggregated objective measure

$$F_{\mathrm{agg}}(\theta) = \left[\sum_{i=1}^{4}[F_i(\theta) + A_i]^2\right]^{1/2} \tag{5}$$

where A_i are transformation constants, reflecting the priorities given to the different objectives.

The objective function in Eq. (5) is optimised automatically using the SCE global search algorithm developed by Duan et al. (1992). In brief, the SCE algorithm involves the following steps:

1. *Initialisation*. An initial sample of parameter sets is randomly generated from the feasible parameter space.
2. *Partitioning into complexes*. The sample is partitioned into several complexes based on the objective function values of the evaluated parameter sets.
3. *Evolution*. Each complex is evolved independently according to the simplex method (Nelder and Mead, 1965).
4. *Complex shuffling*. The evolved complexes are shuffled to enable sharing of information and new complexes are formed, cf. step (2).

For a more detailed description of the algorithm the reader is referred to Duan et al. (1992).

In this method, the user specifies the priorities to be given to certain objectives, depending on the specific model application being considered. In the present test, the balanced aggregated objective function suggested by Madsen (2000a) was applied. In this case, the transformation constants in Eq. (5) are automatically calculated from the randomly generated initial population in the SCE algorithm so that all $(F_i(\theta) + A_i)$ have about the same distance to the origin near the optimum. For the SCE algorithmic parameters, recommended values given by Duan et al. (1994) were used. As suggested by Kuczera (1997), to reduce the chance of premature termination of the search algorithm, the number of complexes was set equal to the number of calibration parameters.

3.2. Clustering and simulated annealing

This method is based on a combination of clustering, simulated annealing, and multi-objective optimisation. Four basic objective functions are considered:

1. Overall volume error, cf. Eq. (1).
2. Relative RMSE of all flow events

$$F_2(\theta) = \left[\frac{1}{N}\sum_{i=1}^{N}\left[\frac{Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)}{Q_{\mathrm{obs},i} + Q_{\mathrm{sim},i}(\theta)}\right]^2\right]^{1/2} \tag{6}$$

3. Weighted RMSE of peak flow events

$$F_3(\theta) = \left[\frac{1}{N}\sum_{i=1}^{N} w_{\mathrm{obs},i}[Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)]^2\right]^{1/2} \tag{7}$$

4. Weighted RMSE of low flow events

$$F_4(\theta) = \left[\frac{1}{N}\sum_{i=1}^{N} w_{\mathrm{obs},i}[Q_{\mathrm{obs},i} - Q_{\mathrm{sim},i}(\theta)]^2\right]^{1/2} \tag{8}$$

In Eqs. (7) and (8), $w_{\mathrm{obs},i}$ is a weighting function of the peak and low flow instances, respectively.

In this methodology a solution is found using clustering followed by simulated annealing of a single objective, which is followed by multi-criteria optimisation of two objectives. In brief, the algorithm involves the following steps:

1. *Initialisation*. An initial sample of parameter sets is randomly generated from the feasible parameter space.
2. *Cluster analysis*. The sample is partitioned into clusters based on the relative RMSE objective function values of the evaluated parameter sets. The best solution in each cluster is taken as a potential solution for further refinement.
3. *Simulated annealing*. Each potential solution is evolved independently following the simulated annealing (SA) algorithm into an improved solution using any one of the single objective functions (except volume error).
4. *Multi-criteria optimisation*. The improved solutions are further improved in the peak-flow and low-flow multi-criteria 2D space by perturbation. Each improved solution becomes a point on the Pareto surface.
5. *Final selection*. The Pareto surface is plotted and the investigator (a human) selects which point on the front (and the corresponding parameters) is 'optimal' by trading-off objectives.

Note that the water-balance objective is only used in the final selection step as an independent check on the possible 'optimal' solutions.

3.3. Knowledge-based expert system

The knowledge-based calibration procedure seeks to reflect the course of a trial-and-error calibration of an experienced hydrologist, focusing on different process descriptions, but replacing the arbitrary (often visual) goodness-of-fit measures by appropriate objective functions to be optimised numerically. The procedure involves a preliminary recession analysis, followed by an iterative scheme where individual parameters are adjusted according to different response modes of the hydrograph.

3.3.1. Recession analysis

In NAM, baseflow is described as the output from a linear reservoir. Thus, the time constant CK_{BF} can be estimated from a linear regression analysis of the log-transformed observed discharges in the low-flow seasons. Similarly, recession analysis is used for an initial estimation of the time constant CK_{12} that determines the peak flow recession. In certain cases, recession analysis also allows initial estimation of the interflow time constant CK_{IF} by analysing the intermediate flow recessions.

3.3.2. Storage parameters

The parameters U_{MAX} and L_{MAX}, representing the storage capacities of the upper zone and the lower zone storage, respectively, control the overall water balance. These two parameters can be optimised by comparing the simulated and the observed total runoff volume during the calibration period. If the simulated volume is larger than the observed volume, U_{MAX} and L_{MAX} should be increased, and vice versa.

3.3.3. Overland flow runoff coefficient

The overland flow runoff coefficient, CQOF, controls the distribution of excess rainfall between overland flow and infiltration. The runoff coefficient mainly affects the volume of peak flows, and can thus be determined by comparing the distribution of the simulated and the observed flow. This is numerically done by comparing the moments of the hydrographs about the origin. If the moment of the simulated hydrograph is larger than the moment of the observed hydrograph, CQOF should be reduced, and vice versa.

3.3.4. Time constant for overland flow

The time constant CK_{12} is fine-tuned by comparing the observed and the simulated peak flows. If peak flows tend to be underestimated, CK_{12} should be increased, and vice versa.

3.3.5. Threshold values

Rainfall occurring after a dry spell will in many cases not generate any flow since all the water infiltrates and fills up the root zone. In NAM, this is modelled by allowing generation of overland flow, interflow and recharge only if the soil moisture content of the lower zone storage is larger than a threshold value (TOF, TIF and TG, respectively). The threshold values can be identified by comparing the observed and the simulated hydrograph in the beginning of the wet seasons. Numerically, the beginning of a wet season is defined as periods where the moisture content in the lower zone is rising and is smaller than a specified level.

Table 2
Calibrated parameters using SCE, cluster analysis and simulated annealing (Cluster + SA), and expert system

Parameter	SCE	Cluster + SA	Expert system
U_{MAX} (mm)	10	9.5	25
L_{MAX} (mm)	390	290	350
CQOF	0.70	0.89	0.92
TOF	0.73	0.78	0.70
TIF	0.83	0.42	0.60
TG	0.71	0.74	0.64
CK_{IF} (h)	790	1070	650
CK_{12} (h)	32	43	24
CK_{BF} (h)	1820	1890	2670

3.3.6. Time constant for interflow

In many cases, interflow is not a dominant flow component, implying that the interflow time constant CK_{IF} does not need to be fine-tuned. For calibration of CK_{IF} the simulated and the observed runoff is compared for days where interflow is the dominant flow component as compared to overland flow.

4. Results

The parameter estimates obtained by the different calibration methods are shown in Table 2, and the evaluated performance statistics are shown in Table 3. The overall performance measures, water balance error and coefficient of determination R^2, are calculated from the entire 5-year validation period. The observed and simulated hydrographs for this period obtained by the SCE multi-objective methodology is shown in Fig. 2. The peak flow measures include average values of the absolute peak error, the bias, and the RMSE of four selected high flow events. Since the snow routine has not been subject to calibration in the present study, the selected high flow events do not include snowmelt events. The observed and simulated runoff for high flow event No. 2 obtained by the three different calibrations is shown in Fig. 3. The low flow measures include average values of the bias and the RMSE of four selected low flow events. The observed and simulated runoff for low flow event No. 2 obtained by the three different calibrations is shown in Fig. 4.

With respect to the overall performance measures, the SCE calibration has the best performance. It provides virtually an unbiased solution and has the best R^2-value. The calibration obtained by the expert system is also virtually unbiased but has a slightly lower R^2-value. The cluster and SA calibration has both worse water balance error and R^2-value.

For the peak flow performance, none of the methods are superior for all of the measures considered. The SCE calibration has the best performance with respect to the general shape of the peak flow hydrographs (lowest RMSE); the cluster and SA calibration has the lowest bias, whereas the expert system has the minimum peak error. The differences between the three calibrations are mainly caused by the use of different time constants CK_{12} for overland flow routing, which is the dominant parameter controlling the shape of the high flow hydrograph. The cluster and SA calibration has the largest CK_{12}-value, which results in longer recessions and a general phase error for

Table 3
Performance statistics of calibrated models using SCE, cluster analysis and simulated annealing (Cluster + SA), and expert system

	SCE	Cluster + SA	Expert system
Overall measures			
Water balance error (%)	0.2	−7.0	−0.2
R^2	0.85	0.82	0.83
Peak flow measures			
Absolute peak error (m^3/s)	1.10	1.14	1.00
Bias (m^3/s)	0.45	−0.11	0.48
RMSE (m^3/s)	0.67	0.92	0.73
Low flow measures			
Bias (m^3/s)	−0.050	−0.031	−0.027
RMSE (m^3/s)	0.091	0.067	0.078

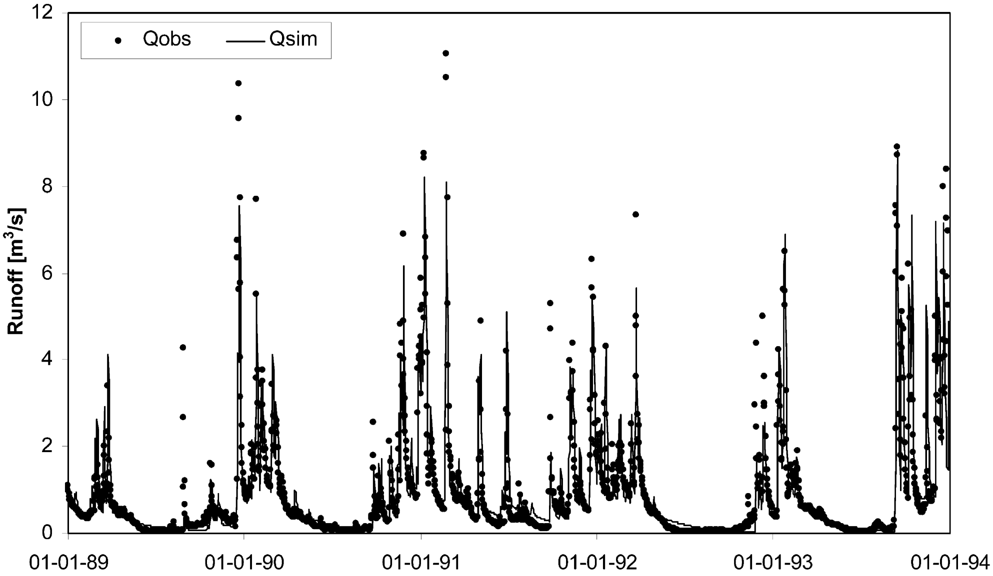

Fig. 2. Observed and simulated hydrographs obtained by the SCE multi-objective methodology.

simulation of the high flow events (see Fig. 3). The expert system has the smallest CK_{12}-value, resulting in more peaky high flow simulations than the other two calibrations.

With respect to the low flow measures, the SCE calibration has the worst performance. The cluster and SA calibration has the smallest RMSE, whereas the expert system has the smallest bias. As can be seen in Fig. 4, the SCE method has problems in simulating the small peaks in the low flow periods, which are better captured by the other two calibrations. Simulation of minor peaks in the low flow period is dominated by the threshold values for interflow and overland flow in relation to the actual water content in the root zone. The higher threshold value for interflow TIF obtained in the SCE calibration requires larger rainfall to simulate a runoff response in the low flow period as compared to the other two calibrations.

5. Discussion

The results show that none of the calibration methods are superior with respect to all the performance measures considered. The different methods put emphasis on different aspects or response modes of the hydrograph. The procedure based on the multi-objective SCE method has the best overall performance but the worst performance for low flow simulations, the clustering and SA based procedure has the worst overall performance but better performance for low flow simulations, and the expert system favours certain aspects for both overall, high flow and low flow performance.

These results clearly illustrate the problem of uniqueness in model calibration. First, in a multi-objective context, there is a multitude of parameter combinations that are equally good according to trade-offs between different objectives (see e.g. Gupta et al., 1998; Madsen, 2000a). Secondly, there are many different parameter combinations that can give acceptable solutions according to one specific objective (see e.g. Beven and Binley, 1992). If one seeks a single solution to the calibration problem, it is important to recognise the non-uniqueness property and put priorities to certain objectives in the calibration process, depending on the model application being considered.

The results also show that calibration based on the use of generic search routines in combination with user-specified calibration priorities compares favourably with the expert system that is designed for the

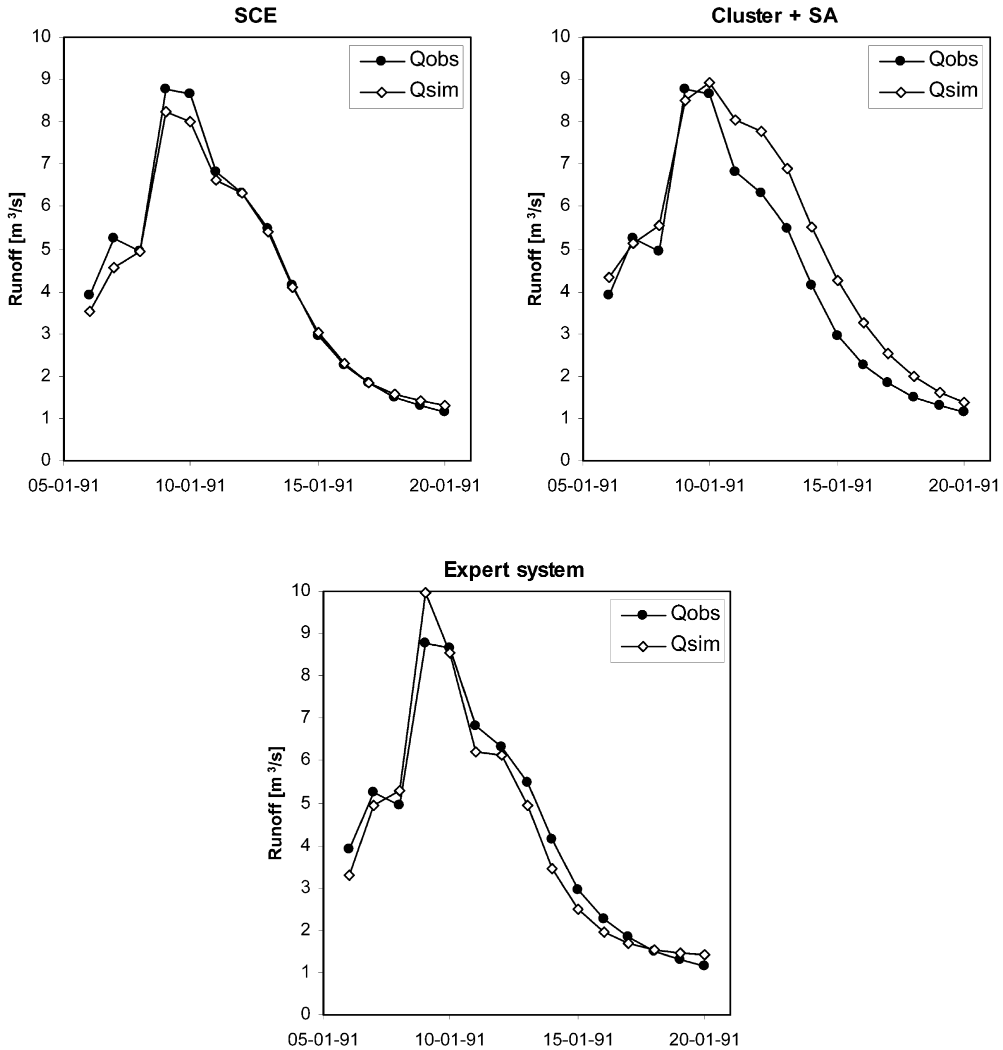

Fig. 3. Observed and simulated runoff in high flow event No. 2 obtained by the three different calibrations.

specific model being considered and requires user intervention during the entire calibration process. In this regard, however, it should be emphasised that automatic calibration based on generic search routines is not an easy push-the-button solution to rainfall-runoff modelling. First, as noted above, user intervention is important for tailoring the model calibration to a specific application. Secondly, critical and sound hydrological knowledge is important to evaluate the quality of data, the model set-up, and the model calibration. Data and model errors should be properly taken into account in order to avoid that the automatic calibration routine provides parameter estimates that compensates for errors rather than optimises the model performance. For instance, due to the inherent uncertainties in measuring the catchment average rainfall and the corresponding runoff, it may be questionable to attempt to obtain a perfect water balance fit. One should rather focus more on other aspects when trading off objectives and then accept a water balance error within certain limits, say 5–10%.

When comparing different calibration algorithms, efficiency is a general aspect to consider. In the present case, the SCE and the cluster and SA methods

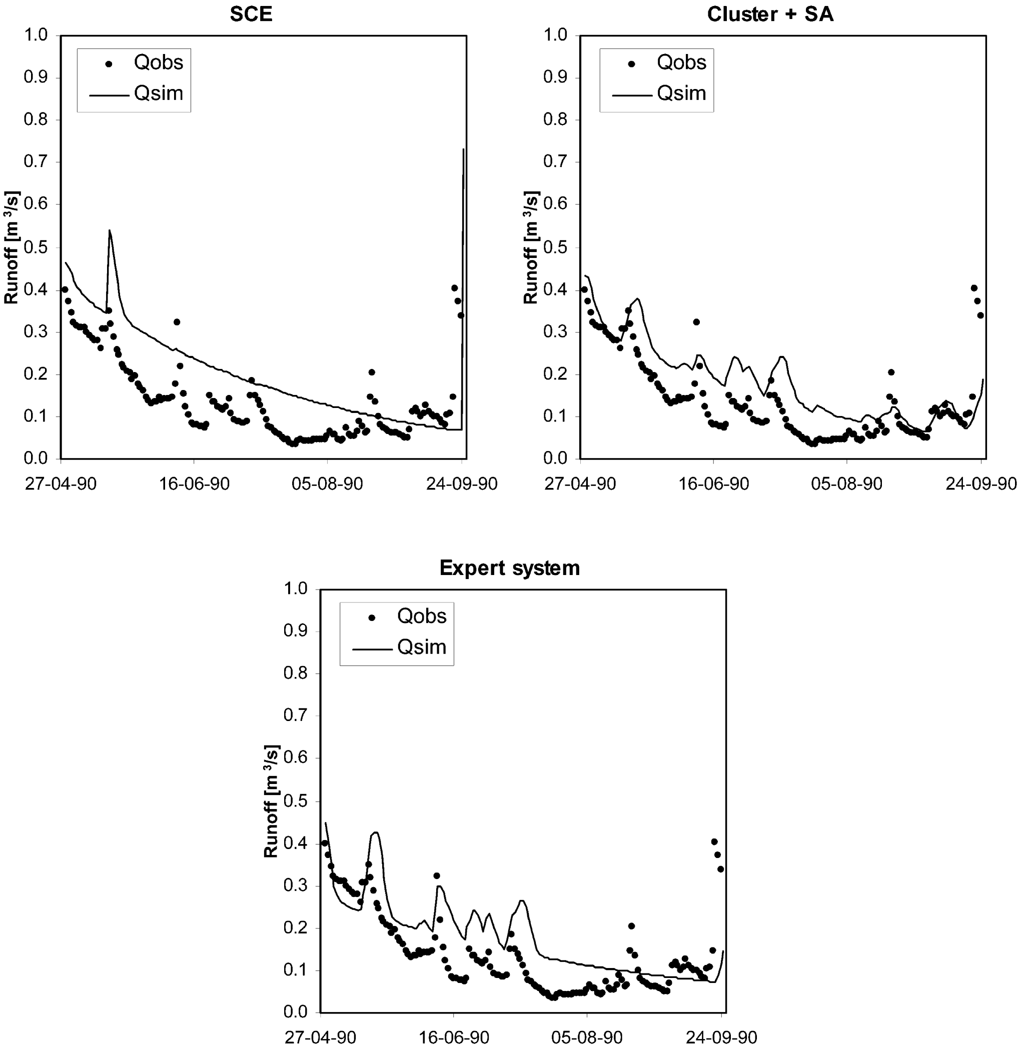

Fig. 4. Observed and simulated runoff in low flow event No. 2 obtained by the three different calibration methods.

used in the order of 1000–10 000 model evaluations whereas the expert system was much more efficient and required only about 50 model evaluations. The considered model, however, is very cheap (in the order of 10^{-2}–10^{-1} CPU seconds for a model evaluation), end hence the efficiency aspect is not crucial in this application. For more expensive models, 1000–10 000 model evaluations may be unacceptable. In such cases, the generic search routines considered herein could still be used but the effectiveness of the routines has to be relaxed. For instance, in the SCE algorithm the number of complexes can be lowered to obtain a more efficient search scheme but at the expense of increasing the chance of locating a local optimum rather than the global one. In general, the use of prior knowledge about the calibration parameters in terms of e.g. specifying realistic limits and correlation structures, and utilising knowledge about the responses of parameter changes will increase the efficiency of an automatic calibration scheme.

Closely related to the efficiency and effectiveness aspects is the problem of parameter insensitivity. Sensitivity analysis such as the Monte Carlo based procedure proposed by Spear and Hornberger (1980) is often used to identify insensitive parameters, and hence reduce the number of free model parameters to be calibrated. The results of such an analysis, however, should be carefully interpreted. The analysis 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 also be noted that inclusion of multiple objectives in the calibration process provides better identifiable parameters and a more well-posed model structure (Madsen, 2000b).

6. Conclusions

Three different automated strategies have been used for calibration of the MIKE 11/NAM rainfall-runoff model. The methods represent automated strategies that allow user intervention on different levels and different stages in the calibration process, including (1) the SCE optimisation algorithm with a user specified objective function that allows simultaneous optimisation of several objectives, (2) clustering and SA optimisation methods that also combines different calibration objectives but requires user intervention at different stages in the calibration process, and (3) an expert system where user intervention is required during the entire calibration process.

The calibrated models have been compared with respect to the overall performance in terms of water balance error and general hydrograph shape, and simulation of high and low flow events. The different methods were seen to put emphasis on different aspects or response modes of the hydrograph, and none of the methods were superior with respect to all performance measures considered. This illustrates the problem of non-uniqueness in model calibration and emphasises the importance of considering multiple objectives in the calibration process.

The automated strategies considered in this paper all reflect the multi-objective nature of model calibration. The use of generic search routines where user intervention is required only for definition of appropriate multi-objective numerical measures is seen to compare favourably with methods that require more user intervention and include subjective rules for trading off objectives.

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

Bergström, S., 1995. The HBV model. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, pp. 443–476.

Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff model. Nordic Hydrol. 4, 147–170.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6, 279–298.

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 (12), 3663–3674.

Burnash, R.J.C., 1995. The NWS river forecast system—catchment modeling. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, pp. 311–366.

Burnash, R.J.C., Ferral, R.L., McGuire, R.A., 1973. A generalized streamflow simulation system—conceptual modeling for digital computers, US Department of Commerce, National Weather Service and State of California, Department of Water Resources.

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

DHI, 1999. Conference Programme and Book of Abstracts, third DHI Software Conference, 7–9 June 1999, DHI Water and Environment, Denmark.

DHI, 2000. NAM Technical Reference and Model Documentation, DHI Water and 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.

Duan, Q., Sorooshian, S., 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., Berra, S., 1998. Global optimization techniques for the calibration of conceptual rainfall-runoff models. Hydrol. Sci. J. 43 (3), 443–458.

Gan, T.Y., 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., Yapo, P.O., 1998. Toward improved calibration of hydrological models: multiple and noncommensurable measures of information. Water Resour. Res. 34 (4), 751–763.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of automatic calibration for hydrological models: comparison with multilevel expert calibration. J. Hydrol. Eng., ASCE 4 (2), 135–143.

Harlin, J., 1991. Development of a process oriented calibration scheme for the HBV hydrological model. Nordic Hydrol. 22, 15–36.

Havnø, K., Madsen, M.N., Dørge, J., 1995. MIKE 11—a generalized river modelling package. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Colorado, pp. 733–782.

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

Lindström, G., 1997. A simple automatic calibration routine for the HBV model. Nordic Hydrol. 28 (3), 153–168.

Liong, S.Y., Khu, S.T., Chan, W.T., 1996. Construction of multiobjective function response surface with genetic algorithm and neural network. Proceedings of the International Conference on Water Resource and Environment Research, October 29–31, Kyoto Japan, vol. II, pp. 31–38.

Liong, S.Y., Khu, S.T., Chan, W.T., 1998. Derivation of Pareto front with accelerated convergence genetic algorithm, ACGA. In: Babovic, V., Larsen, L.C. (Eds.). Hydroinformatics 1998. Balkema, Rotterdam, The Netherlands, pp. 889–896.

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 and Management, Hyatt Regency MN, USA, July 30–August 2, 2000.

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

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

Nielsen, S.A., Hansen, E., 1973. Numerical simulation of the rainfall runoff process on a daily basis. Nordic Hydrol. 4, 171–190.

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

Sumner, N.R., Fleming, P.M., Bates, B.C., 1997. Calibration of a modified SFB model for twenty-five Australian catchments using simulated annealing. J. Hydrol. 197, 166–188.

Thyer, M., Kuczera, G., 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.

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

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrological models. J. Hydrol. 204, 83–97.

Zhang, X., Lindström, G., 1997. Development of an automatic calibration scheme for the HBV hydrological model. Hydrol. Process. 11, 1671–1682.