Hydrology and Earth System Sciences, 4(2), *215–224* (2000) © EGS

Hydrology & Earth System Sciences

Multi-criteria calibration of a conceptual runoff model using a genetic algorithm

Jan Seibert

Oregon State University, Department of Forest Engineering, 213 Peavy Hall, Corvallis, OR 97731, USA
e-mail for corresponding author: jan.seibert@orst.edu

Abstract

Calibration of a model against more than one output variable is important for reliable simulations of internal processes. In this study, a genetic algorithm combined with local optimisation was proposed for automatic single- and multi-criteria calibration of the HBV model, a conceptual runoff model. The model and the optimisation algorithm were applied in two catchments with different geology where, in addition to observed runoff, time series of groundwater level data were available. For a theoretical, error-free test case with synthetic data, the optimisation algorithm was usually able to find the true parameter values. For the real-world case, parameter values varied considerably when calibrating against runoff only. However, parameter values were constrained significantly when calibrating against both runoff and groundwater levels. Furthermore, for one of the catchments, the results of the multi-criteria calibration motivated a modification of the model structure.

Keywords: Multi-criteria calibration; genetic algorithm; parameter uncertainty; conceptual runoff models; HBV model; groundwater levels

Introduction

Calibration and validation of conceptual runoff models usually are limited to comparing simulated with observed streamflow at the basin outlet. Nevertheless, applying a model to make predictions into the unknown as, for instance, extreme floods or effects of climate changes implicitly presupposes that the internal processes are simulated correctly. Therefore, evaluating the capability of a model to simulate various hydrological variables is needed for a more rigorous analysis of model performance. Multi-criteria calibration and validation of conceptual runoff models has become a research topic gaining increasing attention (de Grosbois *et al.*, 1988; Ambroise *et al.*, 1995; Mroczkowski *et al.*, 1997; Franks *et al.*, 1998; Kuczera and Mroczkowski, 1998; Yapo *et al.*, 1998).

In conceptual runoff models, various simple routines represent catchment hydrology in a lumped or semi-distributed way. Usually, at least 10 to 15 parameters are used. These parameters may have a physical basis but, as they are effective parameters on the catchment or sub-catchment scale, almost all of them need to be determined by calibration. The information contained in the rainfall-runoff relationship usually does not allow identification of a unique parameter set. Reducing the number of parameters is unattractive because it would transform the conceptual, grey-box representation of the rainfall-runoff process into a pure black-box description. A more attractive way to reduce parameter uncertainty is the use of additional data. Franks *et al.* (1998) demonstrated that the percentage of saturated areas in the catchment helped to constrain calibrated parameter values and model predictions in an application of TOPMODEL. However, the worth of additional data varies. Using a hydrosalinity model, Kuczera and Mroczkowski (1998) found that groundwater levels helped only a little into reducing the parameter uncertainty, whereas stream salinity data substantially reduced the uncertainties.

In this study, the HBV model (Bergström, 1976, 1992) was used as an example of a typical conceptual runoff model. It has been applied in numerous studies in more than 30 countries (Lindström *et al.*, 1997) and has been found to be capable of simulating runoff in different climatic zones. Comparisons with variables other than observed runoff, however, are relatively rare. Hottelet *et al.* (1994) found snow pack observations to be useful for validation of the model. Bergström and Sandberg (1983) used the HBV model to simulate groundwater levels in three geologically different aquifers. They modified the general model structure and found a good agreement between observed and simulated levels. However, they did not check the simulated runoff against measurements. Furthermore, the model has been modified to simulate the transport of

solutes, which allowed comparing simulations with pH and alkalinity (Bergström *et al.*, 1985) as well as with conservative tracers (Lindström and Rodhe, 1986).

Manual calibration of a model by trial and error is a time-consuming method and results may be subjective. This is particularly true when calibrating against more than one hydrological variable. Therefore, various automatic calibration methods have been developed (Sorooshian and Gupta, 1995). Evolution-based methods have been found to be suitable tools for the optimisation of conceptual runoff models (Wang, 1991; Duan *et al.*, 1992; Franchini, 1996; Kuczera, 1997; Yapo *et al.*, 1998). Genetic algorithms are one class of these methods. The idea of genetic algorithms, originally suggested by Holland (1975/1992), is to mimic evolution. Parameter sets are encoded to chromosome-like strings and different recombination operators are used to generate new parameter sets. The optimisation starts with a population of randomly generated parameter sets. These are evaluated by running the model and those sets that give a better simulation according to some objective function are given more chances to generate new sets than those sets that gave poorer results.

For a general overview of genetic algorithms and discussions on their optimisation capabilities the reader is referred to Goldberg (1989), Davis (1991), Beasley *et al.* (1993a,b), Whitley (1994) and Mitchell (1996). In hydrology, genetic algorithms have been used for the calibration of conceptual runoff models by Wang (1991), Franchini (1996) and Kuczera (1997). Kuczera (1997) compared different probabilistic optimization algorithms including a genetic algorithm and found the shuffled complex evolution (SCE-UA) algorithm (Duan *et al.*, 1992) to be superior to the genetic algorithm. However, Kuczera (1997) used a traditional genetic algorithm and pointed out that "improved genetic algorithm performance is therefore possible" (p. 184). Franchini *et al.* (1998) found only slight differences in the performance between the SCE-UA algorithm and a genetic algorithm combined with a subsequent local optimisation when calibrating a model to a single catchment.

In this study, a genetic algorithm was proposed for multi-criteria calibration of the HBV model. Firstly, the model was calibrated to a synthetic runoff series generated by the model. Thereafter, two catchments with different geology were used to calibrate the model against both runoff and

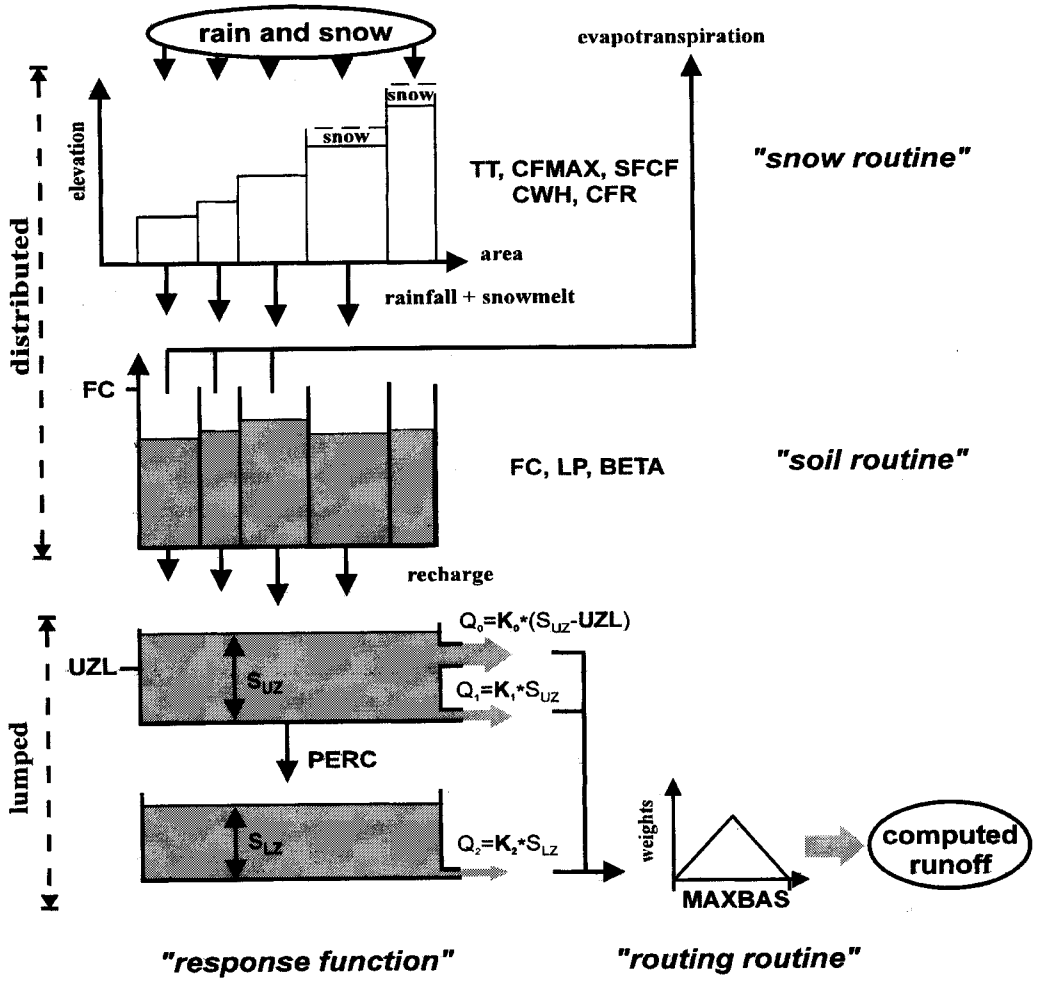

Fig. 1. *Structure of the HBV model.*

groundwater level observations. The aim was twofold, a test of the capability of the genetic algorithm as a tool for multi-criteria calibration and an assessment of the worth of groundwater data for the calibration of a conceptual runoff model.

Materials and methods

HBV MODEL

The HBV model (Bergström, 1976, 1992) is a conceptual model that simulates daily discharge using daily rainfall and temperature, and monthly estimates of potential evaporation as input. The model consists of different routines (Fig. 1, Table 1), where snowmelt is computed by a degree-day method, groundwater recharge and actual evaporation are functions of actual water storage in a soil box, runoff formation is represented by three linear reservoir equations and channel routing is simulated by a triangular weighting function. For both the snow and the soil routine, calculations are performed for each different elevation zone, but the response routine is a lumped representation of the catchment. Further descriptions of the model can be found elsewhere (e.g. Bergström, 1992, 1995; Lindström *et al.*, 1997; Seibert, 1997a). The version of the model used in this study, 'HBV light 1.2' (Seibert, 1997b) corresponds to the original version described by Bergström (1992, 1995).

The agreement between observed (Q_{obs}) and simulated (Q_{sim}) catchment runoff was evaluated by the efficiency, (Nash and Sutcliffe, 1970), here called R_{eff} (Eqn. 1).

$$R_{eff} = 1 - \frac{\sum(Q_{obs} - Q_{sim})^2}{\sum(Q_{obs} - \overline{Q_{obs}})^2} \quad (1)$$

In both catchments used in this study, groundwater levels were measured about fortnightly at a number of groundwater tubes. Because the HBV model simulates the groundwater lumped over the catchment, local observations could not be compared directly to the simulations. The groundwater observations were also spatially averaged i.e. the arithmetic mean was computed from the observations at the different tubes. To allow comparison with the observed mean groundwater level, the storage in the groundwater boxes (S_{UZ} and S_{LZ}) had to be transformed into a groundwater level, z [m a.s.l.]. A linear equation (Eqn. 2) with a slope m, which corresponded to the inverse of the storage coefficient, and an offset c was used. The coefficients were determined by linear regression between the simulated storage and groundwater levels. The performance of the groundwater level simulation was evaluated using the coefficient of determination, r^2, as objective function.

$$z = m(S_{UZ} + S_{LZ}) + c \quad (2)$$

DESCRIPTION OF THE GENETIC ALGORITHM

With a genetic calibration algorithm, optimised parameter sets are found by an evolution of parameter sets using

Table 1. Model parameters and feasible ranges

Parameter	Explanation	Unit	Lower bound	Upper bound
Snow routine				
TT	Threshold temperature	°C	−1.5	2.5
CFMAX	Degree-day factor	mm °C^{-1} d^{-1}	1	10
SFCF	Snowfall correction factor	–	0.5	1.2
CWH	Water holding capacity	–	0	0.2
CFR	Refreezing coefficient	–	0	0.1
Soil routine				
FC	Maximum of SM (storage in the soil)	mm	50	500
LP	Threshold for reduction of evaporation (SM/FC)	–	0.3	1
BETA	Shape coefficient	–	1	6
Response routine				
K_0	Recession coefficient (upper storage)	d^{-1}	0.1	0.5
K_1	Recession coefficient (upper storage)	d^{-1}	0.05	0.3
K_2	Recession coefficient (lower storage)	d^{-1}	0.001	0.1
UZL	Threshold for the K_0-outflow	mm	0	50
PERC	Maximal flow from upper to lower box	mm d^{-1}	0	4
MAXBAS	Routing, length of weighting function	d	1	7

selection and recombination. An initial population of n (set to 50) parameter sets was generated randomly in the parameter space and the 'fitness' of each set was evaluated by the value of the objective function(s). From this population, a new population (generation) was generated by n times combining two of the parameter sets. The two sets were chosen randomly but the chance of being picked was related to the fitness of the parameter set (i.e. the value of the objective function) giving the highest probability to the sets with the highest fitness. A new parameter set was generated from the two parent sets (sets A and B) by applying one of the following four rules for each parameter randomly with certain probabilities, p:

- value of set A ($p = 0.41$)
- value of set B ($p = 0.41$)
- random value between the values of set A and set B (alternatively, if both values were equal, a random value close to this value) ($p = 0.16$)
- random value within the limits given for the parameter (mutation) ($p = 0.02$).

The first two rules preserve the values of the preceding generation, whereas the other two rules provide an amount of random search. A balance between these rules is important for the success of the algorithm. However, within reasonable ranges, adjustments to the probabilities for the different rules had only minor effects on the performance of the algorithm. Subsequently, the fitness of each set in the new population was evaluated and the new generation replaced the old one. The best set was retained if there was no better set in the proceeding generation. This evolution was repeated for a number of generations (the maximum number of model runs was set to 3800).

Wang (1991) mentioned that the combination with a local search method could improve the results of a genetic algorithm. Franchini (1996) used the parameter set found by a genetic algorithm as a starting point for a local optimisation. In addition to this form of subsequent 'fine-tuning' the idea of a local search was implemented in a second way in this study. At a small probability ($p = 0.02$) the new parameter set was generated not by the parameter-by-parameter combination described above but by a one-dimensional optimisation along the line determined by the two parameter sets using Brent's method (Press *et al.*, 1992). The total number of 5000 model runs was divided into 3800 runs for the genetic algorithm and 1200 runs for the subsequent local optimisation. For the latter, Powell's quadratically convergent method as described by Press *et al.* (1992) was used.

The genetic algorithm described above differs from those used by Wang (1991), Franchini (1996) and Kuczera (1997) in several different respects of which the following two are the most important. While the other algorithms used a binary representation of the parameter sets, in this study real numbers were used. The advantage of the latter method is that the parameters are represented directly (Davis, 1991) and it has been found to give faster, more consistent and more accurate results (Janikow and Michalewicz, 1991). Furthermore, in all three studies one- or two-point crossover has been used to combine two strings representing parameter sets. This means that the values of several successive parameters are swapped jointly between the parent parameter sets. For the one-point crossover the parameter strings are divided into two parts at a randomly chosen point and one part is swapped between the two strings. Consequently the location of the various parameters on the string is of importance, e.g. the probability that the values of two parameters are chosen from the same parent decreases with the distance between the parameters on the string. The implementation used in this study, also called uniform crossover ensures independence on the location of the parameters. This is assumed to make the algorithm more robust (Syswerda, 1989; Beasley *et al.*, 1993b) and makes subjective ordering of the parameters unnecessary.

The use of multiple populations can improve genetic algorithms (Whitley, 1994; Punch, 1998). Usually the same criterion are used to define fitness in the different populations, but obviously the use of multiple populations provides a means to extend a genetic algorithm for multi-criteria calibration. In this study, three populations of parameter sets were used in which fitness was defined in different ways: (1) the efficiency, R_{eff}, for the runoff simulations (2) the coefficient of determination, r^2, for the groundwater level simulations and (3) a combined measure (see below). After a certain number of iterations (set to 10 in this study) a number of parameter sets (set to 40) was exchanged between the populations. By this means, characteristics of good parameter sets were transferred and helped to find parameter sets valid for both runoff and groundwater levels.

For the combined measure of model performance, both the efficiency, R_{eff}, of the runoff simulation and the r^2 value for the groundwater level simulations were transformed into fuzzy measures, f_Q and f_{GW}. These transformations were based on the highest value of each objective function obtained by the individual calibration, $R_{eff,\max}$ and $r^2_{\max}$ (Eqns. 3 and 4). The fuzzy measures evaluated the degree of truth of the statement 'this parameter set is the best possible one'. The combined measure, F, was computed as the geometric mean of the two fuzzy measures according to each criterion (Eqn. 5).

$$f_Q = \max\left(0, \frac{R_{eff} - 0.8\ R_{eff,\max}}{0.2\ R_{eff,\max}}\right) \tag{3}$$

$$f_{GW} = \max\left(0, \frac{r^2 - 0.8\ r^2_{\max}}{0.2\ r^2_{\max}}\right) \tag{4}$$

$$F = \sqrt{f_Q \cdot f_{GW}} \tag{5}$$

Table 2. Catchment characteristics

Characteristic	Lilla Tivsjön	Tärnsjö
SMHI[1] station number	42–1920	54–2299
Area [km^2]	12.8	14
Forest percentage [%]	88	85
Lake percentage [%]	2.7	1.8
Range of elevation [m a.s.l.]	246–440	55–105
Mean annual precipitation [mm]	586	729
Mean annual runoff [mm]	262	266
Mean annual temperature [°C]	2.1	5.3

[1] Swedish Meteorological and Hydrological Institute.

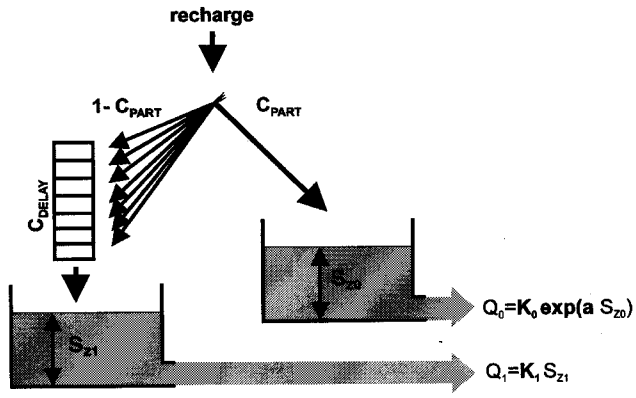

Fig. 2. *Modified model structure of the response routine for the Tärnsjö catchment. The new parameters control the division of the recharge between the two boxes (C_{PART}), the delay of the recharge in one box (C_{DELAY}) and the exponential storage (a).*

STUDY CATCHMENTS

Lilla Tivsjön

The Lilla Tivsjön catchment (Table 2) is a sub-basin of the former International Hydrological Decade representative basin Kassjöån (Waldenström, 1977) in central Sweden, 50 km NW of the city of Sundsvall. The landscape is hilly with elevations ranging from 250 to 440 m a.s.l. The catchment is mainly forested and covered by till soil. The simulation period was September 1971 to August 1981. Precipitation and temperature were measured close to the catchment outlet. Almost complete data series with twice-monthly observations of groundwater levels were available from ten tubes located in different parts of the catchment.

Tärnsjö

The Tärnsjö catchment (Table 2) is located in central Sweden, 50 km NW of the city of Uppsala and about 300 km S of the Lilla Tivsjön catchment. A large esker (ridge of glaciofluvial deposits), rising up to 50 m above the surrounding land, runs through part of the catchment. The remaining part of the mainly forested catchment is covered by till soil. The simulation period was September 1981 to December 1991. Precipitation measurements were available from a station within the catchment whereas temperature was measured about 30 km away from the catchment. Groundwater levels were observed twice monthly at several tubes, from which seven tubes with almost complete data sets were used in this study. It should be noted that some of the tubes were located up to 4 km outside the water divide; however, they were assumed to correspond with the conditions within the catchment.

Initial simulations indicated that the traditional HBV model structure might not be appropriate for the Tärnsjö catchment. Therefore, an alternative model structure was tested as well. The recharge simulated by the soil routine was divided into two parts. A portion C_{PART} [−], related to the portion of the till soil area, was added directly to an exponential storage whereas the remaining recharge generated on one day was evenly distributed over a subsequent period of C_{DELAY} [d] days to a linear storage (Fig. 2). The latter storage was thought to represent the esker in which recharge is delayed because of the large unsaturated zone. A similar approach has been used by Bergström and Sandberg (1983). In the modified model structure, the same number of parameters as in the original version of the model were used. In order to compare the simulations using this modified model structure, with the groundwater observations, the tubes were grouped according to whether they were located on the esker or not and mean time series were computed for both groups. The geometric mean of the r^2 values of the fit for the two series was computed as objective function.

MODEL APPLICATION

The optimisation algorithm was initially tested for calibration against runoff for the theoretical case without any model and data errors. Real precipitation and temperature data were used to generate an 8-year runoff data series by the model. Consequently, for this runoff series a perfect fit was possible and the true parameter values were known. The performance of the algorithm was then tested by calibrating the model 50 times against the synthetic runoff data. To reduce the problem of over-parameterisation the parameters *CWH* and *CFR* were fixed to 0.1 and 0.05 respectively and K_0 and *UZL* were not used, i.e. two instead of three linear equations were used (see Table 1 for parameter definitions). Reducing the number of free parameters from 14 to 10 was assumed not to restrict the general conclusions that could be drawn from this test.

For the theoretical case, the optimisation algorithm was expected always to find the same, i.e. the true, parameter set.

For the real-world case, however, initial calibration tests indicated that a similar goodness of fit was achieved in most calibration trials, but with varying parameter values. The effect of the additional groundwater level data on this parameter uncertainty was investigated using the Lilla Tivsjön catchment. The model was calibrated 25 times to both criteria using three populations of parameter sets as described above. From the 20 best calibration trials the variations of calibrated parameter values were evaluated by the ranges and the standard deviations, and the ratios of those measures of dispersion between the multi-criteria and the single-criterion calibrations were computed for each parameter.

A parameter set found by a multi-criteria calibration will simulate a single variable (in this study runoff or groundwater levels) less accurately than a parameter set which was identified by calibration against this variable only. However, for a good model, which gives an appropriate description of the catchments' hydrological behaviour, this drop in model fit should be small. For both test catchments, the model was calibrated independently to runoff and groundwater levels as well as to the combination of these two criteria. Here, only the best parameter sets found in ten calibration trials were considered.

Results

TEST WITH SYNTHETIC DATA SERIES

For the synthetic data series, only two of the 50 calibration trials failed to provide a good model fit ($R_{eff} < 0.9995$). In the other 48 runs the R_{eff} value was at least 0.99978 with a median of 0.99996. Despite these high R_{eff} values for a few parameters, the true parameter values were not always found. The optimised values for the parameters of the soil routine differed by more than ten per cent from their true values, while these differences were insignificant for the other parameters (Table 3).

PARAMETER UNCERTAINTY

The lowest efficiency obtained after calibrating the model 25 times against runoff was 0.867 and it varied between 0.872 and 0.879 for the 20 best trials. Although the differences in fit between these 20 calibration trials were relatively small, parameter values varied over large ranges. On average these ranges extended over about one third of the feasible ranges as given in Table 1. For the multi-criteria calibration the combined fuzzy measure varied, for the best 20 of 25 trials, between 0.66 and 0.74. The corresponding ranges were 0.821 to 0.834 for R_{eff} and 0.835 to 0.855 for the r^2 of the groundwater levels.

In general, the variations of parameter values found in different calibration trials were considerably smaller when calibrating to both runoff and groundwater levels. The results were similar when comparing ranges and standard deviations (Fig. 3). Most significant was the reduction of parameter uncertainty for some parameters of the response routine, where the variation was only 10 to 30 per cent of the variation of the single-criterion calibration. However, for some parameters (*CFR, FC, MAXBAS*) the variation was between 10 and 80 per cent larger for the multi-criteria calibrations than for the calibrations to runoff alone.

Table 3. Range of calibrated parameter values obtained in the best 48 out of 50 independent calibration trials

Parameter	True value	Calibrated parameter values	
		Min	Max
TT	1	1.00	1.05
CFMAX	3	2.93	3.09
SFCF	0.8	0.78	0.81
FC	250	228	281
LP	0.7	0.61	0.78
BETA	3	2.56	3.58
PERC	0.5	0.47	0.51
K_1	0.3	0.297	0.303
K_2	0.05	0.049	0.053
MAXBAS	2.5	2.48	2.51

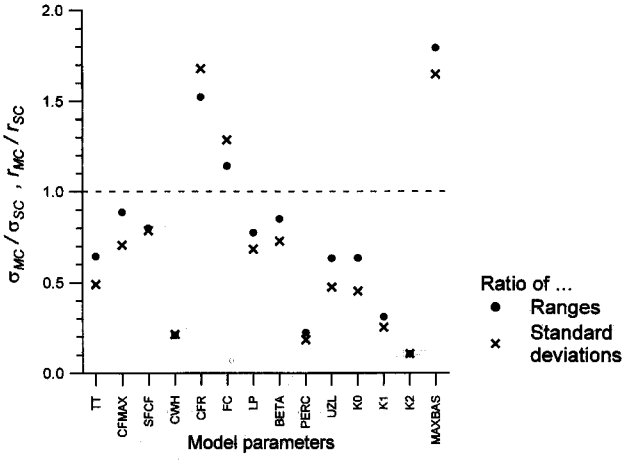

Fig. 3. *Comparison of the variations of calibrated parameter values obtained by single-and multi-criteria calibration of the HBV model in the Lilla Tivsjön catchment. Both ranges and standard deviations were calculated from the 20 best of 25 calibration trials. The ratios were computed by dividing those values for standard deviation and range from calibrations against both runoff and groundwater levels, σ_{MC} and r_{MC}, by the values from calibration against only runoff, σ_{SC} and r_{SC}.*

Table 4. Performance of simulation after calibration according to one and both criteria (measured runoff and groundwater levels). Objective function values which were considered during calibration are written in bold

Model application	Runoff calibration		Groundwater level calibration		Multi-criteria calibration	
	Runoff, R_{eff}	Groundwater levels, r^2	Runoff, R_{eff}	Groundwater levels, r^2	Runoff, R_{eff}	Groundwater levels, r^2
Lilla Tivsjön	**0.879**	0.313	0.649	**0.901**	**0.834**	**0.855**
Tärnsjö	**0.734**	0.412	0.214	**0.837**	**0.677**	**0.720**
Tärnsjö (modified model structure)	**0.762**	0.521	0.435	**0.845**	**0.713**	**0.787**

COMPARISON BETWEEN SINGLE- AND MULTI-CRITERIA CALIBRATIONS

Lilla Tivsjön

A good fit between simulated and observed runoff could be obtained for the Lilla Tivsjön catchment ($R_{eff} = 0.879$). However, when calibrating against runoff, the groundwater level simulations did not fit the observations ($r^2 = 0.313$). The situation was similar when the model was calibrated against groundwater level data, the fit was good for the groundwater levels but poor for runoff (Table 4).

When calibrating the model against both runoff and groundwater simultaneously, the fit between simulations and observations was acceptable for both runoff and groundwater (Fig. 4). The values of the objective functions were about 5 per cent below their values from the single-criterion calibration for both criteria (Table 4). Despite the generally acceptable fit the reproduction of periods with low flow was poor.

Tärnsjö

For the Tärnsjö catchment, the single-criterion calibrations gave poorer fits than in the Lilla Tivsjön catchment (Table 4). Furthermore, the drop in fit when calibrating simultaneously to both runoff and groundwater levels was twice as large as in the Lilla Tivsjön catchment. This difficulty in simulating both runoff and groundwater levels with the same parameter set indicated a major problem in the model structure.

The modified model structure with two parallel groundwater boxes, where one box represented the delayed

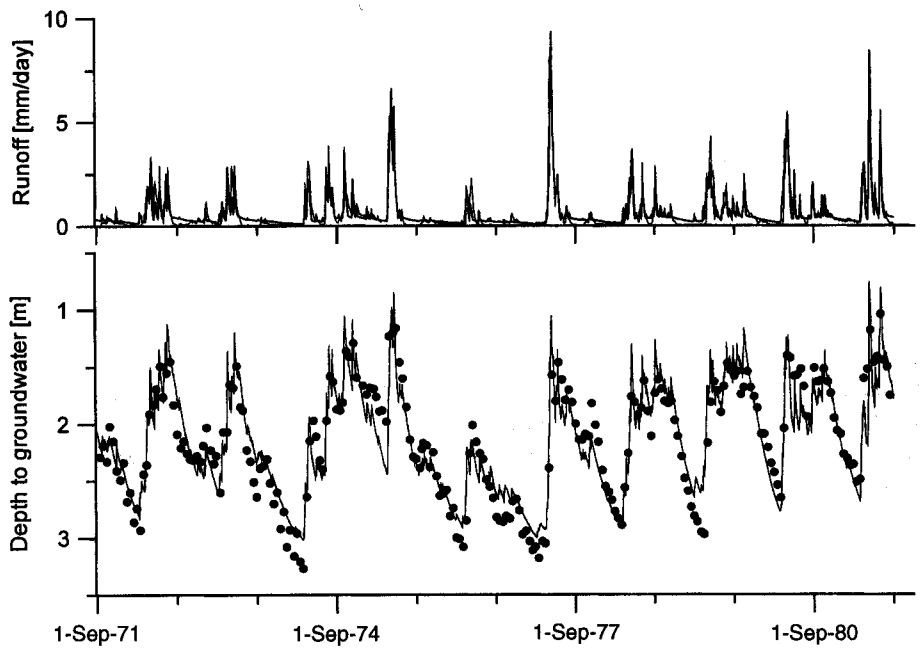

Fig. 4. *Observations and simulations of runoff (thick line = simulated, thin line = observed) and groundwater levels (line = simulated, points = observed) for the Lilla Tivsjön catchment (using the parameter set calibrated to both runoff and groundwater levels).*

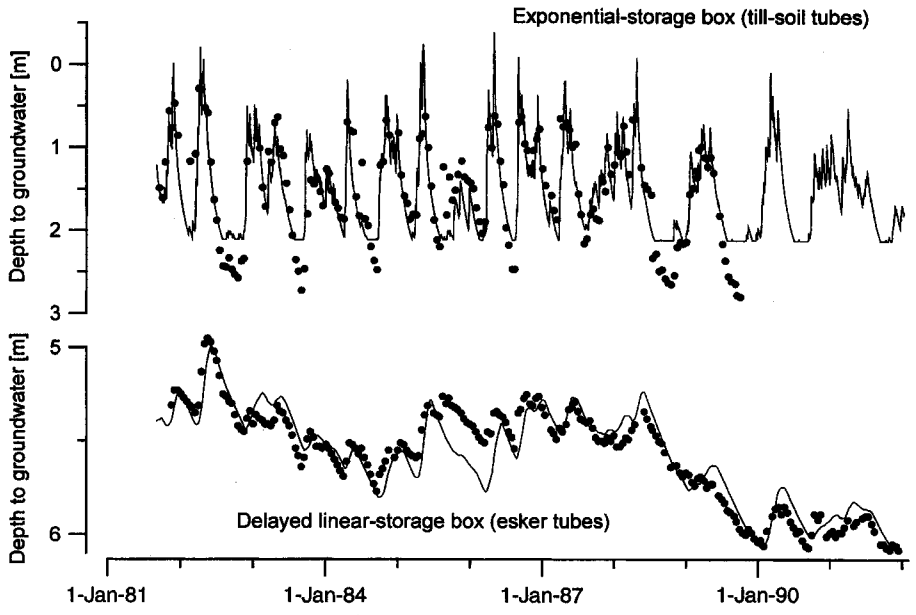

Fig. 5. *Observed (points) and simulated (lines) groundwater levels for the Tärnsjö catchment using the modified model structure (using the parameter set calibrated to both runoff and groundwater levels).*

response of the esker criterion aquifer, gave better results. The fit according to one criterion was slightly better and, most important, the drop of fit in the multi-criteria calibration was smaller and of similar magnitude to the Lilla Tivsjön catchment (Table 4). The groundwater level simulations agreed well with the measurements in the tubes both on the esker and in the till soil (Fig. 5).

There were two apparent deviations between simulated and observed groundwater levels. For the till soil tubes, the model failed systematically to simulate the fall in level during dry conditions, because the simulated storage was already approaching zero at the beginning of these periods. The simulation of levels for the esker tubes that were too high in a half-year period starting in August 1985 was most probably caused by erroneous precipitation data. A large runoff event in August 1985 was also missed totally by the model and did not agree with the precipitation amount used as model input. It is reasonable that a large, but local rainfall event was missed by the measurements because precipitation data were not available from the station at Tärnsjö during this time and stations about 20 km away had to be used instead.

Discussion

For the synthetic data series, the optimisation was in general capable of optimising the parameter values, i.e. the differences between simulated and 'observed' runoff were insignificant. Lidström (1997) proposed an automatic calibration routine for the HBV model, which corresponds in principle with the local optimisation used for the subsequent fine-tuning in this study. He mentioned that for synthetic data an efficiency of 0.998 on average was obtained after several hundred evaluations of the objective function. More model runs were performed using the genetic algorithm, but the deviation from the correct fit ($R_{eff} = 1$) was about 50 times smaller compared to the method of Lindström (1997).

For the synthetic data series, the optimised values of most parameters came very close to the values that were used when generating the series. For the real-world case, optimised parameter values varied over larger ranges although similar values of the efficiency were achieved in most calibration trials. The use of additional data for calibration was expected to reduce this parameter uncertainty. The results obtained in this study agreed with this expectation. The parameter uncertainty decreased for all five parameters of the response routine, which is the part of the model representing groundwater dynamics. Among the other parameters the decrease was less significant and for three parameters the parameter uncertainty even increased. It could be expected that data on snow pack storage or soil moisture might help to constrain the parameters of the snow and soil routine respectively.

Apart from the reduced parameter uncertainty, the multi-criteria calibration is assumed to provide parameter sets that are a more appropriate representation of the catchment than a calibration against runoff alone. During the calibration period, simulated run off will be poor but the internal variables come into much better agreement with the conditions in the catchment. It seems reasonable that this

improved internal consistency is associated with more reliable predictions outside the calibration domain. This assertion has to be tested in future studies using validation periods during which the hydrological conditions differ from those during calibration.

In the case of the Tärnsjö catchment, the additional groundwater data together with the multi-criteria calibration helped to disprove an improper model structure. Obviously, the modification of the model structure could also be motivated by hydrological experience alone. The results of the multi-criteria calibration provided additional and more objective grounds.

The poor reproduction of low flow periods (Fig. 4) is a result of the chosen objective function for evaluation of the agreement between simulated and observed runoff. By using the efficiency, which is based on the sum of squared errors (Eqn. 1), most stress was put on high-flow conditions, whereas deviations during low-flow conditions were of less importance. It would be possible to achieve a better fit for low-flow conditions by using some other objective function (e.g. considering relative, rather than absolute, deviations), but this would cause a poorer fit for high-flow conditions. A combination of different objective functions, which could be implemented similarly to the combination of runoff (R_{eff}) and groundwater levels (r^2) in this study, might be valuable.

The results obtained in this study indicate that the genetic algorithm is capable of optimising the parameters for a conceptual runoff model and that it can easily be extended for multi-criteria calibration. This is supported by Seibert *et al.* (1999) where this algorithm was used to calibrate the HBV model simultaneously with runoff series from nested catchments. Nevertheless, comparisons of the algorithm with other optimisation algorithms remain to be done to assess fully its value in terms of speed and accuracy. Such comparisons are not straightforward because the outcome may depend on the application used for comparison (Whitley *et al.*, 1996). The general assessment of the capabilities of genetic algorithms for optimisation is a topic of active research (Mitchell *et al.*, 1994; Jennison and Sheehan, 1995; Michalewicz and Schoenauer, 1996; Whitley *et al.*, 1996). Usually in hydrological modelling simplified problems are used for comparison. Both Duan *et al.* (1992) and Kuczera (1997) compared the shuffled complex evolution algorithm with other optimisation methods using simplified models with only 6 and 5 parameters respectively. The use of such a model which "... is not intended for use in operational setting" (Duan *et al.*, 1992, p. 1016) is justified since, in models with a larger number of parameters, parameter uncertainty may obstruct the identification of the global optimum. However, results of comparisons between various algorithms may differ in a parameter space of higher dimension. Furthermore, synthetic data series allow for achieving the performance of an optimisation algorithm more directly, but results may differ when using a real-world case.

CONCLUDING REMARKS

In this study, an algorithm for single- and multi-criteria calibration has been proposed for the HBV model which can be implemented easily into other models. It has been demonstrated that the use of additional data, here groundwater levels, can help to constrain the ranges of parameter values. Furthermore, the multi-criteria calibration motivated a modification of the model structure that provided a more realistic representation of the catchment hydrology. In further studies, the genetic algorithm used in this study will be compared to other algorithms and additional data, such as snow cover or soil moisture, will be included in the calibration.

Acknowledgements

Runoff, precipitation and temperature data were measured by SMHI (Swedish Meteorological and Hydrological Institute). The groundwater level data were provided by SGU (Swedish Geological Survey). Allan Rodhe and two anonymous reviewers gave valuable comments on the manuscript.

References

Ambroise, B., Perrin, J.L. and Reutenauer, D., 1995. Multi-criterion validation of a semidistributed conceptual model of the water cycle in the Fecht Catchment (Vosges Massif, France), *Water Resour. Res.*, **31**, 1467–1481.

Beasley, D., Bull, D.R. and Martin, R.R., 1993a. An overview of genetic algorithms: Part 1, Fundamentals. *University Computing*, **15**, 58–69.

Beasley, D., Bull, D.R. and Martin, R.R., 1993b. An overview of genetic algorithms: Part 2, Research Topics. *University Computing*, **15**, 170–181.

Bergström, S., 1976. Development and application of a conceptual runoff model for Scandinavian catchments. *Report No. RHO 7*, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden, 134 pp.

Bergström, S., 1992. The HBV model–its structure and applications. *Report RH No. 4*, Swedish Meteorological and Hydrological Institute (SMHI), Hydrology, Norrköping, Sweden, 35 pp.

Bergström, S., 1995. The HBV model (Chapter 13, pp. 443–476), in: Singh, V.P. (ed.) *Computer models of watershed hydrology, Water Res. Pub.*, Highlands Ranch, Colorado, U.S.A., 1130 pp.

Bergström, S. and Sandberg, G., 1983. Simulation of groundwater response by conceptual models–three case studies, *Nor. Hydrol.*, **14**, 85–92.

Bergström, S., Carlsson, B. and Sandberg, G., 1985. Integrated modelling of runoff, alkalinity, and pH on a daily basis, *Nor. Hydrol.*, **16**, 89–104.

Davis, L., 1991. *Handbook of genetic algorithms*, Van Nostrand Reinhold, New York, 385 pp.

de Groisbois, E., Hooper, R.P. and Christophersen, N., 1988. A multisignal automatic calibration methodology for hydrochemical models: a case study of the Birkenes model, *Water Resour. Res.*, **24**, 1299–1307.

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

Franchini, M., 1996. Using a genetic algorithm combined with a local search method for the automatic calibration of conceptual rainfall-runoff models. *Hydrolog. Sci. J.*, **41**, 21–40.

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

Franks, S., Gineste, Ph., Beven, K.J. and Merot, Ph., 1998. On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resour. Res.*, **34**, 787–797.

Goldberg, D.E., 1989. *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley Publishing Company, Reading, Massachusetts, United States of America, 412 pp.

Holland, J., 1975/1992. *Adaptation in natural and artificial systems*. Cambridge, MA: MIT Press, 211 pp. (First edition 1975, Ann Arbor: University of Michigan Press).

Holland, Ch., Blažková, Š. and Bličik, M., 1994. Application of the ETH Snow Model to three basins of different character in Central Europe, *Nord. Hydrol.*, **25**, 113–128.

Janikow, C.Z. and Michalewicz, Z., 1991. An experimental comparison of binary and floating point representations in genetic algorithms. In: *Proceedings of the Fourth International Conference on Genetic Algorithms*, R.K. Belew and L.B. Booker (eds.) Morgan Kaufmann, 31–36.

Jennison, Ch. and Sheehan, N., 1995. Theoretical and empirical properties of the genetic algorithm as a numerical optimizer. *J. Comput. Graph. Stat.*, **4**, 23.

Kuczera, G., 1997. Efficient subspace probabilistic parameter optimization for catchment models. *Water Resour. Res.*, **22**, 177–185.

Kuczera, G. and Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of data. *Water Resour. Res.*, **34**, 1481–1489.

Lindström, G., 1997. A simple automatic calibration routine for the HBV model. *Nor. Hydrol.*, **28**, 153–168.

Lindström, G. and Rodhe, A., 1986. Modelling water exchange and transit times in till basins using O-18, *Nord. Hydrol.*, **17**, 325–334.

Lindström, G., Johansson, B., Persson, M., Gardelin, M. and Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.*, **201**, 272–288.

Michalewicz, Z. and Schoenauer, M., 1996. Evolutionary algorithms for constrained parameter optimization problems. *Evol. Comput.*, **4**, 32.

Mitchell, M., 1996. *An Introduction to Genetic Algorithms*, The MIT Press, Cambridge, Massachusetts, 205 pp.

Mitchell, M., Holland, J.H. and Forrest, S., 1994. When will a genetic algorithm outperform hill climbing? In: *Advances in Neural Information Processing Systems 6*. J.D. Cowan, G. Tesauro, and J. Alspector (ed), San Mateo, CA: Morgan Kaufmann, 51–58.

Mroczkowski, M., Raper, G.P. and Kuczera, G., 1997. The quest for more powerful validation of conceptual catchment models. *Water Resour. Res.*, **33**, 2325–2335.

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.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T., 1992. *Numerical recipes in FORTRAN: The art of scientific computing*. 2nd edition, Cambridge University Press, Cambridge, Great Britain, 963 pp.

Punch, W.F., 1998. How effective are multiple populations in genetic programming. *Genetic Programming 1998, Proceedings of the Third Annual Conference*, July 22–25, 1998, University of Wisconsin, Madison, Wisconsin. San Francisco, CA, Morgan Kaufmann, 308–313.

Seibert, J., 1997a. Estimation of parameter uncertainty in the HBV model. *Nor. Hydrol.*, **28**((4/5)), 247–262.

Seibert, J., 1997b. *HBV light, version 1.2, User's manual*. Uppsala University, Department of Earth Sciences, Hydrology.

Seibert, J., Uhlenbrook, S., Leibundgut, Ch. and Halldin, S. 2000. Multiscale calibration and validation of a conceptual rainfall-runoff model. *Physics and Chemistry of the Earth*, **25**((1)), 59–64.

Sorooshian, S. and Gupta, V.K., 1995. Model Calibration, In: Singh, V. (Ed.): *Computer models for watershed hydrology*, 23–68, Water Resour. Pub., Highland Ranch, Colorado, USA.

Syswerda, G., 1989. Uniform crossover in genetic algorithms. In J.D. Schaffer (ed.) *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann, 2–9.

Waldenström, A., 1977. Slutrapport över hydrologiska undersökningar i Kassjöåns representativa område. *HB Rapport nr 29*, Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden (in Swedish with English description of the area, English titel: Final report on hydrological studies in the representative basin Kassjöån).

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

Whitley, D., 1994. A genetic algorithm tutorial. *Statistics and Computing*, **4**, 65–85.

Whitley, D., Mathias, K., Rana, S. and Dzubera, J., 1996. Evaluating evolutionary algorithms. *Artif. Intell.* **85**, 245–276.

Yapo, P.O., Gupta, V.K. and Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. *J. Hydrol.*, **204**, 83–97.