Nordic Hydrology, 22, 1991, 15-36
No part may be reproduced by any process without complete reference

Development of a Process Oriented Calibration Scheme for the HBV Hydrological Model

Joakim Harlin

Swedish Meteorological and Hydrological Institute,
Norrköping, Sweden

A process oriented calibration scheme (POC), developed for the HBV hydrological model is presented. Twelve parameters were calibrated in two steps. Firstly, initial parameter estimates were made from recession analysis of observed runoff. Secondly, the parameters were calibrated individually in an iteration loop starting with the snow routine, over the soil routine and finally the runoff-response function. This was done by minimizing different objective functions for different parameters and only over subperiods where the parameters were active. Approximately three hundred and fifty objective function evaluations were needed to find the optimal parameter set, which resulted in a computer time of about 17 hours on a 386 processor PC for a ten-year calibration period. Experiments were also performed with fine tuning as well as direct search of the response surface, where the parameters were allowed to change simultaneously. A calibration period length of between two and six years was found sufficient to find optimal parameters in the test basins. The POC scheme yielded as good model performance as after a manual calibration.

Introduction

Conceptual hydrological models are becoming increasingly used tools in solving practical water resources engineering problems. Advances in computer facilities have enhanced this development and today many models are run on personal computers. Commonly, hydrological models are used for forecasting and for hydrologic design. In Scandinavia for instance, the HBV model (Bergström 1976) is widely spread and has for the last decades been operationally used for flood warning and inflow forecasting to hydro-power reservoirs.

Currently new spillway design guidelines are being adopted in Sweden (Swedish Committee on Spillway Design 1990). The design floods are generated using a hydrological model. Bergström, Lindström and Sanner (1989) and Harlin (1989) describe the methodology and discuss the return periods of the spillway design floods. However, the question of how the model calibration affects the floods has not yet been addressed.

The accuracy of a model output is dependent on the quality of the input data, the model structure and the calibration. In the HBV model, as in all hydrological models, a number of parameters are not directly measurable and have hence to be calibrated. Calibration can be formulated as: to obtain a unique and conceptually realistic parameter set so that the model becomes specific to the system it simulates and performs well. Manual calibration is often a tedious trial and error procedure, whereby the parameters are adjusted by matching the input/output behaviour of the watershed to that of the model. To calibrate the HBV model requires a thorough understanding of the model structure and experience of how the parameters should be changed to achieve an optimal performance. The quality of a manual calibration is often a function of the users knowledge and the time spent calibrating the model.

Unfortunately there has been limited success achieved in relating the parameters of hydrological models to catchment characteristics. Another problem is the lack of representative areal input data. It is also desirable to restrict the number of parameters in a model in order to reduce the data demand and risk of overparameterization. This leads to simplificatons in the description of physical processes and introduces unmeasurable calibration parameters. Even apparently measurable parameters of more complex models often devolve to calibration parameters. It was, for example, found necessary to calibrate the saturated permeability in the ILWAS model when applying it to the Woods Lake catchment in New York, USA (Chen *et al.* 1982). Also in as complex formulations as the SHE model there are inevitably approximations in the representation of the physical processes which lead to calibration parameters (Bathurst 1986). Therefore operational hydrology will have to rely on either manual or automated search techniques for model calibration.

When calibration is done by an automatic algorithm the problem can be formulated as to find those parameter values that maximize or minimize an objective function $OF = f(p1, p2, p3, \ldots, pn)$, where $p1, \ldots, pn$ are the model parameters. OF is a measure of how closely the model-computed runoff compares with the runoff actually measured. In automatic calibration the computational effort is dominated by the cost of evaluating OF for new parameter values, *i.e.* a new model run. Therefore, the strategy is to find the optimum, evalutating OF as few times as possible.

This paper describes a process oriented automatic calibration scheme (POC), developed for the HBV hydrological model. It has been developed using real data on a daily time step. Firstly, a literature review of calibration strategies and commonly experienced problems is given. Secondly, the HBV model and the calibrated parameters are described. Finally, the calibration scheme is presented followed by a discussion of its performance.

Literature Review

Automatic calibration approaches have been extensively discussed in literature. Most commonly the runoff response equations have been studied, but in some cases soil moisture equations have also been included. A popular calibration approach is to use direct search techniques, such as the downhill simplex method (Nelder and Mead 1965), Rosenbrock's coordinate rotation method (Rosenbrock 1960) or Powell's conjugate directions method (Powell 1964).

Ibbitt and O'Donnell (1971) presented a comparison of nine different optimization methods based on experiments performed on the Dawdy and O'Donnell model. They concluded that the decision of the best method depended on what criterion of goodness was used. They selected the Rosenbrock method modified by Ibbitt (1970) as the most efficient one. Johnston and Pilgrim (1976) used the simplex and Fletcher-Powells descent methods in a detailed calibration scheme for the Boughton model. Improvements in the calibration procedure were achieved by modifying the search methods so that the model characteristics were accounted for. Their main problems were: 1) interdependence between model parameters, 2) indifference of the objective function to parameter changes, 3) discontinuities and local optima on the response surface and 4) the ability of the optimization methods to adjust to the response surface being searched. Similar problems were also experienced by Pickup (1977), who tested the efficiency of several calibration algorithms on the Boughton model.

Another calibration strategy is based on trial and error schemes. Sugawara (1979) reported the application of an automatic trial and error calibration method for the Tank model. He used a feedback procedure that evaluated the model performance and divided the total period into subperiods. The subperiods where selected so that the output during each period was governed by one tank. He used volume and shape criteria to adjust the parameters and claimed a high rate of convergence.

Many researchers have developed stochastic estimation procedures, for example Restrepo-Posada (1982), who worked with a simplified version of the NWSRFS model. He pointed out the importance of restricting the parameter number and suggested a modification of the upper zone tension water element so as to make it permanently observable as in the HBV model. Comprehensive studies within this field have also been made by Sorooshian and Dracup (1980), Sorooshian and

Gupta (1983), Gupta and Sorooshian (1985) among others.

Brazil (1989) suggested a three level approach for calibrating the NWS Sacramento Soil Moisture Model. Level one was a guided interactive initial parameter estimator. Level two was an adaptive random search of the parameter space isolating the global optimum. Level three was a fine tuning of the parameters by a pattern search method or a Kalman filtering procedure. He concluded that for most purposes a final calibration result was produced already after level two.

Attempts to automate calibration of the HBV model have also been made. Bergström (1976) used Rosenbrocks method which proved to be able to fit the model rapidly, but he reported on several restrictions which "gave rise to more scepticism than enthusiasm". He listed the following difficulties: choice of objective function, lack of information of the response surface topography and convergence to unrealistic parameters or local optima. Svensson (1977) examined the statistical properties of the residuals of the HBV model in order to guide an automated calibration strategy. He found that the residuals were neither stationary nor independent or normally distributed. After separating the residuals in sets governed by separate processes the autocorrelation was reduced and the residuals became more stationary distributed. These findings together with the long experience of manual calibration at the Swedish Meteorological and Hydrological Institute (SMHI), have formed the base for the calibration methodeology presented in this paper.

Model Structure and Calibrated Parameters

The HBV model was originally developed for use in Scandinavien catchments but has proved to run well in tropical and sub-tropical areas as well, see for example Häggström *et al.* (1990) and Bathia *et al.* (1984). For most applications, the model is run on daily values of rainfall and temperature and monthly estimates of potential evapotranspiration. It consists of routines for snow accumulation and melt, soil moisture accounting, runoff response and, finally, a routing procedure. The model can be used in a distributed mode by dividing the catchment into subbasins. Each subbasin is then divided into zones according to altitude, lake area and vegetation. The snowroutine is based on a degree-day approach and runs separately for each elevation and vegetation zone according to the equation

$$d\mathrm{Melt} = CF\mathrm{MAX}(T-TT) \tag{1}$$

where dMelt is the snowmelt per timestep, CFMAX is the degree-day factor, T is mean air temperature and TT is the threshold temperature for snowmelt and snow accumulation. There is also a general snowfall correction factor (*SFCF*) which adjusts systematic errors in calculated snowfall.

Parameters that were calibrated from the snow routine were *SFCF*, *CF*MAX

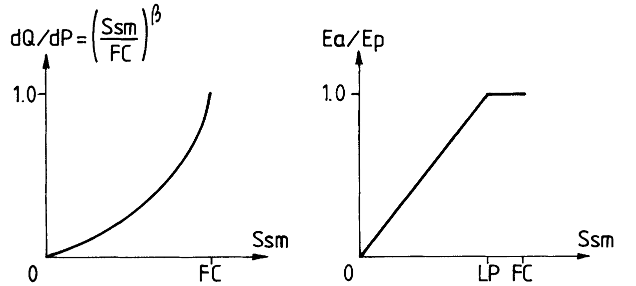

Fig. 1.
Schematic presentation of the soil moisture and evapotranspiration relations in the HBV model.

and *TT*. *TT* was set equal in all vegetation zones, and fixed relations between forest and open area for *SFCF* and *CF*MAX were used.

The soil moisture routine is based on three empirical parameters: β, *FC*, and *LP*, as shown in Fig. 1. β controls the contribution (*dQ*) to the runoff response routine and the increase (1-*dQ*) in soil moisture storage (*Ssm*). *FC* is the maximum soil moisture storage in the model and *LP* is the value of soil moisture, above which evapotranspiration (*Ea*) reaches its potential level (*Ep*). Since mass balance over the soil states: $dQ = dP - dSsm$, soil moisture accounting can be expressed as

$$\frac{dSsm}{dP} = 1 - \left(\frac{Ssm}{FC}\right)^{\beta} \tag{2}$$

The parameters *FC, LP* and β were included in the calibration.

Excess water from the soil is transformed by the runoff-response function. This routine consists of two tanks which distribute the generated runoff in time, so that the quick and the slow parts of the recession are obtained (Fig. 2). The lower tank is a simple linear reservoir representing the contribution to base flow. It also includes the effects of direct precipitation and evaporation over open water bodies in the basin. The lower tank storage (*Slz*) is filled by percolation from the upper tank (*PERC*), and K_2, is the recession coefficient.

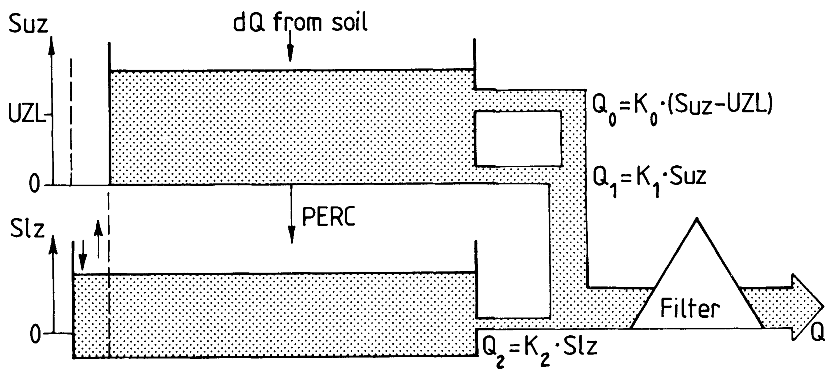

Fig. 2. The runoff-response function of the HBV model.

If the yield (dQ) from the soil moisture routine exceeds the percolation capacity, the upper tank will start to fill. Upper tank storage (Suz) is drained by two recession coefficients K_0 and K_1, separated by a threshold (UZL). This tank models the response at flood periods. Parameters calibrated from the runoff-response function were $PERC$, K_2, K_0, K_1 and UZL.

Finally, runoff is computed independently for each subbasin by adding the contribution from the upper and the lower tank. In order to account for the damping of the flood pulse in the river before reaching the basin outlet, a simple routing transformation is made. This filter has a triangular distribution of weights with the base length MAX*BAS*. Including MAX*BAS* the total number of calibrated parameters amounted to twelve.

Philosophy of the Process Oriented Calibration Scheme (POC)

The philosophy of the calibration scheme has been to utilize the physical representation of the model components and the experience from manual calibrations. This was done by splitting the calibration period into subperiods, within which one specific process dominates the runoff. In this manner the parameters are only evaluated over the subperiods where they are active (Fig. 3). The physical representation of the model components is known, and therefore the calibration should optimize them only when the physical processes they resemble are at hand. This is also an important step in order to avoid the effects of parameter interaction. If the objective function is computed for the whole period, this interaction would create noise on the objective function, with respect to the studied parameter.

By splitting the calibration period, different criterions could be used for different parameters. Since the objective functions are computed only over subperiods, where the current parameters are active a clearer picture of the error caused by each one is received. With this strategy, opposed to direct mathematical optimisation, the final calibration result is not a function of a blunt general fit criterion but a result of sub-optimisation of the different runoff generating procedures. This resembles the strategy of a manual calibration. An experienced hydrologist would calibrate the parameters by visual inspection of the model performance over the different hydrograph components and only use the objective function for the whole period as a guidance.

The different subperiods were compiled by combining the observed temperature and the observed runoff data. This program was made interactive so that the user had the possibility of adjusting the computer-suggested periods. From the duration curve of observed runoff, characteristic high flow (Qh) and baseflow (Qb) discharge limits were estimated. A discharge larger than Qh was regarded as a flood, and a discharge lower than Qb formed base flow. Qh and Qb were found by analyzing the change of slope of the duration curve as illustrated in Fig. 4.

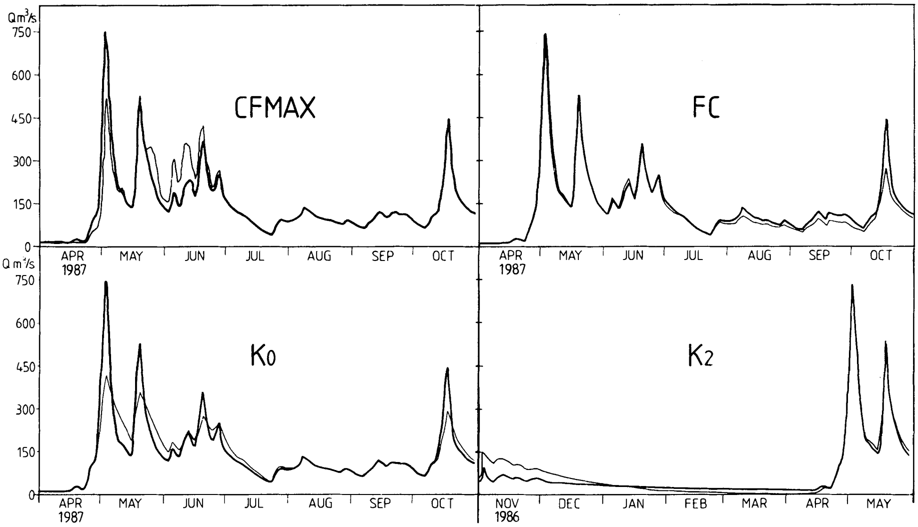

Fig. 3. Example of the active period for the degree-day factor *CFMAX*, the maximum soil moisture storage *FC* and the recession coefficients K_0 and K_2. Thick and thin hydrographs illustrate the effects of changes of the parameter values.

Subperiods dominated by snowmelt floods were found by checking the runoff after cold spells and using *Qb* and *Qh* to follow the floods and define the start and end of them. Warm periods during which runoff was above *Qb* formed the rain flood subperiods. Subperiods dominated by baseflow were compiled by checking when the observed discharge was below *Qb* and so on. Fig. 5 shows an example of how the different subperiods were discriminated.

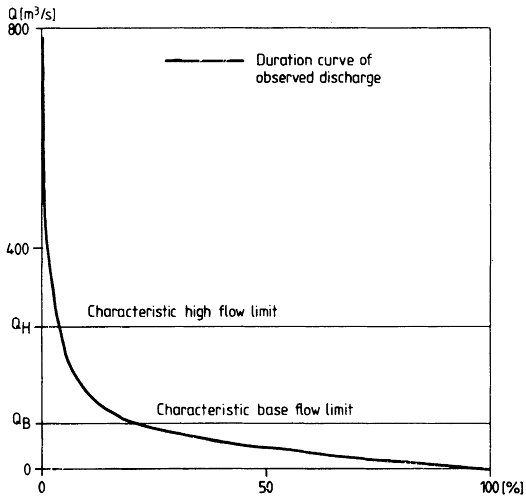

Fig. 4.
Estimation of characteristic high flow and base flow discharge limits from the duration curve.

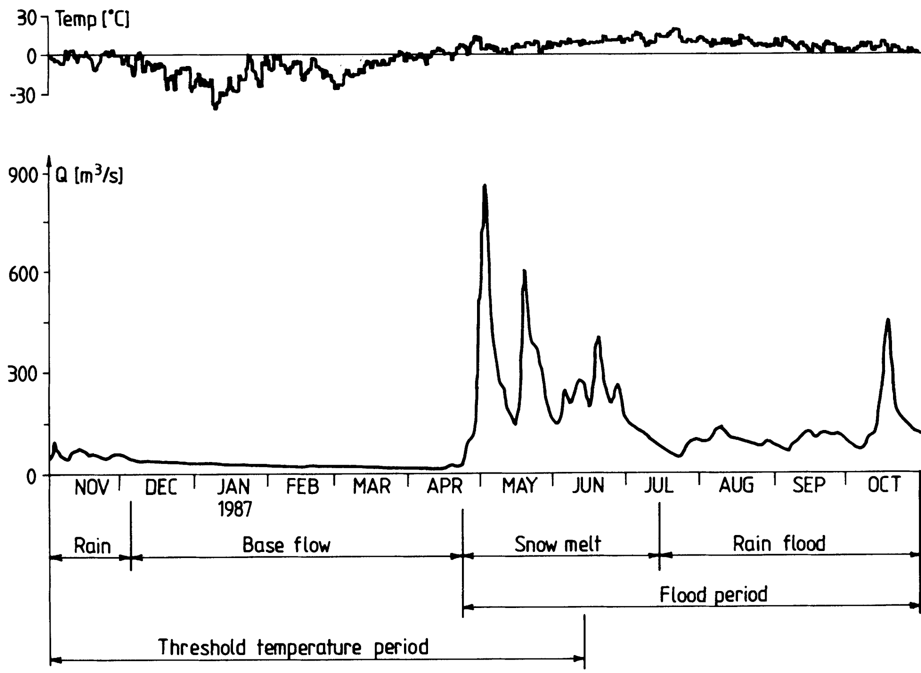

Fig. 5. Splitting of the calibration period in subperiods dominated by one process.

Description of the Process Oriented Calibration Scheme

Step 1 – Initial Parameter Estimation

Initial guesses of the coefficients K_0, K_1 and K_2 were made by recession analysis of the observed runoff larger than Qh, between Qh and Qb and below Qb, respectively, by Eq. (3).

$$K = \frac{1}{dt}\left(\ln Q_t - \ln Q_{t+1}\right) \tag{3}$$

Q_t and Q_{t+1} are discharge values at the time dt apart.

The threshold UZL of the upper tank in the runoff-response function was also estimated by converting the characteristic high flow discharge Qh into a storage level by

$$UZL \equiv \frac{Qh\ 86{,}400}{\text{Area}} \tag{4}$$

where Qh is given in mm^3/s, Area is the catchment area in mm^2 and 86,400 is the number of seconds in 24 hours.

An initial guess of the percolation rate $PERC$ could be computed by the same Eq. (4), only exchanging Qh with Qb. This is motivated by the continuity equation of the lower tank, *i.e.* $dSlz = PERC - Q$. At the beginning of the base flow period $Q = Qb$ and $dSlz$ is zero, $PERC$ equals then Qb.

MAX*BAS* was initially set to one day, and the remaining six parameters were initially set to the middle of their respective ranges found by experience from the large number of manual calibrations of Swedish basins.

Step 2 – Iteration Loop and Criteria of Agreement

An iteration loop was performed over the whole model. The parameters were calibrated one at a time in a set order starting with the snow routine, over the soil routine and finally the runoff-response function. For each subperiod an objective function was computed. These were the mean absolute accumulated volume error, define as

$$MAD \equiv \left|\frac{1}{n}\sum_{t=1}^{n}(Qm(t) - Qo(t))\right| \tag{5}$$

where

n – number of timesteps
Qm – computed discharge
Qo – observed discharge.

This function was used to minimize the volume error of the snowmelt floods. To adjust the phase error of the snowmelt flood start, the mean accumulated absolute error calculated as

$$MABSD \equiv \frac{1}{n}\sum_{t=1}^{n}(|Qm(t) - Qo(t)|) \tag{6}$$

was found most appropriate. Furthermore, the mean square error MSE defined as

$$MSE \equiv \frac{1}{n}\sum_{t=1}^{n}(Qm(t) - Qo(t))^2 \tag{7}$$

and the efficiency criterion R^2 proposed by Nash and Sutcliffe (1970), expressed as

$$R^2 \equiv 1 - \frac{\sum_{t=1}^{n}(Qm(t) - Qo(t))^2}{\sum_{t=1}^{n}(Qo(t) - Qom)^2} \tag{8}$$

where

$$Qom = \frac{1}{n}\sum_{t=1}^{n}Qo(t)$$

were used. MSE and R^2 are equivalent if evaluated over a single period. MSE was used to calibrate parameters active over several subperiods and R^2 was mainly used to evaluate the resulting total model performance.

Normally there were more than one subperiod for each parameter within the calibration period, for example several snowmelt floods. The objective functions were then calculated individually over each subperiod and averaged according to

$$OF = \frac{1}{N} \sum_{i=1}^{N} of_i \tag{9}$$

where

OF – the objective function value for the whole calibration period
N – the number of subperiods
of_i – the objective function value for each individual subperiod.

OF was minimized separately for each parameter with Brents parabolic interpolation method (Brent 1973).

The iteration loop continued until the parameters stabilized, *i.e.* when the R^2 criterion for the whole calibration period stopped changing. Calibration order, objective function and subperiod for each parameter are given in Table 1.

Table 1 – Calibration order, objective functions and subperiods used in the calibration loop

Parameter	Objective function	Subperiods
Snow routine:		
SFCF	MAD	Snowmelt floods
TT	MABSD	Below +2°C*
*CF*MAX	MSE	Snowmelt floods
Transformation function:		
MAX*BAS*	R^2	Whole period
Soil routine:		
FC	MSE	Rain floods
LP	MSE	Rain floods
β	MSE	Rain floods
Upper response tank:		
K_0	MSE	All flood periods
K_1	MSE	All flood periods
UZL	MSE	All flood periods
Lower response tank:		
PERC	MSE	Base flow periods
K_2	MSE	Base flow periods

*) The subperiods for *TT* were all periods where a 14-day moving average of air temperature was below +2°C.

Experiments with Finetuning and Direct Search Methods

Calibration of one parameter at a time has the disadvantage of not taking interdependence between model parameters into explicit consideration. Neither is the information of how parameters, describing the same process, jointly effect the model performance directly utilized. These effects, however, were reduced by calculating the objective function only over subperiods where individual parameters were active and running several loops over all the parameters. The POC yielded parameter values that were considered close to the optimal set. This was checked by a finetuning procedure; a direct search starting from the POC parameters, calibrating several parameters simultaneously.

Powells conjugate directions method (Powell 1964) was chosen since this routine does not require derivatives of the objective function with respect to the parameters and has proved to work well in connection with calibration of hydrological models (Box 1966, Ibbitt and O'Donnell 1971). The method was slightly modified so that the parameter space could be restricted in order to prevent conversion to unrealistic values. Finetuning of all parameters (except MAX*BAS*), simultaneously as well as in subsets for the model routines, was tried.

As an alternative to the POC scheme, experiments with direct search calibration starting from the initial parameter estimates were performed. It was thus possible to compare the accuracy, efficiency and resulting parameter values between the two methods. Direct search calibration was done with Powells method using the mean square error criterion of fit (Eq. (7)).

Results

Calibration Scheme

Fig. 6 gives some key data and examples of the runoff pattern for the three test basins: Torrön, Trängslet and Simlången. These basins represent three different hydrological regimes. Torrön is located in a mountainous region partly above the tree-line. The runoff follows a clear seasonal pattern dominated by large snowmelt floods in spring and rain floods in autumn and sometimes in winter. Trängslet is mainly located in an inland regime and the basin is covered with forest. Also in this basin, runoff has a clear seasonal pattern dominated by large spring floods. The Simlången basin however, belongs to at totally different regime with a humid marine climate producing mainly rain floods. Snowmelt dominated floods seldom occur in this part of Sweden because the winters are shorter and often disturbed by warm periods. Data periods for the test basins were selected so that all changes of the input data stations were avoided.

Model performance in terms of R^2 values, accumulated relative volume errors and volume errors for snowmelt floods only after POC and manual calibration are given in Table 2.

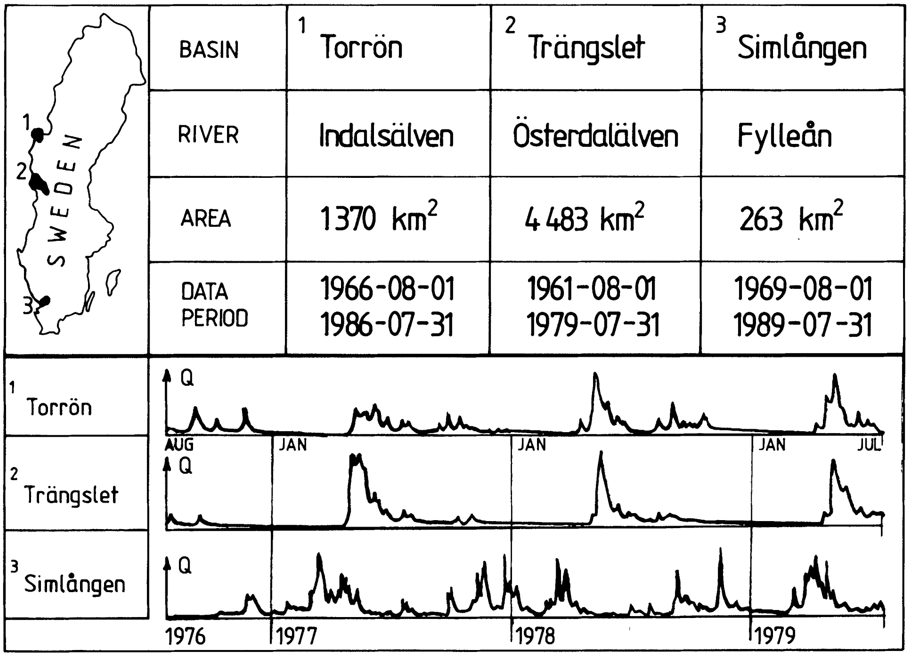

Fig. 6. Key data and examples of the runoff pattern for the test basins.

Table 2 – Model performance after POC and manually calibrated parameters

Basin	Calibration period (POC)	Verification period (POC)	Total period POC	Total period MAN
1. Torrön				
R^2	85.5 (10 years)	84.1 (10 years)	84.8	79.3
VE	0.9	4.2	1.7	3.4
VS	8.0	11.8	10.0	11.8
2. Trängslet				
R^2	94.7 (8 years)	90.7 (10 years)	92.9	92.0
VE	3.5	9.8	4.1	0.8
VS	5.0	10.3	7.8	7.5
3. Simlången				
R^2	89.2 (10 years)	84.3 (10 years)	86.8	83.8
VE	1.9	7.2	4.8	6.6
VS	5.7	11.7	8.7	7.4

R^2 ≡ explained variance (%),
VE ≡ volume error (%), accumulated for all timesteps, and
VS ≡ volume error (%), accumulated over the snowmelt floods.

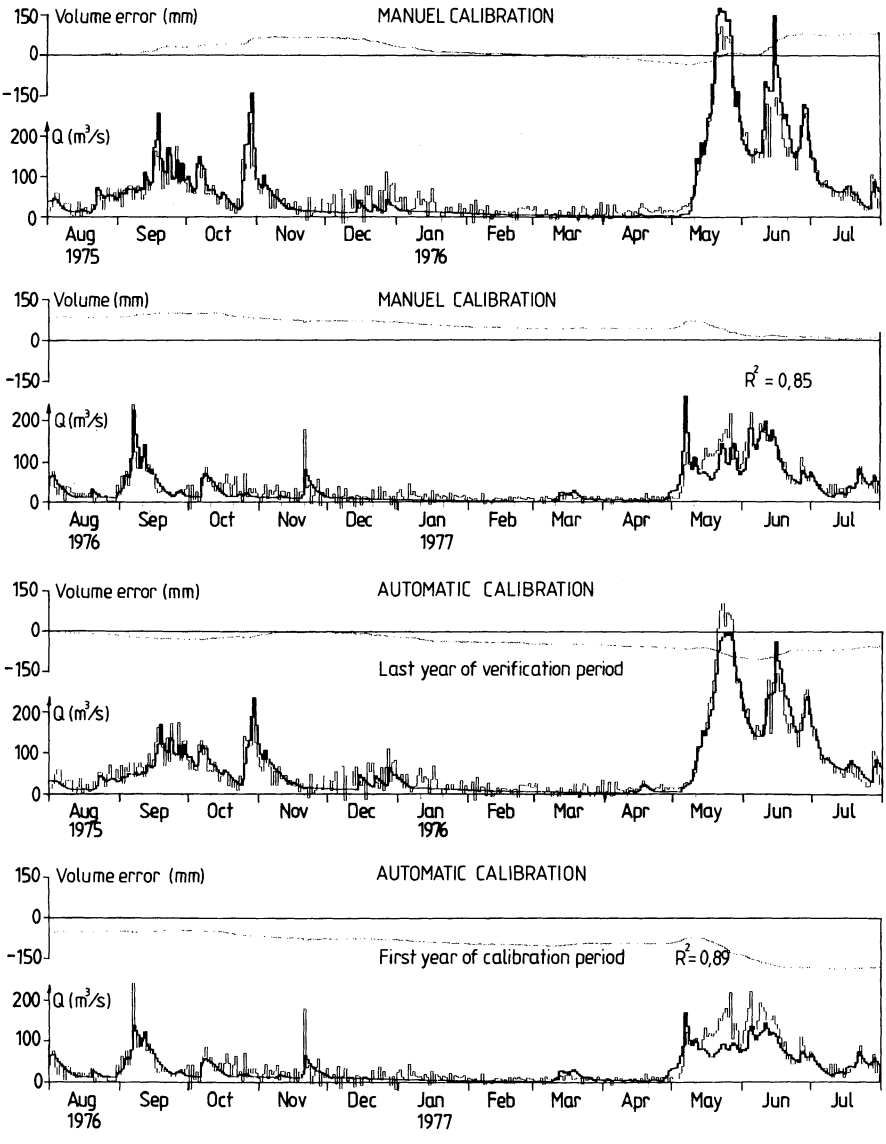

Fig. 7. Example of model performance after manual and automatic calibration (POC), for inflow to the Torrön reservoir. Thick and thin curves show computed and recorded inflow, respectively. The automatic calibration period was 10 years. R^2 values refer to the plotted periods only.

Figs. 7, 8 and 9 show examples of resulting hydrographs after manual and POC calibration. Two hydrological years are depicted for each basin. These figures also illustrate det differences in runoff pattern between the basins.

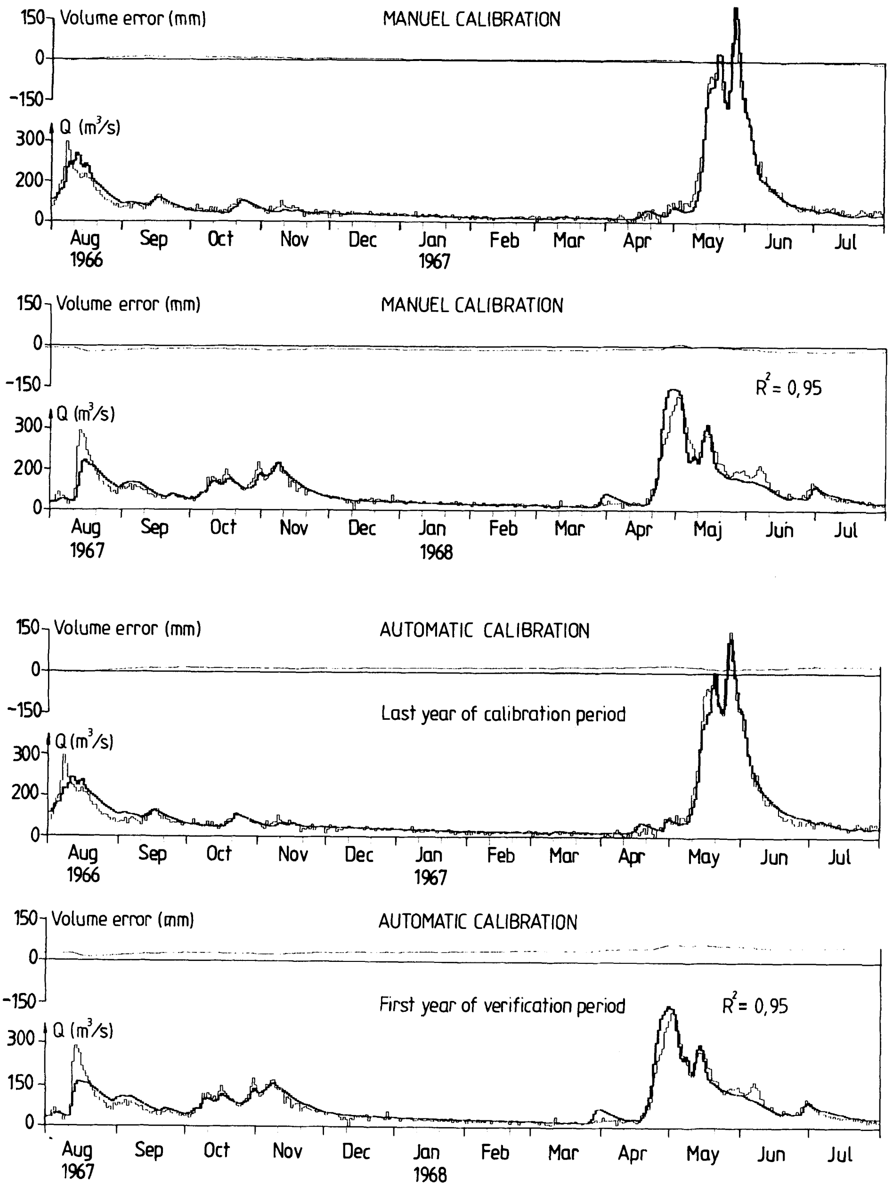

Fig. 8. Example of model performance after manual and automatic calibration (POC), for inflow to the Trängslet reservoir. Thick and thin curves show computed and recorded inflow, respectively. The automatic calibration period was 6 years. R^2 values refer to the plotted periods only.

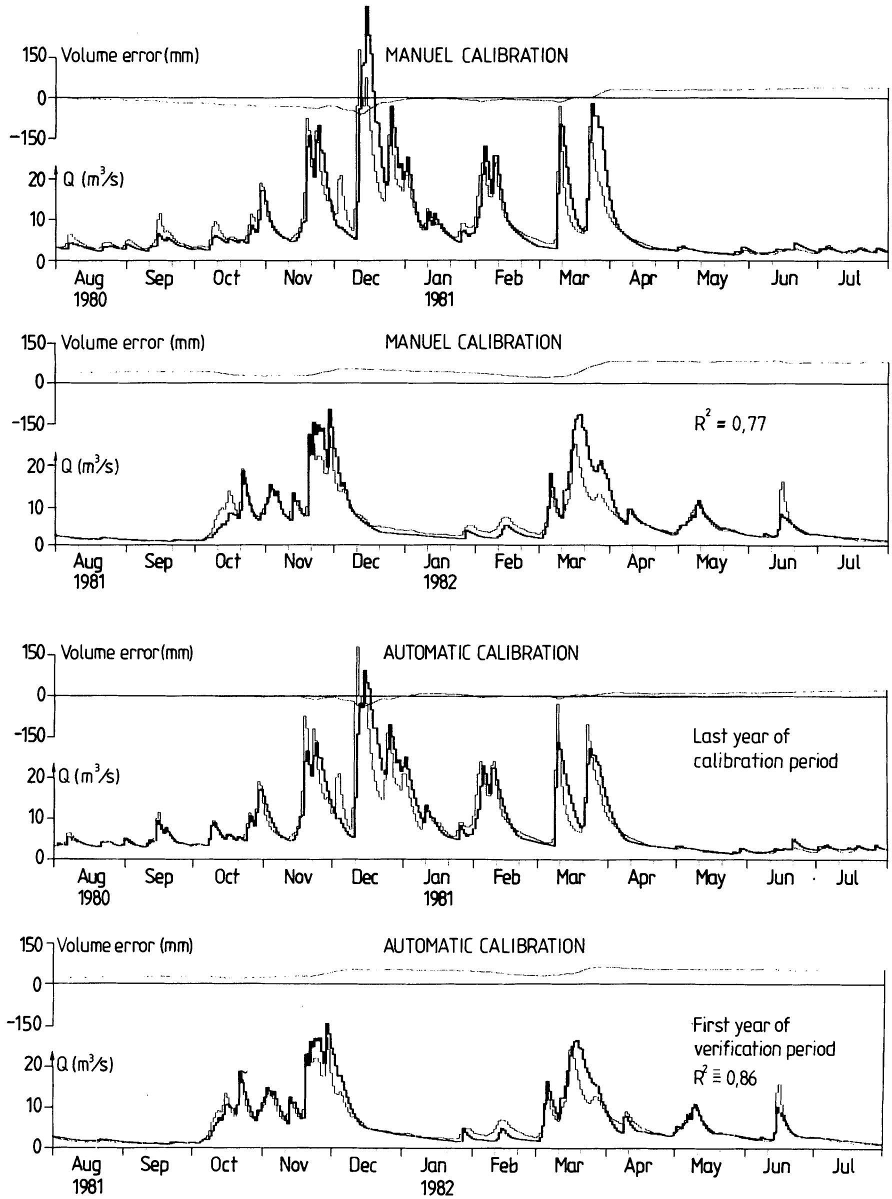

Fig. 9. Example of model performance after manual and automatic calibration (POC), for outflow from lake Simlången. Thick and thin curves show computed and recorded outflow, respectively. The automatic calibration period was 8 years. R^2 values refer to the plotted periods only.

Table 3 – Model performance after calibration by direct search (Powell) and after POC. Performance criteria are compiled for the total data period for each basin; Torrön 20 years, Trängslet 18 years, and Simlången 20 years

Basin	4-year calibration period		6-year calibration period		10-year calibration period	
	Powell	POC	Powell	POC	Powell	POC
1. Torrön						
R^2	83.0	80.0	83.7	83.5	84.0	84.8
VE	2.0	7.3	6.5	3.2	5.0	1.7
VS	9.9	10.4	5.1	12.7	9.8	10.0
2. Trängslet						
R^2	87.2	88.9	89.2	90.9	90.5	92.8
VE	4.8	1.6	5.3	0.4	2.9	4.1
VS	8.1	10.9	7.6	8.6	7.3	7.8
3. Simlången						
R^2	86.9	84.5	89.1	86.3	89.3	86.8
VE	8.0	1.6	0.1	7.7	7.0	4.8
VS	13.2	8.8	6.8	8.2	6.9	8.7

R^2 ≡ explained variance (%),
VE ≡ volume error (%), accumulated for all timesteps, and
VS ≡ volume error (%), accumulated over the snowmelt floods.

Resulting model performances after POC and direct search calibration are given in Table 3.

Computational Speed

Automatic calibration opposed to manual calibration is computer intensive instead of labour intensive. Calibration over a ten-year period would typically take between 15 and 20 hours on a 386 processor PC. For the direct search method, calibration time was about 10 % longer.

Computer time was reduced by looking over the model code and speeding it up, gradually sharpening the termination criteria for each parameter between loops and taking advantage of the model structure. Since the POC calibration order follows the direction of flow through the model, output from calibrated snow parameters could be stored. When calibrating the soil routine these data were read from a file instead of running the snow routine. A similar procedure was employed between the soil routine and the runoff-response function.

Computation speed could have been increased even further by reducing the number of parameters. This would involve reformulation of the model structure. Work along these lines is presently going on for the runoff-response function of the model.

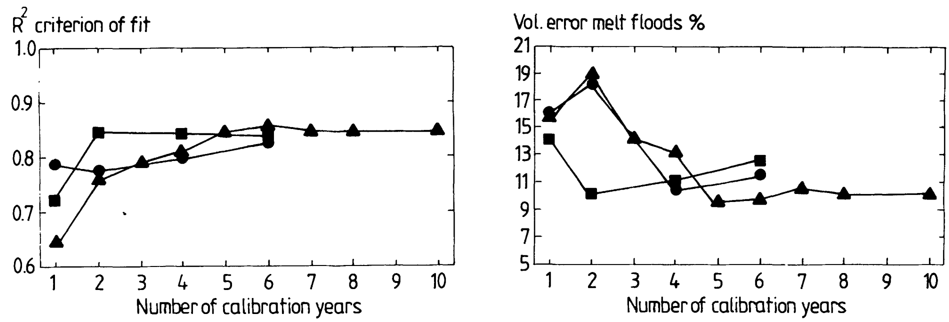

Fig. 10. Model performance after POC, at Torrön over a 20-year period for alternative lengths of the calibration period. The left figure shows the R^2 criterion of fit and the right figure shows relative volume error over snowmelt floods. Squares depict initially wet, circles initially average and triangles initially dry calibration periods.

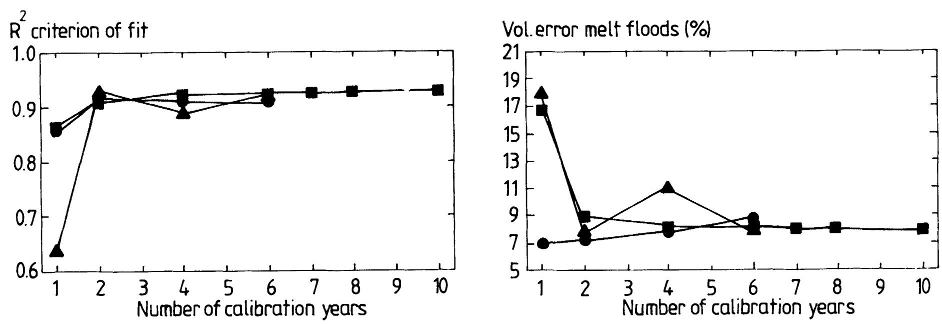

Fig. 11. Model performance after POC, at Trängslet over an 18-year period for alternative lengths of the calibration period. The left figure shows the R^2 criterion of fit and the right figure shows relative volume error over floods with a snowmelt contribution. Squares depict initially wet, circles initially average and triangles initially dry calibration periods.

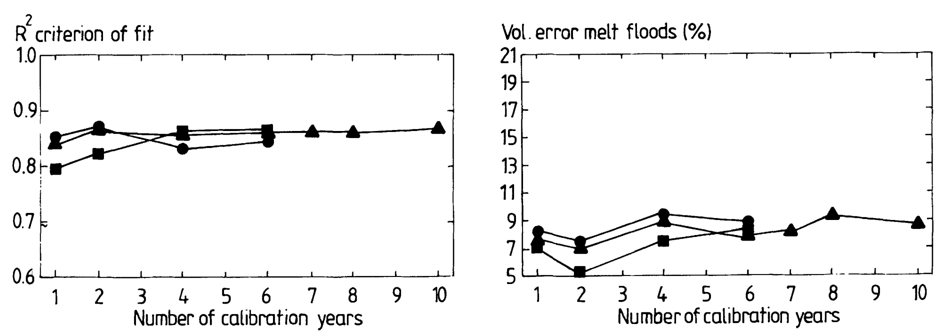

Fig. 12. Model performance after POC, at Simlången over a 20-year period for alternative lengths of the calibration period. The left figure shows the R^2 criterion of fit and the right figure shows relative volume error over snowmelt floods. Squares depict initially wet, circles initially average and triangles initially dry calibration periods.

Optimum Calibration Period Length

An advantage with automatic calibration is that different calibration periods can be tried systematically and objectively. As short calibration period as possible is desirable in order to reduce input data bases and to save computer time. But the shorter the period is, the larger is the risk of overfit, *i.e.* the verification period performance will be poor compared to the calibration period.

In order to find the optimum calibration period length the model was calibrated over alternative periods and verified over the total data period. The total data periods were 20 years for the Torrön and Simlången basins and 18 years for the Trängslet basin. The alternative calibration periods were formed by splitting the total period into three subsets of six years or more, so that one subset started with dry, one with average and one with wet years. To begin with, the first years of each subset were tried, then the following years within the subsets were added to form two-year periods and so on. One 7, 8 and 10-year period was also formed for each test basin.

Figs. 10, 11 and 12 show model performances at different calibration period lengths expressed in R^2 values and relative volume errors over the snowmelt floods. Triangles depict initially dry, circles initially average and squares depict initially wet subsets. After about four years all subsets represent average conditions. Optimum calibration period can be interpreted as the point at which the criterions level out.

Discussion and Conclusions

Calibration Scheme

The most straightforward way to compare calibration results is to compare the quality of the simulations visually. Since the inflow hydrograph contains a large amount of data of a range of types, *e.g.* rain floods, snowmelt floods, baseflow periods *etc.*, it is difficult to find one particular criterion that will objectively show the model performance. The R^2 criterion was chosen, since it shows how much of the initial variance the model explains. The accumulated relative volume error is interesting because it shows the error in water balance over the studied period. The volume error over the snowmelt floods is important when regulating hydropower reservoirs. In most Swedish rivers these floods constitute the majority of the yearly runoff.

The fact that POC gave slightly better model performance than manual calibration in terms of the criterions should not be overemphasized. As was illustrated in Figs. 7, 8 and 9, it is difficult to see the difference in quality between simulations giving different values. The conclusion is rather that POC yields a comparable model performance that is as good as after manual calibration.

Initially guessed parameter values in step one were far from the finally selected but gave fairly good results. The R^2 values from initially guessed parameters over the calibration periods varied between the basins and the length of the periods but were generally in the order of 0.5 to 0.7.

The POC scheme always converged to parameters with an overall good performance. In general three – four loops over the model were sufficient to find the optimal parameter set. In each loop the objective functions were evaluated about 120 times. This behaviour was consistent for all three test basins. The procedure coped well with the errors in inflow data. Observed inflow is normally computed as storage plus release. If the water level is misread or the reservoir oscillates, the storage and therefore also the inflow will be incorrect. These types of errors can clearly be seen in the inflow records for the Torrön reservoir, Fig. 7.

No significant changes in model performance were achieved by the finetuning experiments (less than ± 0.5 % on the R^2 criterion). From this it follows that the POC parameters were very close to an optimum. Unfortunately there are no direct methods of finding the global optimum, if one exist, and a check of the whole parameter space is unfeasible. For example if the twelve parameters that were calibrated could take on only ten values each, and if one evaluation of the objective function only takes one minute, a check of the whole parameter space would take approximately two million years.

The experiments with direct search starting from initially guessed parameters yielded surprisingly good results. The problem of conversion to unrealistic values was overcome by modifying the algorithm, so that the search always stayed within a realistic parameter space. As was shown in Table 3, performance criteria from direct search calibration were generally of the same magnitude as those from POC.

In the Simlången basin however, it was difficult to split the calibration period into subperiods dominated by one process. Perhaps this was the reason why the direct search method performed slightly better in this basin.

Optimal Calibration Period

The intention with trying different calibration periods was to check how efficient the calibration scheme was and to give a rule of thumb of how many years that normally would be needed. Figs. 10, 11 and 12 illustrate when an increase of the calibration period not further contributes to model performance. For the test basins the criterions levelled out at between two and six years. It was also seen that for certain one and two-year periods comparable results to those from ten-year periods were achieved. This indicated that even very short records of streamflow could be very useful in water resources planning if there are longer records of climate data. These records must, however, cover enough hydrological events.

Furthermore, testing different periods and several basins also emphasized the importance of the character of the calibration period and the fact that the amount of information for each parameter is different. The lower tank response function parameters (*PERC* and K_2) were active over long periods of time and very stable. Calibration of snowmelt parameters (*CF*MAX and *TT*), on the other hand, could only be made over a few events each year. This is illustrated by the irregular behaviour of the volume error of snowmelt floods for different calibration periods. Observe that no pure snowmelt floods could be isolated for the Simlången basin (Fig. 12). The volume error depicted for this basin is for floods with a snowmelt contribution.

Some model parameters required a couple of large floods to be correctly tuned, for example the upper tank threshold *UZL* and the highest recession coefficient K_0. At Simlången, floods occur throughout the year thus runoff from this basin contains more hydrological events each year to calibrate against. This is seen by the rapid levelling of the R^2 criterion for increased period lengths.

Parameter Values

In the introduction it was stated that the aim of calibration was to obtain a unique and conceptually realistic parameter set, so that the model becomes specific to the system it simulates. Manual, POC and direct search calibration converged to different parameter combinations, all conceptually realistic and with satisfactory output performance. One should therefore be careful in relating calibrated to measured parameter values. Interesting was that the following behaviour, with respect to parameter values, for all three test basins was observed;

1) Direct search often converged to parameter values close to the initial set. For periods shorter than four years this method sometimes failed to converge.
2) POC was not sensitive to initial parameter values and resulted in values similar to those from manual calibration.
3) In general the most instable parameters were those of the snow and soil routines, in particular the threshold temperature parameter *TT* and the soil parameter β. The upper tank threshold parameter *UZL* and the intermediate recession coefficient K_1, were the most instable parameters of the runoff-response function.
4) Large changes in parameter values were obtained for calibration periods between one and four years. For longer calibration periods the parameter values changed more gradually.

The POC scheme is straightforward, simple and consistently performed well. It was preferred to the direct search method, because it takes advantage of our understanding of the physical system, the model structure and the manual calibration experince. POC offers an automatic objective calibration method which should ensure more homogeneous results in, for example, design flood simulation.

There is always a risk of overemphasizing the aim of matching the model and the catchment response over the calibration period. Calibrating parameters only over time periods, where the physical processes they resemble are dominating, limits the degrees of freedom and reduces the risk of overfit. A model is never perfect, some errors should remain after a properly made calibration!

Acknowledgements

Ideas of how to automate calibraton of the HBV model have developed at SMHI for a long period of time. Almost after every manual calibration these issues are discussed anew. I wish to express my gratitude to my colleagues for sharing their experiences with me. I thank my supervisor, Professor Sten Bergström, who is responsible not only for developing the HBV model but also for many of the ideas behind this paper. Thank you also Göran Lindström, Magnus Persson and Kung Chen-Shan for your advice and constructive criticism.

The financial support from the Swedish River Regulation Enterprises VASO is greatfully acknowledged. Finally I thank all those who have collected and processed the more than 154,000 data observations used in the modelling work of this paper.

References

Bathia, P. K., Bergström, S., and Persson, M. (1984) Application of the distributed HBV-6 model to the upper Narmada basin in India. Report RHO 35, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 40 pp.

Bathurst, J. C. (1986) Physically-based distributed modelling of an upland catchment using the Système Hydrologique Européen, *Journal of Hydrology, Vol. 87,* pp. 79-102.

Bergström, S. (1976) Development and application of a conceptual runoff model for Scandinavian catchments. Report RHO 7, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 134 pp.

Bergström, S., Lindström, G., and Sanner, H. (1989) Proposed Swedish spillway design floods in relation to observations and frequency analysis, *Nordic Hydrology, Vol. 20,* pp. 277-292.

Box, M. J. (1966) A comparison of several current optimization methods, and the use of transformations in constrained problems, *The Computer Jounal, Vol. 9,* pp. 67-77.

Brazil, L. (1989) Multilevel calibration strategy for complex hydrologic simulation models. NOAA Technical Report, NWS 42, U.S. Department of Commerce, 178 pp.

Brent, R. P. (1973) *Algorithms for minimization without derivates,* Prentice-Hall, Inc., Englewood Cliffs, N. J., 195 pp.

Chen, C. W., Dean, J. D., Gherini, S. A., and Goldstein, R. A. (1982) Acid rain model: Hydrologic module, *J. Env. Eng. Div. Am. Soc. Civ. Eng., Vol. 108,* pp. 455-472.

Gupta, V. K., and Sorooshian, S. (1985) The automatic calibration of conceptual catchment models using derivative-based optimization algorithms, *Water Resources Research, Vol. 21, No. 4,* pp. 473-485.

Harlin, J. (1989) Proposed Swedish spillway design guidelines compared with historical flood marks at Lake Siljan, *Nordic Hydrology, Vol. 20,* pp. 293-304.

Häggström, M., Lindström, G., Cobos, C., Martínez, J. R. Merlos, L., Alonzo, R. D., Castillo, G., Sirias, C., Miranda, D., Granados, J., Alfaro, R., Robles, E., Rodríguez, M., and Moscote, R. (1990) Application of the HBV model for flood forecasting in six

Central American rivers. Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 73 pp.

Ibbitt, R. P. (1970) Systematic parameter fitting for conceptual models of catchment hydrology. Ph. D. thesis, Faculty of Engineering, University of London, 400 pp.

Ibbitt, R. P., and O'Donnell, T. M. (1971) Fitting methods for conceptual catchment models, *ASCE proc. Journal of the Hydraulics Division, Vol. 97, No. HY 9,* pp. 1331-1342.

Johnston, P. R., and Pilgrim, D. H. (1976) Parameter optimization for watershed models, *Water Resources Research, Vol. 12, No. 3,* pp. 477-486.

Nash, J. E., and Sutcliffe, J. V. (1970) River flow forecasting through conceptual models. Part I – A discussion of principles, *Journal of Hydrology, Vol. 10,* pp. 282-290.

Nelder, J. A., and Mead, R. (1965) A simplex method for function minimization, *The Computer Journal, Vol. 7,* pp. 308-313.

Pickup, G. (1977) Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models, Hydrological Sciences Bulletin, 22, 2, 6, pp. 257-274.

Powell, M. J. D. (1964) An efficient method for finding the minimum of a function of several variables without calculating derivatives, *The Computer Journal, Vol. 7,* pp. 155-162.

Restrepo-Posada, J. P. (1982) Automatic parameter estimation of a large conceptual rainfall-runoff model: A maximum likelihood approach. Ph. D. thesis, Massachusetts Institute of Technology, 330 pp.

Rosenbrock, H. H. (1960) An automatic method of finding the greatest of least value of a function, *The Computer Journal, Vol. 3,* pp. 303-307.

Sorooshian, S., and Dracup, J. A. (1980) Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, *Water Resources Research, Vol. 16, No. 2,* pp. 430-442.

Sorooshian, S., and Gupta, V. K. (1983) Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resources Research, Vol. 19, No. 1,* pp. 260-268.

Sugawara, M. (1979) Automatic calibration of the tank model, Hydrological Sciences Bulletin, 24, 3, 9, pp. 375-388.

Svensson, S. (1977) A statistical study for automatic calibration of a conceptual runoff model. Report RHO 10, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 74 pp.

Swedish Committee on Spillway Design (1990) Final report, (in Swedish), 103 pp. + 5 appendices.

Received: 21 August, 1990
Accepted: 3 October, 1990

Address:
Swedish Meteorological and Hydrological Institute,
S-601 76 Norrköping,
Sweden.