Journal of Hydrology 257 (2002) 189–211

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Multi-criteria validation of a precipitation–runoff model

Stein Beldring*

Department of Hydrology, Norwegian Water Resources and Energy Administration, P.O. Box 5091, Majorstua, N-0301 Oslo, Norway

Received 2 February 2001; revised 31 August 2001; accepted 12 October 2001

Abstract

The multi-criteria calibration procedure MOCOM-UA was used to evaluate the validity of a physically based precipitation–runoff model by forcing the model to simulate several observed system responses simultaneously. The model is based on kinematic wave approximations to saturated subsurface flow and saturation overland flow at the hillslope scale in a landcape with a shallow layer of permeable deposits overlying a relatively impermeable bedrock. The following results were found; (i) the MOCOM-UA method was capable of exploiting information about the physical system contained in the measurement data time series; (ii) the multi-criteria calibration procedure provided estimates of the uncertainty associated with model predictions and parameters; (iii) multi-criteria calibration constraining the behavior of the precipitation–runoff model to observed runoff and groundwater levels reduced the uncertainty of model predictions; (iv) the multi-criteria method reduced the uncertainty of the estimates of model parameters; (v) the precipitation–runoff model was able to reproduce several observed system responses simultaneously during both calibration and validation periods; and (vi) the groundwater table depths exerted a major control on the hydrological response of the investigated catchment.

Keywords: Precipitation–runoff model; Groundwater; Till deposits; Parameter estimation; Multiple objectives

1. Introduction

Hydrological models are increasingly being used to solve complex problems and synthesize different kinds of information in water resources applications (Sorooshian and Gupta, 1995). Examples include studies of ungauged areas, environmental impacts of land use changes, conjunctive use of groundwater and surface water, and the effects on water resources of anticipated climate change (Refsgaard and Knudsen, 1996). In several of these cases the data required for model calibration may not be directly available, and the required model must perform well under conditions of geographical transposability and non-stationarity as defined by Klemeš (1986). Precipitation–runoff models which combine conceptual descriptions of the flow system with a simplified characterization of the flow domain have proved quite successful when used for operational forecasts of runoff. A severe drawback of these models, however, is that their structure is not directly related to the physical characteristics of the watersheds. Accordingly, it may be expected that their applicability is limited to areas where runoff has been measured for some years and where no significant change of conditions has occurred. Considerable effort has therefore been directed towards the development of physically based models that provide realistic descriptions of the land phase of the hydrological cycle (Refsgaard and Knudsen, 1996). Although these models are probably the best tools presently available for

* Corresponding author. Fax: +47-22-95-92-01.
E-mail address: stein.beldring@nve.no (S. Beldring).

PII: S0022-1694(01)00541-8

complicated tasks (Bathurst and O'Connell, 1992; Refsgaard, 1997), several authors argue that the reliability of physically based hydrological models using effective parameters at the scale of the computational elements can be questioned (e.g. Bergström, 1991; Beven, 1989, 1993). The non-linear nature of the processes involved and the structural heterogeneity of natural systems make it unlikely that the equations of hydrological theories developed at small space and time scale can be generalized to larger scales. It is therefore necessary to evaluate the validity of a model before it is used for purposes, which require that physical processes, state variables and fluxes are correctly described.

Two important aspects of model validation are; (i) scientific evaluation, i.e. the extent to which the model's behavior is consistent with prevailing scientific theory; and (ii) evaluation of operational performance, i.e. the degree to which model-predicted values approach a corresponding set of independently obtained, reliable observations. While general methods of operational evaluation using difference measures are available, scientific evaluation is linked to the specific model or problem at hand and general guidelines are harder to find (Wilmott et al., 1985). In the case of precipitation–runoff models for undisturbed watersheds, scientific evaluation must consider whether the model is able to describe the physical processes which take place in response to rainfall or snowmelt events, and their interactions. This means that model behavior must be examined for each individual sub-system of the watershed for which the model is intended to provide realistic descriptions. A model should therefore only be assumed to be valid, with respect to outputs that have been explicitly validated (Mroczkowski et al., 1997; Refsgaard, 1997). Extended methods of model validation evaluating multiple objectives have been used in several studies. Franks et al. (1998) demonstrated that the percentage of saturated areas in a catchment helped to constrain simulation and parameter uncertainty in an application of TOPMODEL (Beven et al., 1995). Using a hydrosalinity model, Kuczera and Mroczkowski (1998) found that groundwater levels helped little to reduce the parameter uncertainty, whereas stream salinity data substantially reduced the uncertainties. Motovilov et al. (1999) calibrated a precipitation–runoff model using distributed observations of runoff, soil moisture and groundwater levels, and subsequently compared model-simulated values with spatially distributed observations of runoff and evapotranspiration. Gupta et al. (1999) found traditional single-criterion methods to be of limited value when calibrating a complex land surface–atmosphere scheme with many parameters, while forcing the model to match several observed system responses simultaneously constrained the parameter estimates into physically plausible ranges with consequent improvements in model performance.

In order for a precipitation–runoff model to simulate the relationship between input, state variables and output with minimal uncertainty, it is necessary to select appropriate values for the model parameters. Most physically based models have several parameters, which are effective at the scale of the computational elements, and therefore must be calibrated to the observable catchment responses which are the objectives of model predictions (Mroczkowski et al., 1997). In general, there are many combinations of model structures and parameter sets that may be equally good in reproducing observed data, in particular when only one aspect of model performance is considered. This problem arises due to errors in model structure, boundary conditions and observed data (Beven, 1993). If all precipitation–runoff models were scientifically acceptable descriptions of catchment water balance dynamics, one would expect a much smaller set of model structures and parameter sets (Mroczkowski et al., 1997). Since the information contained in the rainfall–runoff relationship is not sufficient to allow identification of complex models, the representation of state variables and fluxes other than runoff must be verified by comparison with the observed data (Jakeman and Hornberger, 1993). Use of state variables such as groundwater levels and soil moisture content during the calibration process may enhance model performance and improve the consistency and stability of parameter estimates (Sorooshian and Gupta, 1995).

The purpose of this study was to evaluate the validity of a precipitation–runoff model which is based on kinematic wave approximations to hydrological processes in hillslopes with a shallow soil cover overlying a relatively impermeable bedrock. The model is formulated at the spatial scale of the physical processes which dominate the conversion of rainfall or snowmelt to catchment runoff (Beldring et al., 2000). The ability of the model to describe hydrological processes and state variables was evaluated by a multi-criteria calibration strategy which constrains the model to several observed system responses simultaneously.

2. Model validation

Precipitation–runoff models which are scientifically valid must employ physical descriptions of the mechanisms of hydrological response that actually occur in a particular watershed (Beven, 1989; Beven and Binley, 1992). For instance, several studies have shown that runoff from small undisturbed catchments in humid temperate environments is dominated by groundwater flow, with a contribution from saturation overland flow and possibly preferential flow during rain or snowmelt events (Rodhe, 1989; Bonell, 1993; Nyberg, 1995). Although a model which applies the concept of infiltration excess overland flow, may predict a particular storm hydrograph using a suitable combination of parameter values, rainfall intensities rarely exceed the infiltration capacity of the soil in these environments, and the model is therefore not based on the relevant mechanisms. Another example of scientifically invalid model structures are those, which describe point values of state variables using lumped, conceptual models. Bergström and Sandberg (1983) were able to reproduce observed groundwater levels with a conceptual model which approximates the soil profile with several linear reservoirs connected through vertical percolation. However, this approach lacks physical realism, as the spatial distribution and downslope flow of subsurface water within the catchments were not described. Neither did this study consider the agreement between model-simulated and observed catchment runoff.

Given that a precipitation–runoff model predicts event response by the correct combination of mechanisms, a second condition of validity is that it accurately predicts all state variables and fluxes which are affected during rainfall or snowmelt events. This implies that the model's operational performance must be evaluated for the relevant aspects of watershed behavior. Klemeš (1986) proposed a hierarchical scheme for validation of operational performance, where a hydrological model is subject to four categories of tests with increasing complexity. The scheme distinguishes between simulations performed for the same catchment used for calibration and for a different catchment, and between stationary and non-stationary conditions. By stationarity it is meant that no significant change in climate, land use or other catchment characteristics occur between the calibration and validation phases (Refsgaard and Knutsen, 1996). Mroczkowski et al. (1997) argued that the lowest level of this scheme (split-sample test) using only streamflow data at the catchment outlet is not an adequate test of model structure or the hypothesis upon which a model is built, while validation based on the model's ability to simulate both catchment runoff and other hydrological processes is a better strategy. The most powerful validation strategy proposed by Mroczkowski et al. (1997) is the use of data from multiple processes in a catchment experiencing a shift in hydrological regime due to disturbance or extreme climatic input.

In order to investigate the validity of a model in terms of its ability to describe watershed behavior, the parameters of the model must be estimated. When the agreement between the model and the real system is close, it may be possible to obtain estimates of some parameters by direct measurements. However, when the agreement is less close, the model parameters can only be viewed as abstract conceptual representations of physical quantities and all that might reasonably be specified is approximate ranges for their values (Gupta et al., 1999). Two approaches to model calibration are in use. One is known as the manual approach, where the values of the model parameters are adjusted in a subjective manner in order to make simulated values of some aspect of watershed behavior, e.g. the shape of the observed hydrograph, resemble observations. The other is known as the automatic approach, where an optimization algorithm is used to find those values of model parameters that minimize or maximize, as appropriate, an objective function or statistic of the residuals between model-simulated output and observed watershed output. Although global optimization methods which search the entire response surface formed by the objective function in the parameter space have been developed (e.g. Duan et al., 1992; Gan and Biftu, 1996), most of these algorithms have difficulty in locating an optimal parameter set due to the existence of multiple local optima (Thyer et al., 1999).

The most commonly used objective criterion of model performance has been some form of weighted function of the residuals between model-simulated output and observed system output (Legates and McCabe, 1999):

$$G(\varphi) = \sum_{t=1}^{n} w_t |O_t - Z_t(\varphi)|^j \tag{1}$$

where O_t is the observed system response at discrete times t, Z_t is the corresponding model output, φ is the vector of model parameters, w_t is the weight at time t, n is the number of data points to be matched and j is a positive integer (usually 1 or 2). If $j = 2$ and all the weights are equal to n^{-1}, G reduces to the mean square error estimator. According to maximum likelihood theory, use of the mean square error estimator is appropriate when the measurement errors of the observations are uncorrelated and have constant variance (Sorooshian and Gupta, 1995; Gupta et al., 1998). The heteroscedastic maximum likelihood estimator has been developed for measuring the closeness between model output and observations when the variance of the measurement errors is assumed to be related to the magnitude of the observations (Sorooshian et al., 1993). Because the structure of the measurement errors is rarely known with observed field data (Freedman et al., 1998), the objective functions selected for this study were not specifically designed for the properties of the errors. In addition, the magnitude of model errors may be equivalent to or even substantially larger than the output measurement errors, and the model errors do not necessarily have any inherent probabilistic structure that can be exploited in the construction of an objective function (Gupta et al., 1998). It has not been proved possible to demonstrate that a particular objective function is better suited for calibration of a hydrological model than some other. Yapo et al. (1996) found that the performance of a precipitation–runoff model was more consistent over all flow ranges when automatic calibration using the heteroscedastic maximum likelihood estimator was performed, whereas the mean square error estimator resulted in better fitting of above-mean flows.

In order to assess model performance more precisely it is useful to consider two or more objective criteria. However, several of these measures are related, suggesting that they measure similar characteristics of the discrepancy between a model and the modeled quantity. The objective functions used in this study were the Nash–Sutcliffe and bias statistics of the residuals, which have a low correlation (Gupta et al., 1998; Węglarczyk, 1998). The Nash–Sutcliffe efficiency criterion ranges from minus infinity to 1.0 with higher values indicating better agreement. It measures the fraction of the variance of observed values explained by the model (Nash and Sutcliffe, 1970):

$$\mathrm{NS} = 1 - \frac{\sum_{t=1}^{n} (O_t - Z_t)^2}{\sum_{t=1}^{n} (O_t - O_{\mathrm{mean}})^2} \tag{2}$$

where O_{mean} is the mean of the observed values. Bias measures the tendency of the model-simulated values to be larger or smaller than their observed counterpart (Yapo et al., 1996):

$$\mathrm{BIAS} = \frac{\sum_{t=1}^{n} (Z_t - O_t)}{\sum_{t=1}^{n} O_t} \tag{3}$$

Since single-criterion calibration strategies consider only one component of watershed behavior, they are unable to find optimal parameter sets for hydrological models with multiple outputs. Because many models employ distributed representations of the watershed, state variables and output fluxes may be simulated and measured at numerous locations. In this case, it is necessary to consider simultaneously the objective measures of several aspects of model performance (Yapo et al., 1998). One classical approach to the problem of using multiple measures of model performance is to make some assumption that permits combining them into a single index, for instance by assigning weights to the individual measures (e.g. Lamb et al., 1998). In general, there is no reasonable basis for the assignment of these weights, and any attempt to convert the objective measures into a single index involves some degree of subjectivity (Gupta et al., 1998).

An algorithm for multi-criteria calibration of physically based models, which is called the multi-objective complex evolution (MOCOM-UA) method, was presented by Yapo et al. (1998) and Gupta et al. (1998, 1999). Consider a model with a vector of parameters φ which is to be calibrated using time series observations collected on r different simulated response variables ($Z_i(\varphi, t)$, $i = 1, \ldots, r$). A separate criterion G_i is defined to measure the difference between each model-simulated response Z_i and the corresponding observations O_i. The purpose of multi-criteria calibration is to find the values for φ within the feasible set of parameters Φ that simultaneously optimize all criteria. An important characteristic of the multi-criteria problem is that in general, it does not, have a unique solution. Because of errors in the model structure and data, it is usually not possible to find a single point φ at which all the criteria have their optimal value. Instead, it is common to have a set of solutions, with the property that moving from one solution to another results in the improvement of one criterion, while causing deterioration in another. Within the set S of solutions to the multi-criteria optimization problem, no point is superior to any other. A particular point may be superior to others for one or more criteria, but it is inferior to them for at least one other criterion. On the other hand, every point within S is superior to all points outside S for all G_i. The set S of solutions is called the non-dominated set or Pareto set. The Pareto set represents the minimal uncertainty that can be achieved for the parameters via calibration, without subjectively assigning relative weights to the individual model responses. The MOCOM-UA method begins by uniformly sampling the feasible parameter space Φ at a number of locations and then uses a multi-criteria population evolution strategy to drive this population of sample points toward the Pareto set. The final solution therefore consists of a set of randomly distributed points, which approximately represents the Pareto set.

The size and characteristics of the Pareto space of parameters and of model output provide useful information about the limitations of the model. A systematic nature of discrepancy between the Pareto space of model-simulated output and the observed system output might suggest a deficiency in the model structure, while a large Pareto range in some of the parameters might suggest that the deficiency lies primarily in the structural representation of the model associated with those parameters. Any parameter set chosen from within the Pareto set is a good solution in the sense that it provides a certain trade-off between the objectives used to measure model performance. Any parameter set chosen from outside the Pareto set is a bad solution in the sense that it will have worse values for all objectives than any point within the Pareto set. The Pareto solution space translates into an uncertainty range in model predictions, where simulations based on different parameter sets matches the observed data in different ways (Yapo et al., 1998; Gupta et al., 1998).

In order to use the MOCOM-UA method, it is necessary to select the criteria that are used to extract the information contained in the different observed time series and transform it into estimates for the model parameters. In this study, the Nash–Sutcliffe statistic is used as the objective function of MOCOM-UA during model calibration, and as a test of its performance, while the bias statistic is only used as a test of model performance. This approach ignores the statistical properties of the measurement errors, while attempting to drive the residuals between simulated and observed system output as close to zero as possible. Unlike several studies which undertake some form of Monte Carlo sampling (Melching, 1995) to generate a large number of parameter sets (e.g. Freer et al., 1996), a subjective threshold in the objective function values is not required to separate acceptable and unacceptable model simulations.

3. Precipitation–runoff model

The most general way to model a catchment's response to rainfall or snowmelt events is to use the complete equations of saturated and unsaturated subsurface flows, overland flow and open channel flow. This involves specification of the governing laws of mathematical physics, the geometry of the system, sources and sinks and initial and boundary conditions. In general, for any water resources system the governing equations are the law of conservation of mass and a flux law (Singh, 1996). In addition, a description of the various hydrological and radiative processes at the land surface–atmosphere interface is necessary in order to include evapotranspiration and snow storage in the model.

Table 1
Parameters of the precipitation–runoff model

Optimized parameters	Description	Lower bound	Upper bound	Unit
a	Determines the rate of decrease of $K(u)$ with depth u	-17.0	-0.1	m^{-1}
C_{act}	Actual evapotranspiration factor	0.1	0.99	
C_{pot}	Potential evaporation factor	10^{-5}	10^{-4}	m/(h hPa)
I_{max}	Interception storage capacity	5×10^{-4}	3×10^{-3}	m
K_0	Saturated hydraulic conductivity at soil surface	10^{-6}	10^{-2}	m/s
m	Overland flow kinematic wave exponent	1.0	1.9	
z_r	Depth of root zone	0.05	0.7	m
β	Overland flow kinematic wave friction parameter	20	140	m/h
δ	Partitions E_{act} between saturated and unsaturated zones	-17.0	-0.1	m^{-1}
ϵ	Storage coefficient of saturated zone	0.03	0.2	
λ	Relates soil moisture content to groundwater level	-17.0	-0.1	m^{-1}

Fixed parameters	Description	Value
A	Albedo of snow surface	0.9
C_{freeze}	Refreeze factor of meltwater in snow	0.05
C_{rad}	Radiation melt factor of snow (inverse of latent heat of fusion)	$(3.34\times10^5\ \text{J/kg})^{-1}$
C_{temp}	Temperature melt factor of snow	5×10^{-5} m/(h °C)
D	Thickness of soil profile, measured orthogonal to bed	0.8 m
L	Hillslope length	300 m
R_{liq}	Relative water holding capacity of the snow	0.08
T_{acc}	Threshold temperature of snow accumulation	0 °C
T_{melt}	Threshold temperature of snow melt	0 °C
α	Hillslope angle	9.4°
θ_{sat}	Saturation volumetric water content at the soil surface	0.7
θ_{wp}	Volumetric water content at the wilting point of vegetation	0.06

Kirkby (1988) suggested that satisfactory event models of small catchments could be developed by considering vertical unsaturated flow and downslope saturated subsurface flow and saturation overland flow on a two-dimensional hillslope strip. A precipitation–runoff model based on these simplifications was presented by Beldring et al. (2000). Kinematic wave approximations were used for describing saturated subsurface flow and saturation overland flow at the hillslope scale in a landscape with a shallow layer of permeable deposits overlying a relatively impermeable bedrock. The model assumes that water infiltrating through the soil surface reaches the groundwater table as soon as the soil moisture deficit in the root zone is replenished, saturated subsurface flow occurs as potential flow parallel to the sloping bed, while saturation overland flow develops due to water input from precipitation or snowmelt when the entire soil profile is saturated. Preferential and return flows are not considered, neither is the downslope unsaturated flow. The governing equations of saturated subsurface flow and saturation overland flow are solved using the method of characteristics (Singh, 1996). The displacement of points on the groundwater table or the overland flow profile due to spatially uniform water input i are described along characteristic curves in the three-dimensional space of length coordinate, time and saturated depth. The precipitation–runoff model has 23 parameters, 12 of these were fixed at specified values, while 11 were determined using the MOCOM-UA algorithm. Table 1 defines all model parameters and their upper and lower bounds if they were subject to optimization, or fixed values. All other symbols used in Eqs. (4)–(27) below are defined in Table 2.

Saturated hydraulic conductivity decreases with depth measured from the soil surface and orthogonal to the impermeable bed:

$$K(u) = K_0\, e^{au} \tag{4}$$

Table 2
Symbols other than parameters used for describing precipitation–runoff model in Eqs. (4)–(27)

Symbols	Description	Unit
E_{act}	Actual evapotranspiration	m/s
E_{pot}	Potential evaporation	m/s
f_{sat}	Fraction of evapotranspiration consumed from the saturated zone	
f_{uns}	Fraction of evapotranspiration consumed from the unsaturated zone	
F	Refreeze rate of liquid water in snow	m/s
h	Vertical depth to the groundwater table	m
i	Lateral inflow to saturated zone flow or overland flow	m/s
$K(u)$	Saturated hydraulic conductivity at depth u below soil surface	m/s
l	Length coordinate along hillslope	m
M	Snowmelt rate	m/s
p	Overland flow discharge per unit width of hillslope	m^2/s
P_{def}	Vapour pressure deficit of the air	hPa
$q(s)$	Saturated zone discharge per unit width of hillslope	m^2/s
Q_{glsw}	Global shortwave radiation	W/m^2
s	Saturated zone depth, measured from and orthogonal to bed	m
t	Time	s
T	Air temperature	°C
u	Depth below soil surface, measured orthogonal to bed	m
V_{def}	Soil moisture deficit of root zone	m
x	Length coordinate of overland flow	m
y	Overland flow depth	m
θ_0	Volumetric water content at the soil surface	
θ_{e0}	Equilibrium volumetric water content at the soil surface	

Saturated subsurface flow is described by Eqs. (5)–(11). The time, length coordinate, saturated depth and discharge of a characteristic curve as it starts traversing the hillslope are given by t_0, l_0, s_0 and q_0.

$$s(t) = s_0 + \frac{i}{\epsilon}(t - t_0) \tag{5}$$

$$s(l) = D - \frac{1}{a}\ln\left[e^{a(D-s_0)} - \frac{ai}{K_0 \sin\alpha}(l - l_0)\right] \tag{6}$$

$$q(s) = \frac{K_0}{a}\sin\alpha\, e^{aD}(1 - e^{-as}) \tag{7}$$

$$q(t) = q_0 + \frac{K_0 \sin\alpha}{a}e^{a(D-s_0)}\left[1 - \exp\left(\frac{-ai}{\epsilon}(t - t_0)\right)\right] \tag{8}$$

$$q(l) = q_0 + i(l - l_0) \tag{9}$$

$$l(t) = l_0 + \frac{K_0 \sin\alpha}{ai}e^{a(D-s_0)}\left[1 - \exp\left(\frac{-ai}{\epsilon}(t - t_0)\right)\right] \tag{10}$$

When $i = 0$, Eq. (10) is replaced by:

$$l(t) = l_0 + \frac{K_0 \sin\alpha}{\epsilon}e^{a(D-s_0)}(t - t_0) \tag{11}$$

Saturation overland flow is described by Eqs. (12)–(18). The time, length coordinate, overland flow depth and discharge of a characteristic curve as it starts traversing the saturated part of the hillslope are given by t_0, x_0, y_0 and p_0.

$$y(t) = y_0 + i(t - t_0) \tag{12}$$

$$y(x) = \left[\frac{i}{\beta}(x - x_0) + y_0^m\right]^{1/m} \tag{13}$$

$$p(y) = \beta y^m \tag{14}$$

$$p(t) = p_0 + \beta[y_0 + i(t - t_0)]^m - \beta y_0^m \tag{15}$$

$$p(x) = p_0 + i(x - x_0) \tag{16}$$

$$x(t) = x_0 + \frac{\beta}{i}[y_0 + i(t - t_0)]^m - \frac{\beta}{i}y_0^m \tag{17}$$

When $i = 0$, Eq. (17) is replaced by:

$$x(t) = x_0 + \beta m y_0^{m-1}(t - t_0) \tag{18}$$

Precipitation accumulates as snow when the air temperature is below a threshold value T_{acc}. The rate of snowmelt for air temperatures exceeding T_{melt} is calculated by:

$$M = C_{temp}(T - T_{melt}) + (1 - A)C_{rad}Q_{glsw} \tag{19}$$

Meltwater is retained in the snowpack until the amount of liquid water exceeds the relative water holding capacity of the snow, R_{liq}. When the temperature is below T_{melt}, liquid water in the snowpack will refreeze at a rate:

$$F = C_{freeze}C_{temp}(T_{melt} - T) \tag{20}$$

Snow storage and snowmelt are assumed to be uniform along the hillslope. The fraction of precipitation lost to interception storage evaporates at the potential rate:

$$E_{\mathrm{pot}} = C_{\mathrm{pot}} P_{\mathrm{def}} \tag{21}$$

During dry periods without interception storage, soil evaporation and transpiration from vegetation occurs at a rate:

$$E_{\mathrm{act}} = E_{\mathrm{pot}} \qquad \text{for } \theta_0 \geq C_{\mathrm{act}}(\theta_{\mathrm{sat}} - \theta_{\mathrm{wp}}) \tag{22}$$

$$E_{\mathrm{act}} = E_{\mathrm{pot}} \frac{\theta_0 - \theta_{\mathrm{wp}}}{C_{\mathrm{act}}(\theta_{\mathrm{sat}} - \theta_{\mathrm{wp}}) - \theta_{\mathrm{wp}}} \tag{23}$$

for $\theta_0 < C_{\mathrm{act}}(\theta_{\mathrm{sat}} - \theta_{\mathrm{wp}})$

When saturation from below occurs, evaporation is consumed from the overland flow profile at the potential rate before water is extracted through soil evaporation and transpiration. Between precipitation events, soil moisture in the unsaturated zone is assumed to be in a state of hydrostatic equilibrium, which is determined by soil characteristics and depth to the groundwater table. Due to soil evaporation and extraction of water by plants, a deficit relative to the equilibrium state develops in the root zone. Neglecting the tension-saturated zone, a simple expression for the equilibrium water content at the soil surface is:

$$\theta_{\mathrm{e}0} = \theta_{\mathrm{sat}}\, \mathrm{e}^{\lambda h} \qquad \text{for } \lambda < 0 \tag{24}$$

Assuming, that the difference between the equilibrium water content and the actual water content is constant through the entire root zone, the soil moisture deficit is:

$$V_{\mathrm{def}} = (\theta_{\mathrm{e}0} - \theta_0) z_{\mathrm{r}} \tag{25}$$

The water raised to the soil surface by soil evaporation and transpiration from vegetation is consumed from both the unsaturated and the saturated zones, depending on root water uptake and capillarity. As long as the groundwater table is within the root zone, water is extracted from the saturated zone only, and the soil moisture deficit is zero. Below the root zone, the fraction of water consumed from the saturated zone decreases with increasing depth to the groundwater table, while the fraction of water extracted from the unsaturated zone increases:

$$f_{\mathrm{sat}} = \mathrm{e}^{\delta(h - z_{\mathrm{r}})} \qquad \text{for } h > z_{\mathrm{r}},\ \delta < 0 \tag{26}$$

$$f_{\mathrm{uns}} = 1 - \mathrm{e}^{\delta(h - z_{\mathrm{r}})} \text{ for } h > z_{\mathrm{r}},\ \delta < 0 \tag{27}$$

The precipitation–runoff model treats evaporation from interception storage as uniform along the hillslope, while a representative value of evaporation from the overland flow profile, soil evaporation and transpiration from plants is calculated by integrating the contributions from the different parts of the hillslope. With the groundwater table below the soil surface, precipitation or snowmelt in excess of interception loss is assumed to infiltrate. The soil moisture deficit must be zero before water percolates to the saturated zone. When saturation from below occurs, all precipitation or snowmelt in excess of interception loss contributes to overland flow. Hillslope average inflow i used in Eqs. (5)–(18) is given by the mean of the contributions from the different parts of the hillslope.

4. Study area

Data from the Sæternbekken catchment in southeast Norway were used in this study. It has an area of 6.32 km^2 and covers altitudes ranging from 110 to 422 m above sea level. Approximately 90% of the catchment is covered by coniferous forest, mostly Norway spruce. The remaining 10% consists of bogs and a small fraction of cultivated areas. There are no lakes of significant size in the catchment. The dominating surface deposits are glacial tills with a thickness which varies from approximately one meter in the valley bottoms to nearly zero on the top of ridges where exposed bedrock is often found. Some small areas with glaciofluvial sediments are located near the catchment outlet. The bedrock is dominated by intrusive and extrusive igneous rocks of Permian age (Erichsen and Nordseth, 1985). The shallow surface deposits of the catchment combined with the small storage coefficient of the tills lead to rapid responses of groundwater levels and runoff to rain or snowmelt events. The till deposits are unable to maintain permanent groundwater storage during dry conditions, and in summer the saturated zone may be

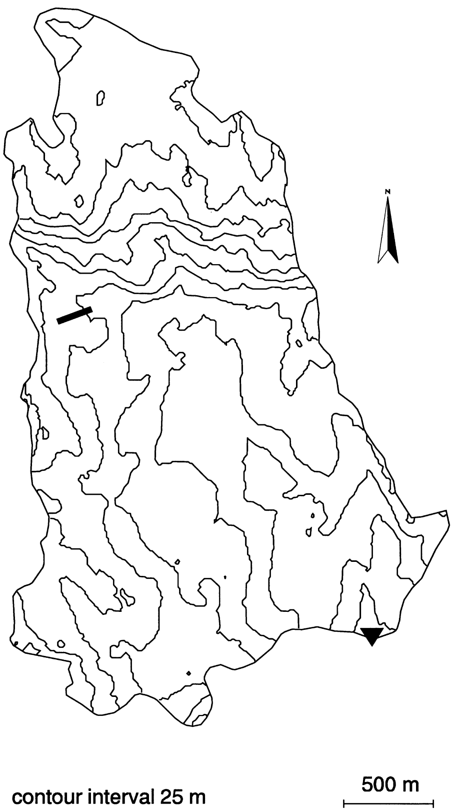

Fig. 1. The Sæternbekken catchment. The triangle shows the catchment outlet (59°57′N, 10°34′E). The hillslope where groundwater levels and meteorological data were measured is shown by a line.

absent in almost the entire catchment (Udnæs, 1991; Myrabø, 1997).

Mean annual precipitation measured at an altitude of 205 m above sea level within the catchment is 850 mm, with maximum occurring during the months August–October and minimum in February and March. Maximum monthly mean temperature is +17 °C in July and minimum is −5 °C in January. Precipitation increases with altitude within the catchment, while temperature decreases (Erichsen and Nordseth, 1985). The river flow regime of the catchment is characterized by spring snowmelt and autumn rain high flows and winter and summer low flows (Gottschalk et al., 1979).

Hourly measurements of hydrological and meteorological data in the Sæternbekken catchment were performed by the Hydrology Department, Norwegian Water Resources and Energy Directorate. Runoff at the catchment outlet was calculated from the water stage measured upstream of a V-notch weir. Depth to the groundwater table was measured in four piezometers installed to a depth of approximately 0.7 m below the ground surface at different positions along a hillslope; two were located in the valley bottom and two in the middle of the hillslope. An automatic meteorological station, located in a small clearing in the forest adjacent to the hillslope with the piezometers provided recordings of precipitation, air temperature, relative humidity and global shortwave radiation. Precipitation was measured using a Geonor vibrating-wire strain gauge (Bakkehøi et al., 1985). Fig. 1 shows a map of the Sæternbekken catchment with the sites used for data collection.

5. Results and discussion

5.1. Data

Research into data requirements has led to an understanding that the information contained in the data is far more important than the amount used for model calibration. The data should be representative of the various phenomena experienced by the watershed (Yapo et al., 1996). Klemeš (1986) argued that three–five years of daily data should be used for calibration, while Sorooshian and Gupta (1995) suggested that time series of at least 500–1000 data points with large hydrological variability are necessary to activate all the operational modes of a model, resulting in reliable parameter estimates. This implies that two months of data, including both wet and dry conditions should be sufficient when calibrating a model with hourly time steps. The precipitation–runoff model was calibrated using data from the period 15 September to 17 November 1996, while the period 29 August–7 November 1997 was used for the validation of model performance. The value of accumulated precipitation for the calibration period was 307 mm and for the validation period 279 mm. Moisture conditions in the catchment varied substantially during these periods. In both cases, an initial period of 20 days with low discharge and a deep ground water table was used to allow the model to adjust to observed conditions. Initial conditions at the start of this spin-up period were zero soil moisture deficit and a wedge-shaped saturated zone, rising from zero to 20% of the soil profile from the upper to the lower extension of the hillslope.

In order to compare model-simulated groundwater levels to observations, some assumptions about the location of the piezometers relative to the hillslope described by the model had to be made. This was based on a subdivision of the landscape into hydrological response units, i.e. landscape elements with broadly similar responses in terms of runoff production and evapotranspiration (Beven, 1995). Krasovskaia (1985) and Beldring et al. (1999) used a classification of catchment topography based on Hack and Goodlett (1960) to differentiate between three types of hydrological response units in a catchment: (i) nose, the driest part, including the ridge crest and the nearby slopes where the contours are convex outward; (ii) hollow, the central part of the basin along the stream with favorable moisture conditions, an area in which the contours are concave outward; (iii) slope, the zone between nose and hollow with transitional moisture conditions where the contours are straight or nearly so. Following this classification, the two piezometers located in the valley bottom were assumed to represent the hollow topographical unit, and a time series of mean groundwater levels observed in these were compared to simulated values at the lower extension of the model's hillslope. The two piezometers located in the middle of the hillslope were assumed to represent the slope topographical unit, and mean values of groundwater levels observed in these were compared to simulated values halfway between the upper and lower ends of the model's hillslope.

5.2. Calibration strategies

Multi-criteria parameter estimation with the MOCOM-UA method was performed with the aim of optimizing the Nash–Sutcliffe statistics of runoff and groundwater levels. Two different multi-criteria calibration studies were performed; both used catchment runoff and groundwater levels from one type of hydrological response unit. In the first case groundwater data from the hollow topographical unit was used, the other used groundwater data from the slope topographical unit. A population size of 250 parameter sets was used to estimate the Pareto solution space with respect to the 11 model parameters. Yapo et al. (1998) used a population size of 500 points to approximate the Pareto space, however, solving the multi-criteria problem for populations larger than 250 points appeared to be extremely demanding in terms of computing time with the precipitation–runoff model and data used in this study. In order to eliminate any dependence of the results on the initial sample of points in the parameter space, the MOCOM-UA procedure was run 20 times with different initial values. This resulted in 5000 parameter sets for each of the two calibration cases. Although each of the 20 solutions is a non-dominated set, it does not necessarily follow that the entire population of 5000 parameter sets constitutes a Pareto set.

Results from the multi-criteria calibration studies were compared to results from a single-criterion calibration procedure where the Nash–Sutcliffe statistic of catchment runoff was used as a measure of model performance. A Monte Carlo procedure was used to generate parameter sets by drawing values of each of the 11 parameters randomly from uniform distributions between the lower and upper bounds, in a manner similar to Beven and Binley (1992). The model was executed and the Nash–Sutcliffe statistic of runoff was calculated. This procedure was repeated 2×10^6 times and the 5000 best parameter sets in terms of model performance were retained, while the remaining were rejected. This resulted in values of the Nash–Sutcliffe statistic of runoff in the range 0.80–0.91 for the calibration period. Values obtained for the Nash–Sutcliffe statistic of runoff for the same period during multi-criteria calibration were in the range 0.77–0.83 when groundwater levels in the hollow topographical unit were used, and in the range 0.75–0.84 when groundwater levels in the slope topographical unit were used.

The parameters chosen for optimization were those which were considered most critical for the performance of the model with regard to describing runoff and groundwater levels. The ranges of values for parameters which were subject to optimization and the fixed parameter values were based on field investigations and samples from the till deposits of the Sæternbekken catchment (Udnæs, 1991; Tallaksen et al., 1996; Myrabø, 1997), previous experience with the precipitation-runoff model (Beldring et al., 2000), and a digital elevation model with horisontal resolution 10 m by 10 m.

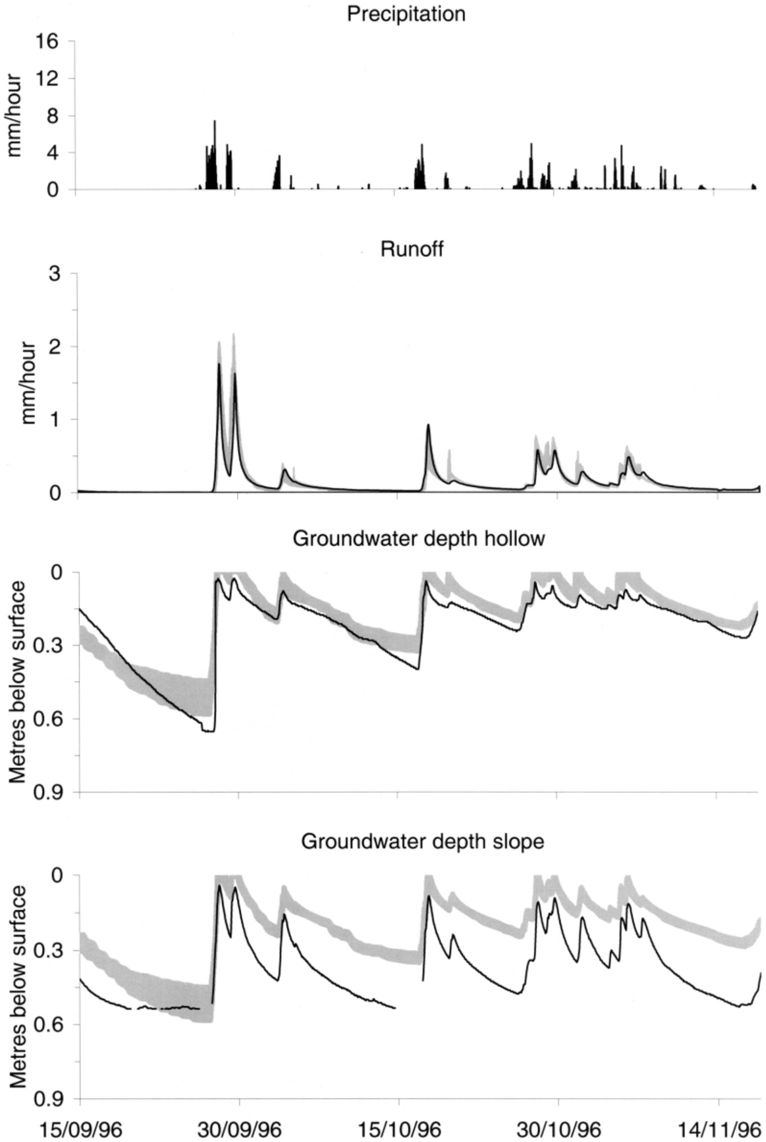

Fig. 2. Observed data (black) and the range of model-simulated values (gray) for the calibration period 15 September–17 November 1996. Multi-criteria calibration using catchment runoff and groundwater levels in hollow topographical unit.

5.3. Model performance

Figs. 2–4 show observed data and ranges of model-simulated values during the calibration period. Results from multi-criteria calibration using catchment runoff and groundwater levels in the hollow topographical unit are shown in Fig. 2, results from multi-criteria calibration using catchment runoff and groundwater levels in the slope topographical unit are shown in Fig. 3, while results from single-criterion calibration using catchment runoff are shown in Fig. 4. Results from applying the parameter sets determined from the three calibration procedures to the validation period are shown in Figs. 5–7. It is important to note that the ranges of model-simulated values, presented as sequences of minimum and maximum

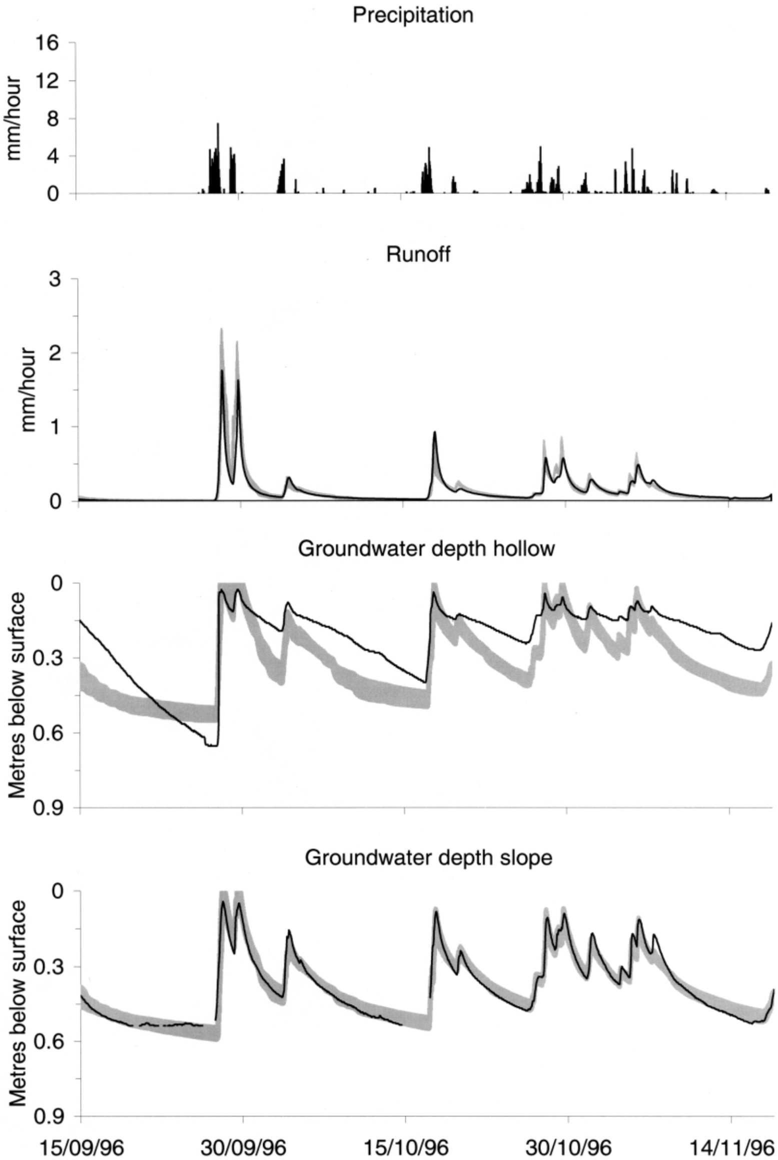

Fig. 3. Observed data (black) and the range of model-simulated values (gray) for the calibration period 15 September–17 November 1996. Multi-criteria calibration using catchment runoff and groundwater levels in slope topographical unit.

values, are not the results of some particular parameter set, but represents the extreme values at each time step.

The general impression is that the precipitation–runoff model reproduced observed runoff during the calibration period (Figs. 2–4) in a realistic manner with all parameter sets from all the three calibration procedures, while model performance deteriorated during the validation period (Figs. 5–7), in particular for parameter sets based on single-criterion calibration. Although observations were generally enclosed between the minimum and maximum of model-simulated values, the ranges of model predictions based on multi-criteria calibrations were constrained to narrower intervals than the ranges of predictions based on single-criterion calibration. This was

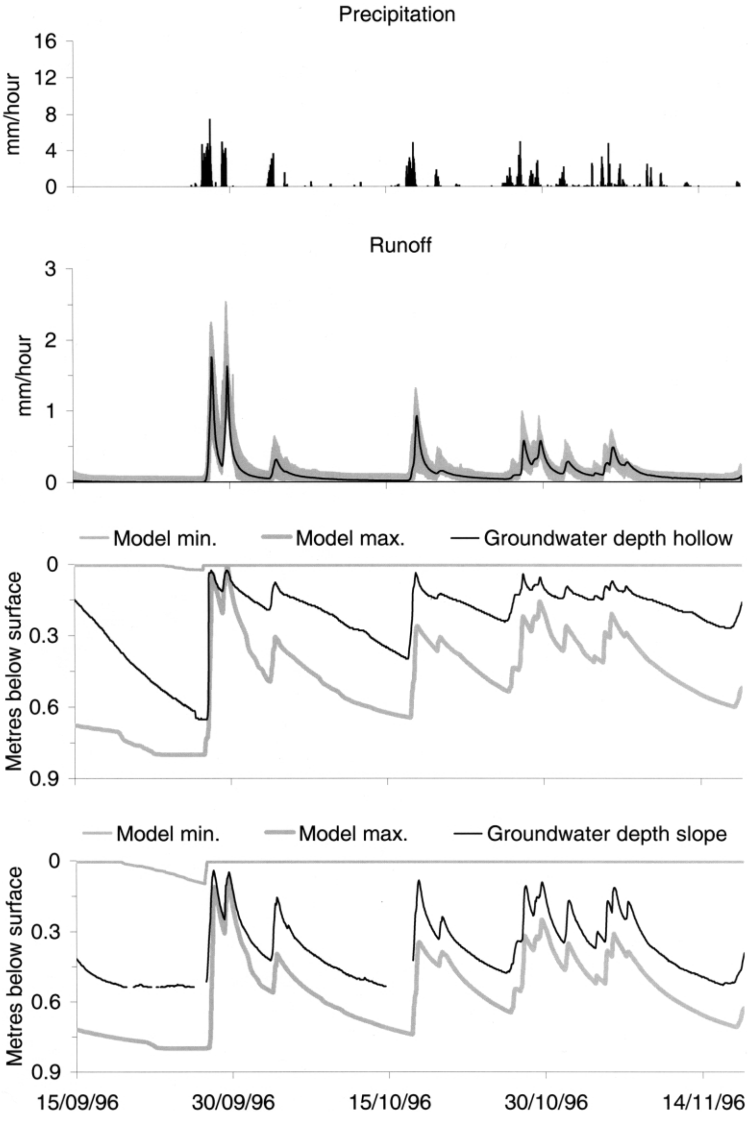

Fig. 4. Observed data (black) and the range of model-simulated values (gray) for the calibration period 15 September–17 November 1996. Single-criterion calibration using catchment runoff.

particularly evident during the validation period. The performance of the physically based model with regard to runoff was best for the parameter sets based on the multi-criteria calibration strategies, indicating that groundwater levels exert a major control on the hydrological response of the investigated catchment. This result is in agreement with previous experience from small undisturbed catchments in humid temperate environments (e.g. Rodhe, 1989; Bonell, 1993; Nyberg, 1995), thereby supporting the validity of the model structure.

Model-simulated and observed groundwater levels agreed well for the topographical units used for multi-criteria calibration during the calibration period (Figs. 2 and 3), although the ranges of simulated values showed a tendency of systematic deviations from the observations. There were larger discrepancies between the levels of the simulations and the

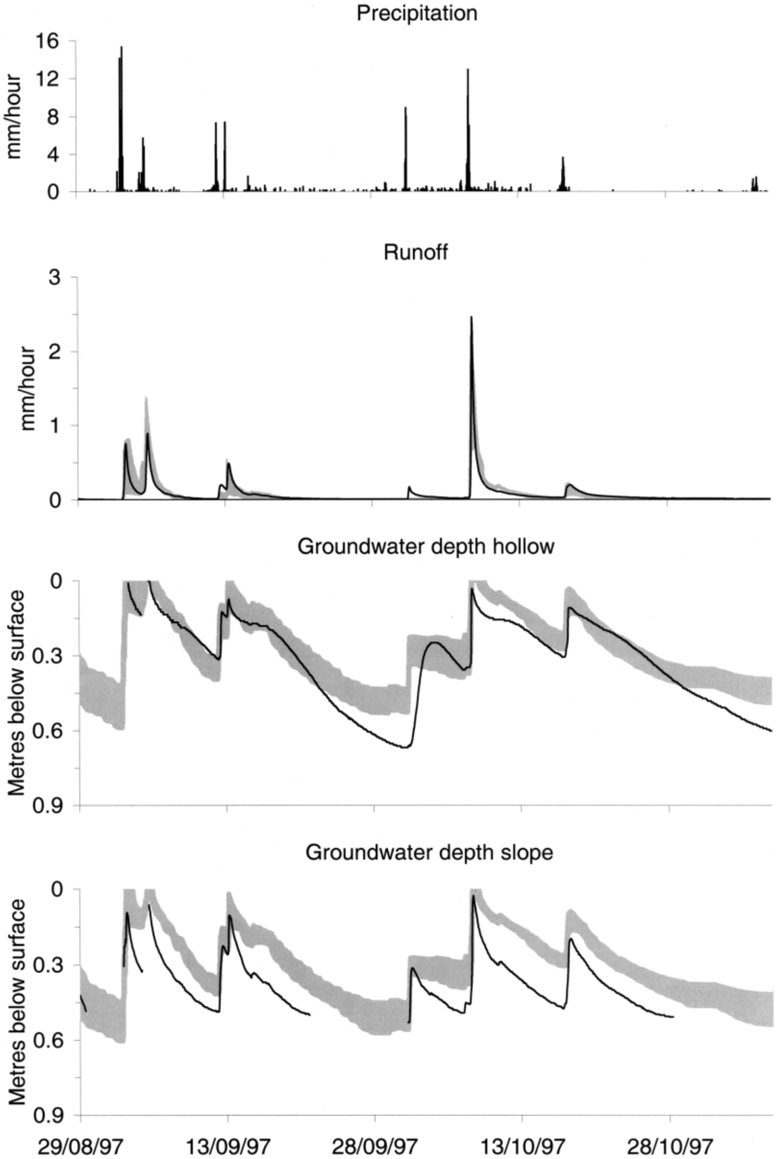

Fig. 5. Observed data (black) and the range of model-simulated values (gray) for the validation period 29 August–7 November 1997. Model parameters determined by multi-criteria calibration using catchment runoff and groundwater levels in hollow topographical unit.

observations from the topographical units not used for calibration. Results from the validation period (Figs. 5 and 6) were similar, however, the performance of the model was not as good. In particular, there was one event in the beginning of October 1997 where a rapid rise of the groundwater table was not described by the parameter sets determined by multi-criteria calibration using data from the slope topographical unit. The ranges of model predictions using the parameter sets from single-criterion calibrations were too wide, both during the calibration and the validation periods (Figs. 4 and 7). Nevertheless, simulated values followed the oscillations of observed data, indicating that the physically based model must describe some of the dynamics of groundwater levels in order to predict the runoff correctly. Although the minimum values of

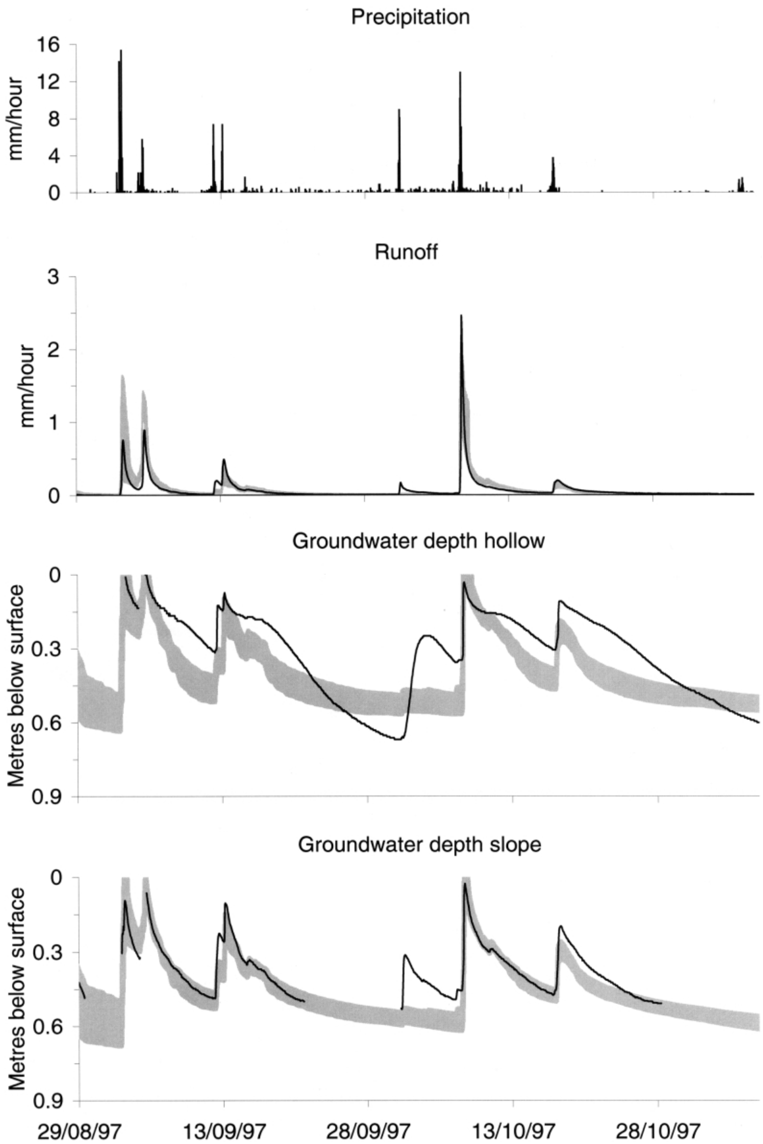

Fig. 6. Observed data (black) and the range of model-simulated values (gray) for the validation period 29 August–7 November 1997. Model parameters determined by multi-criteria calibration using catchment runoff and groundwater levels in slope topographical unit.

groundwater levels in Figs. 4 and 7 are close to zero during both the calibration and validation periods, individual simulations were more dynamic.

5.4. Sensitivity of model performance

Cumulative distribution functions of the Nash–Sutcliffe and bias statistics of catchment runoff and groundwater levels are presented in Figs. 8–10. Results from all three calibration procedures for both calibration and validation periods are shown. The cumulative distribution functions of the Nash–Sutcliffe statistic of runoff did not indicate a difference in model performance, while the cumulative distribution functions of the bias statistic showed a tendency of larger deviations

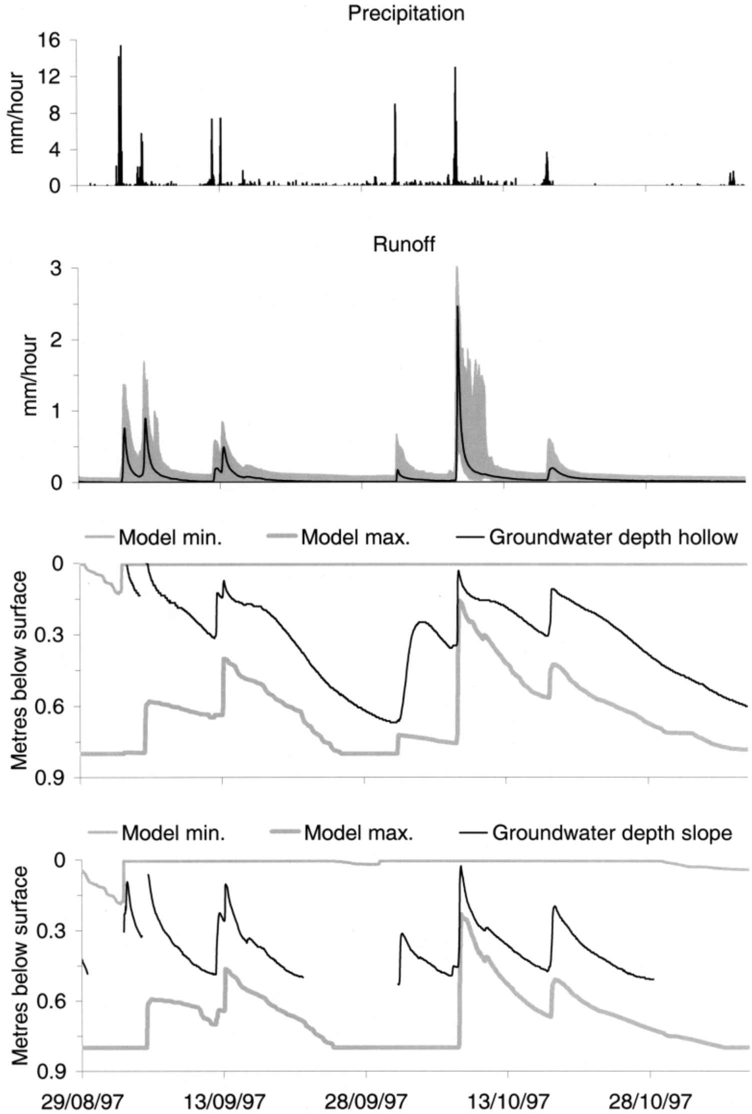

Fig. 7. Observed data (black) and the range of model-simulated values (gray) for the validation period 29 August–7 November 1997. Model parameters determined by single-criterion calibration using catchment runoff.

between model-simulated and observed runoff for parameter sets based on single-criterion calibrations. The cumulative distribution functions of the Nash–Sutcliffe statistic of groundwater levels were generally steeper and shifted further to the right for the simulations using parameter sets from the multi-criteria calibrations, indicating reduced sensitivity to selection of parameter sets and better model performance. Although the cumulative distribution functions of the bias statistic of groundwater levels showed the same tendency of discrepancy between model-simulated and observed groundwater levels as the time series plots, these results are also indicative of better performance for the parameter sets determined by the multi-criteria calibrations.

The behavior of a hydrological model is determined by the performance of each combination of parameter

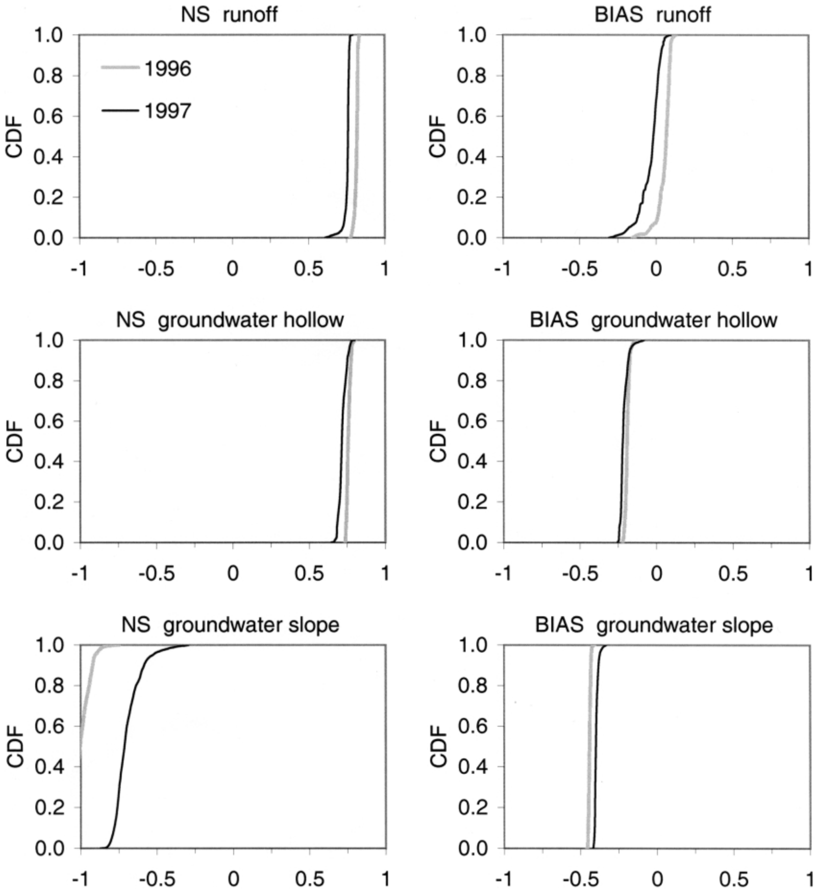

Fig. 8. Cumulative distribution functions of Nash–Sutcliffe (NS) and bias (BIAS) statistics of catchment runoff and groundwater levels. Model parameters determined by multi-criteria calibration using catchment runoff and groundwater levels in hollow topographical unit.

values. Several studies have demonstrated that interactions between individual parameters result in acceptable simulations of runoff over a wide range of values for each parameter for a particular model structure (e.g. Freer et al., 1996; Uhlenbrook et al., 1999). This was also the result of the single-criterion calibration procedure applied in this study; parameters varied approximately uniformly over the initial ranges of values given in Table 1. In contrast, the multi-criteria calibration procedures constrained most parameters to intervals, which were small compared to the initial ranges. This is shown in Fig. 11 which presents cumulative distribution functions of parameter values determined by the three calibration strategies. The results are given as normalized values relative to the initial ranges between the lower and upper bounds. A parameter can be described as well-defined or sensitive with respect to the chosen measures of model performance if increasing information in the calibration data results in steeper cumulative distribution functions covering a small range of values (Freer et al., 1996; Yapo et al., 1996). The sensitivity of model performance was particularly evident for some of the parameters which describe groundwater levels and downslope saturated subsurface flow and saturation overland flow; m, ϵ, K_0, a, λ and z_r. Parameter uncertainty was largest for I_{max} and C_{act}, most likely because no information was available for separating evaporation of intercepted water, soil evaporation and transpiration from plants, resulting in parameter interactions. The two multi-criteria calibration studies resulted in different cumulative distribution functions

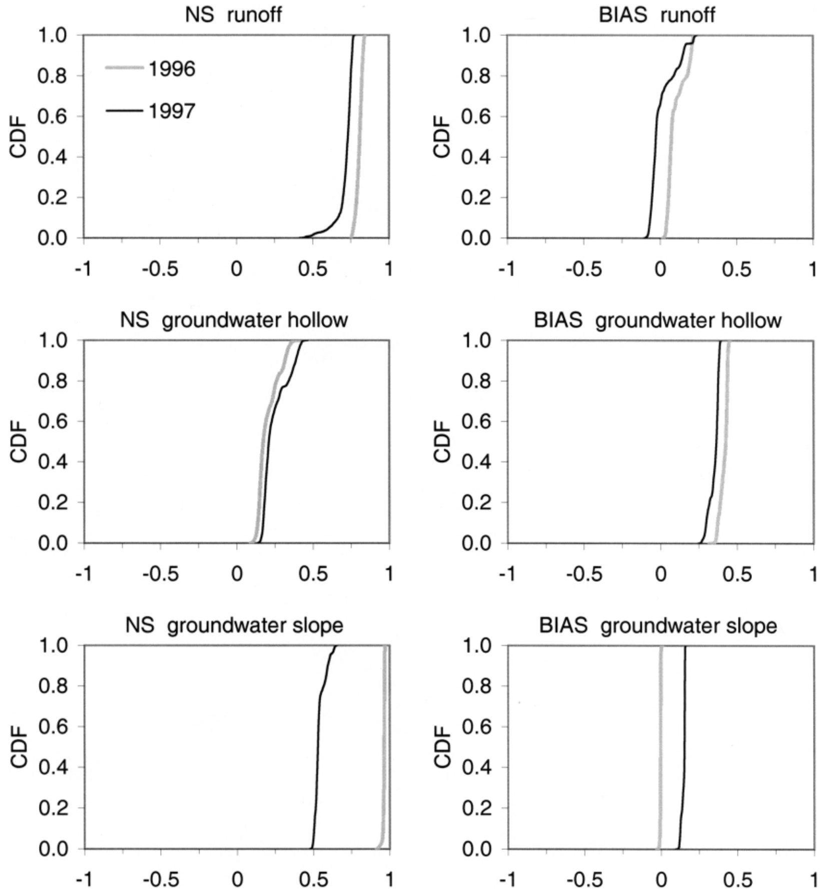

Fig. 9. Cumulative distribution functions of Nash–Sutcliffe (NS) and bias (BIAS) statistics of catchment runoff and groundwater levels. Model parameters determined by multi-criteria calibration using catchment runoff and groundwater levels in slope topographical unit.

for several parameters. This was most likely due to uncertainty about the location of the piezometers used for groundwater observations relative to the model's hillslope, forcing model behavior to match the observed data in different ways. Beldring et al. (2000) showed that model results should be interpreted as the spatial distribution of moisture conditions in a catchment and not as the exact values at specific points. This being the case, the multi-criteria calibration procedure should be based on comparing model results with one or more statistics summarizing the characteristics of observations from a net of piezometers located in different types of hydrological response units in the catchment. Although this would require an extensive data collection program, it is the only way to obtain realistic estimates of groundwater table depths in catchments in forest environments.

5.5. Model validation

The confidence that can be placed in model simulations depends largely on the uncertainty remaining after the model has been calibrated (Freedman et al., 1998). In the context of multiple measures of model performance, the Pareto solution space represents the minimal uncertainty that can be achieved for the parameters via calibration, without subjectively assigning relative weights to the individual model responses (Bastidas et al., 1999). Since it is not possible to select a specific parameter set as being superior to any other, this parameter uncertainty translates into a trade-off

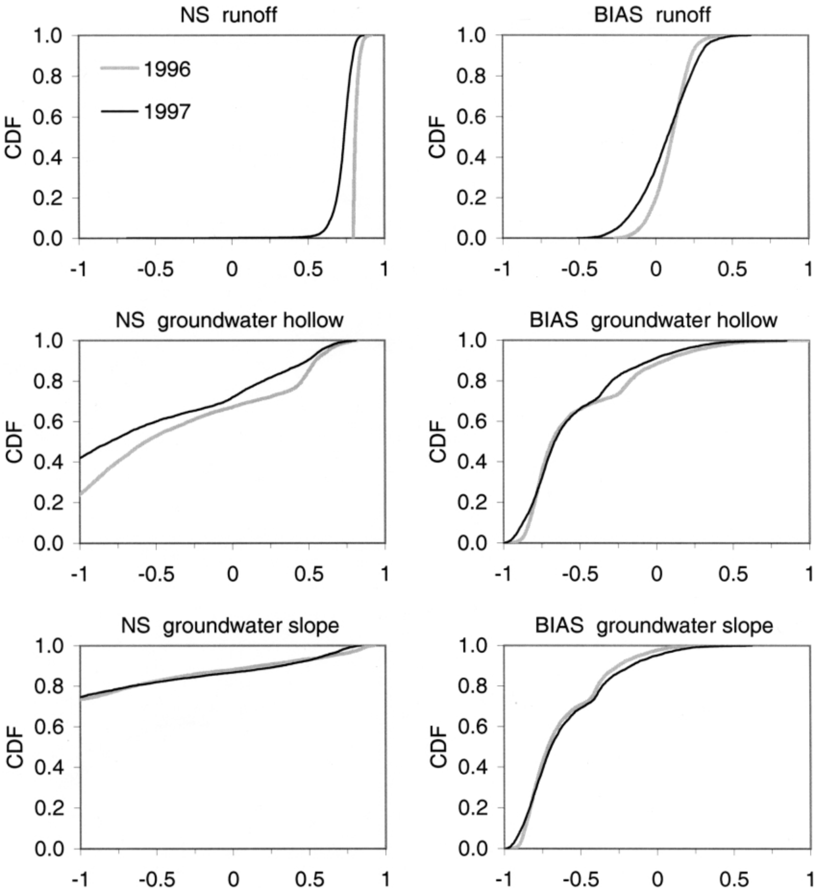

Fig. 10. Cumulative distribution functions of Nash–Sutcliffe (NS) and bias (BIAS) statistics of catchment runoff and groundwater levels. Model parameters determined by single-criterion calibration using catchment runoff.

range in the model predictions (Gupta et al., 1998). Although the 5000 parameter sets determined for each calibration problem cannot be expected to constitute a non-dominated set, they are an approximation to the Pareto solution space. The ranges of model predictions are therefore indications of model uncertainty. Although some systematic differences between model-simulated output and observed system output were found, results from the multi-criteria calibration strategies generally supported the validity of the model structure. This conclusion can also be drawn from the result that parameters describing processes included in the multi-criteria calibration procedures were well-defined.

The precipitation–runoff model described in this study is based on simplified physical descriptions of the characteristic features of runoff formation in shallow till deposits overlying a relatively impermeable bedrock. Three flow processes are considered; vertical unsaturated flow, saturated subsurface flow parallel to the impermeable bed, and saturation overland flow. The model does not explain flow paths and residence times of water particles in detail, and its structure includes simplified boundary conditions and several conceptual descriptions. In spite of these limitations, results from this work and a previous study by Beldring et al. (2000) have shown that the model provides realistic predictions of the spatial and temporal variations of depth to the phreatic zone. The results confirm the important role of soil moisture and groundwater conditions in controlling the dynamic nature of hydrological processes in undisturbed

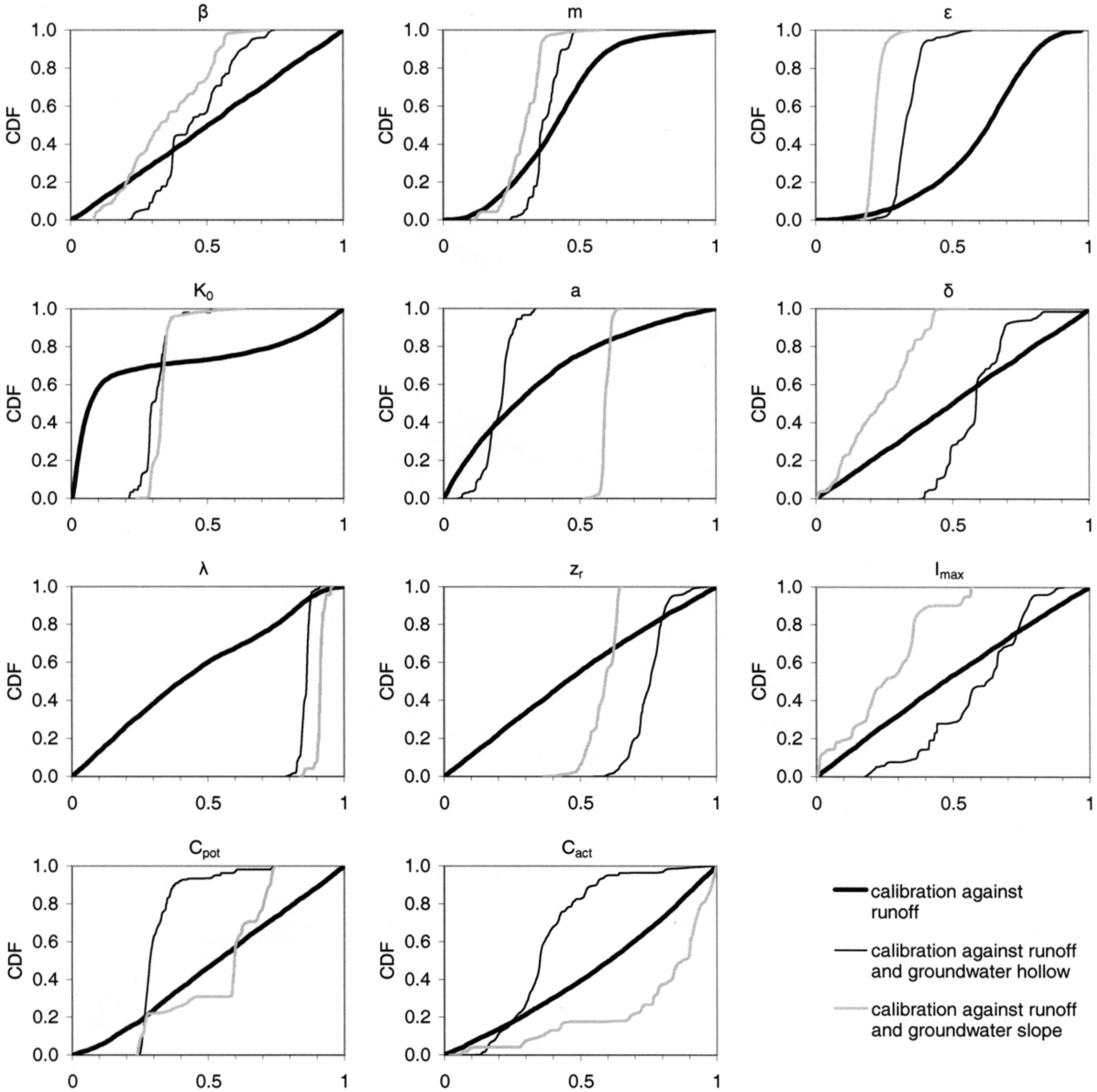

Fig. 11. Cumulative distribution functions of model parameter values determined by the multi-criteria and single-criterion calibration strategies. Results are presented as normalized values relative to the ranges between the lower and upper bounds given in Table 1.

catchments with shallow till deposits. Characterizing the distribution of hydrological state variables and responses in the landscape is a prerequisite for realistic predictions of discharge and evapotranspiration fluxes within heterogeneous terrain (Beven, 1995).

The strategy of model validation performed in this study can be classified as 'undisturbed-catchment multiple-response split-sample'-tests in the hierarchy proposed by Mroczkowski et al. (1997). The purpose of this strategy is to demonstrate that a model is capable of making accurate predictions of multiple processes for periods outside the calibration period. Stream-flow data at the catchment outlet, are supplied with time series data observed at different locations within a catchment, e.g. soil moisture, groundwater levels, chemical stream loads, or subcatchment streamflow. Refsgaard and Knudsen (1996) used a similar definition of model validation, although they argued that only site specific validation is possible. Nevertheless, the following facts support the idea that the precipitation–runoff model is generally valid in small undisturbed catchments with shallow till deposits overlying a relatively impermeable bedrock; (i) it is based on physical descriptions of the processes linking precipitation and runoff; (ii) it has been demonstrated that constraining the model to both runoff and ground water levels during calibration improved model performance; (iii) the model was able to reproduce multiple time series during periods not used for calibration in a satisfactory manner; and (iv) previous results from Beldring et al. (2000) have demonstrated the geographical transposability of the model.

6. Conclusions

In order to describe the hydrological processes in a watershed, a suitable model structure must be selected and values for the parameters must be specified so that the model closely simulates the behavior of the study site. However, model results are only as reliable as the model assumptions, inputs and parameter estimates (Sorooshian and Gupta, 1995). Any precipitation–runoff model, no matter how sophisticated, is a simplified representation of the physical structure of a watershed, and therefore cannot be expected to provide a perfect fit to observed data. Even if the model was a perfect representation of the system, errors in observations of input and output data would prevent the residuals between model-simulated and measured system output from being zero (Gupta et al., 1998). In order to have confidence in a physically based hydrological model, it must be based on a description of the relevant mechanisms linking precipitation and runoff in a particular environment, its operational performance with regard to all relevant aspects of model behavior should be validated, and model parameters should be well-defined (Sorooshian and Gupta, 1995).

This study has demonstrated that:

- Given a physically based hydrological model that simulates several aspects of watershed behavior, the multi-criteria calibration procedure MOCOM-UA was capable of exploiting information about the physical system contained in the measurement data time series.
- The multi-criteria calibration procedure provided estimates of the uncertainty associated with model predictions and parameters.
- Multi-criteria calibration constraining the precipitation–runoff model to observed runoff and groundwater levels simultaneously reduced the uncertainty of model predictions compared to single-criterion calibration constraining the model to observed runoff only.
- The multi-criteria calibration procedure reduced the uncertainty of the estimates of model parameters.
- The investigated precipitation–runoff model was able to reproduce several observed system responses simultaneously during both calibration and validation periods.
- Groundwater table depths exert a major control on the hydrological response of an undisturbed catchment in a landscape with shallow till deposits overlying a relatively impermeable bedrock.

The precipitation–runoff model considered in this study is based on a simplified physical characterization of runoff formation in undisturbed catchments in a landscape with shallow till deposits overlying a relatively impermeable bedrock. Results from applying the model in this and the previous study (Beldring et al., 2000) were found to be consistent with the physical properties of the environments in the experimental catchments, thereby supporting the validity of the model formulation. Although an objective framework for model validation was used, the results were obviously influenced by the specific test conditions, including the particular climate, catchment characteristics, data quality and availability, as well as the selected measures of model performance. To arrive at firm conclusions regarding a model many validations would usually be required.

Acknowledgements

Knut Schult and Hans-Christian Udnæs, Hydrology Department, Norwegian Water Resources and Energy Directorate were responsible for collecting and controlling the data used in this study.

References

Bakkehøi, S., Øien, K., Førland, E.J., 1985. An automatic precipitation gauge based on vibrating-wire strain gauges. Nordic Hydrology 16, 193–202.

Bastidas, L.A., Gupta, H.V., Sorooshian, S., Shuttleworth, W.J., Yang, Z.L., 1999. Sensitivity analysis of a land surface scheme using multi-criteria methods. Journal of Geophysical Research 104 (D16), 19481–19490.

Bathurst, J.C., O'Connell, P.E., 1992. Future of distributed modelling: the System Hydrologique European. Hydrological Processes 6, 265–277.

Beldring, S., Gottschalk, L., Seibert, J., Tallaksen, L.M., 1999. Distribution of soil moisture and groundwater levels at patch and catchment scales. Agricultural and Forest Meteorology 98–99, 305–324.

Beldring, S., Gottschalk, L., Rodhe, A., Tallaksen, L.M., 2000. Kinematic wave approximations to hillslope hydrological processes in tills. Hydrological Processes 14, 727–745.

Bergström, S., 1991. Principles and confidence in hydrological modelling. Nordic Hydrology 22, 123–136.

Bergström, S., Sandberg, G., 1983. Simulation of groundwater response by conceptual models—three case studies. Nordic Hydrology 14, 71–84.

Beven, K., 1989. Changing ideas in hydrology—the case of physically based models. Journal of Hydrology 105, 157–172.

Beven, K., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16, 41–55.

Beven, K., 1995. Linking parameters across scales: subgrid parameterizations and scale dependent hydrological models. In: Kalma, J.D., Sivapalan, M. (Eds.). Scale Issues in Hydrological Modelling. Wiley, Chichester, pp. 263–281.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.

Beven, K.J., Lamb, R., Quinn, P., Romanowicz, R., Freer, J., 1995. TOPMODEL. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, pp. 627–668.

Bonell, M., 1993. Progress in the understanding of runoff generation dynamics in forests. Journal of Hydrology 150, 217–275.

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

Erichsen, B., Nordseth, K., 1985. Sæternbekken øvings-og forsøksfelt (in Norwegian). Norwegian National Commitee for Hydrology, Report No. 18, Oslo, 75 pp.

Franks, S.W., Gineste, P., Beven, K.J., Merot, P., 1998. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas in the calibration process. Water Resources Research 34, 787–797.

Freedman, V.L., Lopes, V.L., Hernandez, M., 1998. Parameter identifiability for catchment–scale erosion modeling: a comparison of optimization algorithms. Journal of Hydrology 207, 83–97.

Freer, J., Beven, K., Ambroise, B., 1996. Bayesian estimation of uncertainty in runoff production and the value of data: an application of the GLUE approach. Water Resources Research 32, 2161–2173.

Gan, T.Y., Biftu, G.F., 1996. Automatic calibration of conceptual rainfall-runoff models: optimization algorithms, catchment conditions, and model structure. Water Resources Research 32, 3513–3524.

Gottschalk, L., Jensen, J.L., Lundquist, D., Solantie, R., Tollan, A., 1979. Hydrologic regions in the nordic countries. Nordic Hydrology 10, 273–286.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and non-commensurable measures of information. Water Resources Research 34, 751–763.

Gupta, H.V., Bastidas, L.A., Sorooshian, S., Shuttleworth, W.J., Yang, Z.L., 1999. Parameter estimation of a land surface scheme using multi-criteria methods. Journal of Geophysical Research 104 (D16), 19491–19503.

Hack, J.T., Goodlett, J.C., 1960. Geomorphology and forest ecology of a mountain region in the Central Appalachians. US Geological Survey Professional Paper 347, 00 66 pp.

Jakeman, A.J., Hornberger, G.M., 1993. How much complexity is warranted in a rainfall-runoff model? Water Resources Research 29, 2637–2649.

Kirkby, M., 1988. Hillslope runoff processes and models. Journal of Hydrology 100, 315–339.

Klemeš, V., 1986. Operational testing of hydrological simulation models. Hydrological Sciences Journal 31, 13–24.

Krasovskaia, I., 1985. To the problem of joint soil moisture and groundwater analysis in basin response studies. Catena 12, 241–254.

Kuczera, G., Mroczkowski, M., 1998. Assessment of hydrologic parameter uncertainty and the worth of multi-response data. Water Resources Research 35, 1481–1489.

Lamb, R., Beven, K., Myrabø, S., 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall–runoff model. Advances in Water Resources 22, 305–317.

Legates, D.R., McCabe, G.J., 1999. Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233–241.

Melching, C.S., 1995. Reliability estimation. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, pp. 69–118.

Motovilov, Y.G., Gottschalk, L., Engeland, K., Rodhe, A., 1999. Validation of a distributed hydrological model against spatial observations. Agricultural and Forest Meteorology 98–99, 257–277.

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

Myrabø, S., 1997. Temporal and spatial scale of response area and groundwater variation in till. Hydrological Processes 11, 1861–1880.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models. Part I—a discussion of principles. Journal of Hydrology 10, 282–290.

Nyberg, L., 1995. Water flow path interactions with soil hydraulic properties in till soil at Gårdsjön, Sweden. Journal of Hydrology 170, 255–275.

Refsgaard, J.C., 1997. Parameterisation, calibration and validation of distributed hydrological models. Journal of Hydrology 198, 69–97.

Refsgaard, J.C., Knudsen, J., 1996. Operational validation and intercomparison of different types of hydrological models. Water Resources Research 32, 2189–2202.

Rodhe, A., 1989. On the generation of stream runoff in till soils. Nordic Hydrology 20, 1–8.

Singh, V.P., 1996. Kinematic Wave Modeling in Water Resources. , Surface-Water Hydrology. Wiley, New York 1399 pp.

Sorooshian, S., Gupta, V.K., 1995. Model calibration. In: Singh, V.P. (Ed.). Computer Models of Watershed Hydrology. Water Resources Publications, Highlands Ranch, pp. 23–68.

Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall–runoff models: application of global optimization to the Sacramento soil moisture accounting model. Water Resources Research 29, 1185–1194.

Tallaksen, L.M., Schunselaar, S., van Veen, R., 1996. Comparative model estimates of interception loss in a coniferous forest stand. Nordic Hydrology 27, 143–160.

Thyer, M., Kuczera, G., Bates, B.C., 1999. Probabilistic optimization for conceptual rainfall–runoff models: a comparison of the shuffled complex evolution and simulated annealing algorithms. Water Resources Research 35, 767–773.

Udnæs, H.-C., 1991. En beskrivelse av grunnvannsmagasinets egenskaper og markvannstrømningen i et lite skogsfelt (in Norwegian). Thesis, Department of Geophysics, University of Oslo, 97 pp.

Uhlenbrook, S., Seibert, J., Leibundgut, C., Rodhe, A., 1999. Prediction uncertainty of conceptual rainfall–runoff models caused by problems in identifying model parameters and structure. Hydrological Sciences Journal 44, 779–797.

Weglarczyk, S., 1998. The interdependence and applicability of some statistical quality measures for hydrological models. Journal of Hydrology 206, 98–103.

Wilmott, C.J., Ackleson, S.G., Davis, R.E., Feddema, J.J., Klink, K.M., Legates, D.R., O'Donnel, J., Rowe, C.M., 1985. Statistics for the evaluation and comparison of models. Journal of Geophysical Research 90 (C5), 8995–9005.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181, 23–48.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology 204, 83–97.