Hydrological Sciences—Journal—des Sciences Hydrologiques, 44(5) October 1999, 779–798.

Prediction uncertainty of conceptual rainfall-runoff models caused by problems to identify model parameters and structure

STEFAN UHLENBROOK[1], JAN SEIBERT[2], CHRISTIAN LEIBUNDGUT[1], ALLAN RODHE[2]

[1] Institute of Hydrology, University of Freiburg, Fahnenbergplatz, D-79098 Freiburg, Germany; email: uhlenbro@ uni-freiburg.de

[2] Uppsala University, Department of Earth Sciences, Hydrology, Villavägen 16, S-75236 Uppsala, Sweden; email: jan.seibert@hyd.uu.se

Abstract The uncertainties arising from the problem of identifying a representative model structure and model parameters in a conceptual rainfall-runoff model were investigated. A conceptual model, the HBV model, was applied to the mountainous Brugga basin (39.9 km^2) in the Black Forest, southwestern Germany. In a first step, a Monte Carlo procedure with randomly generated parameter sets was used for calibration. For a ten-year calibration period, different parameter sets resulted in an equally good correspondence between observed and simulated runoff. A few parameters were well defined (*i.e.*, best parameter values were within small ranges), but for most parameters good simulations were found with values varying over wide ranges. In a second step, model variants with different numbers of elevation and land use zones and various runoff generation conceptualisations were tested. In some cases representation of more spatial variability gave better simulations in terms of discharge. However, good results could be obtained with different and even unrealistic concepts.

The computation of design floods and low flow predictions illustrated that the parameter uncertainty and the uncertainty of identifying a unique best model variant have implications for model predictions. The flow predictions varied considerably. The peak discharge of a flood with a probability of 0.01 yr^{-1}, for instance, varied from 40 to almost 60 mm d^{-1}. It was concluded that model predictions particularly in applied studies should be given as ranges rather than as single values.

L' incertitude de prédiction d'un modèle conceptuel pluie-débit due à l'identification de des paramètres et de la structure

Résumé Cet article étudie les incertitudes dues à l'identification d'une structure représentative et à la détermination des paramètres d'un modèle conceptuel pluie-débit. Le modèle conceptuel HBV a été appliqué au bassin versant montagneux de la Brugga (39.9 km^2), situé au sud-ouest de l'Allemagne, dans la Forêt Noire. Dans un premier temps, une procédure de Monte Carlo avec un jeu de paramètres générés aléatoirement a été appliquée pour la calibration. Pour une période de calibration de dix ans, plusieurs jeux de paramètres ont fourni une bonne correspondance entre les débits observés et simulés. Seuls quelques paramètres ont été bien définis, c'est-à-dire qu'ils prennent des valeurs dont l'étendue est faible. Pour la plupart des autres paramètres, de bonnes simulations ont également été obtenues mais avec des valeurs très variables. Dans un

second temps, des variantes de modèle ont été testées en faisant varier,le nombre des zones d'altitude et d'occupation du sol et les concepts de génération de crue. Dans certains cas, une plus grande variabilité spatiale fournit de meilleures simulations en terme de débit. Il faut noter toutefois que de bons résultats peuvent être obtenus avec des concepts différents et même irréalistes.

La détermination des crues et la prédiction des débits d'étiage ont illustré le fait que l'incertitude attachée au paramètre et celle dû à l'identification d'une seule structure de modèle ont des implications pour les prédictions des modèles. La prédiction des débits a varié considérablement. Le débit de pointe pour une crue avec une probabilité d'apparition de 0.01 a^{-1} peut varier par exemple de 40 mm j^{-1} jusqu'à pratiquement 60 mm j^{-1}. En conclusion, les prédictions des modèles et plus particulièrement dans les études appliquées devraient être données plutôt sous formes d'intervalles que de valeurs uniques.

INTRODUCTION

Conceptual rainfall-runoff models are widely used tools in hydrology. Contrary to more complex, physically based, distributed models such as the SHE model (Abbott *et al.*, 1986a), the required input data are readily available for most applications. Furthermore, conceptual models are usually simple and relatively easy to use. In spite of the attractiveness of conceptual models they suffer from some fundamental problems (Abbott *et al.*, 1986b; Todini, 1988; Beven, 1989; Bergström, 1991). One fundamental problem is that conceptual models often are overparameterized with intercorrelated model parameters (*e.g.*, Jakeman and Hornberger, 1993; Gaume *et al.*, 1998). Some model parameters have a physical basis, but since they are effective parameters on the catchment scale they are hardly measurable in the field. This makes a model calibration inevitable. However, it is often not possible to find one unique 'best' parameter set, *i.e.*, different parameter sets give similar good results during a calibration period (*e.g.*, Mein & Brown, 1978; Beven & Binley, 1992; Duan *et al.*, 1992; Beven, 1993; Freer *et al.*, 1996; van der Perk & Bierkens, 1997; Seibert, 1997a). Parameter uncertainty makes simulations for periods outside the calibration period less reliable (Melching *et al.*, 1990; Harlin & Kung, 1992). In addition, 'model uncertainty' may exist, *i.e.*, an uncertainty in which model to choose (Beck, 1987; Melching *et al.*, 1990, Piñol *et al.*, 1997). During model development, concepts of catchment hydrology are implemented into the model as a simplified representation of real processes. The user has to decide which one of the many existing models is the suitable and choose the spatial delineation as, for instance, the number of elevation or land use zones. Usually, for most of these decisions there is a lack of objective criteria (Mroczkowski *et al.*, 1997).

The HBV model (Bergström, 1976) is a conceptual model of catchment hydrology originally developed for Scandinavian catchments. During the last two decades it has been applied in more than 30 countries world-wide (Bergström, 1992). The HBV model has been used for different hydrological tasks, for instance, to compute spillway design floods or flood forecasting (Bergström, 1992) and to study the effects of changes in climate (Saelthun, 1996) and land use (Brandt *et al.*, 1988). Different attempts have been made to relate the parameters of the HBV model to catchment characteristics for regionalization purposes (Bergström, 1990; Braun & Renner, 1992;

Seibert, 1998; Uhlenbrook *et al.*, 1998). The code of the model has been rewritten in several versions. Its different versions provide examples of different decisions during the model development. In the original Swedish version, for instance, computations for each elevation-land use zone are performed separately in both the snow and the soil routine (Bergström, 1992), while separate computations are made only in the snow routine in the Swiss version where the soil routine is lumped (Braun & Renner, 1992).

The aim of this study was to investigate the uncertainties arising from different model structures and poorly defined model parameters when using a conceptual rainfall runoff model. The HBV model was selected for this study as it is considered to be a typical representative of such models. The parameter uncertainty was examined using a Monte Carlo procedure with the classical HBV model structure (Bergström, 1992). Furthermore, different model settings (*e.g.*, number of elevation or land use zones) and modified model structures (different runoff generation routines) were tested. When testing the model structures, the same modeling framework was used and the number of parameters were kept constant in most model variants. This made it possible to compare the results directly.

It may be argued that the problem of identifying a unique parameter set and model variant is not an issue for practical model applications, *i.e.*, if different parameter sets and model variants were equally suitable to simulate runoff during a calibration period, any one of these may be applied (Lindstöm, 1997). However, for model predictions outside the calibration period these parameter sets and model variants can not be expected to give similar results. To evaluate this important implication of the identifiability problems, the uncertainty of simulated discharge was addressed in this study. The simulation of different hydrological events (floods and low flows) using equally good parameter sets or model variants were compared. This comparison was used as a simple method to quantify the prediction uncertainty caused by identifiability problems when using a conceptual rainfall runoff model.

MATERIAL AND METHODS

The HBV model

The HBV model (Bergström, 1976) is a conceptual model of catchment hydrology which simulates daily discharge using as input variables daily rainfall and temperature and monthly estimates of potential evaporation. The model consists of different routines representing snow accumulation and melt by a degree-day method, groundwater recharge and actual evaporation as functions of actual water storage in a soil box, groundwater by three linear reservoir equations and channel routing by a triangular weighting function (Fig. 1). Further descriptions of the model can be found elsewhere (Bergström, 1992, 1995; Harlin & Kung, 1992; Seibert, 1997a, 1998). The version of the model used in this study, 'HBV light 1.2' (Seibert, 1997b), corresponds to the version HBV-6 described by Bergström (1992).

Study Site

The study was performed in the Brugga basin (39.9 km^2), located in the Southern Black Forest in southwestern Germany. It is a mountainous catchment with elevation ranging from 450 to 1500 m a.m.s.l. and a nival runoff regime. The mean annual precipitation amounts to 1750 mm generating a mean annual discharge of approximately 1200 mm. About two thirds of the annual precipitation in the upper part and one third in the lower part of the catchment falls as snow. The bedrock consists of gneiss and anatexists, covered by soils and drift of varying depths (0.5-10 m). The basin is widely forested (75 %) and the remaining area is pasture; urban land use is below 2 %. Tracer investigations indicate that fast runoff components are generated on saturated areas and on mainly steep, highly permeable slopes, where macropore flow and pipe flow occur, and perched water tables may spread (Güntner *et al.*, 1998; Lindenlaub *et al.*, 1997; Mehlhorn *et al.*, 1998). In addition, ^{18}O investigations showed that there was a considerable old water component in storm runoff, indicating that soil and groundwater displacement take place. The slower runoff components are mainly generated in the deeper weathering zone and the fractured hard rock aquifer (Lindenlaub *et al.*, 1997).

Based on data from two meteorological stations, Freiburg (269 m a.m.s.l., located 10 km outside the catchment) and Feldberg (1493 m a.m.s.l., located at the highest part of the catchment), change in temperature and precipitation with elevation was estimated as a temperature decrease of 0.6 ^{o}C per 100 m and a relative increase of precipitation of 6 percent per 100 m. Monthly long-term mean values of the potential evapotranspiration were computed using the formula of Jensen & Haise (1963). For each day, these values were corrected based on the deviations of the temperature from its long-term mean as proposed by Lindström & Bergström (1992). A 10-year record (1975-84) was used for initialization (11 months) and calibration of the model. According to general experiences (*e.g.*, Bergström, 1992; Sorooshian & Gupta, 1995; Yapo *et al.*, 1996) the length of this period was considered to be sufficient.

Monte Carlo procedure

The hypothesis that very different parameter sets can produce almost equally good fits between simulated and observed runoff was tested by using the following Monte Carlo procedure. For each parameter, wide ranges of possible values were set based on a range of calibrated values from other model applications (*e.g.,* Bergström, 1990). Constant values were used for two less sensitive parameters in the snow routine (CWH and CFR) throughout the study. For the Monte Carlo simulations, 400 000 parameter sets were generated using random numbers from a uniform distribution within the given ranges for each parameter (Table 1). For each parameter set the model was run and the efficiency as proposed by Nash and Sutcliffe (1970), here called R_{eff}, was computed as objective function.

Investigated model variants

In a second step, the identifiability of the model structure was examined. The standard version of the model (Fig. 1, referred to as variant I) was applied using 1, 2, 5, 11 and 20 elevation zones of equal vertical extent. The area of each elevation zone was determined from a digital elevation model. For the model computations the mean elevation within an elevation zone was used. Parameter values were not allowed to vary for the different elevation zones, so the number of parameters (13) was equal for all these cases. As another variant (variant II), the elevation zones (1 to 20) were subdivided into two land use classes (forest and open land). In these cases the number of parameters increased to 18 since parameter values were allowed to vary between the two classes for the snow (CFMAX, SFCF, but not TT) and soil routine (FC, LP, BETA). In the remaining model variants (variant III-VII) five elevation zones with a single land use class were used.

For each elevation zone, the precipitation and the temperature input were computed from the measured data according to elevation. Model computations for the different zones were performed separately for some of the routines (see below) before water was mixed together and the subsequent computations were performed as a lumped approach. Altogether three variants with varying degree of distributed model calculations were studied (Fig. 2): lumped computations in all routines (*i.e.*, one elevation zone), separate computations for each elevation zone only in the snow routine (variant III, Swiss version, Braun & Renner, 1992) and in the snow and soil routines (variant I, Swedish version; Bergström, 1992). As a new model variant (variant IV) the upper box in the response function was treated separately for each elevation zone in addition to the separate computations in the snow- and soil routines (Fig. 2). The number of parameters was equal for all cases because parameter values were restrained for the different elevation zones in any of the variants.

Furthermore, three alternative model structures of the response routine were tested. The first was a response function using only one box, where the upper two outflows were active only if the storage was above certain threshold values, UZL_1 and UZL_2 [mm] (variant V, Fig. 3a). The second variant used three boxes with a linear storage-outflow relation and a maximum flow rate down to the next box of $PERC_1$ and $PERC_2$ [mm d^{-1}] (variant VI, Fig. 3b). The third alternative was to divide the recharge generated from the soil routine into two parts. The portion PART [-] was added directly to a linear storage whereas the remaining recharge generated on one day was evenly distributed over a subsequent period of DELAY [d] days and added to a second linear storage (variant VII, Fig. 3c). The same number of parameters were used in the one and three box variant as in the standard version, whilst for the delay variant, one parameter could be eliminated.

The different variants may reflect different concepts of runoff generation. The one-box variant (variant V) may represent the transmissivity feedback concept (Bishop, 1991) where runoff increase is explained by a large increase in the transmissivity when groundwater levels rise into layers with high hydraulic conductivity near the ground surface. The concept behind the three-box variant (variant VI) consists of separate aquifers, one upon the other, and the development of perched water tables. The standard version may be seen as a combination of both concepts. The delay variant (variant VII) may be seen as a representation of the situation where runoff is the sum of a fast response, *e.g.*, a shallow groundwater, and a damped, slow response, *e.g.*, a deeper hard rock aquifer.

In summary, the following model settings and structures were tested and compared. For the model variants III to VII, five elevation zones with one land use class were used.

I Standard version with 1, 2, 5, 11 and 20 elevation zones and *one* land use class

II Standard version with 1, 2, 5, 11 and 20 elevation zones and *two* land use classes (forest and open)

III Distributed computations only in the snow routine, lumped soil and response

routine
IV Distributed computations for snow and soil routine and for the upper box of the response function, only the lower box and routing routine were lumped
V Response function using one box with three linear outflow equations
VI Response function using three boxes, each with one linear outflow
VII Response function using two parallel boxes and a delayed inflow into one of them

For each case, the model was calibrated using the efficiency, R_{eff}, as objective function. The optimal parameter set for each model variant was identified using a trial and error procedure supported by Monte Carlo simulations, *i.e.*, starting with wide parameter ranges and, in iterations, confining these ranges more and more. Finally a manual fine-tuning followed this procedure.

Uncertainty of simulated discharge

The uncertainty of the simulated discharge was addressed by comparing the simulations of different hydrological scenarios by using various parameter sets of equal quality for model variant I and different model structures and settings (variants I to VII). The quality of the parameter sets was classed as ‘good’ and ‘very good’ for the model variant I, when the computed efficiency for the simulation of the 10 year record was higher than 0.850 and 0.860, respectively. These numbers were fixed with regard to the best parameter set with an efficiency 0.867 (see results). The different model structures were defined as equally good when the efficiency of the best parameter set was higher than 0.860.

These parameter sets and model variants were used to simulate different hydrological events. Extreme rainfall and snowmelt events and low flow periods were investigated. Rainfall events were examined with a 14-day synthetic precipitation sequence (SPS). For the 14 days the same rainfall pattern was used as proposed by Harlin & Kung (1992). However, the daily values were multiplied by a factor so that the 3 days with the maximum precipitation agreed with the design precipitation for 3 days recommended by the German Meteorological Survey for different probabilities (Landesanstalt für Umweltschutz & Deutscher Wetterdienst, 1978). The rainfall sequence was applied in spring, when the soil storage was saturated after snowmelt and in autumn after a dry period in summer. Floods in connection with snowmelt were examined by the simulation of a real event (December 1992). The uncertainty of low flow simulations was investigated for the longest rainless period (18 days) and for the period with the lowest measured runoff for a week within a 20-year record. All these periods were outside the calibration period.

The used methodology can be interpreted as a special, simplified case of the Generalised Likelihood Uncertainty Estimation (GLUE) procedure as proposed by Beven & Binley (1992). In the GLUE procedure as applied , for instance, by Freer *et al.* (1996) all parameter sets providing an model efficiency above some threshold value are assigned a likelihood depending on the efficiency. These parameter sets with certain likelihood values are then used to compute uncertainty bounds. In this study all parameter sets received the same likelihood, but the rejection criteria was much stricter (0.85 and 0.86 compared to 0.3, as used by Freer *et al.* (1996)). Furthermore, it has to be emphasized that in this study a large number of model runs (*i.e.*, 400 000 runs) were performed.

RESULTS

Parameter identifiability

A good agreement between observed and simulated runoff was obtained by the HBV model (variant I), simulations with a R_{eff} greater than 0.80 were obtained for several thousand of the 400 000 parameter sets. The highest value for R_{eff} was 0.867. A number of 38 parameter sets resulted in a R_{eff} value greater than 0.860 (referred as ‘very good’ parameter sets in the further description) and 380 parameter sets resulted in R_{eff} values greater than 0.850 (‘good’ parameter sets). Good simulations were found within a wide parameter range for most of the parameters. Some parameters were better defined and varied within smaller ranges, as shown by a distinct peak when plotting the parameter value against the efficiency (Fig. 4). It should be noted that every parameter value could, of course, result in poor simulations due to the values of the other parameters.

The parameters of the snow routine TT and CFMAX, the soil parameter LP and the routing parameter MAXBAS were, more or less, well defined (Fig. 5). The other parameters were badly defined, *e.g.*, an efficiency, R_{eff} , of more than 0.850 (‘good’ parameter sets) could be achieved with values for FC ranging from 160 to 550 mm. Another example is K_2, for which good results could be obtained for parameter values from 0.01 to 0.00012 d^{-1}. These values correspond to a turnover time of the dynamic reservoir ranging from 100 days to approximately 23 years. In a comparable study in southern Sweden (Seibert, 1997a, 1998) other parameters were found to be well or badly defined. For instance, the FC parameter, which was badly defined in the Brugga catchment, was a rather well defined parameter in some of the Swedish basins. On the other hand, the degree-day factor CFMAX (well defined in this study) was less well defined in the investigated Swedish basins. This indicated that it is difficult to know in advance, whether a specific parameter is well defined or not.

Model identifiability

After calibration the efficiency of the different model variants varied between 0.825 and 0.876. A fairly good simulation of observed runoff was obtained (R_{eff}=0.830) with a totally lumped model application (variant I, one elevation-land use zone). Dividing the catchment into two elevation zones resulted in a significant increase of the model efficiency (R_{eff}=0.862), whereas further divisions had a limited effect on simulations of runoff (Fig. 6).

The use of two different types of land use (variant II) resulted in higher efficiency values. However, the increase was not large considering that five more parameters were used. The increase was independent of the number of elevation zones (Fig. 6). Distributing only the snow routine (variant III) gave a significant increase of the efficiency compared to the totally lumped variant (Fig. 2), whereas the efficiency increased only slightly when the distributed snow and soil routines were included in the simulation (variant I). The simulations became better when separate computations were performed for each elevation zone in the upper groundwater box (variant IV).

With the delay variant (VII) of the response routine an efficiency of only 0.825

could be obtained, whereas the other response routine variants resulted in almost equally good fits. Simulations using the one-box variant (V) were roughly equal (R_{eff} =0.861) to the standard version (R_{eff} =0.864). In applying the three-box variant (VI) a slightly higher efficiency was obtained (R_{eff} =0.869).

Implications of identifiability problems

The ‘good’ and ‘very good’ parameter sets of model variant I as well as all other variants with an efficiency of more than 0.860 (with their respective best parameter set) were used to simulate the different hydrological scenarios. The different parameter sets and model variants gave large ranges in the runoff predictions for the various scenarios (Table 2).

In general the effects of the parameter uncertainty were somewhat larger than the effects of the different model variants. Using only the ‘very good’ parameter sets for model variant I, for instance, the simulated peak discharge varied from 40 to 58 mm d^{-1} when the synthetic precipitation sequence (SPS) with a probability of 0.01 $yr.^{-1}$ was applied in spring (Fig. 7). The range of peak runoff simulated by the different model variants was 44 – 55 mm d^{-1} (Table 2). Autumn events had a lower peak, because of the drier antecedent moisture conditions, and the range of simulated peak discharges was slightly smaller (Fig. 8). The same order of magnitude was found for the uncertainty of peak discharge of the winter event (Table 2).

The effects of the model and parameter uncertainties were larger for low flow conditions than for the flood simulations (Table 2). For example, within a 20 year period the simulated mean discharge during the week with lowest discharge varied between 0.11 and 0.96 mm d^{-1} using only ‘very good’ parameter sets, and between 0.03 and 1.24 mm d^{-1} when the ‘good’ parameter sets were used. The ranges arising from the different parameter sets and different model variants were similar, *e.g.*, the mean discharge during the rainless period of 18 days ranged from 0.8 to 1.5 mm d^{-1} in both cases.

DISCUSSION

General

It is important to distinguish between an insensitive and an uncertain model parameter. In the first case model output is not sensitive to different values of a parameter. For an uncertain parameter, on the other hand, model output may be sensitive to changes of the parameter value, but these changes can be compensated for by other parameters. The Monte Carlo method showed that good simulations could be achieved over a wide range of parameter values even for sensitive parameters. For instance, the simulations were sensitive to changes in FC when FC was changed independently, whereas good simulation could be achieved over a wide range when other parameters were also varied. The Monte Carlo procedure used in the first part of this study had the advantage that interactions between parameters were implicitly taken into account since complete parameter sets were varied instead of varying individual parameters. It was shown that very different but almost equally good parameter sets exist for the standard version of the HBV model. Comparable results were found by Sorooshian & Gupta (1995), after assessing different automatic calibration procedures for the Sacramento model. They also concluded, that it is difficult, if not impossible, to obtain a unique set of optimal parameters and that multiple optima exist.

The threshold values 0.850 and 0.860 of R_{eff} for 'good' and 'very good' parameter sets of model variant I were set arbitrarily. But they relate to the best parameter set with an R_{eff} of 0.876 which was found for this model version. However, varying these values within reasonable ranges would not alter the results substantially. The chosen thresholds are high compared to an application of the GLUE method in connection with TOPMODEL, where the model efficiency rejection criterion was set to 0.3 (Freer *et al.*, 1996). This emphasises the significance of the obtained results.

When using the Monte Carlo procedure, the intervals of the recession coefficients (K_0, K_1, K_2) overlapped, thus parameter sets with a larger storage coefficient of a lower outlet were included, *i.e.*, K_1 could be larger than K_0 or K_2 could be larger than K_1. These combinations are unusual in model applications, but have been used before (*e.g.*, Braun & Renner, 1992). The contributions of the different storages change with different parameterisations. For instance, with large values for PERC and small values for K_1 the contribution of the lower box increases even during periods of high flow. These parameter sets were also tested, because it was not apparent why one outflow should (not) contribute during certain periods. It should be noted that for all 'good' and 'very good' parameter sets all three outflows contributed significantly to runoff, *i.e.*, all parameters were active.

In the second part of the study, the standard version was applied by using different variants of the model based on different settings and modified runoff generation routines. However, the same modeling framework and input data was applied to the simulation. Also the number of parameters remained constant in most cases. This made the comparison of the different model variants straightforward. To estimate the goodness of the model performance only the simulation of one optimised parameter set was considered, *i.e.*, the parameter uncertainty was investigated only for the standard version. Although the same dimension of parameter uncertainty is expected for all investigated model variants.

In this study only the efficiency as proposed by Nash & Sutcliffe (1970) was used as an objective function to evaluate the goodness of the simulation. Results may differ if other objective functions are applied (Gan *et al.*, 1997). It has to be noted that objective functions as the efficiency give an average value over the simulation period and, thus, give no information about which periods are simulated with more or less success. Nevertheless, the use of the efficiency in this study is justified by the fact that it is the most widely used measure to assess model goodness (Singh, 1995).

Model variants versus physical reality

Delineation into two elevation zones was required, but also sufficient to obtain good model fits for this mountainous catchment. This result, and also the finding that fairly good simulation results were obtained with only one elevation zone, was unexpected because snow processes vary considerably with elevation in the Brugga catchment. The upper parts of the catchment are covered by snow for up to six months of the year while in the lower parts the snow cover is not continuous throughout the winter. However, it was not necessary to simulate these differences in detail to get reasonable simulations of observed runoff. Subdividing the elevation zones according to land use resulted in a more realistic model application, because hydrological processes and storage capacities vary between forested and open areas. The increase of runoff simulation goodness, however, was rather small considering the increase in model parameters. A detailed delineation into elevation and land use zones could not be motivated by these results, but still it may be motivated by hydrological common sense. For an objective motivation, however, additional information as, for instance, data on the extension of saturated areas (Franks *et al*., 1998), distributed snow cover (Cazorzi & Dalla Fontana, 1996) or soil moisture is required.

The efficiency of model simulations increased with an increase in the degree of distribution, *i.e.*, the number of routines that are allowed to have reservoirs with a varying degree of filling for the different elevation zones. The best fit between simulated and observed runoff was obtained with a new variant of the HBV model, which was investigated in this study. In this variant not only the snow and soil routine but also the upper groundwater box was computed separately in each zone (variant IV) and this appeared to be the most logical variant for this mountainous catchment. Shallow groundwater responds and contributes to runoff according to local inputs. Snow melt may occur, for instance, only in the upper part of the basin and will raise groundwater levels and generate runoff in the upper reaches of the catchment without influencing the lower parts. The variant with a distributed upper groundwater box has to be tested in other catchments, but the results of this study were encouraging.

The one-box variant (V) of the response routine does not agree with the concept of runoff generation developed in the Brugga catchment from detailed field investigations including tracer studies (Lindenlaub *et al.*, 1997; Güntner *et al.*, 1998; Mehlhorn *et al.*, 1998). Therefore, the good fit using this variant was somewhat surprising. The three-box variant (VI) and in particular the standard response function conceptualize the developed runoff formation concept best, but their efficiency values were not significantly higher. The only formulation of the response function that could be rejected was the delay variant (VII) with an efficiency of 0.825 for the 10-year calibration period. However, the rejection of this variant could only be done in comparison with the other formulations that gave significantly better results.

Zhang & Lindström (1996) compared the HBV model in detail with another conceptual model, namely the Chinese Xinanjiang model (Zhao, 1992). They found that both models performed well and it was difficult to see any great quality difference in the runoff simulations. They recommended to be cautious when interpreting conceptual models as actual physical descriptions of basin hydrology, and when using one model for studying the impact studies of climate or land use change. The results obtained in this study support these results and recommendations.

Piñol *et al.* (1997) compared different modifications of TOPMODEL (Beven *et al.*, 1995) in Mediterranean catchments in which they started from their perceptual understanding of the catchments. They found some of the modifications to provide qualitatively better simulations, but only a slight improvement in terms of the model efficiency. They concluded, that different measures of the goodness-of-fit and distributed data for additional validation would be needed to assess the model predictions more thoroughly.

Sometimes a successful simulation of observed runoff by a conceptual model is interpreted as a proof that the concept behind the model is true for the investigated basin (e.g., Flügel, 1995; Kirnbauer *et al.*, 1996; Buchtele *et al.*, 1996). This study, however, supported the findings of earlier studies (*e.g.*, Jakeman & Hornberger, 1993) that the information content in a rainfall-runoff record is too low to allow complex model identification. Good fits of simulated and observed runoff could be obtained starting from different and even unrealistic concepts. Therefore, from a good fit it should only be concluded that the model may represent one possible concept with other concepts possible as well.

Implications of identifiability problems

The investigated predictions of floods and low flow periods showed, that in particular the parameter uncertainty has implications for the operational use of the model. Lindström (1997) states that "for most practical applications it should be sufficient to find one good parameter set which gives an almost optimum fit and no volume errors" (p. 167). The results of this study disagree with this statement. The implications for the predicted discharges were considerable not only for extreme events (extrapolation), but also for events of similar magnitude to those during the calibration period. Similar findings are expected for the use of the model during the evaluation of land use or climate change effects. Therefore, in such studies parameter uncertainty should be considered.

The difficult identifiability of the model structure also caused uncertainties in the flow predictions. They were slightly smaller than the implications caused by the parameter uncertainty. However, it should be noted that the different model variants tested in this study were rather similar. The variations of simulation results are expected to increase for predictions with totally different conceptual rainfall-runoff models.

CONCLUDING REMARKS

Conceptual models are a strong simplification of the complex reality. A small number of boxes and flows represent the average behaviour of an infinite number of interacting reservoirs within a catchment. On the one hand, conceptual models are helpful simplifications of the natural complexity, but it is obvious that such representations always are far from reality. Fully distributed, physically based models may look more realistic than simple conceptual models. However, Beven (1989) pointed out that the physical basis of these models often is questionable, because formulations based on small-scale physics are applied on large scales which never can be expected to be as homogenous as the small scales. Given the heterogeneity of nature even the most rigorous mathematical description of a catchment must be viewed as a crude representation of reality (Grayson *et al.*, 1992). On the other hand, the representation of different processes makes conceptual models appear physically realistic. However, unless the internal flow and state variables are checked against measured data, the function of such physical elements may rather provide enough degrees of freedom to fit runoff than to simulate the hydrological processes in a realistic way. In this regard, the model may simulate runoff correctly for the wrong reasons (Klemeš, 1983;

Mroczkowski *et al.*, 1997).

Further studies of runoff generation processes in the modelled basin and additional information, *e.g.*, groundwater levels, distribution of soil moisture or saturated areas, remote sensing data, hydrochemical signals or runoff components determined by tracer techniques may help to disprove and reduce possible modelling concepts and variants (Mroczkowski *et al.*, 1997). Furthermore, additional data may enhance the model performance, because the model can be validated against more than just discharge. This may also reduce the parameter uncertainty as demonstrated by, for instance, Kuczera & Mroczkowski (1998) for a hydrosalinity model using groundwater levels and the stream water salinity for calibration. However, additional data are often measured on scales much smaller than the modelling scale. Therefore, up- or downscaling is required before simulations can be compared with these data. This may cause new sources of errors and uncertainties. In addition, there is a need for multi-objective calibration procedures capable of exploiting all the information contained in the different data. Yapo *et al.* (1998) recently suggested an algorithm for solving this multiple-objective optimisation problem. Another way to reduce parameter uncertainty may be to make more use of the information given in the hydrograph. This can be achieved by using various objective functions, which focus on different aspects of the hydrograph and the disagreement between simulated and observed runoff.

In general, model predictions should be presented as ranges or probability distributions (Melching *et al.*, 1990; Beven & Binley, 1992; Freer *et al.*, 1996) rather than as single values. These ranges are caused by equally suitable model variants and the parameter uncertainty. The Monte Carlo procedure used in this study is a suitable tool to determine prediction ranges caused by the parameter uncertainty. It takes interdependencies between parameters implicitly into account as parameter sets instead of individual parameters are varied. The increased computer power makes this formerly time consuming method attractive even for operational use. It should be noted, however, that ranges may increase if other sources of uncertainty are considered. For instance, Kuczera & Williams (1992) found that the prediction interval increased largely when uncertainty in the areal precipitation during the calibration period was considered.

To conclude, this study showed that when applying a conceptual hydrological model, uncertainties of the model structure and the model parameters, and their impacts on model predictions have to be considered. Future studies are needed to promote recommendations and procedures suitable for operational use.

Acknowledgements The runoff data used in this study were collected by the Gewässerdirektion Waldshut, Germany. The study was partly supported by the 'Förderverein Hydrologie der Universität Freiburg' and the 'Wissenschaftliche Gesellschaft Freiburg'. The authors thank Rachel C. Helliwell (Macaulay Land Use Research Institute, Aberdeen, U.K.) for her helpful linguistic suggestions on the paper. The comments of Prof. Keith J. Beven and one anonymous reviewer helped to clarify the text.

REFERENCES

Abbott, M. B., Barthurst J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. (1986a) An introduction to the European Hydrologic System - Systeme Hydrologique Europeen 'SHE' 2: structure of a physically based, distributed modelling system. *J. Hydrol.* **87**, 61-77.

Abbott, M. B., Barthurst J. C., Cunge, J. A., O'Connell, P. E. & Rasmussen, J. (1986b) An introduction to the European Hydrologic System - Systeme Hydrologique Europeen 'SHE' 1: history and philosophy of a physically based, distributed modelling system. *J. Hydrol*. **87**, 45-59.

Beck, M. B. (1987) Water quality modeling: a review of the analysis of uncertainty. *Wat. Resour. Res.* **23,** 8, 1393-1442.

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

Bergström, S. (1990) *Parametervärden för HBV-modellen i Sverige, Erfarenheter från modelkalibreringar under perioden 1975-1989* (Parameter values for the HBV model in Sweden) (in Swedish). SMHI, No. 28, Norrköping, 35 pp.

Bergström, S. (1991) Principles and confidence in hydrological modelling. *Nordic Hydrol.* **22**, 123-136.

Bergström, S. (1992) *The HBV model - its structure and applications*. SMHI, No. 4, Norrköping, 35 pp.

Bergström, S. (1995) The HBV model. In: Singh, V. P. (ed.) *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, Colorado, U.S.A., 443-476.

Beven, K. J. (1989) Changing Ideas in Hydrology: The case of physically-based models. *J. Hydrol*. **105,** 157-172.

Beven, K. J. (1993) Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Wat. Resour.* **16**, 41-51.

Beven, K. J. & Binley, A. (1992) The future of distributed models: model calibration and uncertainty prediction. *Hydrol. 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, Colorado, pp. 627-668.

Bishop, K. (1991) *Episodic increase in stream acidity, catchment flow pathways and hydrograph separation*. Ph.D. Thesis, University of Cambridge, Department of Geography, UK, 246 pp.

Brandt, M., Bergström, S. & Gardelin, M. (1988) Modelling the effects of clearcutting on runoff - examples from central Sweden. *Ambio* **17,** 15, 307-313.

Braun, L. N. & Renner C. B.(1992) Application of a conceptual runoff model in different physiographic regions of Switzerland. *Hydrol. Sci. J.* **37**, 3, 217-231.

Buchtele, J., Elias, V., Tesar, M. & Hermann, A. (1996) Runoff components simulated by rainfall-runoff models. *Hydrol. Sci. J.* **41**, 1, 49-60.

Cazorzi, F. & Dalla Fontana, G. (1996) Snowmelt modelling by combining air temperature and a distributed radiation index. *J. Hydrol*. **181**, 169-187.

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

Flügel, W. A. (1995) Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrol. Processes* **9**, 423-436.

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 into the calibration process. *Wat. Resour. Res*. **34**, 4, 787-797.

Freer, J., Ambroise, B. & Beven, K. J. (1996) Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Wat. Resour. Res.* **32**, 7, 2161-2173.

Gan, T. Y., Dlamini, E. M. & Biftu, G. F. (1997) Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *J. Hydrol*. **192**, 81-103.

Gaume, E., Villeneuve, J.-P. & Desbordes, M. (1998) Uncertainty assessment and analysis of the calibrated parameter values of an urban storm water quality model, *J. Hydrol*. **210**, 38-50.

Grayson, R. B., Moore, I. D. & McMahon, T. A. (1992) Physically based hydrologic modeling, 2. Is the concept realistic? *Wat. Resour. Res*. **26**, 10, 2659-2666.

Güntner, A., Uhlenbrook, S., Seibert, J., Leibundgut, Ch. (1998): Estimation of saturation excess overland flow areas - comparison of topographic index calculations with field mapping. Accepted by *IAHS Publ.*

Harlin, J. & Kung, C. S. (1992) Parameter uncertainty and simulation of design floods in Sweden. *J. Hydrol.* **137**, 209-230.

Jakeman, A. J. & Hornberger, G. M. (1993) How much complexity is warranted in a rainfall-runoff model. *Wat. Resour. Res.* **29**, 8, 2637-2649.

Jensen, M. & Haise, H. (1963) Estimating evaporation from solar radiation. *J. of Irrigation and Drainage* **89**, 4, 15-41.

Kirnbauer, R., Pirkl, H., Haas, P. & Steidl, R. (1996) Abflussmechanismen - Beobachtung und Modellierung (Runoff Processes - observations and modelling) (in German). *Österreichische Wasser- und Abfallwirtschaft* **48**, 1/2, 15-26.

Klemes, V. (1983) Conceptualization and scale in hydrology. *J. Hydrol.* **65**, 1-23.

Kuczera, G. & Williams, B. J. (1992) Effect of rainfall errors on accuracy of design flood estimates. *Wat. Resour. Res.* **28**, 4, 1145-1153.

Kuczera, G. & Mroczkowski, M. (1998) Assessment of hydrologic parameter uncertainty and the worth of data. *Wat. Resour. Res.* **34**, 6, 1481-1489.

Landesanstalt für Umweltschutz (LfU) & Deutscher Wetterdienst (DWD) (1978) *Verteilungskurven der Niederschläge in Baden-Württemberg* (Distribution functions of precipitation in Baden-Württemberg) (in German), Karlsruhe, 131 pp.

Lindenlaub, M., Leibundgut, Ch., Mehlhorn, J. & Uhlenbrook, S. (1997) Interactions of hard rock aquifers and debris cover for runoff generation. In: Pointet, T.(ed): *Hard rock Hydrosystems*. IAHS Publ. no. 241, 63-74.

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

Lindström, G. & Bergström, S. (1992) Improving the HBV and PULSE models by use of temperature anomalies. *Vannet i Norden* **25**, 1, 16-23.

Mehlhorn J., Armbruster F., Uhlenbrook S. & Leibundgut Ch. (1998) Determination of the geomorphological instantaneous unit hydrograph using tracer experiments in a headwater basin. In: Kovar, K., Tappeiner, U., Peters, N. E. & Craig, R. G. (ed.) Hydrology, Water and Resources and Ecology in Headwaters. IAHS Publ. no. 248, 327-335.

Mein, R. G. & Brown, B. M. (1978) Sensitivity of optimized parameters in watershed models. *Wat. Resour. Res.* **14**, 2, 299-303.

Melching, C. S., Yen, B. C. & Wenzel, Jr. H. G. (1990) A reliability estimation in modeling watershed runoff with uncertainties. *Wat. Resour Res.* **26**, 10, 2275-2286.

Mroczkowski, M., Raper, G. P. & Kuczera, G. (1997) The quest for more powerful validation of conceptual catchment models. *Wat. Resour. Res.* **33**, 10, 2325-2335.

Nash, J. E. & Sutcliffe, J. V. (1970) River flow forecasting through conceptual models; Part I - A discussion of principles. *J. Hydrol.* **10**, 282-290.

van der Perk, M. & Bierkens, M. F. P. (1997) The identifiability of parameters in a water quality model of the Biebrza River, Poland. *J. Hydrol.* **200**, 307-322.

Piñol, J., Beven, K. J. & Freer, J. (1997) Modelling the hydrological response of Mediterranean catchments, Prades, Catalonia. The use of distributed models as aids to hypothesis formulation. *Hydrol.Proc.* **11**, 9, 1287-1306.

Saelthun, N. R. (1996) *The 'Nordic' HBV Model. Description and documentation of the model version developed for the project Climate change and Energy Production.* Norges Vassdrags- og Energiverk, **7**, 26 pp.

Seibert, J. (1997a): Estimation of parameter uncertainty in the HBV model. *Nordic Hydrol.* **28**, 4/5, 247-262.

Seibert, J. (1997b) *HBV light 1.2, user's manual*. Uppsala University, Institute of Earth Sciences, Department of Hydrology, Uppsala, Sweden.

Seibert, J. (1998) Regionalisation of parameters for a conceptual rainfall-runoff model. Accepted by *Agricultural and Forest Meteorology.*

Singh, V. P. (ed.) (1995) *Computer models of watershed hydrology*. Water Resources Publications, Highlands Ranch, Colorado, USA, 1130 pp.

Sorooshian, S. & Gupta, V. K. (1995) Model Calibration, in: Singh, V. P. (Ed.): *Computer models for watershed hydrology*. 23-68, Water Resources Publications, Highland Ranch, Colorado, USA.

Todini, E. (1988) Rainfall-runoff modelling - past, present and future. *J. Hydrol*. **100**, 341-352.

Uhlenbrook, S., Holocher, J., Leibundgut, Ch. & Seibert S. (1998) Using a conceptual rainfall-runoff model on different scales by comparing a headwater with larger basins. In: Kovar, K., Tappeiner, U., Peters, N. E. & Craig, R. G. (ed.) Hydrology, Water and Resources and Ecology in Headwaters. *IAHS Publ*. no. 248, 297-306.

Yapo, O. P., Gupta, H. V. & Sorooshian, S. (1996) Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. *J. Hydrol.* **181**, 23-48.

Yapo, O. P., Gupta, H. V. & Sorooshian, S. (1998) Multi-objective global optimization

for hydrologic model. *J. Hydrol.* **204**, 83-97.
Zhang & Lindström, G. (1996) A comparative study of a Swedish and Chinese hydrological model. *Water Resources Bulletin* **32**, 5, 985-994.
Zhao, R. J. (1992) The Xinanjiang model applied in China. *J. Hydrol.* **135**, 371-381.

FIGURE CAPTIONS

Fig. 1 Schematic sketch of the HBV model structure (modified after Bergström, 1992).

Fig. 2 Model variants with increasing degree of distributed model calculations.

Fig. 3 Different model variants of the response routine: (a) one-box variant, (b) three-box variant and (c) delay variant.

Fig. 4 Results of two model parameters after 400 000 model runs, each point represents one model run: a) CFMAX as an example for a well defined parameter; b) FC as an example for a badly defined parameter.

Fig. 5 Summary of the uncertainty of the model parameters, the standardised range refers to the ranges used for the Monte Carlo procedure (Table 1).

Fig. 6 Model goodness versus number of elevation zones for one and two land use zones.

Fig. 7 Model simulations when a synthetic rainfall sequence(SPS) (probability 0.01 yr^{-1}) was applied in spring after snowmelt with all very good parameter sets ($R_{eff} >$ 0.860).

Fig. 8 Range of peak discharge generated by synthetic rainfall sequences (SPS) of different probabilities applied in spring and autumn simulated with all very good ($R_{eff} >$ 0.860) and good ($R_{eff} >$ 0.850) parameter sets.

TABLES

Table 1 Parameters and their ranges used for the Monte Carlo simulations.

Parameter	Explanation	Unit	Minimum	Maximum
Snow routine				
TT	Threshold temperature	°C	-2	0.5
CFMAX	Degree-day factor	mm $°C^{-1}$ d^{-1}	0.5	4
SFCF	Snowfall correction factor	-	0.5	0.9
CWH [a)]	Water holding capacity	-	0.1	0.1
CFR [a)]	Refreezing coefficient	-	0.05	0.05
Soil and evaporation routine				
FC	Maximum SM	mm	100	550
LP	SM threshold for reduction of evaporation	-	0.3	1
BETA	Shape coefficient	-	1	5
C_{ET}	Correction factor for potential evaporation	$°C^{-1}$	0	0.3
Groundwater and Response routine				
K_0	Recession coefficient	d^{-1}	0.1	0.5
K_1	Recession coefficient	d^{-1}	0.01	0.2
K_2	Recession coefficient	d^{-1}	0.00005	0.1
UZL	Threshold for K_0-outflow	mm	0	70
PERC	Maximal flow from upper to lower GW-box	mm d^{-1}	0	4
Routing routine				
MAXBAS	Routing, length of weighting function	d	1	2.5

[a)] This parameter was not varied.

Table 2 Ranges of simulations (mm d^{-1}) of different events with parameter sets and model variants with R_{eff} > 0.860 during the calibration period.

Scenario	Observed value	Simulations using different parameter sets (model variant I)		Simulations using different model variants	
		Minimum	Maximum	Minimum	Maximum
Peak discharge (SPS 0.01 yr^{-1}spring [a)])	-	39.8	57.6	43.6	54.5
Peak discharge (SPS 0.01yr^{-1} autumn [b)])	-	34.2	48.7	36.3	47.1
Peak discharge (December 92)	27.4	21.1	27.5	22.2	26.4
Mean discharge (rainless period)	1.12	0.84	1.46	0.80	1.45
Mean discharge (week with minimum discharge)	0.57	0.11	0.96	0.11	0.71

[a)]synthetic precipitation sequence (SPS) with a probability of 0.01 yr^{-1}, applied in spring
[b)]synthetic precipitation sequence (SPS) with a probability of 0.01 yr^{-1}, applied in autumn

FIGURES

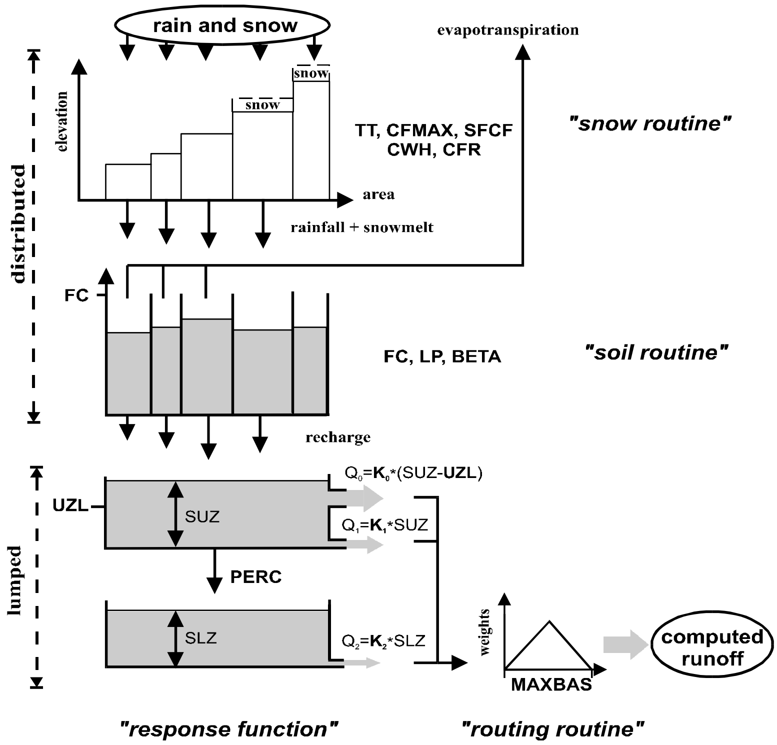

Fig. 1 Schematic sketch of the HBV model structure (modified after Bergström, 1992).

Fig. 2 Model variants with increasing degree of distributed model calculations.
Fig. 3 Different model variants of the response routine: (a) one-box variant, (b) three-box variant and (c) delay variant.
Fig. 4 Results of two model parameters after 400 000 model runs, each point represents one model run: a) CFMAX as an example for a well defined parameter; b) FC as an example for a badly defined parameter.

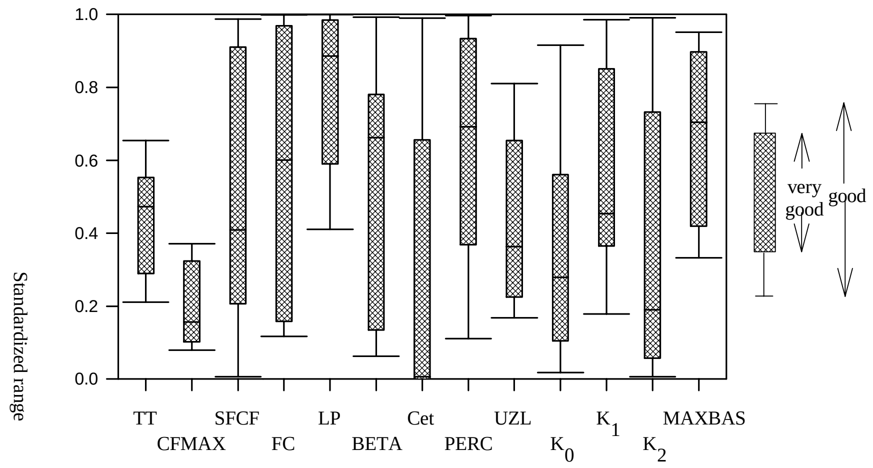

Fig. 5 Summary of the uncertainty of the model parameters, the standardised range refers to the ranges used for the Monte Carlo procedure (Table 1).

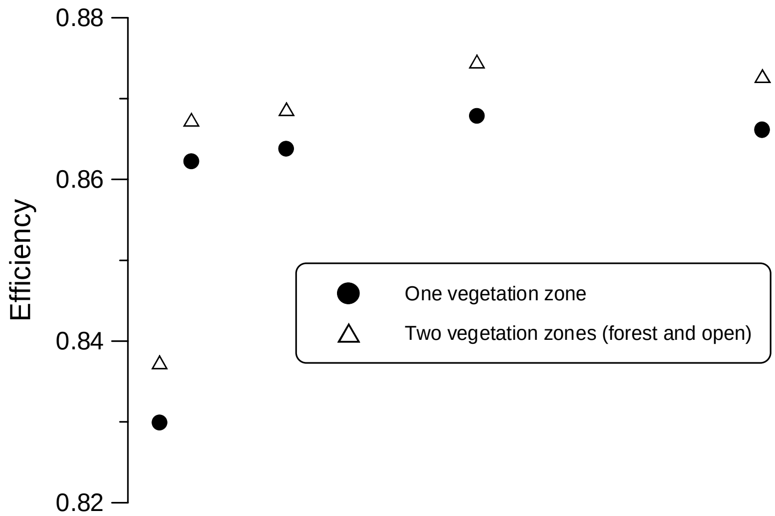

Fig. 6 Model goodness versus number of elevation zones for one and two land use zones.

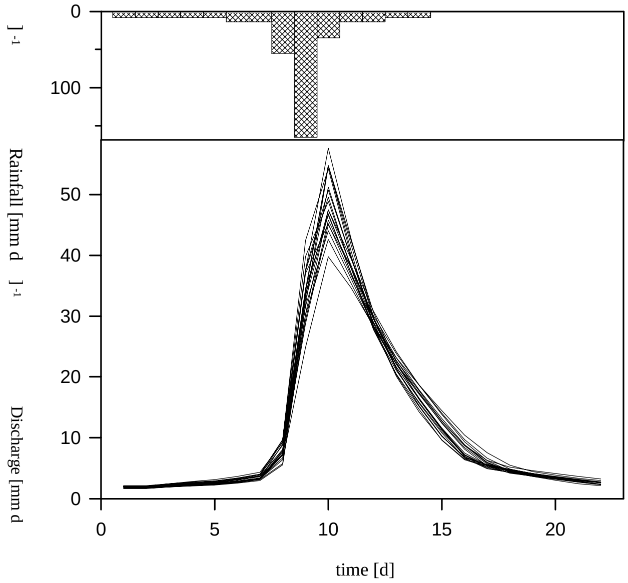

Fig. 7 Model simulations when a synthetic rainfall sequence(SPS) (probability 0.01 yr^{-1}) was applied in spring after snowmelt with all very good parameter sets (R_{eff} > 0.860).

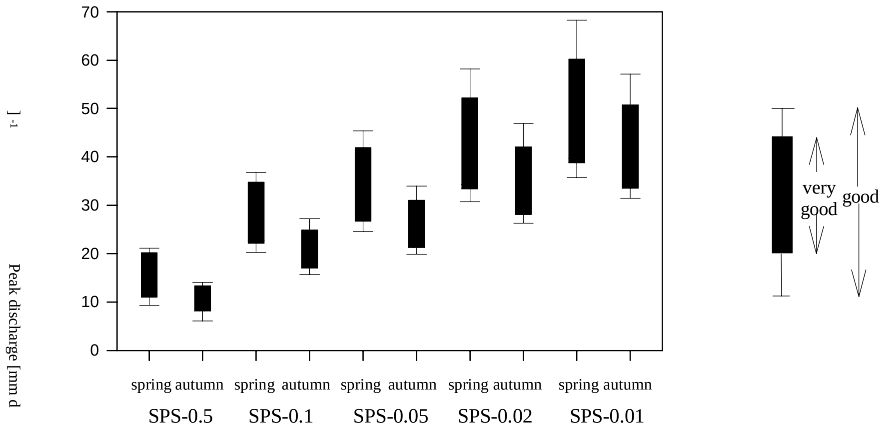

Fig. 8 Range of peak discharge generated by synthetic rainfall sequences (SPS) of different probabilities applied in spring and autumn simulated with all very good (R_{eff} > 0.860) and good (R_{eff} > 0.850) parameter sets.