Comparative evaluation of performances of different conceptualizations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments

Teklu T. Hailegeorgis* and Knut Alfredsen

Department of Hydraulic and Environmental Engineering, NTNU, NO-7491 Trondheim, Norway

***Corresponding author information:**

Name: Teklu Tesfaye Hailegeorgis

Address: Department of Hydraulic and Environmental Engineering, NTNU, NO-7491 Trondheim, Norway.Tel. [+47] 73592411, Mobile [+47] 45069384, Fax [+47] 735 91298 and e-mail addresses: tekhi09@gmail.com,teklu.hailegeorgis@ntnu.no.

ABSTRACT

Unidentifiability and equifinality of parameters pose challenges to calibration and prediction by conceptual precipitation-runoff models. Evaluation of prediction performances of parametrical parsimonious and more complex conceptualizations is lacking for hourly simulation. We conducted a comparative evaluation of four configurations of the distributed (1x1 km^2 grids) HBV runoff response routines for hourly streamflow simulation for boreal mountainous catchments in mid-Norway. The routines include the standard Swedish Meteorological and Hydrological Institute HBV or HBV-SMHI, HBV-Nonlinear (standard soil routine and non-linear reservoirs), HBV-Soil Parsim R (standard soil routine and linear reservoirs) and HBV-Parsim (parsimonious and linear soil routine and reservoirs).

The routines provided simulated hydrographs, flow duration curves and quantile-quantile (QQ) plots, which are marginally different from each other for the study catchments. However, the

HBV-Parsim provided better parameter identifiability and uncertainty, and simulated baseflow that better matches the baseflow separated by filtering techniques. Performances of the HBV-Parsim indicated a potential for application of parametrical parsimonious routines, which would benefit model updating for forecasting purposes. The study revealed strong effects of the soil-moisture parameters on the recharge, percolation and hence the baseflow, which substantiates importance of evaluating the internal simulation (e.g. soil-moisture and baseflow) of the HBV routines against measurements or analytical computations.

Key words: Distributed HBV runoff response routines, hourly runoff simulation, parametrical parsimony, model calibration and spatio-temporal validation, boreal mountainous catchments, baseflow simulation.

INTRODUCTION

Various conceptual runoff response routines are currently used for decision making for operational forecasting though they are not capable of detailed modelling of physical hydrological processes (e.g. preferential flows). Some of the recent works endorsing the utility of conceptual models include Fenicia et al. (2011), and Kavetski and Fenicia (2011) who tested a flexible framework for conceptual model structure for comparison and refinement of alternative hypotheses. Precipitation-runoff models based on the standard HBV (Bergström, 1976) which is named as the HBV-SMHI in the present study and its variants have been widely applied in many countries for real-time forecasting of floods and inflow to storage reservoirs, for design flood estimation and for scenario studies such as the impacts of anticipated climate change (e.g. Harlin, 1992; Harlin and Kung, 1992; Killingveit and Sælthun, 1995; Lindström et al., 1997; Blöschl et al., 2008; Driessen et al., 2010). For the representation of spatial heterogeneity through different levels of spatial discretization, the standard HBV model has evolved through different variants from lumped to ‘fully’ distributed versions e.g. Lindstrom et al. (1997), Hundecha and Bardossy (2004), Beldring et al. (2003), Blöschl et al. (2008), Das et al. (2008) and Li et al. (2014). Wrede et al. (2013) further implemented a 'fully' distributed HBV model complemented by subgrid scale parameterization for distinct land use classes or hydrological response units for a Swedish lowland catchment. Based on the distributed model intercomparison project (DMIP) experiments, Smith et al. (2004) suggested distributed models as complements rather than replacements of the lumped models for operational forecasting purposes.

The HBV model has gained significant applications for daily forecasting, and with recent developments for hourly simulation/forecasting (e.g. Kobold and Brilly, 2006; Shrestha et al., 2008 and Rakovec et al., 2012). There are growing interests for hourly application of the HBV model since hourly flood forecasting is becoming important due to the prevalence of extreme hydrological events driven by intense rainfall events. In addition, for operational hydropower management a reliable hourly inflow prognosis is required. For instance, real-time forecasting of inflow to hydropower reservoirs during hydropeaking (i.e. when hydropower is operating to balance intermittent renewables or non-renewable energy sources for peak demands of electricity) requires runoff simulation for shorter time steps for optimal scheduling and to minimize downstream impacts of releases. In addition, for catchments that are affected by diurnal variations of streamflow due to diurnal variations in snowmelt and evapotranspiration, the hourly simulation is expected to better capture the temporal dynamics of the hydrological processes for reliable prediction/forecasting of streamflow. Also, flood peaks may not be reliably simulated based on a coarse temporal resolution (e.g. daily time step) especially for small basins. Several studies were conducted on time scale dependencies of conceptual model parameters (Littlewood and Croke, 2008; Wang et al., 2009; Merz et al., 2009). Kavetski et al. (2011) thoroughly investigated the time scale dependencies of information content of data, parameter calibration and identifiability, quickflow and hydrograph peak simulation. The considerable loss in performance when parameters calibrated for a longer time step (e.g. daily streamflow) are used for simulation for shorter time step (e.g. hourly) as demonstrated by Bastola and Murphy (2013) showed the introduction of additional uncertainty and hence associated risks in hourly prediction from utilizing the daily calibrated parameters. Therefore, hourly predictions based on parameters calibrated utilizing hourly observations are required for water management purposes.

Depending on their conceptualization, the different versions of the HBV response routines contain different numbers of free parameters while large numbers of free parameters do increase the complexity in parameter calibration and identifiability. Moreover, the reliability of simulation based on different conceptualization may be different. The standard form SMHI (Bergström 1976), the HBV light (Seibert, 1997b; 2002) and the "Nordic" HBV model (Sælthun, 1996) all have two (non-linear upper and linear lower) reservoirs and three outlets (three recession coefficients, upper zone threshold and one percolation parameter). The operational HBV version used by Norwegian hydropower companies (Killingveit and Sælthun, 1995) has two (non-linear upper and linear lower) reservoirs and four outlets (four drainage coefficients, two upper zone thresholds and one percolation parameters). The HBV-96 (Lindström et al., 1997) and the HBV-IWS (Hundecha and Bárdossy, 2004) have two (non-linear upper and linear lower) reservoirs and four parameters (two recession coefficients, non-linearity exponent and percolation parameters). The Tillart (2010) version of HBV contains two (non-linear upper and linear lower) reservoirs with five parameters (upper and lower reservoir recession coefficients, non-linearity exponent, percolation rate and capillary flux).

Reed et al. (2004) in the DMIP noted that model formulation and parameterization could have a bigger impact on simulation accuracy than the spatial computational unit (lumped versus distributed). The authors also pointed out that due to interacting and compensating effects of different factors, studies based on changing one model component at a time will be required to identify reasons for any observed differences among the models. Parsimonious parameterization is important since large numbers of free parameters do not guarantee good model performance (e.g. Michaud and Sorooshian, 1994). The cost of model calibration increases as the number of free parameters increases. The problem of equifinality and poor identifiability related to overparameterization, which potentially happens when complex models are calibrated using data of low information content, were addressed among others by Beven and Binley (1992); Beven and Freer (2001) and Kirchner (2006; 2009). Werkhoven et al. (2009) illustrated the importance of reduction of parameter dimensionality. Each additional parameter represents a whole new dimension of parameter space, so the overparameterization problem grows with the number of free parameters (Kirchner, 2006). Uhlenbrook et al. (1999) investigated the prediction uncertainty of different variants of the HBV model caused by problems in identifying model parameters and structure.

However, studies of relationships between prediction performances and various configurations of the distributed HBV response routines and the numbers of free parameters is lacking from literature. Dependent on the ability of the model to simulate the precipitation-runoff relationships, conceptualization based on less number of parameters is preferable. Despite the pros in allowing better identifiability of parameters, there are also cons against parsimonious models. For example, Kuczera and Mroczkowski (1998) noted that a simple model cannot be relied upon to make meaningful extrapolative predictions, where a complex model may have the potential but because of information constraints may be unable to realize it. Due to the pros and cons related to parsimony and more complexity in hydrological modelling, the focus of the present study is on comparative evaluation of the runoff response routines. Some of the previous attempts to reduce the number of free parameters in the HBV response routine include works by Harlin (1992) and Winsemius et al. (2009). Samuel et al.

(2012) evaluated different configurations of the HBV response routines for baseflow simulation in Canada.

Jakeman and Hornberger (1993) in their ‘principle of parsimony’ claimed that simple conceptual models with four to five parameters provide an adequate fit if only streamflow is available for calibration. Observing a large degree of equifinality in calibration of the nine HBV-96 parameters for four climatologically different river basins, Lidén and Harlin (2000) stated that given the inherent limitations of information in calibration data only a smaller number of parameters can be uniquely identified which calls for a parsimonious model. When calibration is based on different state variables (e.g. ground water level, soil moisture, snow depth) in addition to streamflow, multi-objective calibration gives an opportunity to further exploit the information content of each variable to better constrain the model parameters. However, the challenge is that such a rich data set may not be readily available especially for operational purposes. Hence, the maximum possible exploitation of the information content of the relatively readily available streamflow data is a possible solution.

Kokkonen and Jakeman (2001) while explaining the higher information requirement of a complex model structure stated that the more process complexity one wants to include in the model structure the more types of data are required to estimate the process parameters and to test the model performance. Perrin and Andréassian (2001) compared 19 lumped models for daily simulation in 429 catchments and found that models with a large number of parameters generally yield better calibration results, but were not verified in the validation stage. Hughes (2010) noted that if the model performance is evaluated only by the success of calibration against observed streamflow, certain simpler models would frequently out-perform the more complex models but selection of models for wider objectives such as prediction in ungauged basins is far complex. Better process understanding augmented by field experiments and measurements to conceptualize improved (suitable) model structures for simulation of dominant hydrological processes is indispensable in catchment hydrology but this task was not an objective of the present study.

The main objective of the present study focus on the evaluation of the performance of different configurations of the runoff response routine of the HBV model. The model configurations used in the study is the distributed (1x1 km^2) standard HBV (HBV-SMHI), non-linear storage-discharge relationships HBV (HBV-Nonlinear), standard soil-moisture accounting and linear parsimonious runoff response routine (HBV-Soil Parsim R) and parsimonious and linear conceptualizations of both the soil-moisture accounting and runoff response routines (HBV-Parsim). The main research questions are:

(1) What are the performances of different configurations of the HBV runoff response routines for distributed (1x1 km^2) hourly runoff simulation in boreal mountainous catchments in terms of prediction of various streamflow 'signatures', parameter identifiability and uncertainty?

(2) What are the performances of the routines in terms of spatial and temporal validation when parameters are transferred to internal (interior) subcatchments and among the non-nested catchments inside the study watershed?

THE STUDY REGION

The study region is the mountainous boreal watershed of Gaula in mid Norway. We used streamflow data from four stations, which include Gaulfoss, Eggafoss, Hugdal bru and Lillebudal bru. Hugdal bru, Eggafoss and Lillebudal bru are nested within Gaulfoss, but they are independent of each other. Rainfall mainly occurs from April to October while snowfall occurs from November to March. We used hourly precipitation data from twelve climate stations with an elevation range from 127 to 885 masl. We interpolated the climate inputs on the spatial computational scale of 1x1 km^2 grids by the inverse distance weighing (IDW) method. The maximum length of input records for calibration was two years of hourly temporal scale due to difficulty in finding long series of hourly climate and streamflow data with complete records. However, the length of the hourly data series is expected to provide sufficient information on the rainfall-runoff relationships to constrain the model parameters. Where sub-daily data exist, it would appear to be wise to use the extra information they contain, leading to more accurate calibrated model parameters (Littlewood and Croke, 2013).

The hourly climate data used include precipitation (P), temperature (T), wind speed (W_s), relative humidity (H_R) and global radiation (R_G). The main vegetation/land use types in the catchments are forests (conifers), upland and riparian areas. The dominant loose (soil) formation in the Gaula watershed is glacial tills underlain by impermeable bedrock geology (http://www.ngu.no/no/hm/Kart-og-data). The main surface deposits in the Nordic countries are till soils (Beldring et al., 2000). The main characteristics of the catchments and maps of land use, elevation, locations of climate and streamflow stations for the study catchments are given in Table 1 and Fig.1.

MODELS AND METHODS

The four configurations of the conceptual HBV runoff response routines modelled and evaluated in the present study differ either in the number of conceptual reservoirs (one versus two), form of storage and discharge (S-Q) equations (linear versus non-linear), the soil-moisture accounting routine or the number of free parameters (few/parsimonious versus many/more complex). A summary of the main features of the routines is given in Table 2 and Fig. 2. The lists of the free parameters and ranges of their uniform priors is given in Table 3.

The HBV-SMHI distributed runoff response routine

The distributed runoff response routine based on the standard HBV model used in the present study contains two conceptual reservoirs (Fig. 2a). The upper reservoir contains two outlets (one with non-linear and one with linear drainage equations) and the lower reservoir is linear with only a single outlet. It has five free parameters in the response routine that include very quick and quick upper reservoir recession coefficients (k_2 and k_1 respectively), slow base flow recession coefficient (k_0), percolation rate (PERC) and upper reservoir threshold storage (UZ_t). For the upper reservoir, the runoff generated conceptually represents very quick and quick runoff components, both from overland flow and from groundwater drainage from perched aquifers (Killingtveit and Sælthun, 1995). The storage-outflow relationship for the upper zone is threshold based and the total outflow from the upper reservoir (Q_{UZ}) is the sum of outflows from the lower (Q_{UZ1}) and upper (Q_{UZ2}) outlets. The lower zone conceptually represents the base flow from ground water and the storage-baseflow relationship is linear.

HBV-Nonlinear distributed runoff response routine

The HBV-Nonlinear runoff response routine has two storage reservoirs (upper and lower) conceptually similar to those of the HBV-SMHI routine but with the basic differences in configuration of the structure of the upper conceptual reservoir by a single outlet (Fig. 2b) with an exponent based non-linear storage-outflow relationship and also a non-linear storage-baseflow relationship for the lower reservoir. The total numbers of parameters in the response routine are five (i.e. upper and lower reservoir recession coefficients, two non-linearity parameters and percolation). Simulation over fixed discrete time steps of the hourly data resolution was used in the present study assuming that the input forcings and fluxes are constant over the time step. The fixed-step approaches are commonly used for conceptual models (e.g. Lindström et al., 1997; Blöschl et al., 2008). Focus on the effects of numerical artifacts (e.g. Clark and Kavetski, 2010) in solving the storage-discharge functions were not the objective of the present study.

The HBV-Soil Parsim R distributed runoff response routine

The HBV-Soil Parsim R runoff response routine has two (upper and lower) reservoirs (Fig. 2c) similar to those of the HBV-Nonlinear routine but with the basic differences that both the storage-outflow (upper reservoir) and storage-baseflow (lower-reservoir) relationships are linear. Hence, the total numbers of parameters in the response routine are three, namely the quick reservoir recession coefficient (k_1), slow base flow recession coefficient (k_0) and percolation to the lower reservoir (PERC).

The HBV-Parsim distributed runoff response routine

The HBV-Parsim runoff response routine (Fig. 2d) is similar to the HBV-Soil Parsim R response routine with the only difference in their soil-moisture accounting routines.

The soil moisture accounting routine

The soil moisture accounting routine partitions the effective precipitation in to recharge to the upper zone and contribution to change in the soil moisture storage. For the HBV-SMHI, HBV-Nonlinear and HBV-Soil Parsim R, the soil moisture accounting is based on a non-linear function which partitions the infiltration from rainfall and snowmelt (I) into recharge (R) to upper reservoir and change in soil moisture storage or Δ_{SM} (Bergström 1976). The 'non-linearity parameter' β controls the shape of the partitioning curve. The soil moisture storage (SM) is depleted by evapotranspiration. If the ratio of the actual soil moisture (SM) to the field capacity (FC) or SM/FC exceeds the 'limit for potential evaporation' or evapotranspiration threshold parameter (LP), actual evapotranspiration (AET) is assumed to be equal to the potential (PET). The field capacity represents the maximum soil moisture holding capacity of the soil. However, if the soil moisture is below (LP*FC), the actual evapotranspiration decreases linearly with the decrease in soil moisture storage (i.e. AET/PET = SM/ (LP*FC). Therefore, the soil moisture accounting routine of the HBV-SMHI involves three free parameters namely the field capacity (FC), shape parameter (β) and LP. A capillary flux (flow) from the storage reservoirs to the soil-moisture zone was not considered in the present study.

However, for the HBV-Parsim the soil moisture accounting routine is based on a linear function (i.e. $\beta = 1.0$), LP is also set to a constant value of 0.9 which is a default value of HBV-96 (Booij, 2005). The sensitivity of the HBV model to the FC parameter was studied by Seibert et al. (1999), who reported both well-defined and badly-defined (i.e. insensitive or uncertain) cases respectively for some of Swedish catchments and mountainous catchment in Germany. Abebe et al. (2010) found that the FC parameter to be dominantly affecting the high flow and volume errors for semi-humid catchment in USA. Beldring et al. (2003) obtained FC values of 20 mm to 150 mm for different land cover from simultaneous calibration of a distributed version of the standard HBV model for 141 catchments in Norway. Following the results obtained by Beldring et al. (2003), we assigned field capacity values for different land classes due to our objective of testing a parametrical parsimonious response routine with a total number of five free parameters based on the 'principle of parsimony' by Jakeman and Hornberger (1993). For the Gaula watershed, land cover above timberline (approximate timberline elevation ranges from 800-850 masl) is mainly dominated by sparse vegetation, bedrock/cracked rock with some proportions of lichen, heather and shrubs. The area below the timberline is dominated by coniferous forests with some proportions of deciduous forests and non-forested areas such as farmland and marshes/bogs. Therefore, we used field capacity (FC) maps with values of 150 mm and 50 mm respectively below and above the timberline. Therefore, there is no free parameter in the soil moisture accounting routine (i.e. three free parameters are removed compared to HBV-SMHI, HBV Non-linear and HBV-Soil Parsim R).

The potential evapotranspiration was computed using the Priestley Taylor method (Priestley and Taylor. 1972) for the all response routines:

$$PET = \alpha \frac{\Delta}{\Delta + \gamma} R_n - G \tag{1}$$

,where α is Priestley Taylor constant, Δ is the slope of saturation vapor pressure curve at air temperature at 2m (T2m), Ƴ is the psychometric constant (0.67 $hPaK^{-1}$), R_n is the net radiation = net shortwave radiation (SR_n) + the net longwave radiation (LR_n). Priestley and Taylor (1972) obtained values of α, for diverse well-watered surfaces, between 1.08 and 1.34 with an overall mean of 1.26. Teuling et al. (2010) used α = 1.26 for the Swiss catchment. Gardelin and Lindstrom (1996) determined the value of α by calibration. Following Teuling et al. (2010), we used an alpha value of 1.26 rather than calibrating the parameter due to our objective of testing parametrical parsimonious HBV model and due to expected less effect of fixing the parameter to an average value on the PET calculation for the snow-dominated boreal catchment. The SR_n is computed from the measured global radiation (R_G) and land albedo. The LR_n is computed based on Sicart et al. (2006). The soil/ground heat flux (G) = (0.12)* R_n is assumed following Teuling et al. (2010).

If the grid cells are within a lake, direct precipitation on the lake, evaporation from the lake and outflow are considered for the lower reservoir zone (i.e. there is no soil-moisture accounting routine and the upper zone reservoir). The evaporation from the lake surfaces is assumed to be 40% above the potential evapotranspiration computed from the Priestley Taylor method.

The snow routine

For the study catchments, snow accumulation and melt processes dominate during winter and spring seasons respectively. Snow routines based on temperature index models or degree-day methods are commonly used to simulate snowmelt rates in many variants of the HBV models. However, in the present study we used the gamma distributed snow depletion curve or SDC, which uses radiation for surface layer energy and phase change calculations (Kolberg and Gottschalk, 2006) for all the response routines. The routine uses a mass balance approach to simulate the melt water release (snowmelt runoff) from saturated snow (Qs) and the remaining unmelted snow storage, the snow water equivalent (SWE) (Fig. 2).

The snow depletion curve describes how snow covered area (SCA) reduces gradually through the melt season by relating the fractional snow covered area in a grid cell to the mass balance of a heterogeneous snow cover (Kolberg and Gottschalk, 2010). Kolberg and Gottschalk (2006; 2010) defined the four variables defining the snow pack state in a grid cell (Fig. 2) as the average snow water equivalent or SWE (mm) at the start of melt season, the SWE coefficient of variation cv (-) explaining the subgrid spatial heterogeneity, the fractional snow covered area at the melt start of the melt season (-) and the accumulated melt depth, λ (mm) aggregated from the melt season onset or end winter day. The free parameters in the routine are the TX which is the snow-rain threshold temperature parameter and identifies the form of precipitation (rainfall or snow fall) and wind scale (WS) which is the snow-melt sensitivity to the wind speed or wind turbulence driving heat fluxes. Details of the SDC based snow routine can be found from Kolberg and Gottschalk (2006; 2010).

MODEL CALIBRATION

Different researchers (Harlin, 1991; Lindstrom, 1997; Seibert, 2000; Hundecha and Bárdossy, 2004; Seibert, 1997a; Uhlenbrook, 1999; Seibert, 2003; Das et al, 2008; Bárdossy and Singh, 2008; Lawrence et al., 2009; Shrestha et al., 2009; Sorman et al., 2009; Tillart, 2010; Driessen, 2010; Abebe et al., 2010) applied various algorithms for parameter calibration and identifiability and uncertainty analyses for the HBV model.

In the present study, we used the DREAM algorithm (Vrugt et al., 2009) implemented in Enki hydrological modelling platform (Kolberg and Bruland, 2012) for model calibration and assessment of parameter identifiability and uncertainty. To our knowledge, calibration of the HBV based precipitation-runoff model by the DREAM algorithm had not been pursued so far.

In DREAM, multiple chains with different starting points in the parameter space run simultaneously for global exploration, and automatically tune the scale and orientation of the proposal distribution during the evolution to the posterior distribution (Vrugt and Ter Braak, 2011). The calibration was performed based on a residual based log-likelihood (L-L) objective function:

$$L-L\left(\beta/\sigma_{\varepsilon}^{2},\sum_{i=1}^{n}\left(Qsim_{i}^{(\theta)}-Qobs_{i}^{(\theta)}\right)^{2}\right)=\left(\frac{-n}{2}\log\left(2\pi\right)-\frac{n}{2}\log\left(\sigma_{\varepsilon}^{2}\right)-\frac{\sum_{i=1}^{n}\left(Qsim_{i}^{(\theta)}-Qobs_{i}^{(\theta)}\right)^{2}}{2\sigma_{\varepsilon}^{2}}\right)\times f \quad (2)$$

, where $Qsim^{(\theta)}$ and $Qobs^{(\theta)}$ respectively are the Box-Cox (Box and Cox, 1964) transformed observed and simulated streamflow time series (of length n), β represents model parameters, θ is the Box-Cox transformation parameter, σ_{ε}^{2} is variance of error and f is a fraction of effectively independent observations. The logarithm form was used for simplicity and numerical stability and the function should adequately summarize the statistical properties of the residuals (Vrugt et al., 2013).

The Box-Cox transformation was performed in order to obtain an approximately normal distributed series with homoscedastic residuals (i.e. variance of residuals is independent of streamflow). Homoscedasticity in the residuals provides an advantage that the model residuals can be represented by one single distribution most often Gaussian (Willems et al., 2009). If θ = 0.0 (log-transformation), it corresponds to an assumption of lognormal distributed streamflow and it gives high weights to low flows like the Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) for a log-transformed data (NSELn). If θ = 1.0 (no transformation), it is assumed that the streamflow series is Gaussian and weight of high flows will be much greater than low flows like the Nash-Sutcliffe efficiency or NSE. The main advantages of the DREAM algorithm is that it accepts better parameter proposals (i.e. higher likelihood) and converges to posterior distributions rather than a single optimal parameter set and allow an objective assessment for parameter identifiability and predictive uncertainty. We used the last 50 % of the DREAM accepted marginal posterior parameters after the burn-in iterations (Vrugt et al., 2009) for the evaluation of parameter identifiability, correlation and presentation of minimum and maximum ranges of posterior parameters. But, for a 'fit for purpose' evaluation of the routines in terms of their maximum performance in simulating the hydrographs, we picked the optimal parameter sets which corresponds to the maximum values of the NSE and NSELn performance measures among the whole accepted by the DREAM algorithm. Burn-in iteration refers to discarding an initial portion of the samples to minimize the effect of initial conditions.

Values of θ from 0.25 to 0.3 are commonly used in literature (e.g. Vrugt *et al*., 2002 and references therein; Willens et al., 2009). But, in the present study θ values of -0.14, -0.1, 0.05 and -0.35 respectively for Gaulfoss, Eggafoss, Hugdal bru and Lillebudal bru were estimated/optimized from the hourly observed streamflow series based on the ‘fminsearch’ optimization algorithm in matlab software. The algorithm calls for finding the θ value that maximizes a log-likelihood function (http://www.mathworks.com). Box and Cox (1964) also proposed a maximum likelihood method for estimation of θ that satisfy a normal distributed and homoscedastic transformed series. Details on how the ‘fminsearch’ optimization algorithm works can be found from http://www.mathworks.com. Optimizing the Box-Cox transformation provides as close as normal distributed series and homoscedastic residuals, but when the objective of simulation focuses on prediction of high flows, it may be preferable to use higher values of θ (for instance, $\theta = 0.3$ is common in literature). We used the same transformation for all the compared runoff response routines and this specific issue related to selection of θ was not our focus in the present study.

The fraction of effectively independent observations, *f* was introduced to address problems related to correlation of residuals. The amount of information obtained from the data is much less than the nominal number of observation suggests due to a serial correlation (independence) in the hourly streamflow series. Setting the f value to 1.0 implies the observations are considered independent. We computed the fraction of effectively independent observations from the autoregressive or AR (1) model of error covariance (Zięba, 2010). Further details of the DREAM algorithm can be found from Vrugt et al. (2009).

POST CALIBRATION ANALYSES (COMPARATIVE EVALUATION)

Model validation aims to validate the model's robustness and ability to describe the catchment's hydrological response, and further detect any biases in the calibrated parameters (Gupta et al., 2005). A set of calibrated model parameters are expected to provide reasonable performance when transferred in time to a separate data set for the same catchment. We used the split sample test (Klemes, 1986) for temporal validation of the routines and the proxy basin test for spatial transfer of parameters to 'proxy ungauged' catchments (for spatial validation). We also transferred the parameters in both space and time for spatio-temporal validation of the routines.

Transferability of model parameters in space and consistency of model performance for multi-sites for the internal subcatchments builds our confidence of the predictive performance of the model. The main objective of transferring parameters to other catchments within the watershed in the present study is for spatio-temporal validation of the model and to test the possibility of simulating at interior locations based on parameter set calibrated for the streamflow at the catchment outlet that was one of the research question by the DMIP (Smith et al., 2004). We comparatively evaluated the routines based on transferability of their respective calibrated parameter set to 'proxy ungauged' internal or neighboring catchments with in the study watershed. Transfer of calibrated parameters to another watershed in the region for prediction in ungauged basins (PUB) was not an objective in the present study due to limited number of catchments. In addition, study on the effects of landuse and climate change was not an objective due to limited length of data. The NSE and NSELn performance measures were used for evaluations as discussed earlier. We used the parameter sets among the whole calibration that correspond to maximum values of the NSE and NSELn for comparisons of observed versus simulated hydrographs and to test temporal and spatial transferability of parameters with in the watershed.

Different frameworks for assessment and quantification of parameter identifiability and uncertainty are available in literature (see Uhlenbrook et al., 1999; Wagener et al., 2006; Pechlivanidis et al., 2011; Vrugt and Ter Braak, 2011). Uhlenbrook et al. (1999) used plots of objective function versus the parameters and Vrugt and Ter Braak (2011) used plots of posterior density of the parameters to assess parameter identifiability and uncertainty. In this work, plots of posterior parameters versus corresponding log-likelihood (L-L) objective function and linear correlation coefficient matrix of the posterior parameters (Schoups and Vrugt, 2010; Moreda et al., 2006) were presented to express parameter identifiability and uncertainty. The DREAM calibration algorithm converges as the Gelman-Rubin convergence (Gelman and Rubin, 1992) comes below 1.2. Therefore, it is not prone to subjective fixing of the number of simulations, which is one of the limitations of the pure MonteCarlo calibration methods. Among the whole DREAM generated and evaluated parameter sets, we filtered the parameter vectors that are accepted by the DREAM algorithm. As discussed earlier, we used the last 50 % of the DREAM accepted marginal posteriors after the burn-in iterations for parameter identifiability and uncertainty.

RESULTS

Observed versus simulated streamflow hydrographs for Gaulfoss from parameter set corresponding to the maximum NSE are presented in Fig. 3a and Fig. 3b. Simulated baseflow from parameter set corresponding to the maximum NSE and NSELn performance measures are presented in Fig. 3b and Fig. 3c respectively. The simulated base flow index (BFI) which is the ratio of simulated baseflow to the simulated streamflow are given in Table 4 for Gaulfoss and Eggafoss. For spatial and temporal validation of the model through transfer of calibrated parameters in space (among internal and neighboring catchments inside the Gaulfoss watershed) and also in time, maximum NSE/NSELn values for the calibrated catchments corresponding to the calibrated (optimal) parameter sets and the NSE/NSELn values obtained from transfer of the calibrated parameter sets to 'proxy ungauged' catchments are given in Table 5.

The observed and simulated streamflow hydrographs showed satisfactory agreements (Fig. 3a and b for Gaulfoss) for three of the four catchments. For Gaulfoss, Eggafoss and Hugdal bru, the NSE/NSELn performance measures corresponding to the calibrated parameters range from 0.68 to 0.87, the NSE/NSELn for spatially and spatio-temporally transferred parameters range from 0.65 to 0.84 and from 0.47 to 0.90 respectively (Table 5). The NSE values for the calibration period for Eggafoss could be raised to 0.70 (only by an order of 0.02) by choosing a higher Box-Cox transformation parameter (i.e. $\theta = 0.3$ which is common in literature) rather than optimizing it. Low NSE performances for Eggafoss for the split sample tests may be attributed to the effects of unrepresentativeness or low quality hydro-climatic data for the validation period.

For further comparative evaluations, reliability of predictions was presented for Gaulfoss in Fig. 4a to Fig. 4d in terms of quantile-quantile (QQ) plots to test the consistency of the predictive distribution and the observed data (Kavetski and Fenicia, 2011). We presented the QQ plots in terms of the empirical cumulative distribution functions (CDF) or probability of non-exceedance of the observed and simulated streamflow, and departures of the plots from the theoretical uniform distribution (i.e. the 1:1 diagonal line) indicate the discrepancy between the predictive distribution and the observed data (Fig. 4a to Fig. 4d). In addition, we further evaluated the routines for their performance in simulating the temporal variability of streamflow in terms of flow duration curves, FDCs (Fig. 5).

Table 5 shows the results of split-sample tests (temporal validation), 'proxy ungauged' basin test (spatial-validation) and spatio-temporal transfer of parameters among the four catchments in the Gaulfoss watershed. The DREAM calibrated parameter sets corresponding to the maximum NSE for the Gaulfoss catchment were validated for temporal, spatial and spatio-temporal transferability. Spatial validation and spatio-temporal validation for Eggafossen and Hugdal bru (when parameters are transferred to Gaulfoss catchment) also performed well though temporal validation for Eggafossen resulted in less performances. Good quality streamflow data was available only from 2010 to 2011 for Hugdal bru catchment and hence we did not perform temporal validation for the catchment. But, calibration and temporal validation for the smallest catchment (Lillebudal bru) resulted in low performance measures. The spatio-temporal transfer of calibrated parameters for Lillebudal bru provided better results for the NSE performance measures though it gave very poor performances for the NSELn (low flows). Lillebudal bru catchment is dominated by a mountainous topography (Fig. 1) where large portions of the catchment are at elevations above the climate stations from which data were used in the present study. In addition, the precipitation stations are far from Lillebudal catchment and hence their representativeness for spatially interpolated areal precipitation fields may be low. We also tested the effects of the quality of streamflow data for Lillebudal bru by calibrating the routines for the validation data (2010-2011) but the result showed no improvement in the calibration performances. Nevertheless, contingent on the quality and representativeness of input climate data, calibration of the conceptual HBV runoff response routines provided predictions which were validated by transferability of parameter sets in space and time to the internal (interior) and nearby 'proxy ungauged' catchments.

We presented the results of the DREAM log-likelihood calibration in terms of plots of posterior parameters versus the log-likelihood objective function in Fig. 6a-d for Gaulfoss to evaluate parameter identifiability from the last 50 % of the DREAM accepted parameter vectors after the burn-in iterations until the DREAM converges. The numbers of simulations (iterations) plotted are 3223 (Fig. 6a), 3411 (Fig. 6b), 2857 (Fig. 6c) and 914 (Fig. 6d) for the HBV-SMHI, HBV-Nonlinear, HBV-Soil Parsim R and HBV-Parsim respectively. Ranges of the prior and posterior parameters, and optimal parameters corresponding to maximum NSE and NSELn were given in Table 6. The linear correlation matrix among the posterior parameters in the soil-moisture accounting and runoff response routines for Eggafoss catchment was presented in Table 7.

DISCUSSION

A comparative evaluation of parsimonious versus more complex configurations of distributed conceptual HBV response routines were performed in terms of prediction of different runoff 'signatures' (hydrographs, flow duration curves and baseflow), spatial and temporal validation of the routines and parameter identifiability and uncertainty. The results indicated that there are only marginal differences among the total streamflow predictions by the parsimonious and more complex routines. It comply with the advantages of the 'principle of parsimony' which was illustrated by Jakeman and Hornberger (1993). However, by testing the internal simulation of the routines in terms of the baseflow, we could observe differences in performances among the routines in simulation of the baseflow contributions to the streamflow.

Simulated hydrographs, baseflow, Q-Q plots and FDCs

Plots of simulated streamflow hydrographs corresponding to the maximum NSE performance measure (Fig. 3a and Fig. 3b) versus the observed streamflow hydrograph indicated only marginal differences in the performances of the routines in terms of reproducing the time series of streamflow. All the four cases of conceptualization of the routines gave low simulated streamflow (Qsim) compared to the observed streamflow (Qobs) during summer high flow events. The soil moisture routine parameters particularly influence the discharge volume and the fast flow routine parameters particularly affect the shape of the hydrograph and extreme discharges (Booij, 2005). Abebe et al. (2010) noted the field capacity (FC) as a dominant parameter affecting both high flow series and runoff volume. In the soil moisture accounting routine, the field capacity (FC) and the evapotranspiration threshold parameter (LP) control the amount of evapotranspiration while β controls partitioning of infiltration (I) in to recharge to upper reservoir (R) and change in soil moisture (Δ_{SM}). At any relative saturation (SM/FC) of the soil moisture zone a larger proportion of the infiltrated water will replenish the soil moisture zone as β increases. However, the nonlinear effect of varying soil saturation become more pronounced for larger β and there is a rapid increase in the amount of infiltrated water that recharges the upper zone reservoir as relative saturation increases. But, the HBV-SMHI routine which has three free parameters in the soil moisture accounting routine (FC, LP and β) and also the very quick recession coefficient (k_2) in the response routine shows no marked better performance for simulation of high flows during summers. Though it does not contain the very quick runoff component, the simulated peak flow from HBV-Parsim was indistinguishable from those of the other routines. These findings may indicate compensation effects among model parameters, the crucial importance of the quality of input precipitation data to improve simulation of peak events and the potential for equivalent performances of the parsimonious routines.

The main reasons for lower simulated peak flow during summer periods may be attributed to less representativeness of the input precipitation data, particularly in capturing intense localized rainfall events. In addition, low quality streamflow data especially during flood events may contribute to the problem. The effect of Box-Cox transformation for simulation of high flows was checked for Eggafoss catchment but only an increase of NSE values by 0.02 were obtained for $\theta = 0.3$. The conceptual nature of the model structures could also be another explanation. Uncertainty due to parameter estimation and model structure alone cannot address the uncertainty in the prediction and hence assessments of uncertainties in the input data need to be included to improve decision making under uncertainty.

Also, prediction based on the NSELn performance measure that gives higher weightage to the low flows are indistinguishable for the three configurations (Table 5). However, in terms of the quantity of simulated baseflow from the lower (ground water) reservoir, the HBV-Parsim provided considerably high contribution of baseflow (Fig. 3c and Fig. 3d) based on both performance measures compared to the HBV-SMHI, HBV-Nonlinear and HBV-Soil Parsim R. Simulated BFI corresponding to the maximum NSE/NSELn performance measures for the HBV-Parsim are 0.62/0.55 and 0.70/0.52 for Gaulfoss and Eggafoss respectively (Table 4). However, BFI values of greater than 0.8 were estimated from observed streamflow based on a Web based Hydrography Analysis Tool (WHAT) baseflow separation of the United States Geological Survey (Lim et al., 2005) but the validity of the filtering equations and parameters in 'WHAT' for the boreal catchments needs further study. Also, previous studies such as Beldring et al. (2000 and references therein) reported significant groundwater contribution to the total streamflow for boreal/humid-temperate catchments. Therefore, the HBV-Parsim presumably provided reliable simulation of baseflow but tailor-made study is required on suitable baseflow separation techniques (e.g. Willems, 2009) for the boreal catchments, which was not an objective in the present study.

The HBV-Parsim and HBV-Soil Parsim R have similar linear storage-outflow and storage-baseflow model structure in their response routines but differs in their soil-moisture accounting routines i.e. the latter is based on soil-moisture accounting routine of Bergström (1976). Therefore, the differences in the baseflow simulation results of the HBV-Parsim are most likely related to its parsimonious parameterization of the soil-moisture accounting routine. Abebe et al. (2010) found the strong effects of the soil moisture accounting parameters on the runoff volume and the FC as a dominant parameter affecting both high flow series and runoff volume. The present study showed a strong negative correlation between the FC and percolation parameter, PERC (Table 7). The marked negative correlation between FC and PERC is due to a decrease in the recharge to the upper reservoir as FC increases or relative saturation decreases. Reduced recharge results in reduced percolation rate. This reduction in percolation rate in turn potentially decreases the baseflow contribution, which was observed for the three routines with the FC as a free calibration parameter. Samuel et al. (2012) based on the soil-moisture accounting routine of Bergström (1976) found that the configuration which is similar to the HBV-SMHI but with non-linear storage-discharge relationships in the lower reservoir outperformed for baseflow simulation than configurations which are the same as the HBV-SMHI and HBV-Nonlinear. The authors used the Thornthwaite method for computation of potential evapotranspiration and the recursive digital filter for separation of baseflow from observed streamflow series (for comparison against the simulated baseflow) based on the assumption of filtering out high-frequency signals (quick flow) to separate low-frequency baseflow (Nathan and McMahon, 1990).

The conceptual very quick flow from the upper outlet of the upper reservoir (Q_{UZ2}) for the HBV-SMHI corresponding to the optimal parameters for NSE and NSELn for Gaulfoss and Eggafoss range from 0.30 % to 17 % of the total simulated streamflow which showed slight exceedance of the threshold in the upper reservoir and small proportion of the very quick flow contributing to the total streamflow. Fig. 6 shows less identifiability of the threshold parameter (UZt) which may be related to its low sensitivity and hence its influence on the other parameters is not clearly observed.

The QQ plots for evaluating the reliability of prediction indicated significant over and under predictions from low to high ranges of streamflow that are indistinguishable among the three runoff response routines (Fig. 4a to d). There are also similar marginally different performances in prediction of flow duration curves (FDCs) which are mainly characterized by underestimation of high flows for all of the routines (Fig. 5). Therefore, further evaluations based on additional runoff 'signature' and reliability criteria also do not suggest any clear superiority of the more complex routines.

Model validation (temporal, spatial and spatio-temporal)

Testing if the distributed models that are calibrated with basin outlet streamflow information provide meaningful hydrologic simulation at internal catchments was one of the science question tested by the DMIP (Smith et al., 2004). In the present study, marginal better performances from transfer of parameters from Gaulfoss to its internal catchments Eggafoss and Hugdal bru were observed for the HBV-Nonlinear and HBV-Parsim routines (NSE from 0.71 to 0.72 and NSELn from 0.76 to 0.84 in Table 5). Therefore, calibration results of the distributed routines indicated possibility for hourly prediction also at ungauged interior locations within the watershed through transfer of parameters. For some of the routines, equivalent or slight improvements in the NSE for the transferred parameter set from the watershed outlet (Gaulfoss) than the explicit local calibration at the interior locations of Eggafoss and Hugdal bru were occurred probably due to the differences in the representativeness of climate stations and hence the accuracy of areal precipitation, and the quality of streamflow data used for the model calibration. Both spatial transfer of parameter sets and explicit calibration performed poorly for the smallest catchment of Lillebudal bru especially for NSELn (gives weightage for simulation of low flow). Less performance in low flow may also indicate less performance in simulation of baseflow that may also suggest problems with stage-discharge curve during low flows at Lillebudal and subsurface peculiarities of Lillebudal catchment in addition to the less representative precipitation input.

Parameter identifiability and uncertainty

Fig. 6b, Fig. 6c and Fig. 6d show maximum log-likelihood (L-L) objective function for narrow ranges of parameters of the HBV-Nonlinear, HBV-Soil Parsim R and HBV-Parsim response routines. There is better likelihood of obtaining narrow predictive uncertainty of streamflow from posterior parameters yielding narrow maximum objective function. However, for the HBV-SMHI with ten free parameters the posterior ranges of parameters in the threshold based non-linear upper reservoir k_2 and UZ_t (Fig. 6a) are not much narrower than their corresponding ranges of uniform prior (Table 6) which indicate their poor identifiability. Uhlenbrook et al. (1999) attributed the non-identifiability of parameters to uncertainty in model parameters or lack of sensitivity of model output to the change in parameters. Therefore, prediction uncertainty of a response routine with insensitive and unidentifiable parameters is expected to be high.

Narrower ranges of posterior parameters compared to the prior ranges for the HBV-Soil Parsim R and the HBV-Parsim compared to the others indicated less parameter uncertainty for the parsimonious runoff response routines (Table 6). For instance, for the HBV-Soil Parsim R routine for maximum NSE for the Gaulfoss catchment, the posteriors ranges of k_1 (0.11-0.20), k_0 (0.01-0.03) and PERC (0.65-3.37) were obtained. Also, for the HBV-Parsim routine for maximum NSE of the Gaulfoss catchment, the posterior ranges of k_1 (0.12-0.25), k_0 (0.03-0.04) and PERC (2.36-5.06) were obtained. The uniform prior ranges (Table 3) were k_1 (0.001-1.5), k_0 (0.0005-0.5) and PERC (0-6).

Though the presupposed uniform priors overlap for the recession coefficients k_1 and k_0, the ranges of their DREAM calibrated posteriors do not overlap for the HBV-Soil Parsim R and HBV-Parsim and do not overlap significantly for the other routines. This indicated that the

DREAM calibration was able to constrain the model parameters to satisfy the conceptual representation of the quick runoff that is related to k_1 and the slow base flow that is related to k_0. From the pure Monte Carlo method (MC), Seibert (1997a) and Uhlenbrook et al. (1999) found an overlap in k_1 and k_0 for some MC generated parameter sets, which contradicts with the model conceptualization. Even if there is a slight overlap between calibrated posterior ranges of k_2 and k_1 for the HBV-SMHI and k_1 and k_0 for the HBV-Nonlinear, the parameter set corresponding to the maximum NSE and maximum NSELn comply with the conceptualization of very quick runoff corresponding to k_2, quick runoff corresponding to k_1 and slow baseflow corresponding to k_0 or $k_2 > k_1 > k_0$ (Table 6).

The strong negative correlations between the FC and the PERC parameters (Table 7) shows the marked influence of parameterizations of the soil-moisture accounting on the runoff response parameters. As the FC increases, the relative saturation (SM/FC) decreases and hence the recharge decreases (based on the equation for the recharge in Table 2) which in turn reduces the percolation rate (related to the PERC parameter). For the HBV-Nonlinear, the negative correlations are more pronounced among the recession (drainage) coefficients and the non-linearity exponents of the outflow and baseflow equations showing existence of strong compensation or interaction among the recession coefficients and the non-linearity parameters. But, the results indicated positive correlation between the FC and the upper reservoir non-linearity coefficient. Generally, positive correlations were exhibited among the recession coefficient parameters and the PERC except for the HBV-SMHI, which showed very slight negative correlation between k_0 and PERC. Increase in the k_1 compensates for reduced storage in the upper reservoir as the PERC increases for simulation of outflow from the upper reservoir. Increase in PERC to increase the storage in the lower reservoir accompanied by an increase in k_0 favors generation of baseflow according to the baseflow equation.

CONCLUSIONS

We performed comparative evaluation of the four different configurations of the HBV runoff response routines (HBV-SMHI, HBV-Nonlinear, HBV-Soil Parsim R and HBV-Parsim). The results for the boreal catchments in mid Norway indicated that the parsimonious conceptualization (HBV-Parsim) with only five free parameters could extract the information content of the data efficiently with improved parameter identifiability. Also, better baseflow simulation were observed which complies with the high groundwater contribution to the streamflow for the study catchments as demonstrated in previous studies (Beldring et al., 2000) and as guided by the baseflow separation techniques in the present study.

The findings of the study revealed the importance of comprehensive evaluation of parametrical parsimonious (i.e. small number of free parameters) simple linear storage-discharge relationships and the more complex HBV based response routines. Further parsimony without compromising the quality of streamflow prediction could also be possible for snow free or snow non-dominated catchments by excluding the snow routine. However, the more complex routines with large number of reservoirs and parameters and linear and non-linear conceptualizations like the HBV-SMHI and HBV-Nonlinear may offer more flexibility for scientific research purposes (e.g. process understanding). However, increases in the numbers of parameters increases degrees of freedom and hence parameter unidentifiability problems, which in turn has, negative implications for predictive uncertainty for operational purposes.

Usually only streamflow observation is readily available for parameter calibration, which is also usually the case for prediction for water management purposes. However, evaluations based on only the total streamflow is highly affected by the compensation and interacting effects among the storages and fluxes (outflows) of the upper and lower reservoirs. Therefore, the validity of the results of the conceptual HBV based routines need to be evaluated against their conceptualization. Reliable prediction by the conceptual HBV models calls for evaluation of the internal simulation of the routines such as baseflow and possibly the soil-moisture state. For instance, the conceptualized simulated baseflow from the lower (ground water) reservoir should be compared against the baseflow estimated from other methods e.g. baseflow separation techniques. To this end, development of proper baseflow separation techniques for the catchments would be necessary to evaluate the baseflow simulation of the HBV based routines.

The findings should provide new information for the HBV users and the hydrological modelling community regarding the performances of different configurations of the conceptual HBV model. Equivalent performances of the HBV-Parsim indicated a potential for application of parametrical parsimonious models, which would benefit model updating for forecasting purposes. Further evaluation of the routines based on larger number of catchments (e.g. regional modelling) and for different climate regimes (e.g. Lidén and Harlin, 2000), which requires availability of data and separate study, would be expected to provide further insights.

ACKNOWLEDGMENTS

The research was supported by the Center for Environmental Design of Renewable Energy or CEDREN (http://www.cedren.no). We obtained the climate data from the Norwegian Meteorological Institute (http://met.no/Klima), Statkraft, TrønderEnergi and bioforsk (http://www.bioforsk.no). We obtained the streamflow data from the Norwegian Water Resources and Energy Directorate (http://www.nve.no).

REFERENCES

Abebe, N. A., Ogden, F., L. & Pradhan, N.R. 2010 Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: Implications for parameter estimation. *J. Hydrol.* 389, 301-310.

Bastola, S. & Murphy, C. 2013 Sensitivity of the performance of a conceptual rainfall-runoff model to the temporal sampling of calibration data. *Hydrology Research* 44 (3), 484-494.

Beldring, S., Gottschalk, L., Rodhe, A. & Tallaksen, L.M. 2000 Kinematic wave approximations to hillslope hydrological processes in tills. *Hydrol. Process.* 14, 727-745.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R. & Vøkso, A. 2003 Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrol. Earth. Syst. Sc.* 7(3) 304-316.

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

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

Beven, K. J. & Freer, J. 2001 Equifinality, data assimilation, and un-certainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.* 249, 11-29.

Blöschl, G., Reszler, C. & Komma, J. 2008 A spatially distributed flash flood forecasting model. *Environmental Modelling & Software* 23, 464-478.

Box, G. E. P. & Cox, D. R. 1964 An analysis of transformations. *Journal of the Royal Statistical Society,* Series B 26, 211-252.

Booij, M.J. 2005 Impact of climate change on river flooding assessed with different spatial model resolutions. *J. Hydrol.* 303, 176-198.

Bárdossy, A. & Singh, S. K. 2008 Robust estimation of hydrological model parameters Hydrol. *Earth Syst. Sci.* 12, 1273-1283.

Clark, M. P. & Kavetski, D. 2010 The ancient numerical demons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res*. 46, W10511, doi:10.1029/2009WR008894.

Das, T., Bárdossy, A., Zehe, E. & He, Yi. 2008 Comparison of conceptual model performance using different representations of spatial variability. *J. Hydrol.* 356, 106-118.

Driessen, T. L. A., Hurkmans, R. T. W. L. Terink, W. Hazenberg, P., Torfs, P. J. J. F. & Uijlenhoet, R. 2010 The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model. *Hydrol. Earth Syst. Sci.* 14, 651-665.

Fenicia, F., Kavetski, D. & Savenije, H.H.G. 2011 Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resour. Res.* DOI:10.1029/2011WR010748.

Gardelin, M. & Lindstrom, G. 1997 Priestley-Taylor Evapotranspiration in HBV-Simulations. *Nord. Hydrol.* 28 (4-5), 233-246.

Gelman, A. & Rubin, D. B. 1992 Inference from iterative simulation using multiple sequences. *Stat. Sci. 7*, 457-472.

Gupta, H.V., Beven K. & Wagener T. 2005 Model calibration and uncertainty estimation. *Encyclopedia of Hydrological Sciences*, 11(131), 1-17.

Harlin, J. 1991 Development of a Process Oriented Calibration Scheme for the HBV Hydrological Model. *Nord. Hydrol.* 22 (1), 15–36.

Harlin, J. 1992 Modelling the Hydrological Response of Extreme Floods in Sweden. *Nord. Hydrol.* 23, 227-244.

Hundecha, Y. & Bárdossy, A. 2004 Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model. *J. Hydrol.* 292, 281–295.

Jakeman, A. J. & Hornberger, G. M. 1993 How Much Complexity Is Warranted in a Rainfall-Runoff Model?. *Water Resour. Res.* 29 (8), 2637-2649.

Kavetski, D. & Fenicia, F. 2011 Elements of a flexible approach for conceptual hydrological modeling: 2. Application and experimental insights. *Water Resour. Res.* DOI: 10.1029/2011WR010748.

Kavetski, D., Fenicia, F. & Clark, M. P. 2011 Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment. *Water Resour. Res*. 47, W05501.

Killingtveit, A. & Sælthun, N.R. 1995 Hydropower development: Hydrology. *Technical Report*, Norwegian Institute of Technology, Oslo.

Kirchner, J.W. 2006 Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* 42, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W. 2009 Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res.* 45, W02429, doi:10.1029/2008WR006912.

Klemes, V. 1986 Operational testing of hydrological simulation models. *Hydrolog. Sci. J.* 31, 13-24.

Kobold, M. & Brilly, M. 2006 The use of HBV model for flash flood forecasting. *Nat. Hazards Earth Syst. Sci.* 6, 407-417.

Kokkonen, T. S. & Jakeman, A. J. 2001 A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. *Water Resour. Res.* 37 (9), 2345-2352.

Kolberg, S. A. & Bruland, O. 2012 ENKI - An Open Source environmental modelling platform. *Geophysical Research Abstracts* 14, EGU2012-13630,EGU General Assembly.

Kolberg, S. A. & Gottschalk, L. 2006 Updating of snow depletion curve with remote sensing data. *Hydrol. Process.* 20(11), 2363–2380.

Kolberg, S. A., Gottschalk, L. 2010 Interannual stability of grid cell snow depletion curves as estimated from MODIS images. Water Resour. Res., 46, W11555, doi:10.1029/2008WR007617.

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

Lawrence, D., Haddeland, I. & Langsholt, E. 2009 Calibration of HBV hydrological models using PEST parameter estimation. *Norwegian Water Resources and Energy Directorate Report no. 1-2009*.

Li, H., Beldring, S. & Xu, C-Y. 2014 Implementation and testing of routing algorithms in the distributed HBV model for mountainous catchments. *Hydrology Research* 45(3), 322-333.

Lidén, R. & Harlin, J. 2000 Analysis of conceptual rainfall-runoff modelling performance in different climates. *J. Hydrol.* 238, 231-247.

Lim, K.J., Engel, B.A., Tang, Z., Choi, J., Kim, K., Muthukrishnan, S. & Tripathy, D. 2005 Web GIS-based Hydrograph Analysis Tool, WHAT. *Journal of the American water resources association* 41(6), 1407-1416.

Lindstrom, G. 1997 A Simple Automatic Calibration Routine for the HBV Model. *Nord. Hydrol*. 28 (3), 153-168.

Littlewood, I. G. & Croke, B. F. W. 2008 Data time-step dependency of conceptual rainfall-streamflow model parameters: An empirical study with implications for regionalization, *Hydrol. Sci. J.* 53(4), 685-695.

Merz, R., J. Parajka & Blöschl, G. 2009 Scale effects in conceptual hydrological modeling, *Water Resour. Res*. 45, W09405, doi :10.1029/2009WR007872.

Michaud, J., Sorooshian, S. 1994 Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. *Water Resour. Res.* 30, 593-605.

Moore, R.J. 1985 The probability-distributed principle and runoff production at point and basin scales. *Hydrolog. Sci. J*. 30 (2), 273-297.

Moreda, F., Koren, V., Zhang, Z., Reed, S. & Smith, M. 2006 Parameterization of distributed hydrological models: learning from the experiences of lumped modeling. *J. Hydrol.* 320, 218-237.

Nash, J. E., Sutcliffe, J. V. 1970 River flow forecasting through conceptual models, I. A discussion of principles. *J. Hydrol.* 10, 228-290.

Pechlivanidis, I.G., Jackson, B.M., Mcintyre, N.R. & Wheater, H.S. 2011 Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global NEST Journal* 13 (3), 193-214.

Perrin, C., Michel, C. & Andréassian, V. 2001 Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* 242, 275-301.

Priestley, C.H.B. & Taylor, R.J. 1972 On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* 100, 81-82.

Rakovec, A., Weerts, H., Hazenberg, P., Torfs, P. J. J. F. & Uijlenhoet, R. 2012 State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy. *Hydrol. Earth Syst. Sci*. 16, 3435-3449.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F. & Seo, D.-J. 2004 DMIP Participants. Overall distributed model intercomparison project results. *J. Hydrol.* 298, 27-60.

Samuel, J., Coulibaly, P. & Metcalfe, R. A. 2011 Estimation of continuous streamflow in Ontario ungauged basins: Comparison of regionalization methods. *J. Hydrol. Eng*. 16(5), 447-459.

Schoups, G. & Vrugt, J.A. 2010 A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resour. Res.* 46, W10531, doi:10.1029/2009WR008933.

Seibert, J. 1997 Estimation of parameter Uncertainty in the HBV Model. *Nord. Hydrol.* 28 (4-5), 247-262.

Seibert, J. 1997 HBV light version 1.2, *User's manual*, Uppsala University, Dept. of Earth Science, Hydrology, Uppsala.

Seibert, J. 2000 Multi-criteria calibration of a conceptual runoff model using a genetic algorithm. *Hydrol. Earth Syst. Sci.* 4 (2), 215-224.

Seibert, J. 2002 HBV light version 2, *User's manual*, Environmental Assessment, SLU, Uppsala.

Seibert, J. 2003 Reliability of Model Predictions Outside Calibration Conditions. *Nord. Hydrol.* 34(5), 477-492.

Seibert, J., Uhlenbrook, S. Leibundgut, C. & Halldin, S., 2000 Multiscale Calibration and Validation of a Conceptual Rainfall-Ruoff Medel. *Phys. Chem .Earth (B)* 25 (1), 59-64.

Shrestha, D.L. & Solomantine, D.P. 2008 Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. *Intl. J. River Basin Management* 6 (2), 109-122.

Shrestha, D. L., Kayastha, N. & Solomatine, D. P. 2009 A novel approach to parameter uncertainty analysis of hydrological models using neural networks. *Hydrol. Earth Syst. Sci.*13, 1235-1248.

Sicart, J. E., Pomeroy, J. W., Essery, R. L. H. & Bewley, D. 2006 Incoming longwave radiation to melting snow: observations, sensitivity and estimation in northern environments. *Hydrol. Process.* 20, 3697-3708.

Smith, M.B. Seo, D.-J., Koren, V. I. Reed, S.M., Zhang, Z., Duan, Q., Moreda, F. & Cong, S. 2004 The distributed model intercomparison project (DMIP): motivation and experiment design. *J. Hydrol*. 298, 4-26.

Sorman, A. A., Sensoy, A., Tekeli, A. E., Sorman, A. & Akyurek, U, Z. 2009 Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey *Hydrol. Process.* 23, 1031-1040.

Sælthun, N.R. 1996 The "Nordic" HBV model. Description and documentation of the model version developed for the project climate change and energy production. *Publication No. 7*. Norwegian water resources and energy administration, Hydrology department, Oslo.

Teuling, J. Lehner, I., Kirchner, J. W. & Seneviratne, S. I. 2010 Catchments as simple dynamical systems: Experience from a Swiss prealpine catchment. *Water Resour. Res*. 46, W10502, doi:10.1029/2009WR008777.

Tillaart, S. P.M. 2010 Influence of uncertainties in discharge determination on the parameter estimation and performance of a HBV model in Meuse sub basins Enschede, April 2010 *Master Thesis*, Water Engineering & Management University of Twente, The Netherlands.

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(5), 779-797.

van Werkhoven, K., Wagener, T., Reed, P. M. &Tang, Y. 2009 Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Advances in Water Resources* 32 (8), 1154-1169.

Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M. & Higdon D. 2009 Accelerating Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. J*ournal of Nonlinear Sciences and Numerical Simulation* 10(3), 273-290.

Vrugt, J. A. & Ter Braak, C.J.F. 2011 DREAM(D): an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.* 15, 3701-3713.

Vrugt, J.A., Ter Braak, C.J.F., Diks, C.G.H. & Schoups, G. 2013 Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications. *Advances in Water resources* 51, 457-478.

Wagener, T., Freer, J., Zehe, E., Beven, K., Gupta, H. & Bárdossy, A. 2005 Towards an uncertainty framework for predictions in ungauged basins: The Uncertainty Working Group. In: Predictions in Ungauged Basins: Promise and Progress (Sivapalan, M., Wagner, T., Uhlenbrook, S., Liang, X., Lakshim, V., Kumar, P., Zehe, E. & Tachikawa, Y., eds), *IAHS Publication* 303, ISNN 0144-7815.

Wang, Y., B. He, & K. Takase 2009, Effects of temporal resolution on hydrological model parameters and its impact on prediction of river discharge, *Hydrol. Sci. J*. 54(5), 886-898.

Willems, P. 2009 A time series tool to support the multi-criteria performance evaluation of rainfall-runoff models. *Environmental modelling and software* 24, 311-321.

Winsemius, H. C., Schaefli, B., Montanari A. & Savenije, H. H. G. 2009 On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological Information. *Water Resour. Res*. 45, W12422, doi:10.1029/2009WR007706.

Wrede, S., Seibert, J. & Uhlenbrook, S. 2013 Distributed conceptual modelling in a Swedish lowland catchment: a multi-criteria model assessment. *Hydrol. Res.* 44 (2), 318-333.

Zięba, A. 2010 Effective number of observations and unbiased estimators of variance for autocorrelated data an overview. *Metrol. Meas. Syst.* XVII (1), 3-16.