Journal of Hydrology 527 (2015) 858–871

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

How much can we gain with increasing model complexity with the same model concepts?

Hong Li [a], C.-Y. Xu [a,*], Stein Beldring [b]

[a] Department of Geosciences, University of Oslo, Norway
[b] Norwegian Water Resources and Energy Directorate, Norway

ARTICLE INFO

Article history:
Received 17 January 2015
Received in revised form 20 May 2015
Accepted 23 May 2015
Available online 28 May 2015
This manuscript was handled by Konstantine P. Georgakakos, Editor-in-Chief, with the assistance of Yu Zhang, Associate Editor

Keywords:
Distributed models
HBV
Internal variables
Lumped model
Model comparison

SUMMARY

The main purpose of this study is to test the hypothesis that, with appropriate structures, increasing model complexity with the same model concepts would lead to an increase in the model efficiency in simulating either runoff or internal variables. Five variants of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model, i.e. a lumped model (LWhole), a semi-distributed model (SBand), a grid-model without routing (GRZero), a grid-model with hillslope routing (GROne), and a grid-model with both hillslope and channel routing (GRTwo) are compared in a cold and mountainous catchment in central southern Norway. The five models are compared with respects to (1) runoff simulation at the catchment outlet and the interior points, and (2) simulations of internal variables, i.e. evapotranspiration, snow water equivalent and groundwater depth. The results show that the models with higher complexity can improve the runoff simulation both at the catchment outlet and the interior points. However, there is no superiority of complex grid-models over simple grid-models in reproducing internal variables in this study.

1. Introduction

The fundamental work of hydrologists is to quantify relationship of precipitation over a catchment area and resulting runoff. In the last three decades, a large amount of hydrological models have been published to describe this relationship (Singh and Woolhiser, 2002). These models are suited to simulate runoff at certain spatial and temporal scales (Praskievicz and Chang, 2009). Basically, the differences among the models come from three aspects (Butts et al., 2004): (1) assumptions about factors influencing hydrology or response functions, (2) numerical solutions, and (3) spatial discretisation, such as sub-catchments or square grids.

There has been an evident trend to develop models with high degree of complexity (Mayr et al., 2013; Perrin et al., 2001). The development of hydrological models has gone through three stages, from input–output black-box models, through lumped conceptual models, to physically-based and/or conceptual distributed models. The physically-based distributed models tend to represent physical processes by partial differential equations at a fine spatial resolution. To develop models with high degree of physical dependence and structural complexity is long-sighted and useful in terms of knowledge gained about hydrological processes from catchment studies. However, it would lead to increasing difficulty in estimating parameters and large parameter uncertainty (Butts et al., 2004).

Hydrologists show high interest in comparing hydrological models of varying complexity. The conclusions are not consistent due to many reasons, for example, the models used in the comparisons, data quality and catchment characteristics, etc. However, if the models differ in the assumptions about factors influencing hydrology or the runoff process, results of comparisons are determined by the basins used.

To focus on the model itself, we assume that the responses of runoff to areal mean precipitation are described by the model concepts, which can be described by a lumped model. The model complexity exists in the way that water balance components are presented. For example, the lumped HBV model formulates all basic concepts of the HBV model. The semi-distributed models with elevation bands or sub-catchments and the grid-distributed models are of a higher degree of model complexity. The routing procedure, such as the Muskingum method also adds an additional degree of complexity, because it links the model elements to each other.

This research examines if increasing degree of model complexity improves the model performance and how much. A review of

* Corresponding author.
E-mail address: chongyu.xu@geo.uio.no (C.-Y. Xu).

http://dx.doi.org/10.1016/j.jhydrol.2015.05.044
0022-1694/

the scientific literature did not provide a clear guidance on this issue. In 2004, the Distributed Model Intercomparison Project (DMIP) compared distributed and lumped versions of twelve models with radar precipitation data at a 4 km spatial resolution and hourly temporal resolution (Smith et al., 2004). Twenty one events in eight catchments with areas ranging from 65 to 2484 km^2 were selected in the comparative studies. This project results showed that overall, lumped models outperformed distributed models in terms of discharge simulation, and some distributed models showed comparable results to lumped models in many basins (Reed et al., 2004). The results depended on the basin's shape, orientation, soil and climatic characteristics (Butts et al., 2004; Reed et al., 2004).

Different results are reported by other studies. Atkinson et al. (2003) demonstrated that the best simulation of runoff was obtained by a fully distributed model in a small catchment at an hourly time step. Han et al. (2014) and Yan and Zhang (2014) respectively examined effects of watershed subdivision on modelling runoff. They found that with a larger number of sub-catchments, the model gave a better performance and there was a threshold level, beyond which no significant improvements could be obtained with increasing number of sub-catchments. This confirms the results by comparing two watershed subdivision schemes by Varado et al. (2006). Michaud and Sorooshian (1994) compared the lumped and distributed Soil Conservation Service (CSC) models on a semi-arid catchment of 150 km^2 and the results showed that neither of the models could accurately simulate the peak flows or runoff volumes from 24 sever thunderstorms at an one-minute time step. Using the same model, Boyle et al. (2001) found improvements were related to spatial distribution of model input and streamflow routing and no additional benefits could be obtained when the number of watersheds were more than three.

A well-known way for a better model comparison is involving internal points and state variables in model evaluation (Lindström et al., 1997; Uhlenbrook et al., 1999; Vaché and McDonnell, 2006). Alley (1984) was probably the first to notice that models were similar in runoff simulation while substantial differences existed in other variables. Furthermore, Jiang et al. (2007) found that models differed least in discharge among simulations of discharge, actual evapotranspiration and soil moisture. Varado et al. (2006) applied a conceptual model on the Donga catchment (Benin), and found that the model was only good at simulating runoff, but not at reproducing groundwater table. Vaché and McDonnell (2006) showed that, among the models they evaluated, only the most complex model successfully reproduced both discharge dynamics and residence time. The results of their research indicate that the best objective function value in discharge are obtained during calibration, but with sacrifice of other hydrological variables, where the error resides. In order to achieve deliberate conclusions, comparisons including other hydrological measurements, such as evapotranspiration and groundwater should be adopted (Bookhagen and Burbank, 2010; Lindström et al., 1997; Wagener et al., 2001). This multi-variable comparison will not only contribute to an improved understanding of hydrology processes, but also provide guidance for developing hydrological models.

The aim of this paper is to test the hypothesis that, with appropriate way of organising, increasing complexity leads to an increase in model efficiency of simulating either runoff or other internal variables. A conceptual rainfall–runoff model, the HBV model is selected and modified to five model variants of different complexities, i.e. a lumped model (LWhole), a semi-distributed model with ten elevation bands (SBand), a simple grid-model (GRZero), a grid-model with hillslope routing (GROne), and a grid-model with both hillslope routing and river channel routing (GRTwo). The selection of the models is threefold. Firstly, LWhole is the most widely used HBV model variant for scientific and operational purposes globally. Secondly, SBand is currently used in the Norwegian Water Resources and Energy Directorate (NVE), which is responsible for flood forecasting and water resources administration in Norway. Thirdly, three distributed models are included in the sense of having a higher level of physical realism. The spatial variability and how the runoff routes to the catchment outlet are described in the three grid distributed models at various degrees of complexity.

Lindström et al. (1997) made the first attempt to make a distributed version of the HBV model. They used a typical resolution of 40 km^2 of sub-catchments and each sub-catchment was further divided into elevation bands. This modification significantly improved model performance. In the late 1990s, Uhlenbrook et al. (1999) and Krysanova et al. (1999) respectively compared the effects of spatial distribution on runoff simulation. Uhlenbrook et al. (1999) compared three model variants with different number of elevation bands and land use zone and various runoff generation conceptualisation, on a small mountainous catchment of 40 km^2 in south western Germany. They concluded that the models considering more spatial variability were better than the lumped models when separately computing of the upper zone storage for each model unit. Krysanova et al. (1999) applied the semi-distributed HBV model of elevation bands to a large German catchment of 96,000 km^2. The model with sub-catchments enabled better runoff simulation than without sub-catchment division. However, Das et al. (2008) compared four versions of the HBV model and found that semi-distributed and semi-lumped (a lumped model for each sub-catchment) outperformed the distributed (1 km regular grid) and lumped model structures. The authors suspected that the input data did not reflect the actual spatial variability. The study by Wrede et al. (2013) in a Swedish lowland catchment of 2000 km^2 using the distributed model (250 m regular grid) also showed that the quality of input data was a limitation factor for model performance. For modelling other variables, Mayr et al. (2013) reported that involving glacier mass balance in calibration gave a better prediction in the glacier mass but a slightly worsened discharge prediction. Moreover, they revealed that incorrect snow and ice simulations did not necessarily affect the quality of the runoff simulation.

2. Study area and data

2.1. Study area

The Norsfoss catchment is located in upstream of the longest Norwegian river, the Glomma River in central southern Norway (Fig. 1). This catchment covers an area of 18,932 km^2. The mean altitude is 732 m above the mean sea level (m amsl) ranging from 147 to 2170 m amsl, and approximately 26% of the area is above the potential tree level. The mean slope is 6.7° with a range from 0.0° to 73.2°. Climate varies along the river from upper mountain regions in north to lowlands in south. Additionally, the north-western part is characterised by lower temperature, lower precipitation and longer snow-cover period than the lowland area. Annual precipitation is 849 mm/year, and yearly mean air temperature is −0.62 °C based on the period from 1961 to 1990 with 10.68 °C in July and −11.48 °C in January. More than 60% of the catchment area is covered by forest and marsh, and approximately 20% is covered by bedrock. Floods are usually associated with snow melt, heavy rainfall or their combination (L'Abée-Lund et al., 2009).

2.2. Geography data and processing

Data of elevation, slope and land covers are provided by the Norwegian Mapping Authority, which bears nationwide

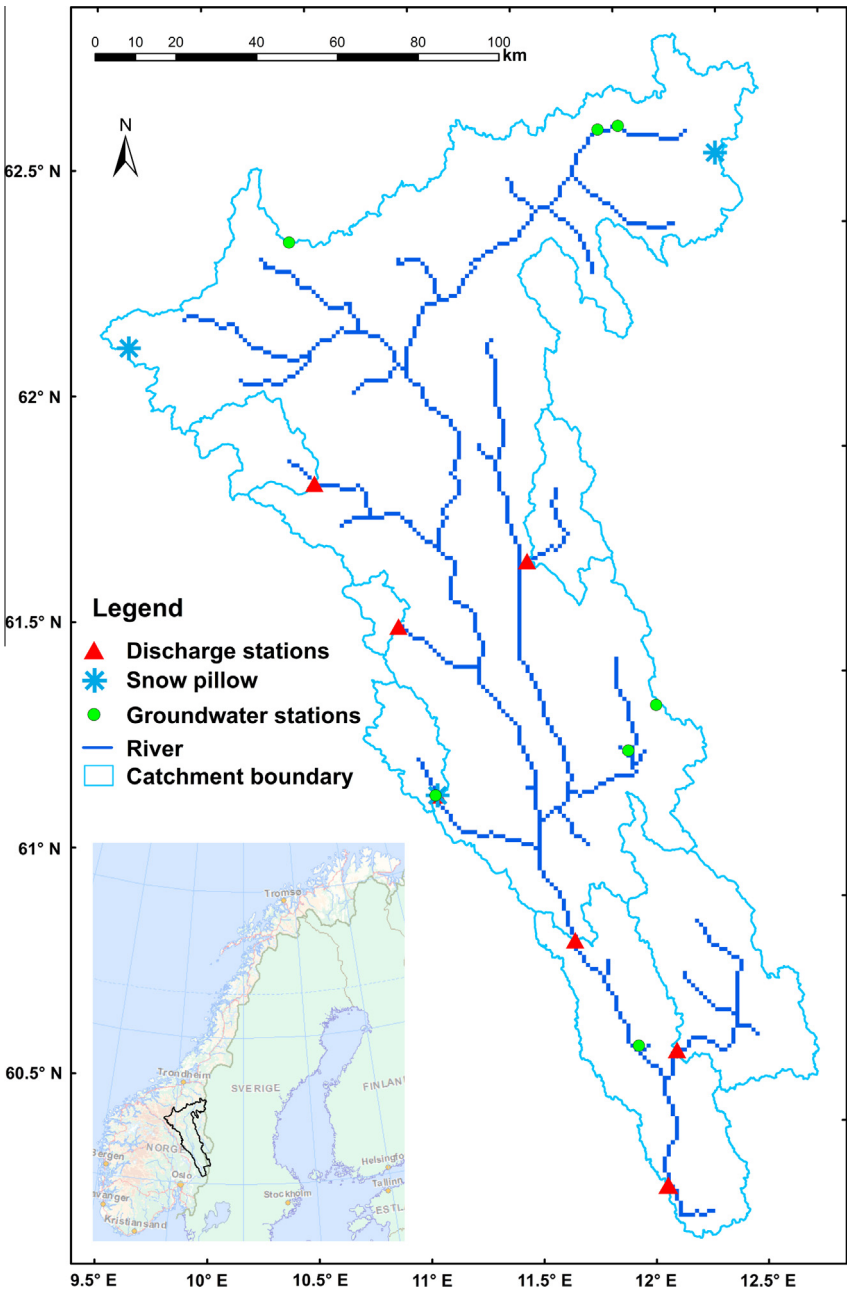

Fig. 1. Location and Digital Elevation Models (DEMs) of the Norsfoss catchment, Norway. The locations of snow pillows, groundwater stations and discharge stations are also shown; more detailed information is given in Table 1. Note that only the discharge at the catchment outlet (the lower right corner discharge station) is used in calibration.

responsibility for geographical information. These datasets are at a spatial resolution of 25 m regular grid.

For elevation and slope, the original datasets are aggregated at 1 km. There are eight land cover types, i.e. lake, bog, forest, bedrock, heather, alpine forest, open land and glacier. Their fractions in each 1 km grid are computed from their respective numbers of small grids at 25 m within 1 km grid.

The drainage networks delineation is constructed in three steps. Firstly, the DEMs at 25 m are aggregated at 100 m. The reason for the aggregation is that there are many small lakes and bogs in the study area. This aggregation removes these local pour points. Secondly, flow direction data are built based on the 100 m DEMs according to the D8 (deterministic eight-node) method. The D8 method developed by O'Callaghan and Mark (1984) assumes that water flows to only one of the eight nearest neighbours based on the steepest slope. Lastly, the drainage networks at a resolution of 1 km are created by using the river Network Scaling Algorithms (Fekete et al., 2001; Gong et al., 2009). The drainage networks are also visually checked according to the National River Network Database (ELVIS) (NVE, 2015) made in the year 2011, and manually modified if necessary. More details can be found in Li et al. (2014).

Estimating the channel cross-sectional size is difficult and essential in channel routing. In this study, 111 cross-sections are measured from digital maps; however, these measurements cannot be utilised without pre-processing by distributed models for three reasons. Firstly, the digital river network sometimes does not really match most natural rivers, which may lead to usage of only a small number of the measured river cross-sections. Secondly, in Norway, some lakes, such as Storsjøen, are very long

and crossed by rivers. The widths of these long lakes are much wider than rivers and should not be counted in the river routing. Thirdly, in some grids, there are more than one measured cross-sections. A simple method assuming that widths along a river section are similar, is developed to get river width of river grids in two steps. Firstly, the digital river network is divided into different sections according to the Strahler system (Strahler, 1957). Secondly the width of one river section is assigned the mean of the widths of all river cross-sections measured in the respective local sub-catchment.

2.3. Climate data

Daily maps of precipitation and temperature at a resolution of 1 km for mainland Norway are used in this study. These datasets are produced by the Norwegian Meteorological Institute (met.no) using daily observations of 24-h mean temperature and accumulated precipitation measured at meteorological stations.

For temperature, daily mean observations of 150 stations are first projected to the mean sea level by regression analysis based upon monthly mean temperature data from 1152 stations in Norway, Sweden, Denmark and Finland. Interpolated temperature by the residual kriging method is readjusted to terrain elevation using a lapse rate that varies among different seasons (Mohr, 2008).

For precipitation, observations of 630 stations are corrected for systematic under-catch due to station's exposure to wind. To interpolate precipitation, a method of triangulation (Mohr, 2008) is used with correction for the altitude of grid point using a vertical precipitation gradient of 10% per 100 m difference below 1000 m amsl and a vertical precipitation gradient of 5% per 100 m difference above 1000 m amsl (Mohr, 2008; Vormoor and Skaugen, 2013). These datasets have been evaluated and used in many studies covering Norway, such as hydrology modelling (Beldring, 2002a; Li et al., 2014), glacier mass estimation (Engelhardt et al., 2012; Engelhardt et al., 2013), permafrost evolution (Gisnås et al., 2013) and snow depth estimation (Mohr, 2008; Vormoor and Skaugen, 2013; Bruland et al., 2015). These studies have shown that the datasets are generally of high quality. In the Norsfoss catchment, there are 64 stations roughly uniformly distributed, which means a station density of 3.4 per 1000 km^2. The high density of stations ensures a good quality of the climate data in this study.

2.4. Observations

Discharge, snow water equivalent (SWE) and groundwater depth are used to calibrate or validate the models.

Discharge is transformed from measured stage using the Bayesian Rating Curve Fitting method (Petersen-Øverleir et al., 2009). Time series of daily discharge in the period from 1981 to 1990 at the Norsfoss gauging station are used in calibration and the period 1991–2000 are used for validation.

SWE is measured by three snow pillows, which are in the national network of Norway. The measurements of the national snow pillows have been re-verified by extensive sampling on a monthly basis during the winter of the two years, 1998 and 1999.

Groundwater depth is measured by seven piezometers. All the piezometers are located in sparse vegetation areas and the measured aquifers are shallow glacier tills. The Gbr Vika and Abrahamsvoll stations are very close to lakes. Generally the measurements are of high quality; however there is a large uncertainty in peak groundwater measurements due to manual measurement at weekly time step (Fleig, 2013). Locations of the piezometers are depicted in Fig. 1 and tabulated in Table 1.

3. Methodology

3.1. Model description

The HBV model concepts were initially developed for usage in the Scandinavia in the 1970s (Bergström, 1976) and the model has been modified into many versions (Sælthun, 1996; Beldring et al., 2003; Ehret et al., 2008; Hailegeorgis and Alfredsen, 2015). The model contains routines for snow accumulation and melting, soil moisture accounting and runoff generation. The HBV model has been applied in more than 80 countries and is currently used as a standard tool for flood forecasting and simulating inflow to hydropower reservoirs in many areas. In NVE, which provides flood forecasting services for the entire Norway, the semi-distributed HBV model with ten elevation bands (SBand) is currently used for flood forecasting.

Bergström (1976) and Lindström et al. (1997) have explicitly described the model algorithms and the scientific bases. Uhlenbrook et al. (1999) further demonstrated the plausibility of the model structure and process conceptualizations by using the GLUE approach. Therefore, only the equations of the studied components, namely runoff, evaporation, snow storage and groundwater are given in this paper.

Snow accumulation and melting are calculated based on temperature according to Eqs. (1) and (2) (Lindström et al., 1997). Snow is a porous media and some rain and melt water can be retained in the pores. In the models, porosity of 8% of the snowpack water equivalent is assumed. Melt water can be released only after pores are filled up. Hence, SWE is the sum of snow and contained liquid water.

$$SF = P \quad T < TACCT \tag{1}$$

$$SMelt = SMELTR \times (T - MELTT) \quad T > TACCT \tag{2}$$

where SF is snow fall; P is precipitation; T is air temperature; $SMelt$ is snow melt rate. $ACCT$, $SMELTR$ and $MELTT$ are parameters. In this study, $ACCT$ is fixed as zero and other parameters are calibrated.

Evaporation is calculated based on the potential evaporation rate (PE) and available water for evaporation. PE is computed based on a parameter for land surface and temperature. Lakes evaporate at PE.

$$PE = EPOT \times T \quad T > 0 \tag{3}$$

Table 1
Types, names and locations of the observing stations. The areas (km^2) of the catchments are shown in the parentheses following the names of the discharge stations. The elevations (in m amsl) of the snow pillows and ground piezometers are in the parentheses following the name.

Type	Name	Longitude (E)	Latitude (N)
Discharge	Søndre Imssjøen (158)	10.6618	61.5479
	Kvarstadseter (377)	10.8933	61.1785
	Atnasjø (463)	10.2221	61.8519
	Mistra Bru (550)	11.2419	61.7111
	Knappom (1,646)	12.0471	60.6412
	Elverum (15,447)	11.5607	60.8742
	Norsfoss (18,933)	12.0355	60.339
Snow pillow	Kvarstadseter (665)	10.8933	61.1785
	Fokstugu (980)	9.2971	62.1188
	Vauldalen (830)	12.0334	62.6411
Groundwater	Haslemoen (169)	11.8738	60.6467
	Kvarstadseter (698)	10.8837	61.1781
	Gbr Vika (436)	11.7595	61.3019
	Stenerseter (625)	11.8782	61.4065
	Settalbekken (994)	10.0159	62.3828
	Glåmos (700)	11.4609	62.6789
	Abrahamsvoll (710)	11.5576	62.6897

where *EPOT* is a parameter for potential evaporation rate.

The rate of actual evaporation (*AE*) is a function of *PE* and available soil moisture (*SM*):

$$AE = \begin{cases} \frac{PE \times SM}{FC \times FCD} & SM < FC \times FCD \\ PE & SM \geqslant FC \times FCD \end{cases} \quad (4)$$

where *FC* is the field capacity; *FCD* is a parameter that takes values between 0 and 1. It is fixed as 1 in this study.

The infiltration is controlled by a infiltration capacity, *INFMAX*. The part of the water input exceeding this parameter goes directly to the upper zone of the groundwater storage (*UZ*) as well as a fraction of the input water.

$$UZ = Inf_e + InSoil \times \left(\frac{SM}{FC}\right)^{BETA} \quad (5)$$

where Inf_e is the input water exceeding *INFMAX*. *InSoil* is the water getting into the soil. *BETA* is a parameter describing shape of soil moisture release curve.

Deep percolation from the upper zone to the lower zone is a product of a soil parameter, *PERC* and *UZ*. Total runoff, *Q*, is sum of the runoff from the upper zone and lower zone (Eq. (6)). This conceptualisation is based on that runoff is mainly generated due to a rising groundwater table and Horton overland flow rarely occurs due to a large hydraulic conductivity of the glacial till deposits (Beldring, 2002b). The runoff from the upper zone represents the quick flow and the runoff from the lower zone represents the base flow.

$$Q = KUZ \times UZ^{ALFA} + KLZ \times LZ \quad (6)$$

where *LZ* is storage of the lower zone; *KUZ* and *KLZ* are recession parameters. *ALFA* is a parameter introduced by Lindström et al. (1997) to handle the over-parametrisation problem. *ALFA* replaces two parameters of the response function of the upper zone and improved the model efficiency (Lindström et al., 1997). This non-linear response reflects the rapidly decreased conductivity of surface deposits in the Nordic countries (Beldring, 2002b). The groundwater storage is the sum of *UZ* and *LZ*.

3.2. Five models

Five models based on the same HBV concepts are compared, namely, LWhole, SBand, GRZero, GROne, and GRTwo. These five models are presenting increasing levels of spatial discretisation and process details. All the models are run at the daily time step with the same input data.

The lumped model, LWhole, is a simple HBV model using the areal mean precipitation and temperature over a catchment with a uniform characteristic. All types of land covers and soil are sharing the common parameter values.

The semi-distributed model with elevation bands, SBand, is the most widely used model setting in Norway. The catchment is divided into ten elevation bands with equal area according to the hypsometry curve. For each zone, time series of mean precipitation and temperature are inputs. Calibrated parameters vary from different type of land cover and soil. The sub-catchments share the same elevation band division and input data but with different weighting of each band.

For the grid-model without routing, GRZero, runoff generation is performed in every grid. Discharge at outlet is sum of runoff from all catchment grids. This model is of high spatial representation in forcing data and in parameter values.

For the grid-model with hillslope routing, GROne, the runoff generation subroutine runs on the most upslope grid, and the generated runoff is added to the groundwater storage of its downslope landscape grid. The downslope grid performs the groundwater routine with input equal to sum of the runoff from its upslope grids and the local net precipitation. This procedure runs from the upstream area to the downstream area until the runoff discharges into the river grids. Then the model runs for next time step.

For the grid-model with both hillslope routing and river flow routing, GRTwo, runoff is summed at respective river grid and routed by the Muskingum-Cunge method between river grids. River channel cross-sections are assumed to be rectangular. When the Courant number is greater than one, output discharge is equal to input discharge, which means no routing. Details of the Muskingum-Cunge method can be seen from, e.g. Todini (2007).

Among the five models, the lumped model, LWhole is the simplest. The semi-distributed model, SBand is suggested to represent the important features of catchment, while at the same time this model does not require high demand of data and computation. The first grid-model, GRZero, is a large assembly of LWhole with high spatial representation. The second grid-model, GROne describes hillslope routing. The third grid-model, GRTwo describes both hillslope routing and channel routing.

3.3. Parameters and calibration

Model parameters are calibrated by a model-independent non-linear parameter estimation and optimization package called PEST (Parameter ESTimation), which is frequently used for model calibration in different research fields. PEST reads files of input parameters and target output, thus calibration with any arbitrary model can be easily set up. The algorithms are based on the implementation of the Gauss–Marquardt–Levenberg algorithm, which combines advantages of the inverse Hessian matrix and the steepest gradient method to allow a fast and efficient convergence towards to the best value of the objective function. It is the minimum of a weighted least square sum of the discrepancies between simulated and observed discharge where similar weights are assumed. Within specified ranges of parameters, PEST approaches to an optimised parameter set (Doherty and Johnston, 2003; Wrede et al., 2013).

The Gauss–Marquardt–Levenberg method can efficiently find local objective function minima (Skahill and Doherty, 2006) with a small number of model runs. In contrast, other algorithms such as the Shuffled Complex Evolution algorithm (Duan et al., 1992) are much more likely to find the global objective function minimum with the cost of a much greater number of model runs (Coron et al., 2012). PEST is selected due to its good performance for Norwegian catchments (Lawrence et al., 2009) as well its sophistication after a long term development (Doherty, 2005). The optimal global parameter set is verified on the basis of different initial parameter sets.

The sensitivity and uncertainty of model parameters of the HBV model have been widely studied in Sweden and Norway. Bergström (1976) mapped the mean square error function of streamflow by the trial and error method. In snow routine, *MELTT* and *SMELTR* in Eq. (2) were sensitive and the sensitivity of *SMELTR* increased with lower *MELTT*. In the soil moisture accounting routine, *BETA* (in Table 2) was sensitive and its sensitivity decreased with increased *FC* in Eq. (4). In the dynamic response, *KUZ* in Eq. (5), *PERC* (in Table 2) and *KLZ* in Eq. (6) were greatly influenced by each other and the parameters in other routines. Using the Monte Carlo method, Harlin and Kung (1992) demonstrated that *SMELTR* and *KUZ* were sensitive, and *BETA* and *KLZ* were moderate. Furthermore, Seibert (1997) found that the sensitivity was hard to be described quantitatively, since the sensitivity changes greatly with different parameter values. *MELTT* and *KUZ* were most sensitive near the optimised value (Seibert, 1997). To achieve good performance in streamflow, only *MELTT* and *SMELTR* were in narrow range (Uhlenbrook et al., 1999). By similar

Table 2
Optimised values of the selected parameters. The parameters for SBand and the grid-based models are land use dependent; therefore the areal mean is given.

Group	Name	Meaning (unit)	LWhole	SBand	GRZero	GROne	GRTwo
Evaporation	*EPOT*	Potential evaporation capacity (m/day/°C)	5.02E−03	1.31E−02	1.31E−02	8.50E−03	7.14E−03
	INT_MAX	Interception storage (m/day)	2.91E−04	2.87E−04	2.86E−04	2.71E−04	2.32E−04
Snow	*MELTT*	Melting temperature of snow (°C)	−1.98	−0.84	−0.84	0.92	0.92
	SMELTR	Temperature index of snow melting rate (m/day/°C)	2.27E−03	3.93E−03	3.89E−03	8.06E−03	7.91E−03
Soil	*FC*	Field capacity (m)	1.293	0.738	0.751	0.318	0.335
	BETA	Shape coefficient of soil moisture	0.396	1.473	1.775	16.009	15.933
Runoff	*KUZ*	Recession coefficient of the upper zone (1/day)	0.133	0.536	0.541	0.567	0.694
	ALFA	Non-linear drainage coefficient of the upper zone	0.976	2.392	2.293	1.379	1.364
	PERC	Percolation from upper zone to the lower zone	0.013	0.019	0.020	0.006	0.005
	KLZ	Recession coefficient of the lower zone (1/day)	0.035	0.074	0.074	0.019	0.014

approach, Seibert (1999) found *MELTT* was the most constrained parameter by the objective function. In addition to studies in Nordic countries, Abebe et al. (2010) concluded that *FC* and *BETA* were sensitive to volume error and *KUZ* and *KLZ* were sensitive to high flow series in semi-humid watershed located in the sub-tropical coastal plain of south-eastern United States.

In the Nordic countries, glacial till deposits over an impermeable bedrock are the main aquifers. The subsurface conditions heavily depend on the land cover (Beldring, 2002b). To simplify, we assume that parameters vary among different land cover (Beldring et al., 2003; Wrede et al., 2013). To reduce the difficulty caused by the larger number of parameters in calibration, some parameters are tied, which means that they are simply maintaining a constant ratio to their parent parameter, which are optimised (Doherty, 2005). The calibrated parameters are the lake parameters, land cover parameters and soil parameters for bog, forest and alpine and *KLZ* for other land covers. Other parameters in the land cover and soil groups are tied to forest.

Initial parameter values were selected on the basis of previous best manual calibration trials. To achieve global optimised parameter values, we did start calibrations with different initial values and different parameter ranges. The meaning and optimised value of every parameter are shown in Appendix A and their areal means are summarised in Table 2.

3.4. Evaluation criteria

The models are evaluated with respect to the comparison between simulated and observed runoff both at the catchment outlet and internal points, and evapotranspiration, SWE and groundwater table are also compared among the models. These variables are selected because they are important components of water balance or can be validated by in situ measurements. The relative mean error (*RME*), the Nash–Sutcliffe efficiency (*NSE*) (Nash and Sutcliffe, 1970), the inverse Nash–Sutcliffe efficiency (*InNSE*) (Pushpalatha et al., 2012) and the correlation coefficient (*R*) were used as criteria for model performance.

$$RME = \frac{\sum_{i=1}^{n}(S_i - O_i)}{\sum_{i=1}^{n} O_i} \times 100 \quad (7)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O})^2} \quad (8)$$

$$InNSE = 1 - \frac{\sum_{i=1}^{n}\left(\frac{1}{S_i} - \frac{1}{O_i}\right)^2}{\sum_{i=1}^{n}\left(\frac{1}{O_i} - \frac{1}{\overline{O}}\right)^2} \quad (9)$$

$$R = \frac{\sum_{i=1}^{n}(S_i - \overline{S})(O_i - \overline{O})}{\sqrt{\sum_{i=1}^{n}(S_i - \overline{S})^2}\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2}} \quad (10)$$

where O_i is the observed series; S_i is the simulated series; n is the length of series; $\overline{O}$ and $\overline{S}$ are the mean of observed series and simulated series. The perfect values (dimensionless) are 1 except *RME* being 0.

RME is for water balance error; *NSE* is for match of simulations and observations with large weight on high flow; *InNSE* is also for closeness of simulations and observations with large weight on the low flow (Pushpalatha et al., 2012). *R* is used in evaluating internal variables. Therefore, combination of these four criteria can give an appropriate evaluation of model performance.

4. Results

In this study, the model comparison is partly done according to runoff simulation either at the catchment outlet or interior points. The better a model is in runoff simulation, the more reliable the model is. Additionally, the differences of model simulations in evapotranspiration, SWE and groundwater storage are also analysed.

4.1. Discharge estimation at the catchment outlet

Results of *NSE* and *InNSE* are shown in Fig. 2. Differences existing in LWhole, SBand and GRZero show effects of finer spatial representation of the input data; differences existing in GRZero, GROne and GRTwo show effects of finer spatial representation of hydrologic processes. The figure shows that (1) The values of *NSE* increase from LWhole to the fully distributed model, GRTwo. (2) The most significant improvement occurs from LWhole to SBand and then from GRZero to GROne. (3) In simulation of the low flow, GROne and GRTwo are still able to give a fair estimation with relatively high *InNSE* values whereas the *InNSE* values of the other models fall below zero. (4) Implementing the Muskingum-Cunge method does not add additional value to runoff simulation at a daily time step in the study area as reflected by the unchanged value of *NSE* and *InNSE* between GROne and GRTwo.

The runoff seasonality is of high concern by hydrologists for water management. The monthly mean runoff is shown in Fig. 3(a) and the deviations of the five models from the observed values are shown in Fig. 3(b). It is seen that the model errors of LWhole and SBand exhibit a stronger seasonal pattern than GROne and GRTwo. Less water is predicted in pre-flooding seasons and more floods are predicted to occur in high flow seasons. In consequence LWhole and SBand have a tendency to amplify risks of drought and flooding and to hint a depressing situation for water resources. Although all five model structures have good total *RME* values of less than 3%, GROne and GRTwo are more reliable in overall evaluation.

4.2. Discharge estimation at interior points

One of big benefits of using distributed models is that they can explicitly account for spatial variability inside a basin and have the

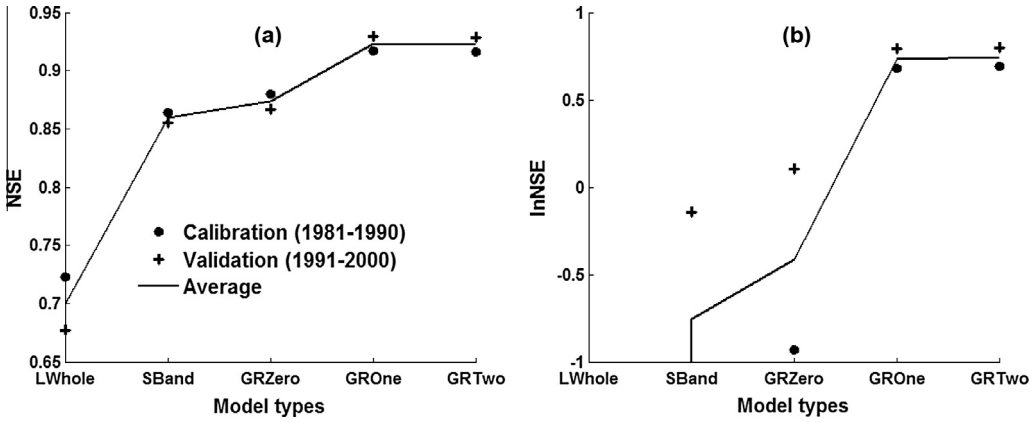

Fig. 2. The *NSE* (a) and *lnNSE* (b) values of different spatial discretisation schemes for calibration and validation periods at the outlet. Average is the mean of the *NSE* values for the calibration and the validation. In (b), the values, which are not shown, −46.12 for calibration, −13.32 for validation, −29.72 as the average of LWhole, and −1.37 for calibration of SBand.

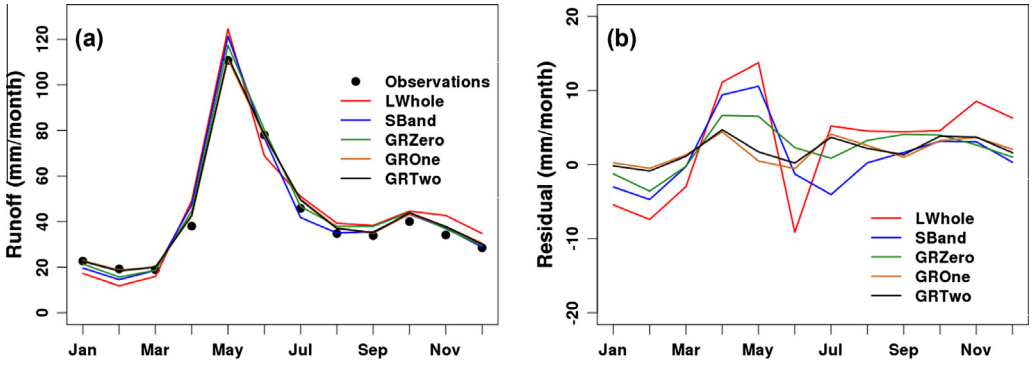

Fig. 3. Monthly mean runoff for the validation period at the outlet. (a) Comparison of simulations by the five models and the observations. (b) The residuals of the simulations.

ability to produce simulations at interior points without explicit calibration at these points. However, by transferring the parameter values and areal mean inputs, the lumped models can also estimate runoff at interior points. Then all the models can be evaluated at interior points; therefore we treat the performance of LWhole as a benchmark for the assessment of the other models, because we simply run LWhole in the sub-catchments with the parameter values calibrated at the catchment outlet and the time series of areal mean inputs. The forcing data in the sub-catchments for LWhole are the same as in the entire Norsfoss catchment.

There are six sub-catchments in total and results are presented in Fig. 4 and Table 3. It is noteworthy that only the discharge of the Norsfoss station was used in the calibration. Generally, the *NSE* values of all models increase with increasing catchment area. In the sub-catchments larger than 1646 km^2 (less than 10% of the area of the Norsfoss catchment) the *NSE* values of all the models except LWhole are larger than 0.6.

The mean and coefficient of variance (*CV*) of the *NSE* values of the five models are shown in Table 3. High mean with low *CV* indicates a good model performance. As expected, LWhole gives the lowest efficiency among the five models. GRTwo is the best with the largest mean *NSE* and the smallest *CV*. GROne and GRTwo are more likely to be capable than other three models.

4.3. Internal variables

Involving other variables in model evaluation is essential to make a breakthrough in model structure identification and process

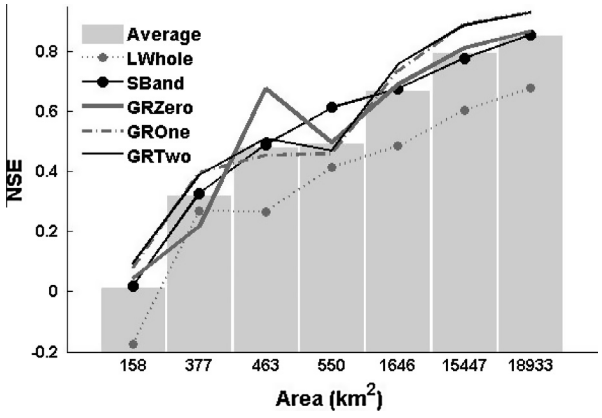

Fig. 4. The *NSE* values of six sub-catchments and the entire study area of five model structures for the validation period. Average is the mean of the *NSE* values of different model structures. The catchment with the largest area (18,933 km^2) is the entire study area.

Table 3
The mean value (mean) and *CV* of *NSE* on six sub-catchments and the entire study area in the validation period.

	LWhole	SBand	GRZero	GROne	GRTwo
Mean	0.363	0.536	0.543	0.564	0.576
CV	0.722	0.498	0.527	0.499	0.482

understanding. Simulations of evapotranspiration, SWE and groundwater are compared either among the five models or with observed data.

4.3.1. Evapotranspiration

It can be seen from Fig. 5 that the simulations of evapotranspiration by the five models are very close. Slight differences occur during the summer time, especially at the peaks in July. In the winter time, all the simulated evapotranspiration values are zero or close to zero. In humid area, evapotranspiration highly depends on the available energy for evaporation and vegetation growth. Thus evapotranspiration increases from January to July, when it is the warmest month and biological activity is intense (Beldring et al., 2003), and decreases from July to December. The calculated evapotranspiration is 352 mm/year and accounts for 40% of annual precipitation. This reflects a typical water balance situation at the east side of high mountains in central southern Norway.

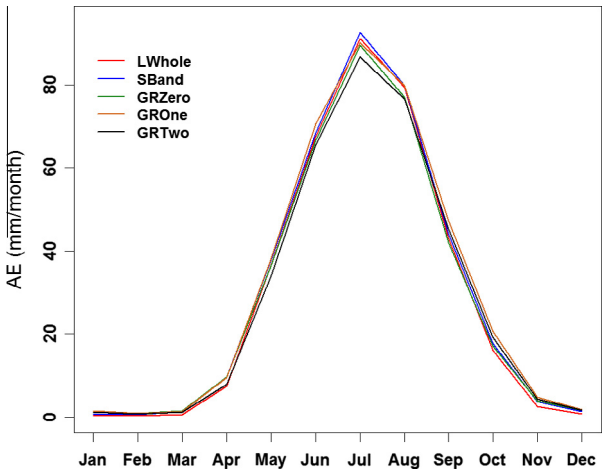

Fig. 5. Monthly mean evaportransipiration simulated by the five models for the validation period for the entire study area.

4.3.2. SWE

The SWE time series measured by three snow pillows and the simulations of three grids where the snow pillows are located by three grid-models are shown in Fig. 6. All the grid-models can give reasonable pattern of snow storage both by visual inspection and by the R values in Table 4. The models capture the time of accumulation and melting. However, all the three models do not accurately estimate the SWE amount. At the Kvarstadseter snow pillow site, all the models compute more snow than the observations; however, the results of the models are inconsistent at the Vauldalen snow pillow site. GRZero gives the highest R value at the three sites.

The areal mean monthly SWE simulations are presented in Fig. 7. It shows that the models can be roughly grouped into three classes. LWhole is the first class; SBand and GRZero are in the second class; and GROne and GRTwo are in the third class. There is more SWE calculated by the models of the third class than others.

4.3.3. Groundwater

In the HBV model, groundwater is simulated as areal mean volume of groundwater storage, whereas in practice, groundwater level is measured by piezometers. However, the storage and depth are highly correlated. Assuming a homogeneous aquifer and horizontal water table, the storage of the aquifer can be calculated by Eq. (11).

$$S_a = (H - H_0) \times \phi = H \times \phi - H_0 \times \phi \tag{11}$$

where S_a is the storage of aquifer per unit area in depth; H is the groundwater level; H_0 is the reference level, which is usually thought as the lowest groundwater level; ϕ is the effective porosity in relative volume.

The R between observed groundwater level and simulated groundwater storage is used to evaluate groundwater models in this study. The seven groundwater piezometers as presented in Table 1 and Fig. 1 are used. Weekly observations are available for the period from 1981 to 2000 with some missing values. In total, there are 5702 observations. All the measured aquifers are

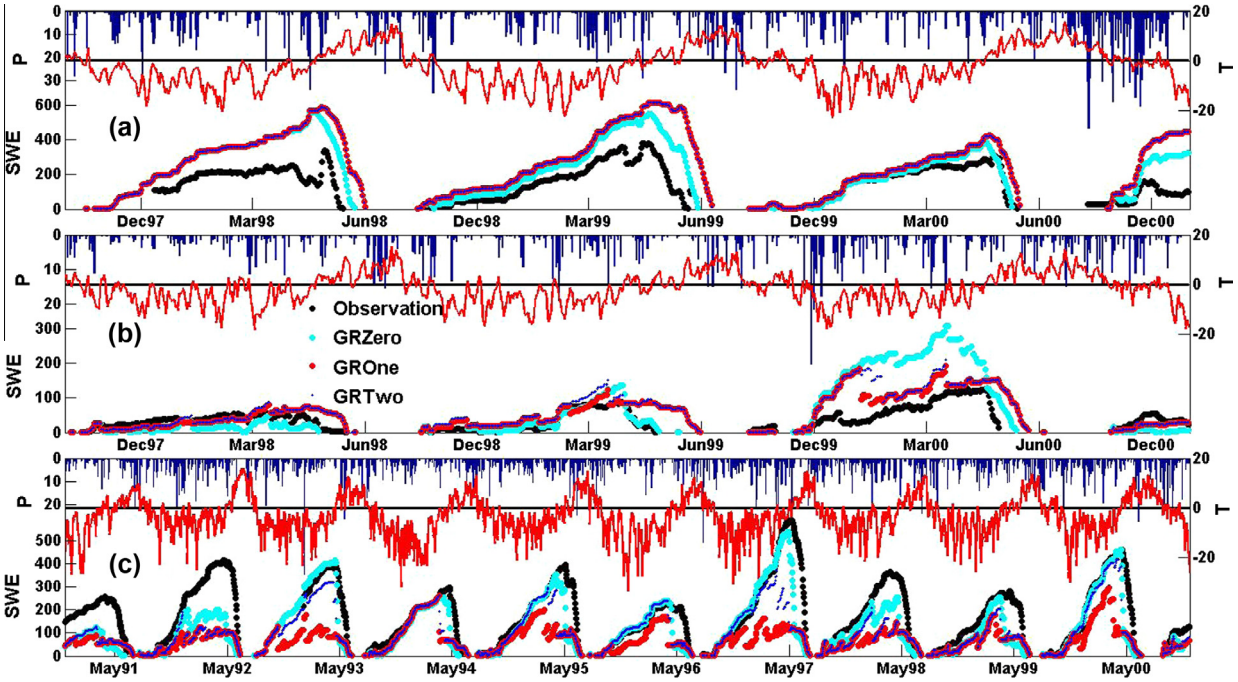

Fig. 6. Observations of SWE by the snow pillows, Kvarstadseter (a), Fokstugu (b) and Vauldalen (c) and the simulations of SWE of the grids where the snow pillows are located. P is the daily precipitation in mm/day. SWE is the daily observation and simulations in mm. T is the daily temperature in °C.

Table 4
The *R* values between snow pillow measurements and grid-model simulations.

Snow pillow	GRZero	GROne	GRTwo
Kvarstadseter	0.8	0.68	0.68
Fokstugu	0.73	0.7	0.69
Vauldalen	0.77	0.61	0.64
Mean	0.77	0.66	0.67

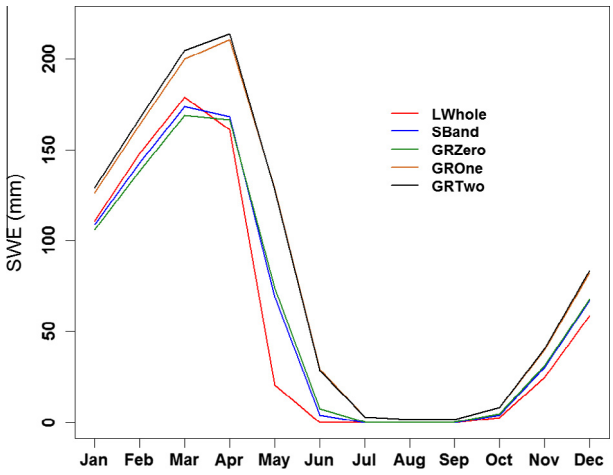

Fig. 7. Monthly mean SWE by the five models for the entire study area.

unconfined. *R* is calculated between the observed time series and the simulated time series of the grid where the piezometer is located.

Fig. 8 presents the results against the mean groundwater depth at the piezometers sites, elevation and slope of the grids where the piezometers are located. Firstly, all the grid-models give lower correlation coefficients at the piezometers with the deepest groundwater depth. Secondly, the patterns against elevation seem quite complex. The correlation coefficients go up with increased elevation and then down. The highest correlation coefficients are at approximately 600 m. Thirdly, there is an up trend against slope with large spread. The mean of all the *R* is 0.53 and the values of *R* of the three grid-models are comparable in magnitude.

5. Discussions

5.1. Runoff simulations

In rainfall–runoff modelling, runoff is indeed the most crucial output. As expected, the model performance improves most from LWhole to SBand and minor improvement occurs from the semi-distributed model, SBand to the grid distributed model, GRZero. This confirms the previous studies by Lindström et al. (1997), Das et al. (2008) and Wrede et al. (2013). The spatial variability in the forcing data, particularly temperature and precipitation must be accounted in hydrological modelling in mountainous and cold catchments. This is quite different from the catchment used by Michaud and Sorooshian (1994) and the DMIP catchments by Reed et al. (2004) and Butts et al. (2004).

The improvements both in the *NSE* and *lnNSE* from GRZero to GROne show that distributed hillslope routing is essential in runoff modelling, especially for simulating the low flow. In the hillslope routing, the water routes from an upstream grid to the outlet rather than directly add to its runoff volume. This process prolongs the response time and more water drains in the low flow seasons than in high flow seasons. This hypothesis is confirmed by Fig. 3. The models except GROne and GRTwo overestimate the high flow and underestimate the low flow.

The low efficiency of channel routing is caused by the high slope and sparse vegetation. Most runoff flows out within two days (Li et al., 2014). It is difficult to include this fast channel response in a daily simulation. The data error in observed discharge and aggregation of precipitation and temperature (Montanari and Baldassarre, 2013) is another possible reason.

The up trend of model efficiency with size of sub-catchments is the same as what was reported by Varado et al. (2006). A primary contributing factor to this may be that smaller basins have less capacity to deal with data error either in forcing data or discharge series. Besides, observations in small basins indeed exhibit more variability and dynamics in responses functions (Varado et al., 2006) than in larger basins, making accurate simulation more difficult (Reed et al., 2004). Another reason is that in this study the models are calibrated for the outlet station, and large sub-catchments share more common features with the one used for calibration.

5.2. Internal variables' simulation

Validating models on internal variables is always difficult and of primary importance. The first reason is that most internal variables, such as soil moisture and groundwater depth are measured at point scale. Simulations by hydrological models, even distributed models at a very fine resolution are areal mean values. The second reason is the less satisfactory length and quality of measurements compared to runoff data. However, this approach is required to assess the general behaviour of hydrological models and identification of model structures, especially in cases when runoff simulations by different models are close.

All grid-simulations are at a regular grid of 1 km. It is the most adopted spatial resolution in catchment hydrology studies and global datasets, for example, Hydro1k (EROS, 1996). These results will

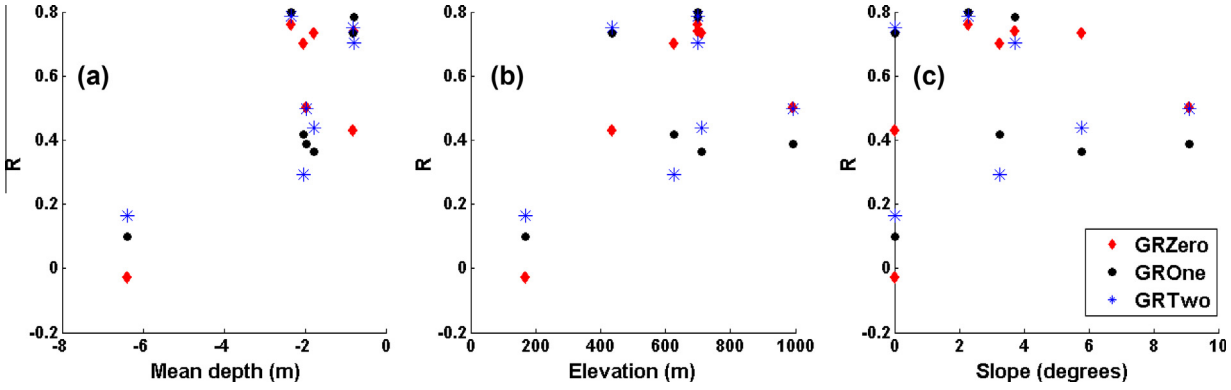

Fig. 8. The *R* values between measured groundwater level series and the simulated groundwater storage series.

give insights of other researchers and will be compared to on-going research.

5.2.1. Evapotranspiration

Evapotranspiration is a crucial term to connect a water balance model and a land surface energy balance model (Xu et al., 2005; Xu and Singh, 2005). The best available technology for observing evapotranspiration is a weighing lysimeter. However, establishing and maintaining a lysimeter for a long time period is very costly (Xu and Chen, 2005). Therefore, evapotranspiration is usually estimated using other methods, such as the Penman-Monteith method, reference evapotranspiration based methods, hydrological models (water balance method) and complementary relationship based methods (Gao et al., 2012). Recently, remote sensing data have been gradually used to estimate actual evapotranspiration (El Hai EI Tahir et al., 2012; Yang et al., 2012; Corbari et al., August 2014). The most conceptually simple tool is the hydrological modelling, especially when regional estimation is required (Xu and Singh, 2005).

Although no observations of evapotranspiration are available, the simulated patterns are considered to be reasonable. The five models produced similar results. This can be explained from the water balance, because all five models used the same precipitation and discharge data and the model reproduced almost same runoff volume. The leftover of precipitation after runoff is the evaporation, if water storage of catchment does not change. Another reason is that the five models used the same temperature based method with the same areal mean temperature.

5.2.2. SWE

In Norway, almost half of the precipitation falls as snow, and it appreciably influences the hydrological responses. The most severe flood in southeast Norway in recent times occurred in the year 1995 and was fed from extensive snow covered high-mountain areas (Sorteberg et al., 2001).

Table 5
Ranks in the descending order of the five most sensitive parameters for each model. Other parameters are found insensitive.

Rank	LWhole	SBand	GRZero	GROne	GRTwo
1	*ALFA*	*SMELTR*	*EPOT*	*ALFA*	*ALFA*
2	*SMELTR*	*EPOT*	*SMELTR*	*KUZ*	*KUZ*
3	*EPOT*	*ALFA*	*ALFA*	*EPOT*	*EPOT*
4	*KLZ*	*KLZ*	*KLZ*	*SMELTR*	*SMELTR*
5	*MELTT*	*FC*	*FC*	*MELTT*	*FC*

The data errors existing in the measurements by snow pillow and in the grid temperature make the model simulation less accurate than other variables. For example, at the Kvarstadseter snow pillow site, there was a melting event on April 4th, 1998 according to the observations (Fig. 6). Contradictorily all models responded as an accumulation. The temperature of this grid was −7.5 °C in April 4th, 1998. Therefore, the models are believed to function well to the forcing data. The error may be investigated from the uncertainty of forcing data and measurements, or the sub-grid variability (Gisnås et al., 2013).

5.2.3. Groundwater

Bergström and Sandberg (1983) have shown that the HBV model is able to model groundwater level of unconfined aquifers after calibration with an objective function to achieve the maximum R between observed groundwater level and simulated groundwater storage. Their research provides confidence in the model and a benchmark in presenting our results. The values of R were around 0.8 and higher in most cases. However, Bergström and Sandberg (1983) used the average values of the stations at the study catchment and closeness of simulated and observed groundwater depth was the objective function of calibration. Varado et al. (2006) gave an example of model efficiency of a conceptual hydrological model (REW-v4.0) when it was only calibrated with discharge data. The model efficiency was very low and they concluded that was caused by the lack of presentation of the processes in the unsaturated zone.

Although our models produce lower R than the model of Bergström and Sandberg (1983), no groundwater data are used in calibration in our study. Additionally, the observed groundwater depth series are individually compared to the model produced series of the grid where they are located. If counting these differences, our models give equivalent performance.

The HBV model is a conceptual rainfall–runoff model. Therefore, it shows less capability in simulating groundwater levels of deeper depth, which are less relevant to runoff peaks and total runoff volume. Additionally, the three stations with low efficiency are located at approximately 200 m amsl, where human effects are stronger than at other stations with higher elevations.

In cold mountainous catchments, groundwater aquifers are recharged by rainfall and snow melt. At higher elevation places, snow melt has large proportion. Less accuracy in snow melt produced by models than the input precipitation is a possible explanation of low R at around the 1000 m height.

The trend of R against slope would be misleading. If the three stations at the 200 m height (R of which are below 0.2) are removed, there is no trend any more.

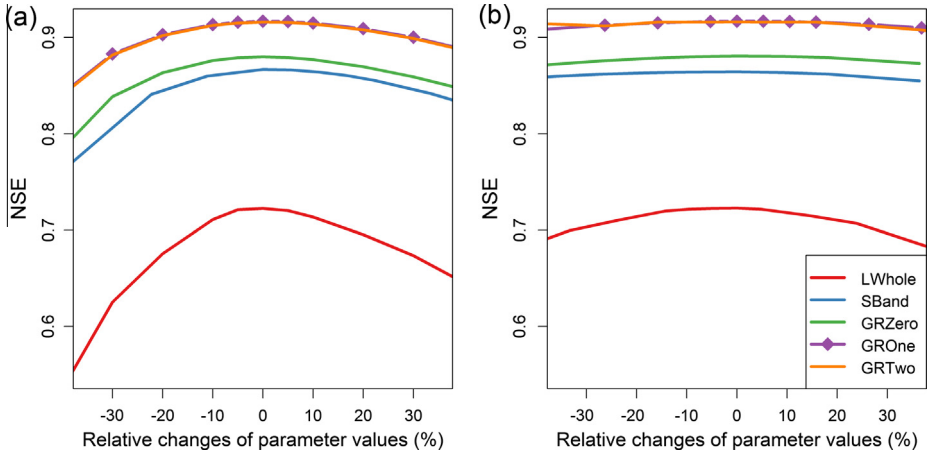

Fig. 9. Sensitivities of *SMELTR* (a) and *MELTT* (b). Other parameters are kept in their optimised values in each model.

5.3. Parameter analysis

Complex models are easily required to prove that their good efficiency is not achieved by adding additional parameters and the best solution should be supported by deliberate measurements of hydrological processes (Vaché and McDonnell, 2006). Otherwise, the optimised parameter values likely provide high evidence of parameter identifiability. As shown in Table 2, the optimised values of the selected parameters are considered to be in their plausible ranges. There are interdependence between certain parameters. *MELTT* and *SMELTR* are strongly correlated and positively associated. However, the soil parameters *FC* and *BETA* are negatively associated.

Parameter sensitivity is difficult to quantify due to the non-linear response of runoff to parameters and parameter interactions (Seibert, 1997; Tang et al., 2007). The sensitivities are impacted by selection of analysis method and the physical characteristic of study area (Tang et al., 2007). Thereby, we define the parameter sensitivity as the changes of *NSE* against the relative changes of the parameter values. Shown in Table 5, *SMELTR*, *ALFA* and *EPOT* are among the five most sensitive parameters for every model. It can be concluded that the ranks of parameter sensitivity for the five models are similar. For an illustrative purpose, the sensitivities of a sensitive parameter (*SMELTR*) and a less sensitive parameter (*MELTT*) are plotted in Fig. 9(a) and (b), respectively.

6. Conclusions

The aim of this paper is to test the hypothesis that with appropriate structures, increasing complexity would increase the model efficiency either in runoff simulation or other internal variables. Five variants of the HBV model, i.e. a lumped model (LWhole), a semi-distributed model with elevation bands (SBand), a grid-model without routing (GRZero), a grid-model with hillslope routing (GROne), and a grid-model with both hillslope and channel routing (GRTwo) are compared in a cold and mountainous catchment in central southern Norway. The five models are compared according to runoff simulation either at the catchment outlet or interior points as well as internal variables, i.e. actual evapotranspiration, snow water equivalent and groundwater depth. The following conclusions are drawn from this study.

According to the Nash–Sutcliffe efficiency (*NSE*) and the inverse Nash–Sutcliffe efficiency (*InNSE*) of the outlet discharge, the rank of model efficiency in the descending order is GRTwo, GROne, GRZero, SBand and LWhole both in the calibration and validation mode. Applying the parameter values calibrated at the outlet to the six sub-catchments within the study area, GRTwo is still the best but with less superiority than in producing the outlet discharge. The *NSE* values increase with the areas of the sub-catchments.

Three internal variables are compared with observations measured at several sites and among models. The areal mean actual evapotranspiration simulations are very similar. Considerable differences exist among areal mean monthly SWE. The *R* values between the measurements by three snow pillows and grid-simulations show that GRZero is slightly better than GROne and GRTwo. Moreover, the three grid-models are comparable in simulations of groundwater levels; model performances are affected by the groundwater depth and elevation of ground piezometers.

It is worth noting that this study is performed only in one catchment. Robustness of the results needs to be tested by, for example, performing the study in other catchments and/or applying the models in non-stationary conditions. The latter can be done by calibrating and validating the models in different seasons or under different climate conditions. This kind of research will be done in the ongoing research.

Acknowledgements

This study is funded by the Research Council of Norway through the research project-JOINTINDNOR (project number 203867). The provision of the meteorological data by met.no, and the discharge and the spatial data by NVE are gratefully acknowledged. The authors are also grateful to the editor and the reviewers for their valuable comments and suggestions, which eventually lead to a substantial improvement of the paper.

Appendix A

see Table A1.

Table A1
Meanings of the parameters and the calibrated values. The *MannR* was only used in the Muskingum-Cunge channel routing method and calibrated for each land cover in calibration. Parameters are subscripted with *b* for the bog, *f* for the forest, *a* for the alpine, *r* for the rock, *h* for the heather, *o* for the open land, *g* for the glacier. The parameters of LWhole for the other land and soil were the same as those for the open land.

Group	Parameters	Meaning	Unit	LWhole	SBand	GRZero	GROne	GRTwo
Lake	*EPOT_L*	Potential evaporation capacity	m/day/°C	0.000001	0.000097	0.000010	0.000971	0.000980
	KLAKE	Rating curve constant	–	0.006224	0.035225	0.015674	0.031472	0.066361
	NLAKE	Rating curve exponent	–	0.560846	1.619736	1.025064	0.512362	0.164502
	MannR	Manning roughness coefficient	–	–	–	–	–	0.030000
Land	INT_MAX_o	Maximum interception storage	m/day	0.005017	0.009845	0.009943	0.002957	0.002488
	$EPOT_o$	Potential evaporation capacity	m/day/°C	0.000291	0.000334	0.000331	0.000196	0.000204
	$MELTT_o$	Melting temperature of snow	°C	−1.977322	−1.399138	−1.400558	1.517349	1.542897
	$SMELTR_o$	Temperature index of snow melting rate	m/day/°C	0.002267	0.002491	0.002488	0.003125	0.003190
	$MannR_o$	Manning roughness coefficient	–	–	–	–	–	0.050000
	INT_MAX_b	Maximum interception storage	m/day	–	0.000528	0.000282	0.007291	0.000733
	$EPOT_b$	Potential evaporation capacity	m/day/°C	–	0.000040	0.000036	0.000175	0.000105
	$MELTT_b$	Melting temperature of snow	°C	–	2.313019	2.300974	−1.470627	−1.890260
	$SMELTR_b$	Temperature index of snow melting rate	m/day/°C	–	0.008284	0.008275	0.008171	0.006826
	$MannR_b$	Manning roughness coefficient	–	–	–	–	–	0.040000
	INT_MAX_f	Maximum interception storage	m/day	–	0.009845	0.009943	0.002957	0.002488
	$EPOT_f$	Potential evaporation capacity	m/day/°C	–	0.000334	0.000331	0.000196	0.000204
	$MELTT_f$	Melting temperature of snow	°C	–	−1.399138	−1.400558	1.517349	1.542897
	$SMELTR_f$	Temperature index of snow melting rate	m/day/°C	–	0.002491	0.002488	0.003125	0.003190
	$MannR_f$	Manning roughness coefficient	–	–	–	–	–	0.010000

Table A1 (*continued*)

Group	Parameters	Meaning	Unit	LWhole	SBand	GRZero	GROne	GRTwo
	INT_MAX_a	Maximum interception storage	m/day	–	0.000600	0.000159	0.000108	0.000100
	$EPOT_a$	Potential evaporation capacity	m/day/°C	–	0.000267	0.000308	0.002000	0.001202
	$MELTT_a$	Melting temperature of snow	°C	–	−1.324652	−1.402145	−2.000000	−0.974982
	$SMELTR_a$	Temperature index of snow melting rate	m/day/°C	–	0.015644	0.014929	0.100000	0.100000
	$MannR_a$	Manning roughness coefficient	–	–	–	–	–	0.025000
	INT_MAX_h	Maximum interception storage	m/day	–	0.009845	0.009943	0.002957	0.002488
	$EPOT_h$	Potential evaporation capacity	m/day/°C	–	0.000334	0.000331	0.000196	0.000204
	$MELTT_h$	Melting temperature of snow	°C	–	−1.399138	−1.400558	1.517349	1.542897
	$SMELTR_h$	Temperature index of snow melting rate	m/day/°C	–	0.002491	0.002488	0.003125	0.003190
	$MannR_h$	Manning roughness coefficient	–	–	–	–	–	0.050000
	INT_MAX_r	Maximum interception storage	m/day	–	0.009845	0.009943	0.002957	0.002488
	$EPOT_r$	Potential evaporation capacity	m/day/°C	–	0.000334	0.000331	0.000196	0.000204
	$MELTT_r$	Melting temperature of snow	°C	–	−1.399138	−1.400558	1.517349	1.542897
	$SMELTR_r$	Temperature index of snow melting rate	m/day/°C	–	0.002491	0.002488	0.003125	0.003190
	$MannR_r$	Manning roughness coefficient	–	–	–	–	–	0.030000
	INT_MAX_g	Maximum interception storage	m/day	–	0.009845	0.009943	0.002957	0.002488
	$EPOT_g$	Potential evaporation capacity	m/day/°C	–	0.000334	0.000331	0.000196	0.000204
	$MELTT_g$	Melting temperature of snow	°C	–	−1.399138	−1.400558	1.517349	1.542897
	$SMELTR_g$	Temperature index of snow melting rate	m/day/°C	–	0.002491	0.002488	0.003125	0.003190
	$MannR_g$	Manning roughness coefficient	–	–	–	–	–	0.030000
Soil	FC_o	Field capacity	m	1.292544	0.854174	0.852845	0.149419	0.170029
	$BETA_o$	Shape coefficient of soil moisture	–	0.395717	0.632809	0.621915	10.000000	9.905496
	KUZ_o	Recession coefficient of the upper zone	1/day	0.133252	0.031759	0.000075	1.000000	1.000000
	$ALFA_o$	Non-linear drainage coefficient of the upper zone	–	0.976138	2.500000	2.374286	1.208677	1.217343
	$PERC_o$	Percolation from upper zone to the lower zone	–	0.012520	0.022951	0.023132	0.001013	0.001000
	KLZ_o	Recession coefficient of the lower zone	1/day	0.035208	0.089434	0.089466	0.020157	0.011136
	FC_b	Field capacity	m	–	0.195487	0.264188	1.300000	1.300000
	$BETA_b$	Shape coefficient of soil moisture	–	–	6.254378	8.313165	50.000000	50.000000
	KUZ_b	Recession coefficient of the upper zone	1/day	–	0.973313	0.998324	0.871973	0.770876
	$ALFA_b$	Non-linear drainage coefficient of the upper zone	–	–	1.784086	1.796010	1.979247	1.826975
	$PERC_b$	Percolation from upper zone to the lower zone	–	–	0.003896	0.004053	0.001000	0.001000
	KLZ_b	Recession coefficient of the lower zone	1/day	–	0.007228	0.006526	0.014653	0.026939
	FC_f	Field capacity	m	–	0.854174	0.852845	0.149419	0.170029
	$BETA_f$	Shape coefficient of soil moisture	–	–	0.632809	0.621915	10.000000	9.905496
	KUZ_f	Recession coefficient of the upper zone	1/day	–	0.005303	0.217543	1.000000	1.000000
	$ALFA_f$	Non-linear drainage coefficient of the upper zone	–	–	2.500000	2.374286	1.208677	1.217343
	$PERC_f$	Percolation from upper zone to the lower zone	–	–	0.022951	0.023132	0.001013	0.001000
	KLZ_f	Recession coefficient of the lower zone	1/day	–	0.089434	0.089466	0.020157	0.011136
	FC_a	Field capacity	m	–	0.444265	0.544096	0.050000	0.050000
	$BETA_a$	Shape coefficient of soil moisture	–	–	0.538706	0.568572	10.000000	10.000000
	KUZ_a	Recession coefficient of the upper zone	1/day	–	0.277815	0.322954	0.132547	0.018076
	$ALFA_a$	Non-linear drainage coefficient of the upper zone	–	–	2.500000	2.500000	2.477734	2.500000
	$PERC_a$	Percolation from upper zone to the lower zone	–	–	0.008084	0.007916	0.122702	0.084411
	KLZ_a	Recession coefficient of the lower zone	1/day	–	0.007617	0.008469	0.022254	0.013849
	FC_h	Field capacity	m	–	0.854174	0.852845	0.149419	0.170029
	$BETA_h$	Shape coefficient of soil moisture	–	–	0.632809	0.621915	10.000000	9.905496
	KUZ_h	Recession coefficient of the upper zone	1/day	–	1.000000	0.146038	1.000000	1.000000
	$ALFA_h$	Non-linear drainage coefficient of the upper zone	–	–	2.500000	2.374286	1.208677	1.217343
	$PERC_h$	Percolation from upper zone to the lower zone	–	–	0.022951	0.023132	0.001013	0.001000
	KLZ_h	Recession coefficient of the lower zone	1/day	–	0.089434	0.089466	0.020157	0.011136
	FC_r	Field capacity	m	–	0.854174	0.852845	0.149419	0.170029
	$BETA_r$	Shape coefficient of soil moisture	–	–	0.632809	0.621915	10.000000	9.905496
	KUZ_r	Recession coefficient of the upper zone	1/day	–	1.000000	0.312415	0.294833	1.000000
	$ALFA_r$	non linear drainage coefficient of the upper zone	–	–	2.500000	2.374286	1.208677	1.217343
	$PERC_r$	Percolation from upper zone to the lower zone	–	–	0.022951	0.023132	0.001013	0.001000
	KLZ_r	Recession coefficient of the lower zone	1/day	–	0.089434	0.089466	0.020157	0.011136
	FC_g	Field capacity	m	–	0.854174	0.852845	0.149419	0.170029
	$BETA_g$	Shape coefficient of soil moisture	–	–	0.632809	0.621915	10.000000	9.905496
	KUZ_g	Recession coefficient of the upper zone	1/day	–	0.466374	1.000000	0.014564	0.048685
	$ALFA_g$	Non-linear drainage coefficient of the upper zone	–	–	2.500000	2.374286	1.208677	1.217343
	$PERC_g$	Percolation from upper zone to the lower zone	–	–	0.022951	0.023132	0.001013	0.001000
	KLZ_g	Recession coefficient of the lower zone	1/day	–	0.089434	0.089466	0.020157	0.011136

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. http://dx.doi.org/10.1016/j.jhydrol.2010.06.007.

Alley, W.M., 1984. On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models. Water Resour. Res. 20, 1137–1149. http://dx.doi.org/10.1029/WR020i008p01137.

Atkinson, S., Sivapalan, M., Woods, R., Viney, N., 2003. Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. Adv. Water Resour. 26, 219–235. http://dx.doi.org/10.1016/S0309-1708(02)00183-5.

Beldring, S., 2002a. Multi-criteria validation of a precipitation runoff model. J. Hydrol. 257, 189–211. http://dx.doi.org/10.1016/S0022-1694(01)00541-8.

Beldring, S., 2002b. Runoff generating processes in boreal forest environments with glacial tills. Nord. Hydrol. 33, 347–372. http://dx.doi.org/10.2166/nh.2002.021.

Beldring, S., Engeland, K., Roald, L.A., Slthun, N.R., Voks, A., 2003. Estimation of parameters in a distributed precipitation-runoff model for Norway. Hydrol. Earth Syst. Sci. 7, 304–316. http://dx.doi.org/10.5194/hess-7-304-2003.

Bergström, S., 1976. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. Technical Report 07 Swedish Meteorological and Hydrological Institute. <http://books.google.no/books?id=vRyeQAAACAAJ>.

Bergström, S., Sandberg, G., 1983. Simulation of groundwater response by conceptual models. Nord. Hydrol. 14, 71–84. http://dx.doi.org/10.2166/nh.1983.007.

Bookhagen, B., Burbank, D.W., 2010. Toward a complete Himalayan hydrological budget: spatiotemporal distribution of snowmelt and rainfall and their impact on river discharge. J. Geophys. Res.: Earth Surf. 115, F03019. http://dx.doi.org/10.1029/2009JF001426.

Boyle, D.P., Gupta, H.V., Sorooshian, S., Koren, V., Zhang, Z., Smith, M., 2001. Toward improved streamflow forecasts: value of semidistributed modeling. Water Resour. Res. 37, 2749–2759. http://dx.doi.org/10.1029/2000WR000207.

Bruland, O., Frevåg, Å., Steinsland, I., Liston, G.E., Sand, K., 2015. Weather SDM – a model for estimating snow density with high precision using snow depth and local climate. Hydrol. Res. http://dx.doi.org/10.2166/nh.2015.059 (in press).

Butts, M.B., Payne, J.T., Kristensen, M., Madsen, H., 2004. An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. J. Hydrol. 298, 242–266. http://dx.doi.org/10.1016/j.jhydrol.2004.03.042.

Corbari, C., Mancini, M., Su, Z., Li, J., August 2014. Evapotranspiration estimate from water balance closure using satellite data for the Upper Yangtze River basin. Hydrol. Res. 45, 603–614. http://dx.doi.org/10.2166/nh.2013.016.

Coron, L., Andréassian, V., Perrin, C., Lerat, J., Vaze, J., Bourqui, M., Hendrickx, F., 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. Water Resour. Res. 48, W05552. http://dx.doi.org/10.1029/2011WR011721.

Das, T., Bádossy, A., Zehe, E., He, Y., 2008. Comparison of conceptual model performance using different representations of spatial variability. J. Hydrol. 356, 106–118. http://dx.doi.org/10.1016/j.jhydrol.2008.04.008.

Doherty, J., 2005. PEST Model-Independent Parameter Estimation User Manual. Watermark Numerical Computing, fifth ed. <http://www.pesthomepage.org/Downloads.php>.

Doherty, J., Johnston, J.M., 2003. Methodologies for calibration and predective analysis of a watershed model 1. J. Am. Water Resour. Assoc. 39, 251–265. http://dx.doi.org/10.1111/j.1752-1688.2003.tb04381.x.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall–runoff models. Water Resour. Res. 28, 1015–1031. http://dx.doi.org/10.1029/91WR02985.

Ehret, U., Götzinger, J., Bárdossy, A., Pegram, G.G., 2008. Radar-based flood forecasting in small catchments, exemplified by the Goldersbach catchment, Germany. Int. J. River Basin Manage. 6, 323–329. http://dx.doi.org/10.1080/15715124.2008.9635359.

EI Hai EI Tahir, M., Wang, W., Xu, C.-Y., Zhang, Y., Singh, V., 2012. Comparison of methods for estimation of regional actual evapotranspiration in data scarce regions: Blue Nile Region, Eastern Sudan. J. Hydrol. Eng. 17, 578–589. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000429.

Engelhardt, M., Schuler, T.V., Anderson, L.M., 2012. Evaluation of gridded precipitation for Norway using glacier mass-balance measurements. Geografiska Ann.: Ser. A, Phys. Geogr. 94, 501–509. http://dx.doi.org/10.1111/j.1468-0459.2012.00473.x.

Engelhardt, M., Schuler, T.V., Andreassen, L.M., 2013. Glacier mass balance of Norway 1961–2010 calculated by a temperature-index model. Ann. Glaciol. 54, 32–40. http://dx.doi.org/10.3189/2013AoG63A245.

EROS, 1996. Hydro1k Elevation Derivative Database. <http:s://lta.cr.usgs.gov/HYDRO1K>.

Fekete, B.M., Vörösmarty, C.J., Lammers, R.B., 2001. Scaling gridded river networks for macroscale hydrology: development, analysis, and control of error. Water Resour. Res. 37, 1955–1967. http://dx.doi.org/10.1029/2001WR900024.

Fleig, A.K., 2013. Norwegian Hydrological Reference Dataset for Climate Change Studies. Technical Report 02 Norwegian Water Resources and Energy Directorate. <http://webby.nve.no/publikasjoner/rapport/2013/rapport201302.pdf>.

Gao, G., Xu, C.-Y., Chen, D., Singh, V., 2012. Spatial and temporal characteristics of actual evapotranspiration over Haihe River basin in China. Stoch. Environ. Res. Risk Assess. 26, 655–669. http://dx.doi.org/10.1007/s00477-011-0525-1.

Gisnås, K., Etzelmüller, B., Farbrot, H., Schuler, T.V., Westermann, S., 2013. CryoGRID 1.0: Permafrost distribution in Norway estimated by a spatial numerical model. Permafrost Periglac. Process. 24, 2–19. http://dx.doi.org/10.1002/ppp.1765.

Gong, L., Widén-Nilsson, E., Halldin, S., Xu, C.-Y., 2009. Large-scale runoff routing with an aggregated network-response function. J. Hydrol. 368, 237–250. http://dx.doi.org/10.1016/j.jhydrol.2009.02.007.

Hailegeorgis, T.T., Alfredsen, K., 2015. Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. Hydrol. Res. http://dx.doi.org/10.2166/nh.2014.051 (in press).

Han, J., Huang, G., Zhang, H., Li, Z., Li, Y., 2014. Effects of watershed subdivision level on semi-distributed hydrological simulations: case study of the SLURP model applied to the Xiangxi River watershed, China. Hydrol. Sci. J. 59, 108–125. http://dx.doi.org/10.1080/02626667.2013.854368.

Harlin, J., Kung, C., 1992. Parameter uncertainty and simulation of design floods in Sweden. J. Hydrol. 137, 209–230. http://dx.doi.org/10.1016/0022-1694(92)90057-3.

Jiang, T., Chen, Y.D., Xu, C.-Y., Chen, X., Chen, X., Singh, V., 2007. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. J. Hydrol. 336, 316–333. http://dx.doi.org/10.1016/j.jhydrol.2007.01.010.

Krysanova, V., Bronstert, A., Müller-Wohlfeil, D.-I., 1999. Modelling river discharge for large drainage basins: from lumped to distributed approach. Hydrol. Sci. J. 44, 313–331. http://dx.doi.org/10.1080/02626669909492224.

L'Abée-Lund, J., Eie, J., Faugli, P., Haugland, S., Hvidsten, N., Jensen, A., Melvold, K., Pettersen, V., Saltveit, S., 2009. Rivers of Europe. Rivers in Boreal Uplands. Academic Press, pp. 577–606. <http://www.sciencedirect.com/science/book/9780123694492> (chapter).

Lawrence, D., Haddeland, I., Langsholt, E., 2009. Calibration of HBV Hydrological Models Using PEST Parameter Estimation. Technical Report 01 Norwegian Water Resources and Energy Directorate. <http://www.nve.no/Global/Publikasjoner/Publikasjoner>.

Li, H., Beldring, S., Xu, C.-Y., 2014. Implementation and testing of routing algorithms in the distributed Hydrologiska Byråns Vattenbalansavdelning model for mountainous catchments. Hydrol. Res. 45, 322–333. http://dx.doi.org/10.2166/nh.2013.009.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201, 272–288. http://dx.doi.org/10.1016/S0022-1694(97)00041-3.

Mayr, E., Hagg, W., Mayer, C., Braun, L., 2013. Calibrating a spatially distributed conceptual hydrological model using runoff, annual mass balance and winter mass balance. J. Hydrol. 478, 40–49. http://dx.doi.org/10.1016/j.jhydrol.2012.11.035.

Michaud, J., Sorooshian, S., 1994. Comparison of simple versus complex distributed runoff models on a midsized semiarid watershed. Water Resour. Res. 30, 593–605. http://dx.doi.org/10.1029/93WR03218.

Mohr, M., 2008. New Routines for Gridding of Temperature and Precipitation Observations for seNorge. no. Technical Report 08 Norwegian Meteorological Institute. <http://met.no/Forskning/Publikasjoner/Publikasjoner2008/filestore/NewRoutinesforGriddingofTemperature.pdf>.

Montanari, A., Baldassarre, G.D., 2013. Data errors and hydrological modelling: the role of model structure to propagate observation uncertainty. Adv. Water Resour. 51, 498–504. http://dx.doi.org/10.1016/j.advwatres.2012.09.007.

Nash, J., Sutcliffe, J., 1970. River flow forecasting through conceptual models part I-A discussion of principles. J. Hydrol. 10, 282–290. http://dx.doi.org/10.1016/0022-1694(70)90255-6.

NVE. The National River Network Database (ELVIS). <http://www.nve.no/en/Water/NVEs-geographic-databases/The-National-River-Network-Database-ELVIS/> (accessed in 02.15).

O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from digital elevation data. Comput. Vis., Graph., Image Process. 28, 323–344. http://dx.doi.org/10.1016/S0734-189X(84)80011-0.

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. http://dx.doi.org/10.1016/S0022-1694(00)00393-0.

Petersen-verleir, A., Soot, A., Reitan, T., 2009. Bayesian rating curve inference as a streamflow data quality assessment tool. Water Resour. Manage 23, 1835–1842. http://dx.doi.org/10.1007/s11269-008-9354-5.

Praskievicz, S., Chang, H., 2009. A review of hydrological modelling of basin-scale climate change and urban development impacts. Prog. Phys. Geogr. 33, 650–671. http://dx.doi.org/10.1177/0309133309348098.

Pushpalatha, R., Perrin, C., Moine, N.L., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. 420–421, 171–182. http://dx.doi.org/10.1016/j.jhydrol.2011.11.055.

Reed, S., Koren, V., Smith, M., Zhang, Z., Moreda, F., Seo, D.-J., Participants, D., 2004. Overall distributed model intercomparison project results. J. Hydrol. 298, 27–60. http://dx.doi.org/10.1016/j.jhydrol.2004.03.031.

Slthun, N.R. 1996. The Nordic HBV model. Technical Report 07 Norwegian Water Resources and Energy Directorate. <http://folk.uio.no/nilsroar/gf247/hbvmod.pdf>.

Seibert, J., 1997. Estimation of parameter uncertainty in the HBV model. Nord. Hydrol. 28, 247–262. http://dx.doi.org/10.2166/nh.1997.015.

Seibert, J., 1999. Regionalisation of parameters for a conceptual rainfall–runoff model. Agric. For. Meteorol. 98–99, 279–293. http://dx.doi.org/10.1016/S0168-1923(99)00105-7.

Singh, V., Woolhiser, D., 2002. Mathematical modeling of watershed hydrology. J. Hydrol. Eng. 7, 270–292. http://dx.doi.org/10.1061/(ASCE)1084-0699(2002)7:4(270).

Skahill, B.E., Doherty, J., 2006. Efficient accommodation of local minima in watershed model calibration. J. Hydrol. 329, 122–139. http://dx.doi.org/10.1016/j.jhydrol.2006.02.005.

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. http://dx.doi.org/10.1016/j.jhydrol.2004.03.040.

Sorteberg, H., Engeset, R., Udnas, H., 2001. A national network for snow monitoring in Norway: snow pillow verification using observations and models. Phys. Chem. Earth, Part C: Solar, Terrestr. Planet. Sci. 26, 723–729. http://dx.doi.org/10.1016/S1464-1917(01)95016-0.

Strahler, A.N., 1957. Quantitative analysis of watershed geomorphology. Trans., Am. Geophys. Union 38, 913–920. http://dx.doi.org/10.1029/TR038i006p00913.

Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrol. Earth Syst. Sci. 11, 793–817. http://dx.doi.org/10.5194/hess-11-793-2007.

Todini, E., 2007. A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach. Hydrol. Earth Syst. Sci. 11, 1645–1659. http://dx.doi.org/10.5194/hess-11-1645-2007.

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. Hydrol. Sci. J. 44, 779–797. http://dx.doi.org/10.1080/02626669909492273.

Vaché, K.B., McDonnell, J.J., 2006. A process-based rejectionist framework for evaluating catchment runoff model structure. Water Resour. Res. 42, W02409. http://dx.doi.org/10.1029/2005WR004247.

Varado, N., Braud, I., Galle, S., Le Lay, M., Seguis, L., Kamagate, B., Depraetere, C., 2006. Multi-criteria assessment of the Representative Elementary Watershed approach on the Donga catchment (Benin) using a downward approach of model complexity. Hydrol. Earth Syst. Sci. 10, 427–442. http://dx.doi.org/10.5194/hess-10-427-2006.

Vormoor, K., Skaugen, T., 2013. Temporal disaggregation of daily temperature and precipitation grid data for Norway. J. Hydrometeorol. 14, 989–999. http://dx.doi.org/10.1175/JHM-D-12-0139.1.

Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrol. Earth Syst. Sci. 5, 13–26. http://dx.doi.org/10.5194/hess-5-13-2001.

Wrede, S., Seibert, J., Uhlenbrook, S., 2013. Distributed conceptual modelling in a Swedish lowland catchment: a multi-criteria model assessment. Hydrol. Res. 44, 318–333. http://dx.doi.org/10.2166/nh.2012.056.

Xu, C.-Y., Chen, D., 2005. Comparison of seven models for estimation of evapotranspiration and groundwater recharge using lysimeter measurement data in Germany. Hydrol. Process. 19, 3717–3734. http://dx.doi.org/10.1002/hyp.5853.

Xu, C.-Y., Singh, V., 2005. Evaluation of three complementary relationship evapotranspiration models by water balance approach to estimate actual regional evapotranspiration in different climatic regions. J. Hydrol. 308, 105–121. http://dx.doi.org/10.1016/j.jhydrol.2004.10.024.

Xu, C.-Y., Widén, E., Halldin, S., 2005. Modelling hydrological consequences of climate change–progress and challenges. Adv. Atmos. Sci. 22, 789–797. http://dx.doi.org/10.1007/BF02918679.

Yan, C., Zhang, W., 2014. Effects of model segmentation approach on the performance and parameters of the Hydrological Simulation Program Fortran (HSPF) models. Hydrol. Res. 45, 893–907. http://dx.doi.org/10.2166/nh.2014.182.

Yang, X., Ren, L., Singh, V., Liu, X., Yuan, F., Jiang, S., Yong, B., 2012. Impacts of land use and land cover changes on evapotranspiration and runoff at Shalamulun River watershed, China. Hydrol. Res. 43, 23–37. http://dx.doi.org/10.2166/nh.2011.120.