Published in "Water Resources Research 51(4): 1939–1958, 2015"
which should be cited to refer to this work.

The value of multiple data set calibration versus model complexity for improving the performance of hydrological models in mountain catchments

David Finger[1,2,3], Marc Vis[3], Matthias Huss[4], and Jan Seibert[3,5]

[1]School of Science and Engineering, Reykjavik University, Reykjavik, Iceland, [2]Icelandic Meteorological Office, Reykjavik, Iceland, [3]Department of Geography, University of Zurich, Switzerland, [4]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland, [5]Department of Earth Sciences, Uppsala University, Sweden

Correspondence to:
D. Finger,
fingerd@gmx.net

Abstract The assessment of snow, glacier, and rainfall runoff contribution to discharge in mountain streams is of major importance for an adequate water resource management. Such contributions can be estimated via hydrological models, provided that the modeling adequately accounts for snow and glacier melt, as well as rainfall runoff. We present a multiple data set calibration approach to estimate runoff composition using hydrological models with three levels of complexity. For this purpose, the code of the conceptual runoff model HBV-light was enhanced to allow calibration and validation of simulations against glacier mass balances, satellite-derived snow cover area and measured discharge. Three levels of complexity of the model were applied to glacierized catchments in Switzerland, ranging from 39 to 103 km^2. The results indicate that all three observational data sets are reproduced adequately by the model, allowing an accurate estimation of the runoff composition in the three mountain streams. However, calibration against only runoff leads to unrealistic snow and glacier melt rates. Based on these results, we recommend using all three observational data sets in order to constrain model parameters and compute snow, glacier, and rain contributions. Finally, based on the comparison of model performance of different complexities, we postulate that the availability and use of different data sets to calibrate hydrological models might be more important than model complexity to achieve realistic estimations of runoff composition.

1. Introduction

The contribution of snow melt, glacier, melt and rain to runoff in mountain streams is of major importance for water resource management as climate variability and change is expected to impact on all three components [e.g., *Crochet*, 2013; *Kumar et al*., 2007; *Rathore et al*., 2009]. While glaciers are retreating worldwide, the snow cover duration in winter becomes shorter and precipitation events are expected to intensify [*IPCC*, 2013]. Besides climate change impact studies, hydrologic models are also used for a wide variety of practical purposes, such as flood forecasting, environmental impact assessments, and seasonal water availability estimations to mention just a few. For such purposes, hydrological models can provide a realistic estimate of the contribution of snow, glacier, and rainfall runoff in mountain streams, provided that they have been calibrated and validated adequately. Yet hydrological modeling faces two main challenges [*Grayson et al*., 2002]: (i) uncertainty in observational data for calibration and validation purposes and (ii) model uncertainty due to the simplification of natural processes expressed in model structure and parameter uncertainty.

Tremendous progress has been achieved over the last decades in making observational data more accurate and reliable. Accuracy of discharge measurements has been improved using remotely controlled gauging stations equipped with current profilers [*Muste et al*., 2004], spatially distributed meteorological patterns are observed with satellites and weather radars [*Borga*, 2002; *Xie and Arkin*, 1996], and catchment characteristics of land cover and soil properties have been mapped worldwide [*McBratney et al*., 2003]. Nevertheless, for remote headwaters, data availability is frequently limited and subject to uncertainty due to icing, intense snow fall, and channel instability at the gauging station. In these areas estimations of snow, glacier, and rainfall, runoff contribution frequently have to be based on limited data availability, sometimes relying on only 1 year of accurate data.

Likewise to the progress in collecting observational data sets intense research has been conducted in optimizing the complexity of hydrological models. Model complexity refers to the level of detail in process representations [e.g., *Grayson et al*., 1992; *Johnson et al*., 2003; *Vrugt et al*., 2002], the spatial discretization of a catchment [e.g., *Kirnbauer et al*., 1994; *Refsgaard and Knudsen*, 1996; *van der Linden and Woo*, 2003], or a combination of these two aspects. Over 20 years ago *Jakeman and Hornberger* [1993] already investigated the level of complexity necessary for accurate rainfall-runoff simulation, basing their analysis mainly on discharge efficiency. A particular challenge during calibration of hydrological models is the equifinality as discussed in numerous studies [e.g., *Beven*, 1996; *Beven and Binley*, 1992; *Shen et al*., 2012]. Accordingly, many authors have investigated possibilities to use additional data sets to constrain model parameters [e.g., *Ambroise et al*., 1995; *Kuczera and Mroczkowski*, 1998; *Refsgaard*, 1997]. Nevertheless, investigations of model performance of rainfall runoff models used in different countries indicate that the calibration method is more important than model complexity [*Gan et al*., 1997]. *Perrin et al*. [2001] concluded that models have been developed with excessive confidence and that model structure is not always able to extract information from available runoff time series. Further attempts to reduce the equifinality by including additional processes into hydrological models have exacerbated the equifinality problem due to the addition of more parameters that require calibration [*Beven*, 2006]. *Kirchner* [2006] argued that scientific progress should focus on the customization of data availability to theory, rather than toward increased model complexity. Only recently *McMillan et al*. [2011] suggested that different sources of field data should be used to optimize the hydrological processes within a hydrological model. Indeed, when lumped calibration strategies are used, semi-distributed models yield higher performance than fully distributed models [*Khakbaz et al*., 2012]. Hence, the discussion on appropriate model complexity for runoff modeling is still ongoing [*Cunderlik et al*., 2013].

Research on estimations of snow, ice, and rain runoff contributions to discharge in mountainous regions has focused on using multiple data sets to enhance the consistent estimation regarding different water sources. *Parajka and Blöschl* [2008] showed that the additional use of satellite snow cover images during calibration can improve both snow cover and discharge simulations. *Fleming et al*. [2010] used glacier equilibrium line observations to constrain snow melt-glacier melt partitioning at high elevations, *Nolin et al*. [2010] constrained a runoff model using stable isotope data and glacier melt observations, and *Schaefli and Huss* [2011] used seasonal point glacier mass balances to calibrate a conceptual model. Along this line of research *Konz and Seibert* [2010] employed annual mass balances, *Jost et al*. [2012a] used repeated glacier mapping and *Mayr et al*. [2013] used seasonal and winter mass balances to constrain the model parameters of a conceptual model. *Koboltschnig et al*. [2008] validated the runoff contribution computed with a conceptual model in a glacierized Alpine catchment using discharge, snow cover images, and glacier mass balances. *Finger et al*. [2011] demonstrated that the combined use of discharge, snow cover images, and seasonal glacier mass balances can constrain model parameters of a physically based fully distributed hydrological model compared to calibration against discharge only, reducing the equifinality significantly. In particular, satellite snow cover images have increasingly been used in recent years and across the world to constrain hydrological models [*Duethmann et al*., 2014; *Finger et al*., 2012; *Franz and Karsten*, 2013; *Pellicciotti et al*., 2012]. Nevertheless, the question as to whether conceptual lumped models can be consistently calibrated regarding glacier mass balances, snow cover images, and discharge to estimate snow, ice, and rain runoff contribution to discharge in mountain streams remains unanswered.

In this study, we assess how 1 year of daily snow cover images, seasonal glacier mass balance data, and daily runoff can be used to improve the estimation of snow, glacier, and rain contribution in mountain streams. By using three levels of complexity of a conceptual lumped hydrological model, we complement the results of a previous study using a physically based, fully distributed hydrological model [*Finger et al*., 2011]. We demonstrate the added value of optimizing model performance with the three data sets in regard to the three model complexity levels. Hence, this study complements previous studies and concludes that more reliable prediction of snow, glacier, and rain contribution to runoff can be achieved if all three calibration data sets are weighted equally during calibration.

2. Study Sites and Data

We chose three Swiss alpine streams to test our modeling approach (Figure 1): (i) Rhone River at the gauging station Gletsch, (ii) Hinterrhein River at the village of Hinterrhein, and (iii) Landquart River close to the

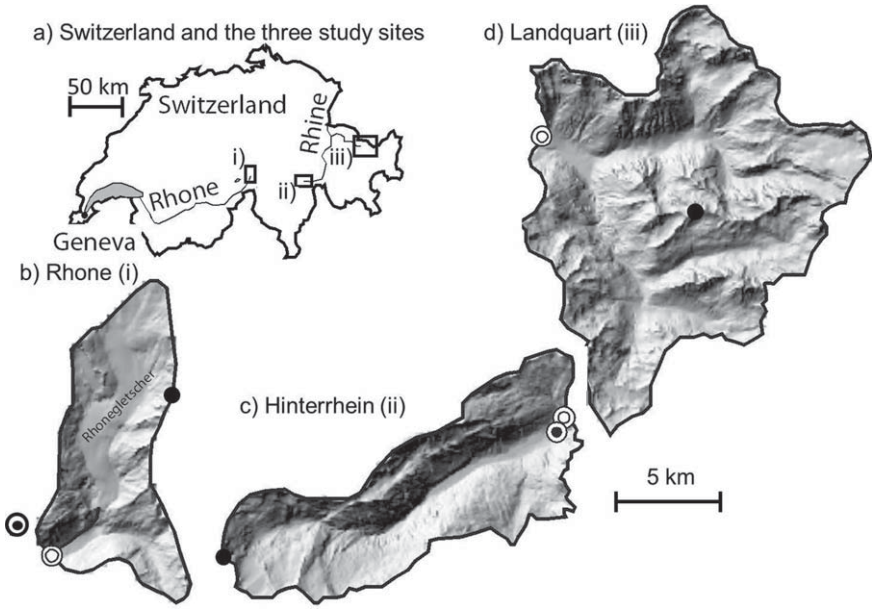

Figure 1. Overview and map of the three study sites: (a) locates the three catchments, Rhone River at Gletsch (i), Hinterrhein River at Hinterrhein (ii), and Landquart River at Klosters (iii) within Switzerland; (b), (c), and (d) give an overview of the three catchments. Black dots locate prominent landmark peaks in the three catchments (in Figure 1b: Galenstock with 3586 m asl; in Figure 1c: Rheinwaldhorn with 3402 m asl; in Figure 1d: Roggenhorn with 2891 m asl), double circles indicate gauging station and circle with black dot locate meteorological station.

village of Klosters. The Rhonegletscher at the source of the Rhone has been investigated since the end of the 19th century making it an ideal case study to test novel modeling approaches [*Finger et al*., 2011; *Huss et al*., 2008; *Klok et al*., 2001; *Verbunt et al*., 2003]. In 2010, the Rhonegletscher and several smaller glaciers covered 42.3% of the catchment area (Table 1). The gauging station at Gletsch operated by the Swiss Federal Office of Environment (FOEN) is located about 2 km southwest of the glacier terminus gauging the discharge from a 38.9 km^2 catchment area. An automatic weather station operated since 1990 by the Federal Office of Meteorology and Climatology (MeteoSwiss) is located about 3.5 km west of the glacier. The second investigated catchment lies in eastern Switzerland and contains the source of the Hinterrhein. The gauging station operated by FOEN gauges an area of 53.7 km^2, of which 6.3% was glacierized in 2009 [*Fischer et al*., 2014]. Meteorological data are available from an automatic weather station operated by MeteoSwiss about 1.7 km southeast of the gauging station (Table 1). The third catchment lies at the border of Switzerland and Austria and contains the Silvrettagletscher, the source of the Landquart River. The gauging station close to

Table 1. Summary of Catchment Characteristics of the Three Study Sites

River	Rhone	Hinterrhein	Landquart
Gauging station	Gletsch	Hinterrhein	Klosters, Auelti
[a](Location CH 1903)	670810 / 157200	735480 / 154680	790480 / 192690
Data availability	1903–present	[b]2001–2004	[b]2001–2004
Catchment Area	38.9 km^2	53.7 km^2	103.0 km^2
Lowest altitude	1761 m asl	1584 m asl	1317 m asl
Mean altitude	2719 m asl	2360 m asl	2332 m asl
Highest elevation	3630 m asl	3402 m asl	3410 m asl
[c]Glacierization	42.3% (2010)	6.3% (2009)	4.5% (2008)
[d]Mean discharge	2.8 $m^3 s^{-1}$	3.4 $m^3 s^{-1}$	5.3 $m^3 s^{-1}$
	2270 mm a^{-1}	1997 mm a^{-1}	1623 mm a^{-1}
Weather station	Grimsel Hospiz	Hinterrhein	Davos
Location	668583/158215	733900/153980	783514/187457
Dist. to stream gauging	2.4 km (outside catchment)	1.7 km	8.7km (outside catchment)
Source for glacier mass balances	*Huss et al.* [2008]	*Huss et al*. [2010]	*Huss et al*. [2009]

[a]Coordinates are given in CH1905 System.
[b]An avalanche damaged the gauging station in Hinterrhein and monitoring was suspended in 2004 at Klosters.
[c]Catchment glacierization according to the latest Swiss Glacier Inventory [*Fischer et al*., 2014]. The corresponding year is given in brackets.
[d]According to the FOEN station data.

Klosters operated by FOEN gauges an area of 103 km^2, whereof 4.5% was glacierized in 2008. The closest automatic weather station operated by MeteoSwiss is located 8.7 km southwest of the gauging station. Due to the high availability of observational data, these study sites are ideal to test and validate our modeling concept. In Table 1, the catchment characteristics of the three study sites are summarized. For the Hinterrhein and the Landquart Rivers, discharge data are only available until 2004, when monitoring was suspended.

Daily satellite snow cover images are available since 2001 from the Moderate Resolution Imaging Spectroradiometer (MODIS, product MOD10A1.5 available at http://nsidc.org/) [*Hall et al*., 2002]. In this study, we used all days with less than 10% cloud cover which is an optimum tradeoff between uncertainty of fractional snow cover observations ($<$ 0.05 uncertainty corresponding to half of the obscured area) and number of available satellite images (on average 84 days per year had less than 10% cloud cover) (Figure 2). Hence, the average period with obscuration due to cloud cover was only about 4.3 days long, making an interpolation between observations acceptable. Details about the interpolation method are described in *Glaus* [2013]. The topography for all three catchments was obtained from a digital elevation model with 250 m grid size provided by the Swiss Federal Office of Topography [*swisstopo*, 2004]. Three vegetation zones (forest, grassland, and areas without vegetation) were identified based on digital land cover maps from the Swiss Federal Statistical Office (FSO).

Glacier mass balance data for the Rhonegletscher were obtained from *Huss et al*. [2008] based on a combination of seasonal direct observations and modeling to extrapolate the variables to the entire glacier. Mass balance of glaciers in the Hinterrhein catchment were generated by extrapolating temporal variability from several nearby glaciers provided by *Huss et al*. [2010] and combining these results with observed decadal ice volume changes of the Hinterrhein glaciers. Mass balance data of the Silvrettagletscher are available from homogenized direct observations [*Huss et al*., 2009]. All glacier mass balance data cover the accumulation season (1 October to 30 April) and the ablation season (30 April to 1 October) and resolve the distribution of accumulation and ablation in 100 m elevation bands.

3. Modeling Approach and Multiple Data Set Calibration

3.1. The HBV-Light Model

The Hydrologiska Byråns Vattenbalansavdelning model (HBV) is a conceptual runoff model originally developed by *Bergström* [1976, 1992]. The HBV model has been widely used in northern Europe [e.g., *Seibert*, 1999; *Steele-Dunne et al*., 2008] and other regions of the world [e.g., *Cunderlik et al*., 2013; *Krysanova et al*., 1999; *Razavi and Coulibaly*, 2013]. Here we use the software implementation HBV-light [*Seibert and Vis*, 2012], which includes a glacier routine described in *Konz and Seibert* [2010]. In this model version, a watershed is represented by the area fractions of aspect and vegetation classes for different elevation zones. For each specific zone, hydrological processes are computed separately. The model simulates catchment discharge with a daily resolution using time series of precipitation, air temperature as well as estimates of monthly long-term potential evaporation rates. Snow accumulation is derived from extrapolated precipitation below a temperature threshold and snow and ice melt are computed by a temperature-index model [e.g., *Hock*, 2003]. Groundwater recharge and actual evaporation are simulated as functions of actual water storage in the soil routine. In the groundwater routine runoff is computed as a function of water storage in two groundwater reservoirs. Runoff from the lumped reservoirs is determined with a triangular weighting function to simulate the effect of channel routing on the arrival of stream flow at the gauging station. All model parameters are summarized in Table 2 and a detailed description of the model parameters and their functions is given in *Seibert and Vis* [2012].

Furthermore, for this study, the glacier routine of the HBV-light was enhanced with a nonlinear discharge coefficient depending on snowpack water equivalent on the glacier so that the glacial water storage-outflow relationship varies over time to represent the seasonal development of the subglacial drainage system [*Stahl et al*., 2008]. Within glacierized areas, 0.1% of the snowpack is converted to ice in every daily time step. Thus, snow not melted during the summer season is transformed into glacier ice within a few years, which corresponds to observations on Alpine glaciers. The output routines of the HBV-light software were updated to generate typical runoff component results as defined in previous studies [*Radic and Hock*, 2014]: (i) total glacier outflow, G_{out} (includes ice melt, snow melt, and rain over glacierized area), (ii) total rainwater infiltration into the soil, I_{soil} (accounting for infiltration only in snow and glacier-free areas), (iii) snow melt,

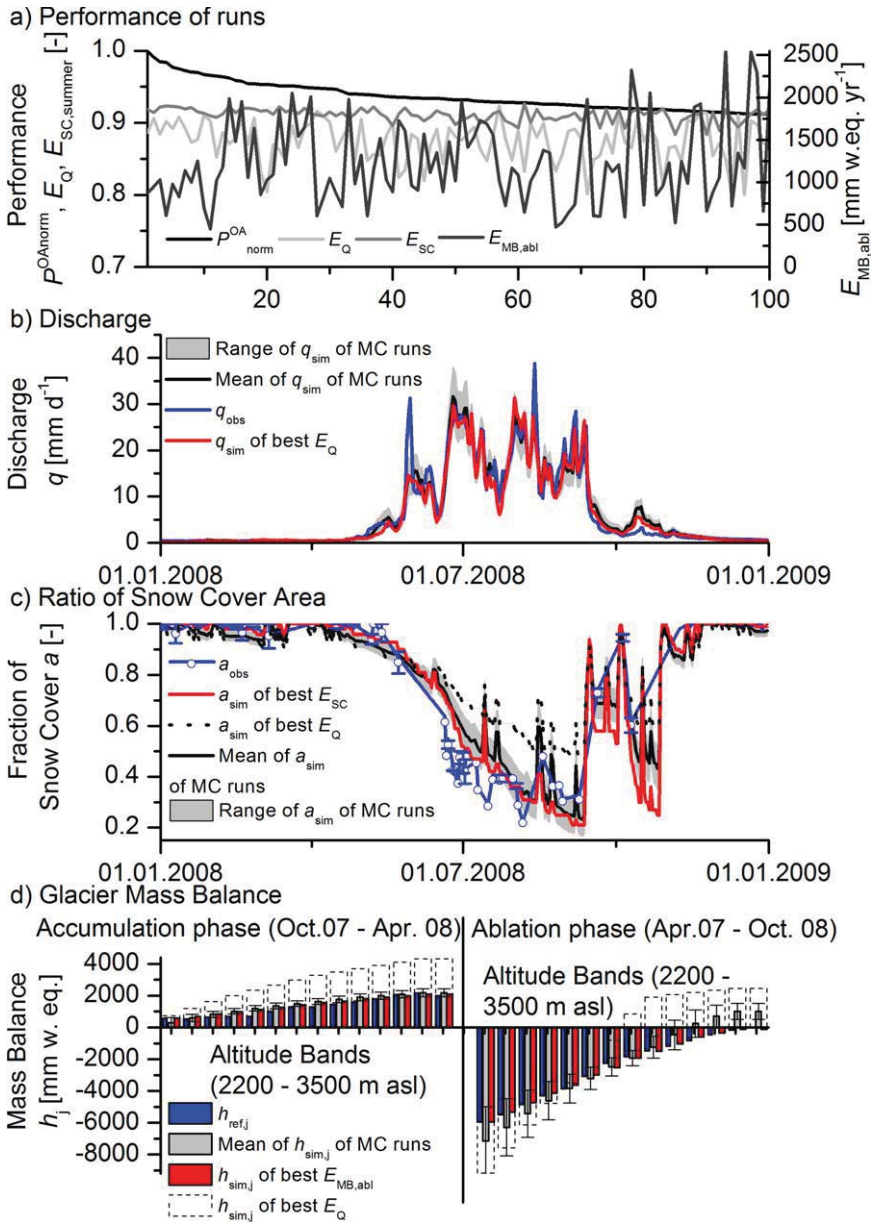

Figure 2. Performance of the 100 best MC runs (N = 10,000) using HBV3 regarding best overall consistency performance, POAnorm. (a) illustrates the efficiencies of the 100 best runs regarding POAnorm, EQ, ESC,summer and EMB,abl. (b), (c), and (d) illustrate simulated and observed discharge, fraction of snow cover area, and glacier mass balances. Black lines and gray bars indicate mean of the best 100 runs, gray area indicates range of the 100 best runs, red lines and bars indicate best simulation within the ensemble, and the dashed line and the bar illustrate runs with best performance regarding Q. In Figure 2b open circles indicate days with less than 10% cloud cover and whiskers illustrate the respective uncertainty due to cloud cover.

Q_{snow} (comprises all snow also on glacierized area), (iv) ice melt, Q_{ice} (comprises only bare-ice melt of glaciers), and rainfall runoff, (v) Q_{rain} (includes rain fall on the ground, on snow and glacier), at the gauging station. These updates allow a quantification of the contribution of the different water sources to the runoff.

3.2. Complexity of Model Setups

To investigate the value of model complexity, we set up the HBV-light model for the Rhone catchment with three levels of complexity (Table 3). In the simplest HBV setup (hereafter labeled HBV1), the Rhone catchment is divided into 18 elevation zones with 100 m vertical spacing, without taking into account aspect or vegetation cover. In the second HBV-light setup (hereafter labeled HBV2), we divided each elevation zone

Table 2. Overview of Model Parameters of the HBV3 Model

Parameter	[a]Description	Units	Min	Max	Rhone Mean	Rhone Std	Hinterrhein Mean	Hinterrhein Std	Landquart Mean	Landquart Std
Rescaling parameters of input data										
P_{PCALT}	Change of precipitation with elevation	% $(100m)^{-1}$	5	15	7.31	1.77	7.27	1.65	12.37	1.91
P_{TCALT}	Change of temperature with elevation	°C $(100m)^{-1}$	0.5	1.5	0.92	0.15	0.81	0.15	0.66	0.12
Snow and ice melt parameters										
P_{TT}	Threshold temperature for liquid and solid precipitation.	°C	3	1	1.68	0.87	0.29	0.81	0.67	0.91
P_{CFMAX}	Degree-day factor	mm $d^{-1}°C^{-1}$	1.5	10	7.05	1.81	8.26	1.27	5.78	1.78
P_{SFCF}	Snowfall correction factor	-	0.8	1.2	0.93	0.09	0.98	0.10	1.07	0.10
P_{CFR}	Refreezing coefficient	-	0.02	0.1	0.06	0.02	0.06	0.02	0.06	0.03
P_{CWH}	Water holding capacity of the snow storage	-	0.1	0.4	0.26	0.09	0.23	0.09	0.26	0.09
$P_{CFGlacier}$	Glacier melt correction factor	-	0.3	3	0.88	0.50	1.19	0.72	0.90	0.41
[b]$P_{CFSlope}$	Slope snow melt correction factor	-	0.3	3	1.30	0.74	1.88	0.65	1.68	0.42
[c]P_{Kgmin}	Minimum value for the outflow coefficient representing conditions with poorly developed glacial drainage systems in late winter	-	0.01	0.2	0.11	0.06	0.09	0.06	0.10	0.06
[c]$P_{RangeKG}$	Range of the annual outflow coefficient variation	-	0.01	0.5	0.25	0.14	0.25	0.13	0.23	0.14
[c]P_{AG}	Calibration parameter defining the sensitivity of the outflow coefficient to changes in the snow storage	-	0	0.1	0.05	0.03	0.05	0.03	0.50	0.31
Soil parameters										
P_{PERC}	Maximum percolation from upper to lower groundwater storage	mm d^{-1}	0	4	2.07	0.99	2.36	1.03	2.24	0.99
P_{K0}	Storage (or recession) coefficient 0	d^{-1}	0.1	0.5	0.29	0.11	0.31	0.11	0.26	0.12
P_{K1}	Storage (or recession) coefficient 1	d^{-1}	0.01	0.2	0.10	0.05	0.10	0.06	0.09	0.06
P_{K2}	Storage (or recession) coefficient 2	d^{-1}	5E-05	0.1	0.04	0.03	0.02	0.02	0.02	0.02
P_{MAXBAS}	Length of triangular weighting function	d	1	2.5	1.71	0.46	1.56	0.39	1.82	0.42
[d]P_{FC}	Maximum soil moisture storage	mm	100	700	376	162	382	166	397	175
					416	175	415	176	377	166
					414	168	386	170	420	165
[d]P_{LP}	Relative soil water storage below which AET is reduced linearly	-	0.3	1	0.69	0.20	0.63	0.21	0.65	0.21
					0.68	0.20	0.65	0.18	0.62	0.20
					0.71	0.21	0.66	0.20	0.63	0.21
[d]P_{Beta}	Shape factor for the function used to calculate the distribution of rain and snow melt going to runoff and soil box, respectively	-	1	5	2.87	1.17	2.94	1.11	3.00	1.17
					3.17	1.14	2.87	1.12	2.99	1.14
					2.97	1.12	3.12	1.12	2.86	1.17

[a]A detailed description of model parameters is given in *Seibert and Vis* [2012].
[b]Slope factor correcting P_{CFMAX} accounting for dependency of melt rates on aspect of topography.
[c]Glacier parameters to according to *Stahl et al*. [2008].
[d]Parameters which are vegetation specific; accordingly for the HBV3 version, these parameters were optimized for the three vegetation zone: (i) forests, (ii) grassland, and (iii) without vegetation.

into three aspect classes, differentiating between south, east-west, and north facing slopes. Accordingly, in the HBV2 setup, an additional slope correction factor, $P_{CFSlope}$, had to be calibrated (Table 3).

Furthermore, an enhanced setup accounting for three vegetation zones (forests, grassland, and areas without vegetation; thereafter named HBV3) was applied to all three catchments. In the HBV3 setup, each vegetation zone is described by specific soil parameters, adding three model parameters per vegetation zone. Accordingly, the maximum soil moisture storage, P_{FC}, soil moisture at maximum actual evaporation, P_{LP}, and the empirical scaling factor, P_{Beta}, had to be calibrated additionally (Table 3).

Some results of the Rhone catchment can directly be compared to a previous study performed with the physically based, fully distributed Topographic Kinematic Approximation and Integration model (TOPKAPI) calibrated with the same data sets [*Finger et al*., 2011]. TOPKAPI was originally developed by the University

Table 3. Summary of the Three Different HBV Model Setups

	HBV1	HBV2	HBV3
Aspect zone	1	3	3
Vegetation zones	1	1	3
[a]Number of parameters	19	20	26
[b]Additional parameters	-	$P_{CFSlope}$	$P_{CFSlope}$, P_{FC}, P_{LP}, P_{Beta}

[a]Number of model parameters to be calibrated.
[b]Additional parameters compared to HBV1. The parameters P_{FC}, P_{LP}, and P_{Beta} are vegetation specific (see Table 2 for details).

of Bologna and is based on the conceptual ARNO model [*Todini*, 1996]. Besides being able to investigate the spatial distribution of the snow cover, the higher complexity of TOPKAPI allowed *Finger et al*. [2011] to perform simulations at hourly resolution rather than daily time steps usually used in the HBV model. This is also reflected by *Micovic and Quick* [2009] who demonstrated that process representations and model parameters that appear unimportant during the long-term simulation have significant effects on the short-term extreme event model simulation.

3.3. Calibration With Discharge, Snow Cover Images and Glacier Mass Balances

To determine an ensemble of model parameter sets that leads to model results adequately simulating observed discharge, snow cover, and seasonal glacier mass balances, we used the same Monte Carlo (MC) calibration procedure as presented by *Finger et al*. [2011]. Accordingly, only a short summary is given here, outlining how the method was adapted to the HBV-light model. As discussed by *Finger et al*. [2011], the selection of 100 runs from 10,000 parameter sets generated using a uniform distribution of the values of each parameter (Table 2) are sufficient to obtain an adequate consent of model stability, model performance, and parameter variability. Hence, 10,000 parameter sets were randomly generated from a uniformly distributed physically constrained range determined by test runs and based on values found in previous studies [*Seibert*, 1999]. These parameter sets were applied to the HBV-light model in order to compute model efficiencies (see Table 4) regarding discharge (Q), snow cover area (SC), and glacier mass balances (MB) during a 1 year calibration period. Hence, for every run the ranking value for all six efficiency criteria, E_c, was computed by dividing the rank regarding a specific efficiency by the total number of runs. As defined by *Finger et al*. [2011], we used the overall consistency performance, P^{OA}, to quantify the simultaneous performance regarding all criteria considered. P^{OA} was obtained for all individual runs by averaging the ranking value of all six efficiencies, assuring equal weighting of all efficiencies considered. For visualization purposes, we normalized the overall consistency performance, P^{OAnorm}, by dividing P^{OA} by the best ranking value of all runs. By ranking the 10,000 runs according to P^{OAnorm}, an ensemble of the 100 best runs regarding all six E_c could be identified. Similar to *Finger et al*. [2011], the value of the three observational data sets (glacier mass balances, snow cover images, and discharge data) was assessed by comparing P^{OAnorm} of the 100 best runs selected according to the consistency performance of a specific criterion. To focus on the calibration during the ablation season, we used all six criteria only for the computation of P^{OAnorm}, but performed all other computations using only efficiency regarding discharge (Q, quantified using E_Q), summer mass balance (MB, quantified using $E_{MB,abl}$), and snow cover during summer (SC, quantified using $E_{SC,summer}$). To allow a consistent illustration of all efficiencies with increasing performance for higher values, $E_{MB,norm}$ was computed by normalizing $E_{MB,abl}$ to mean mass change of the glacier as defined in Table 4.

Table 4. The Six Efficiency Criteria Used to Evaluate Model Performance Regarding the Three Data Sets

Efficiency Criteria	[a]Opt.	Calibration Period	Equation
[b,c]Nash-Sutcliffe of Q, E_Q	Max	1 Jan to 31 Dec	$E_Q = 1 - \frac{\sum_{i=1}^{n}(q_{obs,i}-q_{sim,i})^2}{\sum_{i=1}^{n}(q_{obs,i}-\overline{q_{sim,i}})^2}$
[b,c]Nash-Sutcliffe of log (Q), E_{Qlog}	Max	1 Jan to 31 Dec	$E_{Qlog} = 1 - \frac{\sum_{i=1}^{n}(\log(q_{obs,i})-\log(q_{i,sim}))^2}{\sum_{i=1}^{n}(\log(q_{obs,i})-\overline{\log(q_{obs,i})})^2}$
[b,d]Root mean square error of mass balance, $E_{MB,acc}$	Min	1 Oct to 30 Apr (7 months)	$E_{MB} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(\Delta h_{ref,j}-\Delta h_{sim,j})^2}$
[b,d]Root mean square error of mass balance, $E_{MB,abl}$	Min	1 May to 30 Sep (5 months)	
[e]Normalized MB efficacy $E_{MB,norm}$	Max		$E_{MB,norm} = 1-(E_{MB,abl}/h_{ref,j,mean})$
[b,f]Correctly predicted snow cover area, $E_{SC,year}$	Max	1 Jan to 31 Dec	$E_{sc} = \frac{1}{n}\sum_{i=1}^{n}(1-\lvert a_{sim,i}-a_{obs,i}\rvert)$
[b,f]Correctly predicted snow cover area, $E_{SC,summer}$	Max	1 Apr to 1 Aug	

[a]Indicates if the criterion should be maximized (max) or minimized (min) during calibration.

[b]In order to compute P^{OAnorm} all six criteria were considered; for all other results only E_Q, $E_{MB,abl}$, and $E_{SC,summer}$ were considered.

[c]q_{obs} is observed daily discharge; q_{sim} is simulated daily discharge for time step *i*; to be consistent with *Finger et al* [2011] the calend.ar year was chosen for calibration rather than the water year.

[d]Δh is the combined change in snow and ice height in w. eq. during the indicated period for a specific 100 m altitude band *j*; indices ref and sim designate reference and simulated heights.

[e]$h_{ref,j,mean}$ is the mean of all Δh during the entire validation period.

[f]*a* is the daily area fraction covered by snow; index sim and obs stands, respectively, for estimations based on satellite images; index i stands for the time step and n stands for the number of days considered.

In principle, our calibration technique is comparable to a modified GLUE approach [*Beven and Binley*, 1992; *Beven and Freer*, 2001; *Freer et al*., 1996; *Juston et al*., 2009]. The main difference is that we use the rank of the runs to select the ensemble of acceptable runs, assuring equal weighting of all criteria considered (see discussion), rather than using thresholds and model parameter frequency analysis.

The calibration procedure was applied to the three catchments using data from a typical hydrological and meteorological year with close-to-average discharge, precipitation, and temperature relative to the first decade of the 21st century. Preliminary calibration runs revealed that choosing an average year yields best model performances for the remaining years. For the Rhone catchment, the year 2008 was chosen for calibration purposes. For the Hinterrhein and the Landquart catchment, the years 2002 and 2004, respectively, were chosen for calibration.

4. Results

To assess the value of daily discharge and snow cover images, as well as seasonal glacier mass balances to calibrate models of different complexity levels (HBV1, HBV2, and HBV3), we first present the performance of the HBV3 model in the Rhone catchment during the calibration period and then proceed by comparing the calibration performance of the three HBV model setups (section 5.1). Subsequently, we present the results for the validation periods (section 5.2) and show the computed contribution of snow, glacier, and rain to total runoff (section 5.3). Finally, we demonstrate that the results of the Rhone catchment are consistent with results from two complementary study sites (Hinterrhein and Landquart) with smaller glacier coverage (section 5.4).

4.1. Model Performance During Calibration

The performance of the HBV3 model setup in the Rhone catchment for the 100 best MC runs during the calibration year 2008 is visualized in Figure 2. The values of the efficiency criteria listed in Table 4 of the best 100 runs reveal that the selection of 100 runs from 10,000 MC runs adequately accounts for variability and optimization of efficiency (Figure 2a). The mean E_Q value is 0.88 (standard deviation: $\pm$0.03), mean $E_{SC,summer}$ is 0.91 ($\pm$0.01), and mean $E_{MB,abl}$ is 1226 ($\pm$499) mm water equivalent (w. eq.) over the ablation season (Table 5), indicating an adequate prediction of all three observational data sets. Simulated mean mass balance during ablation season is 2664 mm w. eq. compared to 2747 mm w. eq. according to the observations, indicating that glacier melt is captured with an accuracy of $\sim$3%. Specific discharge reaches 38 mm d^{-1} during the melting season and drops below 2 mm d^{-1} during the low-flow season (Figure 2b). From November to April, the entire catchment was covered by snow, while during July and August snow coverage was reduced to less than 40% of the catchment (Figure 2c). During the accumulation phase between 1 October 2007 and 30 April 2008, the Rhonegletscher gained from 575 mm snow water equivalent (w. eq.) in the lowest to over 2000 mm w. eq. in the highest altitude bands (Figure 2d). During the depletion phase, the glacier lost between 20 mm w. eq. and 6000 mm w. eq. All these observations were adequately predicted by the 100 best simulations as illustrated by the grey areas and bars in Figure 2. Nevertheless, the 100 best runs regarding P^{OAnorm} reveal slightly lower performance regarding discharge than the best performance obtained if the model runs were selected only according to E_Q (Figure 2b).

Table 5. Performance of Rhone Regarding Different Selection Criteria of the 100 Best MC-Runs During Calibration

	Performance Criteria							
	Discharge E_Q [-]		Snow Cover $E_{SC,summer}$ [−]		Mass Balances $E_{MB,abl}$ [mm w. eq.]		Consistency Perf. P^{OAnorm} [−]	
[a]Selection Criteria	Mean	Std	Mean	Std	Mean	Std	Mean	Std
Q	0.912	0.006	0.879	0.026	1907.227	1007.419	0.791	0.109
SC	1.961	3.223	0.925	0.001	10002.875	6122.284	0.573	0.150
MB	0.782	0.087	0.873	0.033	486.965	115.472	0.753	0.122
Q+SC	0.889	0.019	0.915	0.005	1842.336	1021.465	0.906	0.047
Q+MB	0.895	0.015	0.890	0.020	893.392	215.205	0.859	0.080
MB+SC	0.807	0.122	0.916	0.005	950.524	287.872	0.883	0.077
Q+SC+MB	0.875	0.028	0.911	0.009	1225.856	498.622	0.937	0.020

[a]Shaded cells indicate that the data sets relevant for the criterion were used to select the best runs.

In order to calculate the contribution of snow melt, glacier melt, and rain to runoff, it is important to adequately optimize the simulation of the snow cover during the snow melt season (expressed by $E_{SC,summer}$), the glacier volume loss during depletion season (expressed by $E_{MB,abl}$), and the seasonal discharge dynamics during the entire year (expressed by the *Nash and Sutcliffe* [1970] values, E_Q). Accordingly, we compared E_Q, $E_{SC,summer}$, $E_{MB,abl}$, and P^{OAnorm} in the 100 best runs selected according to each of these efficiencies in Table 5. Mean E_Q in the 100 best runs selected with daily discharge data is 0.91, which is higher than in the 100 best runs selected according to P^{OAnorm} which reveal a mean E_Q value of 0.88. However, as also illustrated in Figure 2c and d model results using only Q for calibration yield unsatisfactory results for daily snow cover and seasonal glacier mass balances. The same is true for the mean performance of the 100 best runs regarding P^{OAnorm} or a specific criterion listed in Table 4. Mean $E_{SC,summer}$ and $E_{MB,abl}$ of the runs selected only with discharge reveal efficiencies of 0.88 and 1907 mm w. eq., respectively, which indicates a lower performance than the 100 best runs selected with P^{OAnorm}. The same findings also apply for the 100 best runs selected regarding their efficiencies in snow cover and mass balances (Table 5). Every criterion used to select the best runs yields highest values of the specific efficiency but the remaining criteria are significantly lower. If P^{OAnorm} is used to select the 100 best runs, efficiency regarding all criteria appears to be adequate, as illustrated in Figure 2 but not maximized. This trade-off illustrates the equifinality of the calibration of the HBV-light model. An explicit equifinality can be observed if only SC is used for calibration, as an increase in precipitation can be compensated by enhanced melting rates to produce similar $E_{SC,summer}$ results, leading, however, to unrealistic discharge and mass balance simulations.

As defined by *Finger et al*. [2011], P^{OAnorm} is a numerical value that quantifies the simultaneous performance of a model regarding all efficiency criteria considered relative to the performance of all MC runs performed. In our case, we considered six criteria (Table 4), consisting of three pairs evaluating the performance regarding discharge (Q), snow cover (SC), and glacier mass balances (MB), assuring equal weighting of the three observational data sets. In Figure 3, mean P^{OAnorm} of the 100 best runs regarding seven selection criteria are visualized: (i) regarding Q using E_Q, (ii) regarding SC using $E_{SC,summer}$, (iii) regarding MB using $E_{MB,abl}$, (iv) regarding Q and SC by averaging the ranking values of E_Q and $E_{SC,summer}$, (v) regarding Q and MB by

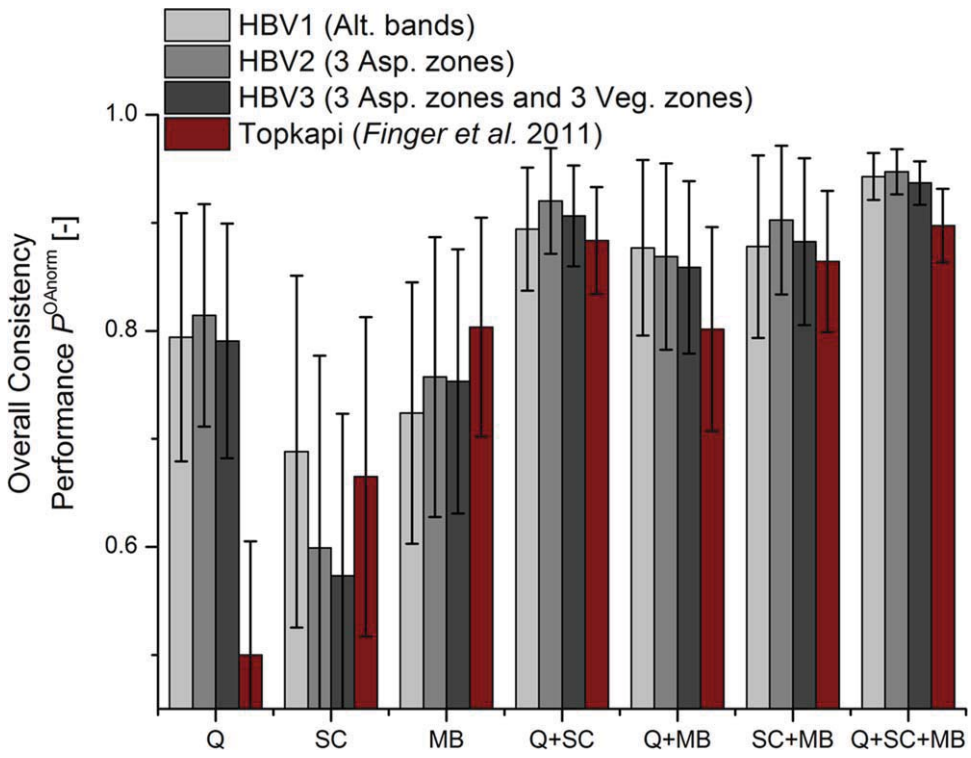

Figure 3. Overall consistency performance, P^{OAnorm}, of the 100 best runs performed with HBV1, HBV2, and HBV3 regarding the selection criteria listed on the abscissa (Q: discharge; SC: snow cover images; MB: glacier mass balance). The whiskers illustrate the standard deviation from the mean. Star indicates results found by *Finger et al*. [2011] using the distributed, physically based TOPKAPI model.

averaging the ranking values of E_Q and $E_{MB,abl}$, (vi) regarding SC and MB by averaging the ranking values of $E_{SC,summer}$ and $E_{MB,abl}$ and finally (vii) regarding Q, SC and MB using all six efficiency criteria. By definition, P^{OAnorm} is highest if all three data sets are used to select the best runs.

Furthermore, values for P^{OAnorm} of HBV1, HBV2, and HBV3 are compared in Figure 3. The results clearly indicate that overall consistency performance is significantly lower (difference is larger than the standard deviation from the mean) if only one observational data set (Q, SC, or MB) is used to select the best runs, independent of model complexity. These results are in line and complementary to the results obtained by *Finger et al*. [2011] using the fully distributed TOPKAPI model. However, the increase in P^{OAnorm} when additional data sets are used for calibration is more emphasized in the results produced by TOPKAPI than HBV (Figure 3), indicating an enhanced exploitation of spatial information from satellite images within TOPKAPI.

All three models achieve substantially higher overall consistency performances if at least two data sets are combined to select the best MC runs. If only two data sets are available (Q+SC, Q+MB, and SC+MB), all three models reach best performances when Q+SC are used to select the best runs. The combinations Q+MB and SC+MB to select the best runs lead to higher performances than if only one data set was used, but remain below the performance obtained with Q+SC. Finally, enhanced model complexity does not lead to a significant increase in overall consistency performance regardless of the observational data sets used for calibration.

4.2. Validation and Model Consistency

Model performance during the calibration period indicates that model consistency is increased if all available data sets are used for calibration, regardless of model complexity. However, this finding has to be validated for an independent validation period characterized by different weather patterns. By applying the parameter sets from the calibration to an 8 year validation period, including the record breaking heat wave in 2003 [*Schär et al*., 2004] which resulted in exceptional glacier melt runoff [*Zappa and Kan*, 2007] and the extreme flood event in 2005 [*Barredo*, 2007], the robustness of our calibration routine can be assessed for different meteorological conditions. As the three HBV-light setups have a similar structure, a direct comparison of their efficiencies during the validation period is possible.

In Figure 4, the model efficiencies for the calibration (2008) and validation period (2001–2007) of the three HBV model complexities regarding Q, SC, and MB are compared to the efficiencies obtained with HBV3 version calibrated only with Q and with Q+SC. Evidently, all 8 years reveal the best Nash-Sutcliffe efficiency if the MC runs are selected only considering Q. Nevertheless, in this case $E_{SC,summer}$ is, as expected, during several years significantly lower. Furthermore, $E_{MB,norm}$ is also lowest when MC runs are only selected regarding their respective E_Q. Nevertheless, $E_{MB,norm}$ is consistently higher if MC runs are selected considering Q and SC than if selected only considering Q. This result is in line with the mean overall consistency performance obtained during the calibration period. Finally, in Figure 4d the simulated accumulated mean snow height in the catchment is illustrated. While the calibration using only Q reveals a continuous increase of mean simulated snow height to values of more than 6 m w. eq. within 8 years, the calibration using additionally SC indicates significantly smaller perennial snow accumulation falling in line with the long-term glacier mass balance observations.

The comparison of the three HBV model setups with different levels of complexity does not reveal a significant change in model performance, neither regarding the overall consistency performance using different data sets for calibration (Figure 3), nor specific efficiency criteria during the 8 year validation period (Figure 4). The comparison of monthly discharge and runoff composition computed with different model complexities was also minimal (Figure 5a), thus indicating the redundancy of the investigated model complexities. Furthermore, these results reveal that the investigated higher model complexity does not lead to a better performance during extreme weather patterns such as the heat wave in 2003 or the flood of 2005 (Figure 4).

4.3. Estimations of Snow, Glacier, and Rainfall Contribution to Runoff

Model estimates of mean monthly total runoff (Q_{tot}), fraction of snow-covered area (A_{SC}), total glacier outflow (G_{out}), and rain infiltration into the soil (I_{soil}) in the Rhone catchment between 2001 and 2008 are presented in Figure 5. Furthermore, Figure 5 compares the estimates of the three HBV setups (HBV1, HBV2, and

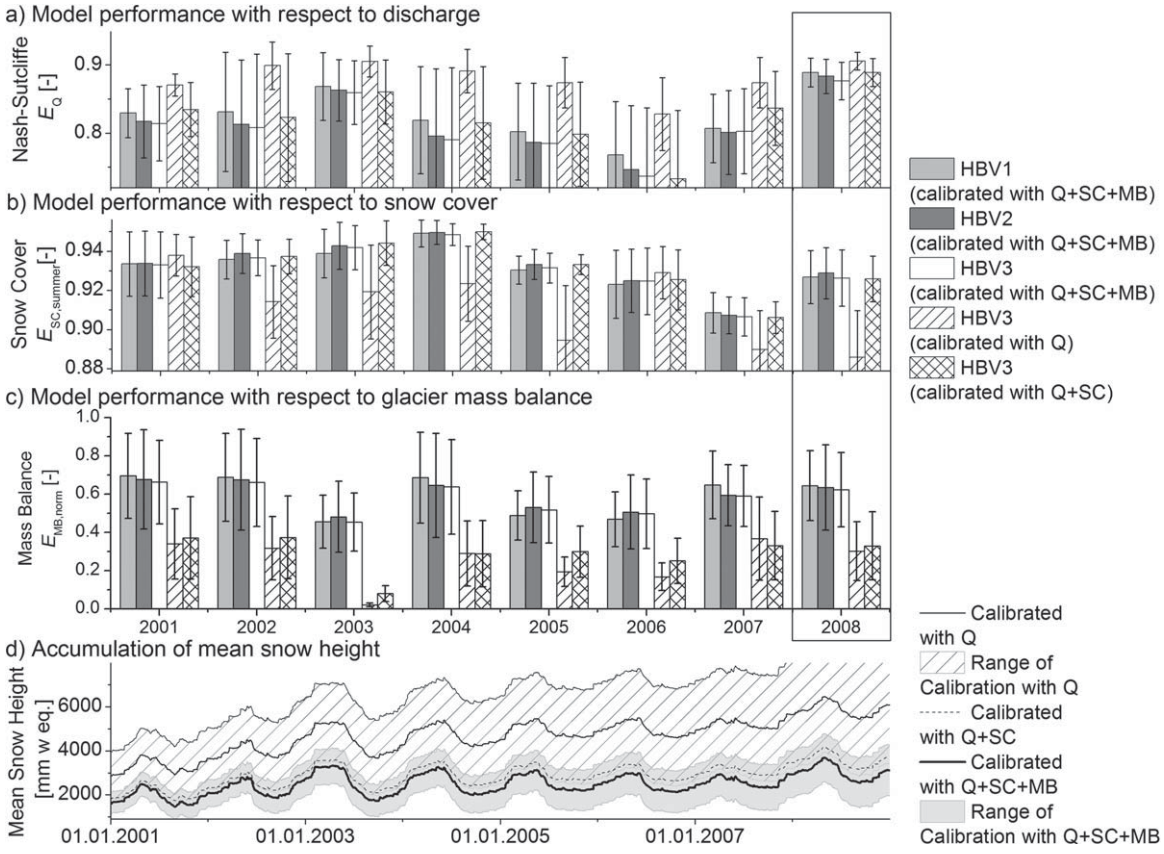

Figure 4. (a,b,c) Mean model performance of the 100 best runs of the three HBV setup versions for the Rhone catchment during a validation period (2001–2008). Light gray, dark gray, and plain white bars illustrate model performance of the HBV1, HBV2, and HBV3 model, respectively, using the model parameter sets obtained during calibration in 2008 (enframed area) with all three observational data sets (Q, SC, and MB). Stripes and cross stripes in bars indicate, respectively, model performance of the HBV3 model using parameter sets selected with (i) discharge and with (ii) Q and snow cover images (Q+SC) combined. The upward and downward whiskers illustrate the standard deviation from the mean. (d) presents the average of simulated accumulated mean snow height in the catchment of the 100 best runs of HBV 3 with different calibrations. The striped and gray area represents, respectively, the standard deviation from the average of calibration using (i) Q and (ii) all three data sets combined.

HBV3) and the three calibration methods (using only Q, using Q+SC, and using Q+SC+MB combined). We limit ourselves here to three ways of calibrating the model, as the combination of Q+SC revealed to contain complementary information [*Duethmann et al*., 2014; *Finger et al*., 2011] and obtained high overall consistency performance (Figure 3). Indeed, MODIS snow cover products are available for most areas of the world and have been increasingly used for calibration purposes [*Duethmann et al*., 2014; *Finger et al*., 2012; *Franz and Karsten*, 2013]. Glacier mass balances are more difficult to acquire but have the highest and most direct information content regarding the contribution of glacier melt to runoff.

If all three data sets (Q+SC+MB) are used for calibration, the seasonal dynamics of snow, glacier, and rain contribution are reproduced by all three model complexities consistently (Figure 5a). The results clearly show that snow contribution dominates discharge until the month of June. Starting in July, snow melt gradually decreases and in August glacier melt is the dominant contributor to runoff. Simulated rainfall runoff reaches its maximum contribution in August consistent with precipitation patterns observed at the nearby weather station, but never becomes a main contributor to total runoff in the Rhone catchment. These results are confirmed by all three model complexities.

Results obtained with HBV3 calibrated with (i) Q, (ii) Q+SC, and (iii) Q+SC+MB, however, partly reveal significant differences in the results (Figures 5b–5d). From June to October, estimates of snow cover using calibration with Q are significantly overestimated compared to observations in satellite snow cover images. Results obtained with calibration of multiple data sets perform significantly better regarding snow cover. Accordingly, if calibration is performed using only Q, the overestimation of snow cover leads to an underestimation of glacier outflow G_{out}, and rain infiltration during June and July (Figures 5c and 5d).

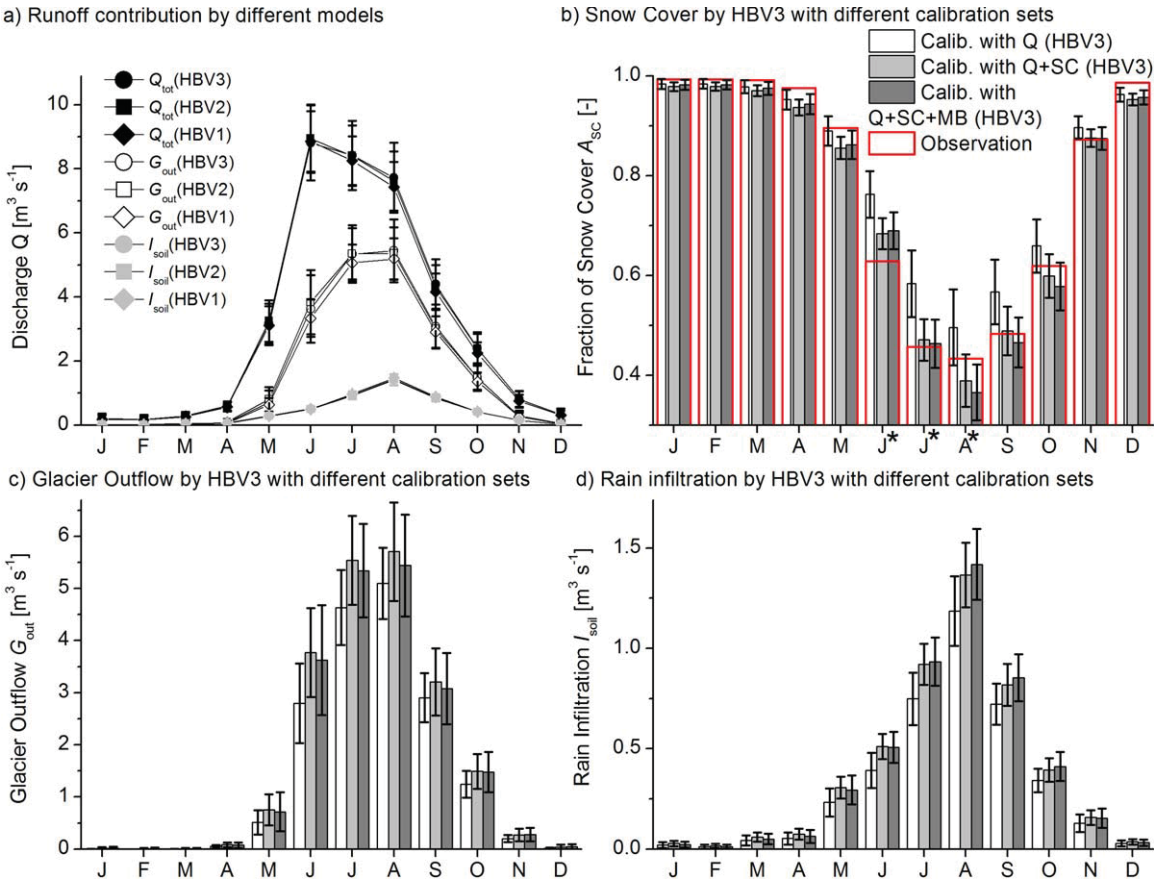

Figure 5. Mean monthly discharge, fraction of snow cover, glacier outflow, and rain infiltration from 2001 to 2008 in the Rhone catchment. (a) Illustrates the results from three different model complexities calibrated with Q+SC+MB. (b), (c), and (d) Compares the results of fractional snow cover, the contributions of glacier to runoff, and the rain infiltration computed by the HBV3 model calibrated with (i) only Q, with (ii) Q+SC, and with (iii) Q+SC+MB. Red open bars in Figure 5b indicate the fractional snow cover estimates based on MODIS satellite images. The whiskers illustrate the standard deviation from the mean. Months with significant differences are labeled with an asterisk (*).

4.4. Results for the Hinterrhein and Landquart Catchments

The results for the Rhone catchment indicate that discharge data and snow cover data are sufficient to adequately estimate snow, glacier, and rain contribution to the runoff from glacierized catchments (see also discussion section). To confirm these findings, we applied the same calibration method to the Hinterrhein (calibration year 2002) and Landquart (calibration year 2004) catchments, both characterized by smaller glacierization. In the following, we present the results for all catchments during the respective validation period using the parameter sets determined during calibration.

In Figure 6, the annual efficiencies for the years 2001–2004 for discharge, snow-covered area, and glacier mass balances in the Hinterrhein and Landquart catchments are compared to the efficiencies obtained for the Rhone catchment. Furthermore, for all three catchments, the efficiencies of simulations selected by using only Q, Q+SC combined, and by using all three observational data sets to select the best runs are illustrated. The result for Hinterrhein should be interpreted with reservation, as the glacier mass balance variability used for calibration is based on a glaciological model rather than being direct observations as is the case for the other study sites.

Mean efficiency regarding Q is best if the model is only calibrated with Q in all three catchments. However, during extreme weather patterns, such as the heat wave in 2003, the model efficiency regarding Q in the Landquart catchment is slightly better when Q+SC+MB are combined to calibrate the model than when only Q is used (Figure 6a). Given that the decrease in efficiency for discharge by optimizing the model performance also using glacier mass balances and snow cover is in most years not significant, the gain in efficiency regarding SC and MB is significant and remarkable (Figures 6b and 6c). Indeed, the efficiency

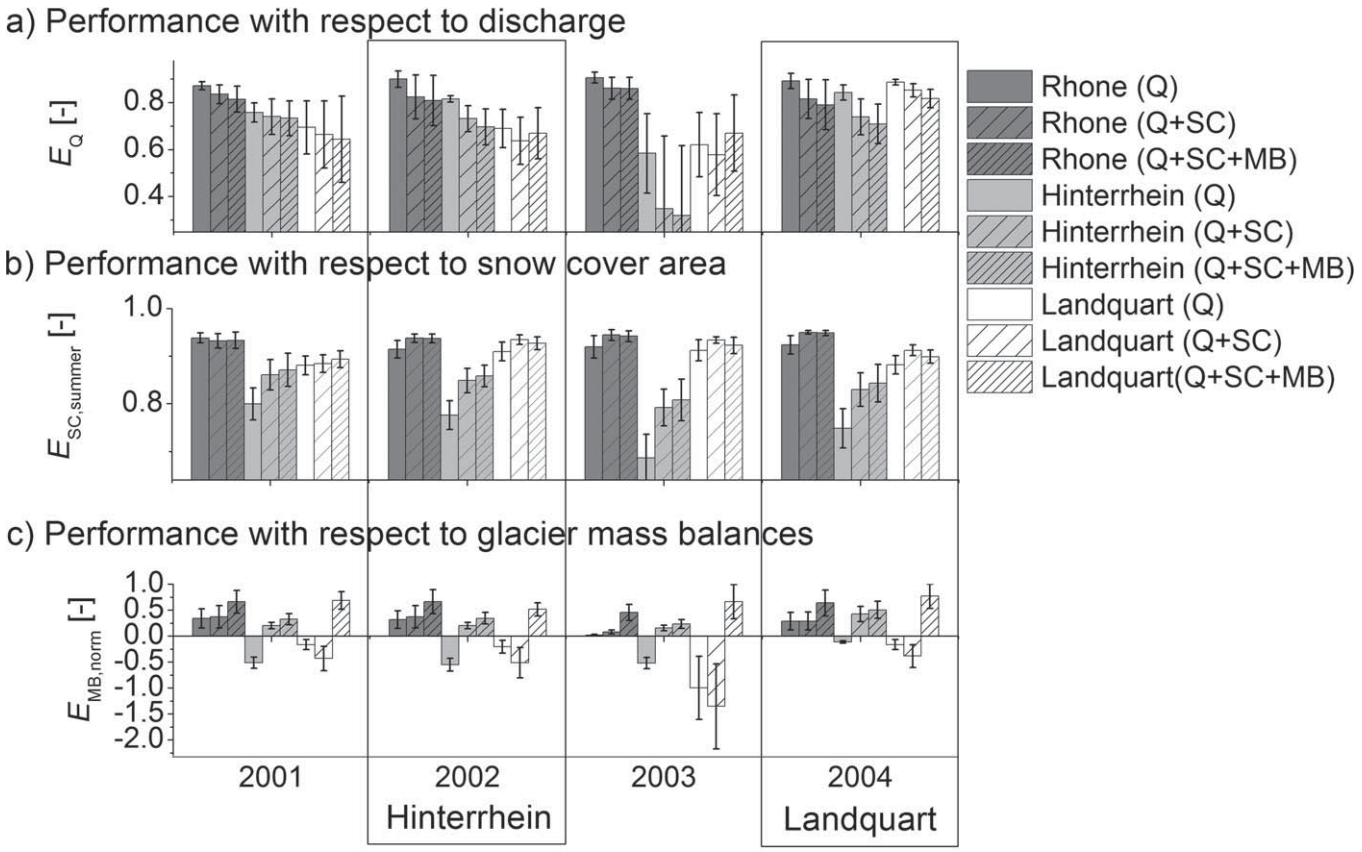

Figure 6. Comparison of model efficiency during the validation period (2001–2004) of Rhone, Hinterrhein and Landquart using the parameters sets from the calibration year (calibration year is enframed and labeled with the respective site, n.b. Rhone was calibrated for 2008). Empty bars illustrate the performance of simulations calibrated only with Q, bars with sparse stripes indicate efficiencies of simulations calibrated with Q and SC combined, and bars with dense stripes indicate efficiencies of simulation using all three data sets combined. The whiskers illustrate the standard deviation from the mean.

regarding SC is consistently increased if SC is used additionally for calibration or if all three data sets are used to calibrate the model. Calibration using only Q reveals significantly lower efficiency for snow cover than the combined use of all data sets for calibration. A similar result is revealed by the efficiency regarding mass balances (Figure 6c). Simulations using only discharge data for calibration indicate a significantly lower efficiency regarding calculated glacier mass balance than calibration using at least two observational data sets.

Indeed, the seasonal evolution of simulated snow and ice melt and rain runoff in the three catchments shows the expected sequence (Figure 7): from May to July intense snow melt and rainfall runoff dominate the discharge in all three catchments, in August and September glacial outflow reaches its maximum, and in October and November the runoff is mainly composed of rainfall runoff. The different calibration routines using only discharge (Q), using discharge and satellite images (Q+SC), and additionally using glacier mass balances (Q+SC+MB) result in the same seasonal patterns. However, as already discussed for the Rhonegletscher, significantly different compositions of total runoff can be observed during specific months if different data sets are used to calibrate the model. In particular in June and July, when melting rates of snow and ice are at maximum, simulation results appear to be sensitive to the multiple data set calibration technique. However, these months are particularly important for water resource management and water users in downstream areas.

The assessment of model performance regarding discharge indicates that the use of multiple data sets does not decrease the efficiency regarding Q below an acceptable level (Figures 2, 4, and 7), but significantly increases the overall consistency performance (Figure 3). In Figure 8, the performance during the validation period of long-term mass balances and monthly snow cover fraction using different data sets for

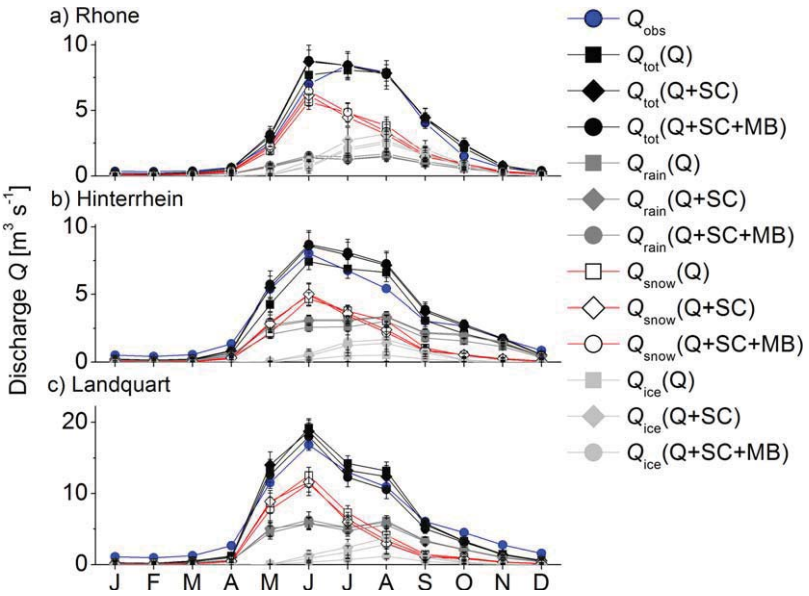

Figure 7. Mean monthly discharge, Q_{rain}, Q_{snow}, and Q_{ice} contribution in the Rhonegletscher (2001–2008 using parameters from the calibration year 2008), Hinterrhein (2001–2004 using parameters from the calibration year 2002), and Landquart (2001–2004 using parameters from the calibration year 2004) catchment computed with the HBV3 model setup using (i) Q, (ii) Q+SC, and (iii) Q+SC+MB for calibration.

calibration of the HBV3 model are presented (see Figure 5b for fractional snow cover for Rhone). Only the use of all three data sets combined yields an acceptable performance for long-term mass balances, as the observations lie within the standard deviations of the simulations (Figures 8a, 8b, and 8d). However, if only Q is used for calibration observations are outside the standard deviation of the simulations in most cases, indicating significant inconsistencies regarding observed glacier mass balances. Similar patterns regarding the calibration methods can be observed in the simulated snow cover fraction (Figures 8c and 8e). Similarly to the Rhone catchment (Figure 5b), fractional snow cover is significantly overestimated during the summer months in the Hinterrhein catchment if only Q is used for calibration (Figure 8c). In the Landquart catchment, the added value of SC for calibration purposes can be observed as well, even though the differences are not as pronounced as in the Rhone and Hinterrhein catchments.

Figure 9 illustrates the mean P^{OAnorm} of the HBV3 model setup using a single observational data set (Q, SC, or MB), combining two observational data sets (Q+SC) and using all three data sets (Q+SC+MB) to calibrate the model. This comparison reveals that, if only one data set is available, discharge data lead to highest overall consistency performance, regardless of the glacierization of the catchment. However, the use of combined observational data sets (Q+SC or Q+SC+MB) significantly increases the mean overall consistency performance in all three catchments. The comparison of the mean overall consistency performance also indicates that MB enhances P^{OAnorm} in particular in smaller catchments with higher glacierization.

5. Discussion

The objective of this study is to evaluate the value of multiple data sets versus model complexity to adequately estimate snow, glacier, and rain runoff contribution in mountain streams. An experimental estimation of the different sources of runoff would require extensive field work, chemical analysis and tracer experiments [*Finger et al*., 2013; *Jansson et al*., 2003; *Taylor et al*., 2001]. While the annual glacier contribution to runoff can be computed by water balance calculations based on total runoff and changes in glacier volume [*Huss*, 2011; *Kaser et al*., 2010], the seasonal dynamics requires an assessment of continuous melt and runoff processes. The processes leading to snow, ice melt, and rainfall runoff are complex, as rain may fall on snow, refreeze during night and eventually be stored subglacially to be released after several days mixed with snow and glacier ice melt water. By validating the model outputs against daily snow cover images,

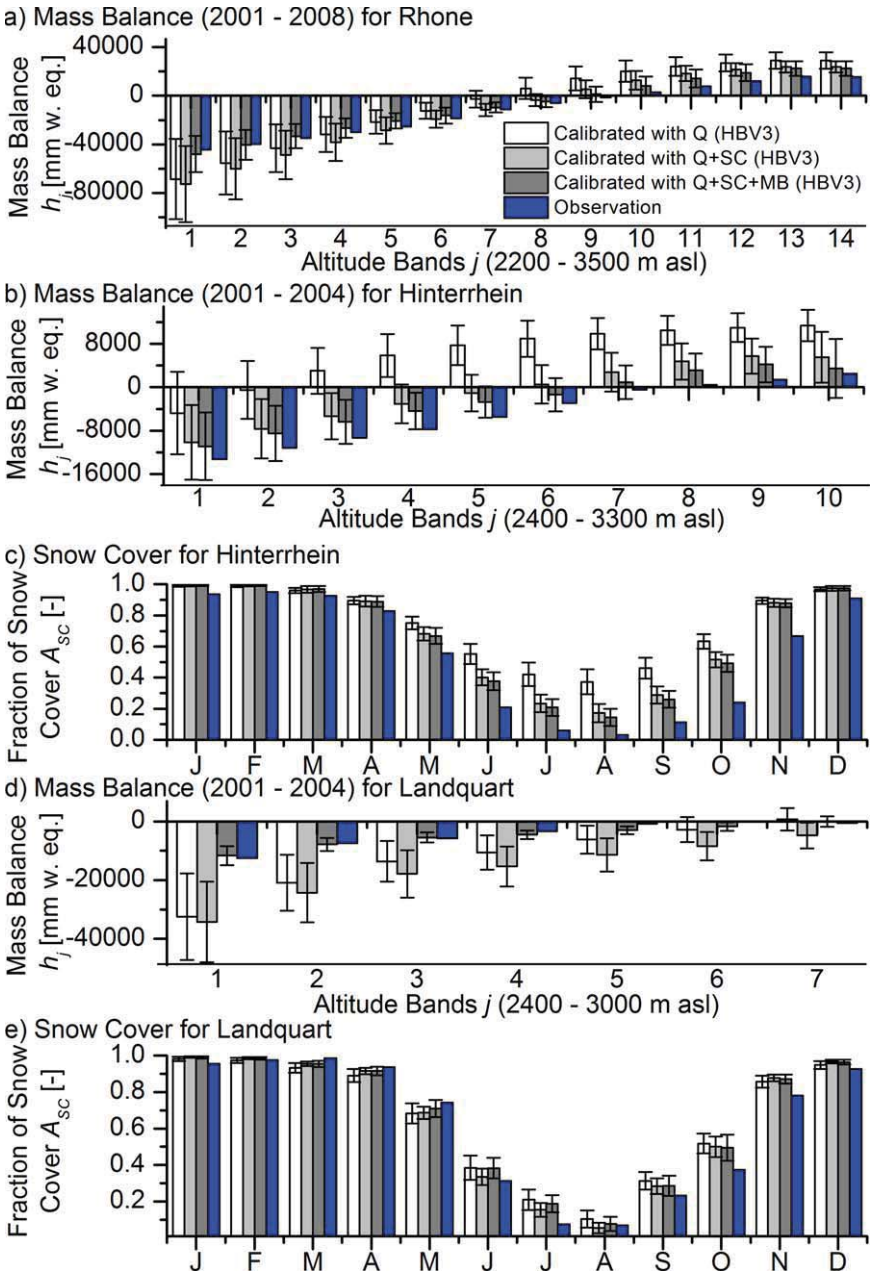

Figure 8. Long-term validation of the HBV3 model regarding glacier mass balances and snow cover. (a), (b), and (d) illustrate mass balances between 2001 and 2004 for the glaciers in Rhone (calibration year: 2008, n.b. validation period for Rhone extends from 2001 to 2008), Hinterrhein (calibration year: 2002), and Landquart (calibration year: 2004), respectively. (c) and (e) show monthly mean fractional snow cover between 2001 and 2004 for Hinterrhein and Landquart.

seasonal glacier mass balances and daily discharge observations, we validate our results with the three relevant observational data sets, leading to increased modeling consistency.

A particular advantage of our approach using satellite snow cover images is reflected in the increased performance regarding snow cover estimations. Our approach reduces the accumulation of simulated mean snow height to acceptable levels for simulations not exceeding several decades (Figure 4d). This supports the argument that calibration using satellite snow cover images yields more realistic results [*Duethmann et al*., 2014; *Finger et al*., 2012; *Parajka and Blöschl*, 2008]. Nevertheless, for a proper consideration of transformation of snow to ice and glacier flow, more detailed corresponding modules would have to be included in the model.

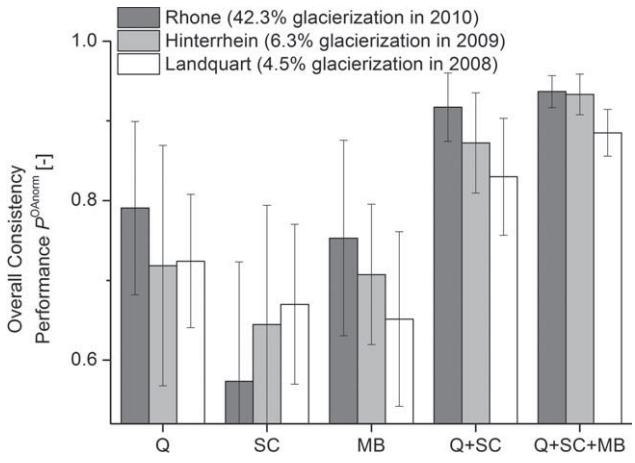

Figure 9. Mean overall consistency performance in all three catchments of the 100 best runs during the calibration year (Rhone: 2008, Hinterrhein: 2002, and Landquart: 2004) selected with the observational data sets indicated on the abscissa: (i) discharge (Q), (ii) Q combined with snow cover (SC), and (iii) Q combined with SC and mass balances (MB).

We limited the number of MC runs to 10,000 runs, being aware that we cannot explore the entire parameter space and that our approach only partially gives a solution to the equifinality problem [*Beven*, 2006; *Beven and Binley*, 1992]. However, model performance stabilizes after 10,000 runs and the best efficiencies regarding discharge, snow cover, and glacier mass balances (Table 5) are comparable to earlier studies [*Finger et al*. 2011, *Konz and Seibert* 2010, *Verbunt et al*. 2003]. By selecting only 100 runs out of all 10,000 MC runs, we implicitly set the threshold to be the best 1% of all runs. The lowest performance value of the 1% best runs can hence be considered as a reference taking into account catchment characteristics and data quality [*Schaefli and Gupta*, 2007; *Seibert*, 2001]. An eminent advantage of not utilizing user-defined benchmarks is that all data sets are weighted equally, while benchmarks have to be chosen carefully to assure equal weighting of efficiency criteria. The model performance regarding all three observational data sets (Figure 2) is reasonable, given the uncertainty in discharge observations [*Sikorska et al*., 2013], snow cover images [*Hall et al*., 2010], and measured glacier mass balances [*Zemp et al*., 2013]. Accordingly, we conclude that our approach of determining an overall consistency performance is an adequate method to obtain consistent model performance for all three observational data sets.

Even though we restricted the calibration period to 1 year, long-term validation yields adequate results for all three observational data sets, making our method also applicable to catchments with limited data availability. Furthermore, the 1 year calibration period reduces computational time of MC calculation to an acceptable level. Nevertheless, longer calibration periods may also help reducing equifinality [*Razavi and Tolson*, 2013], although this was not investigated in this study. As our method improves overall consistency performance using only 1 year of data, this opens new opportunities for water managers to investigate water resources in remote and unexplored areas with limited data availability and short time series.

We compared three levels of HBV-model complexities to assess the value of model complexity on performance. This assessment only provides a limited insight into the wide range of different model complexities, and could easily be extended to e.g., seasonally varying melt parameters and varying snow melt factors for different vegetation zones [*Jost et al*., 2012b]. The selected complexities were not specifically selected based on data availability but rather on most commonly used aspects in hydrological models (e.g., aspect of slopes and vegetation zones). Accordingly, the observational data sets used in this study did only partially constrain the added complexity. Nevertheless, our results demonstrate that model complexity does not necessarily enhance model performance if the available data does not contain the appropriate information to constrain the results. This is an important finding, as complex models are widely used without having the required observational data to constrain model parameters.

Our findings are in line with results obtained with a physically based, fully distributed TOPKAPI model [*Finger et al*., 2011]. The comparison of model consistency performance between HBV and TOPKAPI reveals that the conceptual HBV model yields higher P^{OAnorm} if only discharge is used for calibration, compared to the spatially distributed TOPKAPI model (Figure 3). This indicates that spatial modeling of snow cover (as done by TOPKAPI) can retrieve more constraining information from snow cover images than lumped hydrological modeling (as done by HBV). Nevertheless, a comprehensive comparison of aggregated high-resolution (hourly runoff and gridded snow melt as computed in TOPKAPI) and coarse-resolution (daily runoff and

snow melt from elevation bands as computed in HBV) model performance would require an extensive discussion which is beyond the scope of this study.

Moreover, our results demonstrate that the use of discharge alone can lead to unrealistic glacier mass balances and snow cover evolution, distorting the contribution of snow, and ice melt to runoff (Table 5 and Figure 2). Furthermore, a steady increase of mean simulated snow height (Figure 4d) can lead to systematic errors in long-term simulations. However, if SC is additionally used for calibration, this systematic error can be minimized. For calculations of future scenarios, it is essential that a model result is produced for the correct reasons [*Finger et al*., 2011; *Kirchner*, 2006]. Accordingly, simulations with high Nash-Sutcliffe efficiency but poor snow cover and mass balance efficiency may not be suitable for scenario projections as they reproduce observed runoff for the wrong reasons. Hence, while an increased overall consistence performance might not improve efficiency regarding a specific data set nor decrease uncertainty of the simulations, it certainly quantifies the model performance regarding all considered data sets simultaneously.

Finally, calibration routines should be adapted to the modeling objectives. In our case, we want to achieve highest accuracy regarding snow, glacier, and rain contribution. Thus, the use of all three data sets for calibration purposes seems most appropriate, despite a certain loss in performance regarding discharge. Our results demonstrate that the use of our multiple data set calibration significantly enhances the consistency of runoff prediction in spring, a period of major importance for hydropower production [*Engelhardt et al*., 2014; *Gaudard et al*., 2013; *Kim and Palmer*, 1997; *Sorg et al*., 2012], sediment transport loads [*Finger et al*., 2006; *Riihimaki et al*., 2005], and downstream freshwater ecosystems [*Finger et al*., 2007a; *Finger et al*., 2007b], to name just a few. Accordingly, our modeling approach using multiple data sets for calibration of a hydrological model presents a robust and accurate estimation of runoff contribution, providing important insights for water resources managers.

6. Conclusions

Three levels of complexity of the conceptual hydrological model HBV-light were evaluated with respect to data availability of daily discharge (Q), daily satellite snow cover images (SC), and seasonal glacier mass balance (MB) and their combination to consistently and reliably estimate snow, glacier, and rain contribution to runoff in glacierized Alpine drainage basins. Our results demonstrate (1) that the use of multiple data sets significantly improves the estimation of snow, glacier, and rainfall contribution to runoff compared to calibrations with runoff only, and (2) that the increase in model complexity does not lead to a substantial improvement of modeling performance. Based on the presented results, the following conclusions can be drawn:

1. Our results demonstrate that 10,000 MC runs using randomly generated parameter sets are sufficient to define an ensemble of 100 parameters sets for the HBV-light model with adequate performance regarding daily snow cover, seasonal glacier mass balances, and daily discharge. Metric efficiencies of the 100 best runs were comparable to previous studies, revealing that a threshold to exclude poor performances is not necessary. The omission of thresholds consolidates the multiple data set calibration as it guarantees that all efficiencies are equally represented in the selected runs.
2. The overall consistency performance is increased if different observational data sets are used for model calibration regardless of the complexity of the hydrological model. In particular, the combination of discharge data and satellite-derived snow cover images produces substantially better results. The aggregated spatial information from satellite images to fractional snow cover and the temporal and volumetric information from discharge data are complementary, allowing a realistic model calibration reproducing snow cover, glacier mass balances, and discharge adequately.
3. The increase in model complexity by introducing aspect zones (HBV2) and vegetation zones (HBV3) into the HBV setup does not have a significant influence on model performance regarding snow, glacier, and rain contribution to runoff. This shows that the increasing model complexity is redundant if the available data does not contain specific information to constrain the added complexity.
4. The use of satellite-derived snow cover images to constrain model parameters reduces the overestimation of snow cover during summer months and thus increases the performance regarding long-term

mass balances for the two catchments with higher glacierization (Rhone and Hinterrhein). This indicates that simulated snow accumulation becomes more realistic and is acceptable even for decadal simulation periods.

5. For the three investigated catchments, the value of combining glacier mass balances, snow cover images, and discharge to calibrate a hydrological model increases for the smaller catchments with a high percentage of glacier cover. However, the use of snow cover images particularly increases model performance in the larger catchments with a smaller level of glacierization.

6. In particular during the ablation season (e.g., June and July), the use of multiple data sets to calibrate a hydrological model leads to significantly higher performance regarding the snow, glacier, and rain contribution in runoff. According to our study, snow cover was generally overestimated when only discharge was used for calibration, leading to inaccurate glacier mass balances and unrealistic glacier and rain contribution to runoff.

7. Given that increasing model complexity did not increase model performance significantly, we conclude that it is more important to obtain and use additional data sets to constrain model parameters, rather than enhancing the precision of specific hydrological processes within a model. Hence, in order to increase hydrological model performance, future efforts should focus on the acquisition, processing, publication, and incorporation of multiple data types into standard modeling procedures, rather than enhancing model complexity.

Acknowledgments

This study was part of the project "Runoff amounts from snow and glacial melts in the Rhine River and its tributaries against the background of climate change" funded by CHR (International Commission for the Hydrology of the Rhine basin) and of the Panta Rhei Research Initiative (WG25) of the International Association of Hydrological Sciences (IAHS). The first author was financed by the Icelandic Meteorological Office, Reykjavik University, and the CHR-project. Meteorological data from all weather stations were provided by Swiss Federal Office of Meteorology and Climatology (MeteoSwiss, available at http://www.meteoschweiz.admin.ch). Discharge data from the three gauging stations were provided by the Swiss Federal Office for the Environment (FOEN, available at http://www.hydrodaten.admin.ch). Glacier mass balances were obtain from various studies as mentioned in the text (Table 1) and MODIS snow cover images were obtained from the National Snow and Ice Data Center (NSIDC, available at http://nsidc.org/). Digital elevation maps were provided by the Swiss Federal Office of Topography (Swisstopo, available at http://www.swisstopo.admin.ch/) and land use maps from the Swiss Federal Statistical Office (available at http://www.bfs.admin.ch/). Finally, we thank Philippe Crochet, Massimiliano Zappa, two anonymous reviewers, and the associated editor for valuable comments on an earlier version of this manuscript.

References

Ambroise, B., J. L. Perrin, and D. Reutenauer (1995), Multicriterion validation of semidistributed conceptual-model of the water cycle in the Fecht catchment (Vosges Massif, France), *Water Resour. Res.*, *31*(6), 1467–1481, doi:10.1029/94wr03293.

Barredo, J. I. (2007), Major flood disasters in Europe: 1950–2005, *Nat. Hazards*, *42*(1), 125–148, doi:10.1007/s11069-006-9065-2.

Bergström, S. (1976), Development and application of a conceptual runoff model for Scandinavian catchments, *SMHI Rep. RHO 7*, *134 pp.*, Norrköping.

Bergström, S. (1992), *The HBV Model: Its Structure and Applications*, Swedish Meteorol. and Hydrol. Inst. Norrköping.

Beven, K. (1996), The limits of splitting: Hydrology, *Sci. Total Environ.*, *183*(1–2), 89–97, doi:10.1016/0048-9697(95)04964-9.

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(1-2), 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processess*, *6*(3), 279–298.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, *249*(1-4), 11–29, doi:10.1016/S0022-1694(01)00421-8.

Borga, M. (2002), Accuracy of radar rainfall estimates for streamflow simulation, *J. Hydrol.*, *267*(1-2), 26–39, doi:10.1016/S0022-1694(02)00137-3.

Crochet, P. (2013), Sensitivity of Icelandic river basins to recent climate variations, *Jokull*, *63*, 71–90.

Cunderlik, J. M., S. W. Fleming, R. W. Jenkinson, M. Thiemann, N. Kouwen, and M. Quick (2013), Integrating logistical and technical criteria into a multiteam, competitive watershed model ranking procedure, *J. Hydrol. Eng.*, *18*(6), 641–654, doi:10.1061/(ASCE)HE.1943-5584.0000670.

Duethmann, D., J. Peters, T. Blume, S. Vorogushyn, and A. Guntner (2014), The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, *Water Resour. Res.*, *50*, 2002–2021, doi:10.1002/2013WR014382.

Engelhardt, M., T. V. Schuler, and L. M. Andreassen (2014), Contribution of snow and glacier melt to discharge for highly glacierised catchments in Norway, *Hydrol. Earth Syst. Sci.*, *18*, 511–523, doi:10.5194/hess-18-511-2014.

Finger, D., M. Schmid, and A. Wuest (2006), Effects of upstream hydropower operation on riverine particle transport and turbidity in downstream lakes, *Water Resour. Res.*, *42*, W08429, doi:10.1029/2005WR004751.

Finger, D., M. Schmid, and A. Wüest (2007a), Comparing effects of oligotrophication and upstream hydropower dams on plankton and productivity in perialpine lakes, *Water Resour. Res.*, *43*, W12404, doi:10.1029/2007WR005868.

Finger, D., P. Bossard, M. Schmid, L. Jaun, B. Müller, D. Steiner, E. Schaffer, M. Zeh, and A. Wüest (2007b), Effects of alpine hydropower operations on primary production in a downstream lake, *Aquat. Sci.*, *69*(2), 240–256, doi:10.1007/s00027-007-0873-6.

Finger, D., F. Pellicciotti, M. Konz, S. Rimkus, and P. Burlando (2011), The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model, *Water Resour. Res.*, *47*, W07519, doi:10.1029/2010WR009824.

Finger, D., G. Heinrich, A. Gobiet, and A. Bauder (2012), Projections of future water resources and their uncertainty in a glacierized catchment in the Swiss Alps and the subsequent effects on hydropower production during the 21st century, *Water Resour. Res.*, *48*, W02521, doi:10.1029/2011WR010733.

Finger, D., et al. (2013), Identification of glacial melt water runoff in a karstic environment and its implication for present and future water availability, *Hydrol. Earth Syst. Sci.*, *17*, 3261–3277, doi:10.5194/hess-17-3261-2013.

Fischer, M., M. Huss, C. Barboux, and M. Hoelzle (2014), The new Swiss Glacier Inventory SGI2010: Relevance of using high-resolution source data in areas dominated by very small glaciers, *Arct. Antarct. Alp. Res.*, *46*(4), 933–945.

Fleming, S. W., F. A. Weber, and S. Weston (2010), Multiobjective, manifoldly constrained Monte Carlo optimization and uncertainty estimation for an operational hydrologic forecast model, paper presented at American Meteorological Society 90th Annual Meeting, Atlanta, Ga.

Franz, K. J., and L. R. Karsten (2013), Calibration of a distributed snow model using MODIS snow covered area data, *J. Hydrol.*, *494*, 160–175, doi:10.1016/j.jhydrol.2013.04.026.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, doi:10.1029/95WR03723.
Gan, T. Y., E. M. Dlamini, and G. F. Biftu (1997), Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling, *J. Hydrol.*, *192*(1-4), 81–103.
Gaudard, L., M. Gilli, and F. Romerio (2013), Climate change impacts on hydropower management, *Water Resour. Manage.*, *27*(15), 5143–5156, doi:10.1007/s11269-013-0458-1.
Glaus, N. (2013), Snow cover in the vicinity of Glacier de la Plaine Morte, BSc thesis, 103 pp, Univ. of Bern, Bern.
Grayson, R. B., I. D. Moore, and T. A. McMahon (1992), Physically based hydrologic modeling .2. is the concept realistic, *Water Resour. Res.*, *28*(10), 2659–2666, doi:10.1029/92WR01259.
Grayson, R. B., G. Blöschl, A. W. Western, and T. A. McMahon (2002), Advances in the use of observed spatial patterns of catchment hydrological response, *Adv. Water Resour.*, *25*(8–12), 1313–1334.
Hall, D. K., G. A. Riggs, V. V. Salomonson, N. E. DiGirolamo, and K. J. Bayr (2002), MODIS snow-cover products, *Remote Sens. Environ.*, *83*(1-2), 181–194, doi:10.1016/S0034-4257(02)00095-0.
Hall, D. K., G. A. Riggs, J. L. Foster, and S. V. Kumar (2010), Development and evaluation of a cloud-gap-filled MODIS daily snow-cover product, *Remote Sens. Environ.*, *114*(3), 496–503, doi:10.1016/j.rse.2009.10.007.
Hock, R. (2003), Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*(1-4), 104–115, doi:10.1016/S0022-1694(03)00257-9.
Huss, M. (2011), Present and future contribution of glacier storage change to runoff from macroscale drainage basins in Europe, *Water Resour. Res.*, *47*, W07511, doi:10.1029/2010WR010299.
Huss, M., A. Bauder, M. Funk and R. Hock (2008), Determination of the seasonal mass balance of four Alpine glaciers since 1865, *J. Geophy. Res.*, *113*(F1), F01015, doi:10.1029/2007JF000803.
Huss, M., A. Bauder, and M. Funk (2009), Homogenization of long-term mass-balance time series, *Ann. Glaciol.*, *50*(50), 198–206.
Huss, M., R. Hock, A. Bauder, and M. Funk (2010), 100-year mass changes in the Swiss Alps linked to the Atlantic Multidecadal Oscillation, *Geophys. Res. Lett.*, *37*, L10501, doi:10.1029/2010GL042616.
IPCC (2013), Climate Change 2013, in The Physical Science Basis, Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, WMO/UNEP, Cambridge.
Jakeman, A. J., and G. M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model, *Water Resour. Res.*, *29*(8), 2637–2649, doi:10.1029/93WR00877.
Jansson, P., R. Hock, and T. Schneider (2003), The concept of glacier storage: A review, *J. Hydrol.*, *282*(1-4), 116–129, doi:10.1016/S0022-1694(03)00258-0.
Johnson, M. S., W. F. Coon, V. K. Mehta, T. S. Steenhuis, E. S. Brooks, and J. Boll (2003), Application of two hydrologic models with different runoff mechanisms to a hillslope dominated watershed in the northeastern US: A comparison of HSPF and SMR, *J. Hydrol.*, *284*(1-4), 57–76, doi:10.1016/j.jhydrol.2003.07.005.
Jost, G., R. D. Moore, B. Menounos, and R. Wheate (2012a), Quantifying the contribution of glacier runoff to streamflow in the upper Columbia River Basin, Canada, *Hydrol. Earth Syst. Sci.*, *16*(3), 849–860, doi:10.5194/hess-16-849-2012.
Jost, G., R. D. Moore, R. Smith, and D. R. Gluns (2012b), Distributed temperature-index snowmelt modelling for forested catchments, *J. Hydrol.*, *420*, 87–101, doi:10.1016/j.jhydrol.2011.11.045.
Juston, J., J. Seibert, and P. O. Johansson (2009), Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment, *Hydrol. Processess*, *23*(21), 3093–3109, doi:10.1002/hyp.7421.
Kaser, G., M. Grosshauser, and B. Marzeion (2010), Contribution potential of glaciers to water availability in different climate regimes, *Proc. Natl. Acad. Sci. U. S. A.*, *107*(47), 20,223–20,227, doi:10.1073/pnas.1008162107.
Khakbaz, B., B. Imam, K. Hsu, and S. Sorooshian (2012), From lumped to distributed via semi-distributed: Calibration strategies for semi-distributed hydrologic models, *J. Hydrol.*, *418*, 61–77, doi:10.1016/j.jhydrol.2009.02.021.
Kim, Y. O., and R. N. Palmer (1997), Value of seasonal flow forecasts in Bayesian stochastic programming, *J. Water Resour. Plann. Manage.*, *123*(6), 327–335, doi:10.1061/(asce)0733-9496(1997)123:6(327).
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.
Kirnbauer, R., G. Bloschl, and D. Gutknecht (1994), Entering the eta of distributed snow models, *Nord. Hydrol.*, *25*(1-2), 1–24.
Klok, E. J., K. Jasper, K. P. Roelofsma, J. Gurtz, and A. Badoux (2001), Distributed hydrological modelling of a heavily glaciated Alpine river basin, *Hydrol. Sci. J.*, *46*(4), 553–570, doi:10.1080/02626660109492850.
Koboltschnig, G. R., W. Schoener, M. Zappa, C. Kroisleitner, and H. Holzmann (2008), Runoff modelling of the glacierized Alpine Upper Salzach basin (Austria): Multi-criteria result validation, *Hydrol. Processess*, *22*(19), 3950–3964, doi:10.1002/hyp.7112.
Konz, M., and J. Seibert (2010), On the value of glacier mass balances for hydrological model calibration, *J. Hydrol.*, *385*, 238–246, doi:10.1016/j.jhydrol.2010.02.025.
Krysanova, V., A. Bronstert, and D. I. Muller-Wohlfeil (1999), Modelling river discharge for large drainage basins: From lumped to distributed approach, *Hydrol. Sci. J.*, *44*(2), 313–331, doi:10.1080/02626669909492224.
Kuczera, G., and M. Mroczkowski (1998), Assessment of hydrologic parameter uncertainty and the worth of multiresponse data, *Water Resour. Res.*, *34*(6), 1481–1489, doi:10.1029/98WR00496.
Kumar, V., P. Singh, and V. Singh (2007), Snow and glacier melt contribution in the Beas River at Pandoh Dam, Himachal Pradesh, India, *Hydrol. Sci. J.*, *52*(2), 376–388, doi:10.1623/hysj.52.2.376.
Mayr, E., W. Hagg, C. Mayer, and L. Braun (2013), Calibrating a spatially distributed conceptual hydrological model using runoff, annual mass balance and winter mass balance, *J. Hydrol.*, *478*, 40–49, doi:10.1016/j.jhydrol.2012.11.035.
McBratney, A. B., M. L. M. Santos, and B. Minasny (2003), On digital soil mapping, *Geoderma*, *117*(1-2), 3–52, doi:10.1016/S0016-7061(03)00223-4.
McMillan, H. K., M. P. Clark, W. B. Bowden, M. Duncan, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, *Hydrol. Processess*, *25*(4), 511–522, doi:10.1002/hyp.7841.
Micovic, Z., and M. C. Quick (2009), Investigation of the model complexity required in runoff simulation at different time scales, *Hydrol. Sci. J.*, *54*(5), 872–885, doi:10.1623/hysj.54.5.872.
Muste, M., K. Yu, and M. Spasojevic (2004), Practical aspects of ADCP data use for quantification of mean river flow characteristics; Part 1: Moving-vessel measurements, *Flow Meas. Instrum.*, *15*(1), 1–16, doi:10.1016/j.flowmeasinst.2003.09.001.
Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models Part 1 - A discussion of principles, *J. Hydrol.*, *10*, 282–290.

Nolin, A. W., J. Phillippe, A. Jefferson, and S. L. Lewis (2010), Present-day and future contributions of glacier runoff to summertime flows in a Pacific Northwest watershed: Implications for water resources, *Water Resour. Res.*, *46*, W12509, doi:10.1029/2009WR008968.
Parajka, J., and G. Blöschl (2008), The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, *358*(3-4), 240–258, doi:10.1016/j.jhydrol.2008.06.006.
Pellicciotti, F., C. Buergi, W. W. Immerzeel, M. Konz, and A. B. Shrestha (2012), Challenges and Uncertainties in Hydrological Modeling of Remote Hindu Kush-Karakoram-Himalayan (HKH) Basins: Suggestions for Calibration Strategies, *Mt. Res. Dev.*, *32*(1), 39–50, doi:10.1659/MRD-JOURNAL-D-11-00092.1.
Perrin, C., C. Michel, and V. Andreassian (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, *242*(3-4), 275–301, doi:10.1016/S0022-1694(00)00393-0.
Radic, V., and R. Hock (2014), Glaciers in the Earth's hydrological cycle: Assessments of Glacier mass and runoff changes on global and regional scales, *Surv. Geophys.*, *35*(3), 813–837, doi:10.1007/s10712-013-9262-y.
Rathore, B. P., A. V. Kulkarni, and N. K. Sherasia (2009), Understanding future changes in snow and glacier melt runoff due to global warming in Wangar Gad basin, India, *Curr. Sci.*, *97*(7), 1077–1081.
Razavi, S., and B. A. Tolson (2013), An efficient framework for hydrologic model calibration on long data periods, *Water Resour. Res.*, *49*, 8418–8431, doi:10.1002/2012WR013442.
Razavi, T., and P. Coulibaly (2013), Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods, *J. Hydrol. Eng.*, *18*(8), 958–975, doi:10.1061/(ASCE)HE.1943-5584.0000690.
Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, *198*(1-4), 69–97, doi:10.1016/S0022-1694(96)03329-x.
Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*(7), 2189–2202, doi:10.1029/96wr00896.
Riihimaki, C. A., K. R. MacGregor, R. S. Anderson, S. P. Anderson, and M. G. Loso (2005), Sediment evacuation and glacial erosion rates at a small alpine glacier, *J. Geophys. Res.*, *110*, F03003, doi:10.1029/2004JF000189.
Schaefli, B., and H. V. Gupta (2007), Do Nash values have value?, *Hydrol. Processess*, *21*(15), 2075–2080, doi:10.1002/hyp.6825.
Schaefli, B., and M. Huss (2011), Integrating point glacier mass balance observations into hydrologic model identification, *Hydrol. Earth Syst. Sci.*, *15*(4), 1227–1241, doi:10.5194/hess-15-1227-2011.
Schär, C., P. L. Vidale, D. Luthi, C. Frei, C. Haberli, M. A. Liniger, and C. Appenzeller (2004), The role of increasing temperature variability in European summer heatwaves, *Nature*, *427*(6972), 332–336, doi:10.1038/nature02300.
Seibert, J. (1999), Regionalisation of parameters for a conceptual rainfall-runoff model, *Agric. For. Meteorol.*, *98–99*, 279–293, doi:10.1016/s0168-1923(99)00105-7.
Seibert, J. (2001), On the need for benchmarks in hydrological modelling, *Hydrol. Processess*, *15*(6), 1063–1064, doi:10.1002/hyp.446.
Seibert, J., and M. J. P. Vis (2012), Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, *Hydrol. Earth Syst. Sci.*, *16*(9), 3315–3325, doi:10.5194/hess-16-3315-2012.
Shen, Z. Y., L. Chen, and T. Chen (2012), Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: A case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrol. Earth Syst. Sci.*, *16*(1), 121–132, doi:10.5194/hess-16-121-2012.
Sikorska, A. E., A. Scheidegger, K. Banasik, and J. Rieckermann (2013), Considering rating curve uncertainty in water level predictions, *Hydrol. Earth Syst. Sci.*, *17*(11), 4415–4427, doi:10.5194/hess-17-4415-2013.
Sorg, A., T. Bolch, M. Stoffel, O. Solomina, and M. Beniston (2012), Climate change impacts on glaciers and runoff in Tien Shan (Central Asia), *Nat. Clim. Change*, *2*(10), 725–731, doi:10.1038/nclimate1592.
Stahl, K., R. D. Moore, J. M. Shea, D. Hutchinson, and A. J. Cannon (2008), Coupled modelling of glacier and streamflow response to future climate scenarios, *Water Resour. Res.*, *44*, W02422, doi:10.1029/2007WR005956.
Steele-Dunne, S., P. Lynch, R. McGrath, T. Semmler, S. Y. Wang, J. Hanafin, and P. Nolan (2008), The impacts of climate change on hydrology in Ireland, *J. Hydrol.*, *356*(1-2), 28–45, doi:10.1016/j.jhydrol.2008.03.025.
swisstopo (2004), DHM 25: The digital height model of Switzerland, 15 pp, Fed. Off. of Topogr., Wabern.
Taylor, S., X. H. Feng, J. W. Kirchner, R. Osterhuber, B. Klaue, and C. E. Renshaw (2001), Isotopic evolution of a seasonal snowpack and its melt, *Water Resour. Res.*, *37*(3), 759–769, doi:10.1029/2000WR900341.
Todini, E. (1996), The ARNO rainfall-runoff model, *J. Hydrol.*, *175*(1-4), 339–382.
van der Linden, S., and M. K. Woo (2003), Application of hydrological models with increasing complexity to subarctic catchments, *J. Hydrol.*, *270*(1-2), 145–157, doi:10.1016/S0022-1694(02)00291-3.
Verbunt, M., J. Gurtz, K. Jasper, H. Lang, P. Warmerdam, and M. Zappa (2003), The hydrological role of snow and glaciers in alpine river basins and their distributed modeling, *J. Hydrol.*, *282*(1–4), 36–55, doi:10.1016/S0022-1694(03)00251-8.
Vrugt, J. A., W. Bouten, H. V. Gupta, and S. Sorooshian (2002), Toward improved identifiability of hydrologic model parameters: The information content of experimental data, *Water Resour. Res.*, *38*(12), 1312, doi:10.1029/2001WR001118.
Xie, P. P., and P. A. Arkin (1996), Analyses of global monthly precipitation using gauge observations, satellite estimates, and numerical model predictions, *J. Clim.*, *9*(4), 840–858, doi:10.1175/1520-0442(1996)009<0840:AOGMPU>2.0.CO;2.
Zappa, M., and C. Kan (2007), Extreme heat and runoff extremes in the Swiss Alps, *Nat. Hazards Earth Syst. Sci.*, *7*(3), 375–389.
Zemp, M., et al. (2013), Reanalysing glacier mass balance measurement series, *The Cryosphere*, *7*(4), 1227–1245, doi:10.5194/tc-7-1227-2013.