The value of glacier mass balance, satellite snow cover images, and hourly discharge for improving the performance of a physically based distributed hydrological model

David Finger,[1,2] Francesca Pellicciotti,[1] Markus Konz,[1] Stefan Rimkus,[1] and Paolo Burlando[1]

Received 29 July 2010; revised 31 March 2011; accepted 25 April 2011; published 12 July 2011.

[1] Physically based hydrological models describe natural processes more accurately than conceptual models but require extensive data sets to produce accurate results. To identify the value of different data sets for improving the performance of the distributed hydrological model TOPKAPI we combine a multivariable validation technique with Monte Carlo simulations. The study is carried out in the snow and ice-dominated Rhonegletscher basin, as these types of mountainous basins are generally the most critical with respect to data availability and sensitivity to climate fluctuations. Each observational data set is used individually and in combination with the other data sets to determine a subset of best parameter combinations out of 10,000 Monte Carlos runs performed with randomly generated parameter sets. We validate model results against discharge, glacier mass balance, and satellite snow cover images for a 14 year time period (1994–2007). While the use of all data sets combined provides the best overall model performance (defined by the concurrent best agreement of simulated discharge, snow cover and mass balance with their respective measurements), the use of one or two variables for constraining the model results in poorer performance. Using only one data set for constraining the model glacier mass balance proved to be the most efficient observation leading to the best overall model performance. Our main result is that a combination of discharge and satellite snow cover images is best for improving model performance, since the volumetric information of discharge data and the spatial information of snow cover images are complementary.

Citation: 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.

1. Introduction

[2] An advantage of physically based, fully distributed hydrological models is their ability to account for complex problems, such as environmental impacts of land-use change, effects of climate change and water management in water applications [*Sorooshian and Gupta*, 1995]. However, their higher complexity requires more observational data sets in order to calibrate all the parameters used in the description of the hydrological processes included in the models [*Grayson and Blöschl*, 2000]. Calibration of distributed models is complicated, in particular for large catchments where distributed observations are not available. Traditional simple trial-and-error techniques are subjective and tedious [*Gupta et al.*, 1998], while parameter estimations with automatic calibration techniques depend on the objective function and the initial parameter values used [*Lindstrom*, 1997].

[3] Recent research has focused on multivariable and multicriteria calibration approaches as a way of overcoming the shortcomings of calibration against one single measured variable [*Beldring*, 2002; *Bergstrom et al.*, 2002; *Cao et al.*, 2006]. Several studies have shown the advantage of such an approach over traditional calibration. Observed soil saturation was used to constrain model parameters in an application of TOPMODEL to a small catchment in France [*Franks et al.*, 1998]. The uncertainty of a lumped conceptual model applied to a catchment in western Australia could be significantly reduced using stream salinity data [*Kuczera and Mroczkowski*, 1998]. Distributed observations of runoff, soil moisture and groundwater levels revealed to be sufficient to calibrate a distributed precipitation-runoff model in southern Sweden [*Motovilov et al.*, 1999]. *Gupta et al.* [1999] showed that single criterion methods were of limited value to calibrate a complex land surface-atmosphere scheme applied to a grassland site in Oklahoma, while forcing the model to match sensible and latent heat fluxes, as well as soil moisture and temperature, led to constrained parameter estimates. *Juston et al.* [2009] demonstrated for a catchment in central Sweden that a limited number of streamflow and groundwater gauging stations can provide the same information as continuous time series of discharge. In glacierized mountainous regions, glacier mass balance has been shown to increase the model performance of conceptual hydrologic models [*Konz et al.*, 2007; *Konz and Seibert*, 2010].

[1]Institute of Environmental Engineering, ETH Zurich, Zurich, Switzerland.
[2]Now at Institute of Geography, University of Bern, Bern, Switzerland.

0043-1397/11/2010WR009824

[4] In particular remotely sensed data contain valuable spatial information through which hydrological models can be calibrated and validated [*Grayson et al.*, 2002; *Wagner et al.*, 2009]. *Campo et al.* [2006] demonstrated that remotely sensed soil saturation indexes can be used to calibrate a distributed hydrological model for small study sites in northern Italy. Recently, similar techniques were tested with conceptual semidistributed models in study sites in Austria [*Parajka et al.*, 2009] and in controlled testing sites in the United States [*Crow and Ryu*, 2009].

[5] The observation of snow cover extent is of particular value in headwaters of mountainous regions. Remotely sensed daily snow cover extent improved model performance of the semidistributed conceptual model applied to various catchments in Austria [*Parajka and Blöschl*, 2006, 2008] and southern Turkey [*Sorman et al.*, 2009]. *Koboltsching et al.* [2008] demonstrated that satellite snow cover images, glacier mass balance, and hourly discharge allowed to calibrate a conceptual hydrological model to a 5% glaciated catchment in Austria. *Paul et al.* [2009] calibrated a glacier mass balance model in the Austrian Alps to reproduce snowlines derived from remote-sensing data and *Immerzeel et al.* [2009] demonstrated that remotely sensed snow cover extent can also be used to validate large scale runoff simulations in a Himalayan river basin using a conceptual, degree-day-based model.

[6] Following a similar research line and using a similar approach as Beven, Seibert and Konz [*Konz and Seibert*, 2010; *Seibert and Beven*, 2009] the main objective of this study is to investigate the predictive power of three data sets, namely discharge, glacier mass balance, and remotely sensed satellite snow cover data. In particular the use of satellite snow cover data for calibration purposes may be very valuable for future studies in remote areas with limited data availability. We do this by (1) assessing model performance of the distributed hydrological model TOPKAPI for the Rhonegletscher basin in the Swiss Alps (Figure 1) by means of Monte Carlo simulations in which the parameters of the hydrological model are allowed to vary within a physically plausible constrained range; and (2) validating an ensemble of best calibrated parameter sets against single observational data set and their combinations by means of performance statistics during a 14 year validation period. Hence, while we are not identifying the optimal parameter

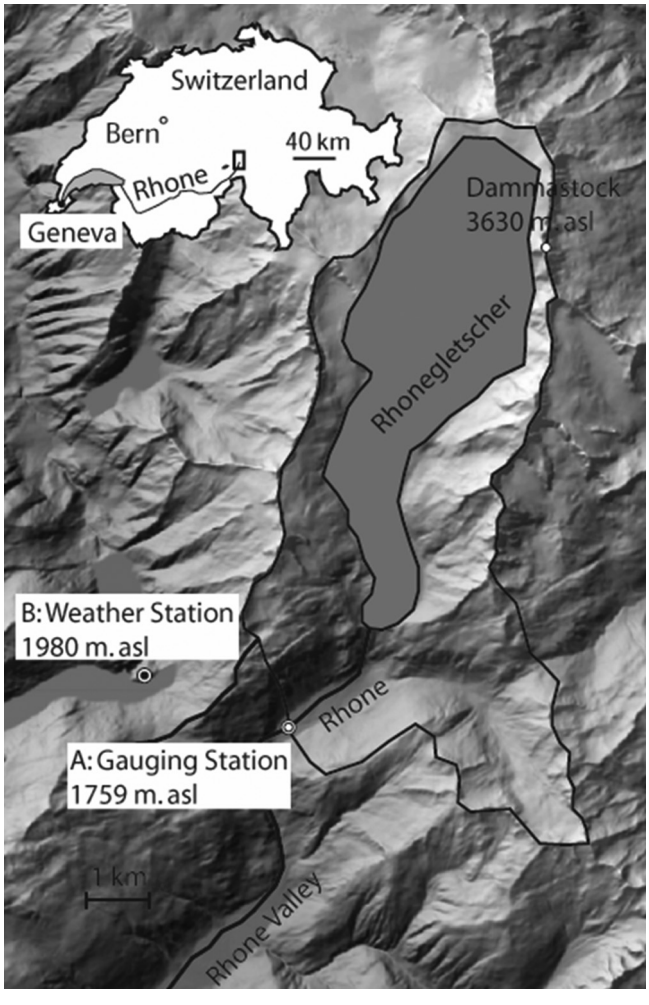

Figure 1. Rhonegletscher and its basin. Circled white dot (indicated by A) indicates the gauging station near Gletsch. Circled black dot (indicated by B) shows the weather station at Grimsel Hospiz where air temperature and precipitation are recorded. The small map in the upper left corner locates the study site in Switzerland.

set, we are demonstrating the statistical value of the three observational data sets for model calibration.

2. Study Site and Data

[7] The Rhonegletscher is located in the Rhone valley in Southern Switzerland (Figure 1). The gauging station downstream of it in Gletsch drains an area of 38.9 km^2. The catchment elevation ranges from 1759 m above sea level (asl) in Gletsch to 3630 m asl at the Dammastock. About 52.2% of the catchment area is covered by the Rhonegletscher, which is a south-facing, medium-size valley glacier that has retreated by about 1300 m during the last 120 years [*Huss et al.*, 2008]. The average annual runoff volume observed at Gletsch is 8.6×10^{-2} km^3 and represents about 1.5% of the total runoff from the Rhone valley as measured at the outlet into Lake Geneva. The Federal Office of Meteorology and Climatology (MeteoSwiss) operates an automatic weather station since 1990 about 3.5 km to the west of the glacier tongue at the Grimsel Hospiz (Figure 1), which provides hourly air temperature and precipitation records. Hourly discharge data (Q) from the gauging station at Gletsch were provided by the Swiss Federal Office of the Environment (FOEN) starting from 1980.

[8] The Rhonegletscher basin is a convenient location for this study, as a long-term time series of glacier mass balance (MB) is available from *Huss et al.* [2008], who reconstructed it using an accumulation model coupled with a distributed temperature index melt model calibrated against observed glacier mass balance and runoff. The model was extensively tested for several Swiss glaciers [*Huss et al.*, 2010] and proved to be able to simulate mass balance variations accurately, with deviations from measured mass balance during the hydrological year 2007–2008 of $\sim$0.06 m w.e. a^{-1} (A. Bauder, personal commmunication). We therefore consider these data as a suitable substitute of direct observations of mass balance to validate our simulations. Finally, snow cover (SC) extent is derived from daily satellite images recorded by the Moderate Resolution Imaging Spectroradiometer (MODIS) since 2001 [*Hall et al.*, 2002]. In this study we used the MODIS product MOD10A1.5 (see http://nsidc.org/).

[9] The catchment topography is described by means of a digital elevation model (DEM) with a 250 m grid size provided by the Swiss Federal Office of Topography (Swisstopo). Although this resolution is suboptimal to describe the complex topography of mountain areas, we chose it as a compromise between the spatial resolution of the observational data sets (SC has a resolution of $\sim$500 m and MB a resolution larger than 700 m) and sufficient detail to capture the variability of the local topographic settings. Soil and geological properties, as well as land cover for the entire catchment were obtained from digital thematic maps available from the Swiss Federal Statistical Office (FSO).

3. Methods

3.1. The Modified Distributed Hydrological Model TOPKAPI

[10] The physically based rainfall-runoff model Topographic Kinematic Approximation and Integration model (TOPKAPI) uses the kinematic wave approach to simulate subsurface flow, overland flow due to saturation excess and channel flow [*Todini and Ciarapica*, 2001]. In its original formulation, TOPKAPI has been applied to various case studies in Italy [*Ciarapica and Todini*, 2002], Switzerland [*Foglia et al.*, 2009], China [*Liu et al.*, 2005], and South Africa [*Sinclair and Pegram*, 2010], encompassing a variety of catchments scales from 0.56 km^2 to more than 10,000 km^2.

[11] The model is raster based, thus requiring a DEM to describe the basin topography, as well as digital soil and land use maps for the spatial variability of land and soil properties. The minimal meteorological input data requirements are air temperature and precipitation time series at one or more locations. The spatial variability is accounted for by the model through established interpolation methods. In this study we use calibrated elevation gradients to extrapolate both air temperature and precipitation across the catchment. The temporal resolution can be set according to simulation needs, either daily or hourly, but is set to hourly for this study.

[12] The original TOPKAPI model has been extensively modified to increase its suitability for high mountains and glacierized catchments. Here we provide only a brief overview of the new components which are employed in this study. A detailed description of the new modules can be found in the auxiliary material.[1] The most important extensions to the model were the implementation (1) of a snow and glacier melt module, including linear reservoirs for the routing of surface melt water; (2) of a new evapotranspiration module; and (3) of a second soil layer.

[13] The modeling of the snow and ice melt processes was implemented based on an enhanced temperature index model which explicitly takes into account the effects of shortwave radiation and albedo [*Pellicciotti et al.*, 2005]. This required, in turn, (1) to model the incoming shortwave radiation by means of a parameterization of the clear-sky global radiation as described by *Corripio* [2003]; (2) to correct the clear-sky global radiation for cloud cover based on a daily cloud transmittance, which depends on minimum and maximum daily temperatures [*Pellicciotti et al.*, 2005, 2011]; and (3) to compute snow albedo assuming an exponential decay controlled by accumulated positive maximum daily temperature [*Brock et al.*, 2000], while glacier albedo is assumed to be constant. Snow and glacier melt routing was implemented using a linear reservoir approach [*Hock and Noetzli*, 1997]. The spatial extent of the glacier and permanent snow fields is static and is an external input in form of a digital map. An increase in precipitation with increasing altitude is included through extrapolation with a constant gradient following specific literature for the Swiss Alps [*Sevruk and Mieglitz*, 2002; *Sevruk*, 2004]. An overview of the melt model parameters is given in Table 1.

[14] The computation of evapotranspiration was changed from the Thornthwaite method [*Thornthwaite*, 1948] to the Makkink equation [*Makkink*, 1957], which is based on incoming shortwave radiation and proved to perform well for hydrological applications at shorter temporal resolution [*De Bruin and Stricker*, 2000]. A second soil layer was implemented, in addition to the single layer of the original

[1] Auxiliary materials are available in the HTML. doi:10.1029/2010WR009824.

Table 1. Summary of the Melt Parameters Required by the Melt Model Component of TOPKAPI

Parameter	Definition	Unit	Value Range
T_t	Threshold air temperature for melt	(°C)	−2 to +2
$T_{t,P}$	Threshold air temperature for precipitation state transition (solid/fluid)	(°C)	−2 to +2
T_{onset}	Threshold temperature for N_{day}[a]	(°C)	−2 to +2
N_{day}	Number of days to exceed T_{onset} to allow melt onset[b]	(d)	1
T_{grad}	Temperature gradient with elevation	(°C m^{-1})	0.002–0.007
T_{mod}	Temperature decrease over glacier area	(°C)	0–2
SRF	Shortwave Radiation Factor (SRF)	(m we m^2 W^{-1} h^{-1})	0–14
TF	Temperature Factor (TF)	(m we C^{-1} h^{-1})	0–400
p_1	First albedo factor		0.7–1
p_2	Second albedo factor		0.1–0.2
α_G	Glacier ice albedo		0–0.4
α_{ground}	Average basin wide ground albedo		0.2–0.3
K_{snow}	Storage constant for snow melt and rain on glaciers	(h)	5–100
K_{ice}	Storage constant for ice melt and channel inflow of glaciers	(h)	5–100
P^{su}_{prec}	Precipitation gradient with elevation from May to Oct	(m^{-1})	0.01–0.06
P^{wi}_{prec}	Precipitation gradient with elevation from Nov to Apr	(m^{-1})	0.055–0.33

[a] T_t, $T_{t,P}$, and T_{onset} were set equal to minimize numbers of parameters to be varied.
[b] N_{day} was kept constant at 1 d.

model, to account for runoff originating from percolation to deeper soil layers and into fractured bedrock, in particular during the winter season, thus allowing to mimic the slow base flow component which characterizes the winter recession. The new enhanced TOPKAPI version was already successfully applied to glacierized basins in the dry Andes of central Chile [*Ragettli*, 2010], and to the Tamor basin in Eastern Nepal [*Normand*, 2010].

[15] Because parameters of the newly implemented model components rely on observations we will allow them to vary in the calibration phase. However, preliminary test runs allowed us to identify parameters to which the model simulations are poorly sensitive in the specific context of the Rhonegletscher basin. For instance, test runs showed that model results are not sensitive to Manning's roughness coefficients controlling overland and channel flows, as well as crop factors governing evapotranspiration. This is consistent with the simplicity of the drainage network and the limited area of vegetation cover in the basin. Accordingly Manning's roughness coefficients and crop factors were assigned constant values based on land use type and channel characteristics. The channel geometry in the catchment, consistently with common assumptions in the literature, is assumed to be rectangular with upstream decreasing width depending on the distance from the outlet of the catchment. The soil parameters and their ranges of variability used in the calibration phase are summarized in Table 2. Values of Manning's roughness coefficients and crop factors are listed in Table 3.

3.2. Monte Carlo Parameter Calibration

[16] In stochastic model calibration no clear strategy exists as to how many runs are enough to guarantee that the entire parameter space is sampled. Several studies have used 10,000 runs to calibrate up to 38 parameters [*Beven and Binley*, 1992; *Seibert and Beven*, 2009; *Sudret*, 2008; *Uhlenbrook and Sieber*, 2005], to name just a few. While these studies have been performed using conceptual models, we are using a physically based model, which allows us to predefine a physically plausible initial range for every model parameter. This should help limiting the number of simulations. Indeed, we could observe that the model performance, as measured through the metrics described later, stabilizes after 10,000 runs, thus suggesting a sufficient sampling from the parameter space. Accordingly, we ran 10,000 Monte Carlo simulations, each one characterized by a parameter set randomly generated from a uniformly distributed physically plausible constrained range indicated in Table 1 and 2. In order to ensure

Table 2. Initial Ranges of Soil Parameters Required by the Runoff Generation Model Component

Parameter	Unit	Glacier	Ridges	Hollows	Steep Slopes
		Soil Layer 1			
Soil depth, d_1	(m)	0.25–1.5	0.15–0.9	2.5–15	0.75–4.5
Horizontal saturated hydraulic conductivity, K_{sh1}[a]	(m h^{-1})	1e^{-5}–0.1	1e^{-5}–0.1	1e^{-7}–1e^{-3}	1e^{-7}–1e^{-3}
Vertical saturated hydraulic conductivity, K_{sv1}[b]	(m h^{-1})	0.2* K_{sh1}	0.2* K_{sh1}	0.2* K_{sh1}	0.2* K_{sh1}
Saturated soil moisture content, θ_{s1}		0.15–0.45	0.1–0.3	0.075–0.225	0.15–0.45
Residual soil moisture content, θ_{r1}[b]		0.2	0.2	0.2	0.2
		Soil Layer 2			
Soil depth, d_2	(m)	0.25–1.5	0.25–1.5	25–150	2.5–15
Horizontal saturated hydraulic conductivity, K_{sh2}[b]	(m h^{-1})	0.1* K_{sh1}	0.1* K_{sh1}	0.1* K_{sh1}	0.1* K_{sh1}
Vertical saturated hydraulic conductivity, K_{sv2}[b]	(m h^{-1})	0.2* K_{sh1}	0.2* K_{sh1}	0.2* K_{sh1}	0.2* K_{sh1}
Saturated soil moisture content, θ_{r2}[a]		0.1–0.3	0.1–0.3	0.1–0.3	0.1–0.3
Residual soil moisture content, θ_{r2}[a]		0.02	0.02	0.02	0.02

[a] K_{sh}-values were generated from a log-uniform distribution for Monte Carlo runs.
[b] In order to minimize the number of parameters to be varied some soil parameters were kept constant or were made dependent on other soil parameters.

Table 3. Summary of the Surface and Vegetation Parameter Required by the Runoff Generation Model Component of TOPKAPI

Land Use Type	Manning's Coefficients	Crop Factor
Alpine Pastures	0.035	0.7
Unproductive Vegetation	0.03	0.7
Surface Without Vegetation	0.025	0.7

that values of saturated hydraulic conductivity, K_{sh}, are uniformly distributed within several magnitudes, K_{sh} values were generated from a log-uniform distribution (see Table 2 for details), and the parameters of the second soil layer were kept dependent on values from the first soil layer.

[17] The initial ranges of all the parameters are based on available data, literature or previous modeling experiences. For instance, the ranges of the melt model parameters were chosen on the basis of the work by *Pellicciotti et al.* [2008] and *Carenzo et al.* [2009]. Values for the storage constant of the linear reservoirs in the glacier routing were taken from literature values [*Hock and Noetzli*, 1997; *Jansson et al.*, 2003; *Koch*, 2009]. As suggested by several authors [*Bandaragoda et al.*, 2004; *Eckhardt and Arnold*, 2001; *Frances et al.*, 2007], the ratios of all soil parameters of the four distinct soil types were kept constant, in order to maintain the spatially variable pattern observed in the available digital thematic maps.

[18] The calibration period for each of the 10,000 parameter sets was selected to be a hydrological year with average runoff, from October 2007 to October 2008, for which all of the three calibration variables, discharge, snow cover, and glacier mass balance, were available from observations. All calibration runs were started on 1 August 2005 in order to initialize the water content in the soil and streams, as well as the seasonal snowpack over a period of 26 months. We limited the calibration period to one representative average year only in order to challenge the model by validating it against the 14 years of observational data, including the extreme heat wave of the year 2003.

3.3. Model Performance Metric

[19] The assessment of the model performance was carried out by computing different efficiency criteria (Table 4), each one appropriate to the nature of the data sets against which the model simulations are compared, and then by using the computed efficiencies to rank the 10,000 simulations, each one corresponding to a different parameter set.

[20] The accuracy of the simulated discharge, Q_{sim}, was evaluated using the Nash-Sutcliffe efficiency criterion [*Nash and Sutcliffe*, 1970], R^2 and R^2_{log} (Table 4), computed respectively for the absolute and the logarithm value of hourly discharge. The simulated glacier mass balance, MB_{sim}, was evaluated against the reference mass balance, MB_{ref}, using the Root Mean Square Error, (RMSE), of both the ablation phase (defined from 1 May to 30 September) and the accumulation period (defined from 1 October to 30 April). To visualize RMSE on the same scale as R^2, we defined a model efficiency MB_{eff} for mass balance as the complement to 1 of the RMSE normalized with respect to the mean absolute mass change of all altitudes bands and all available years (see Table 4 for details). Finally, the models ability to reproduce daily snow cover extent was evaluated by determining the ratio of correctly predicted snow cover area and total area for which satellite images are available (CPSC) for both the entire year and the summer melt period.

[21] We ranked the 10,000 simulations by assigning rank 1 to the best simulation with respect (1) to each single efficiency criterion and subsequently (2) to aggregated criteria. For this purpose, we defined the ranking value, P^i_r, for each run, r, and each efficiency criterion, i, as

$$P^i_r = \frac{(N+1) - R^i_r}{N}, \tag{1}$$

where N is the total number of runs performed and R^i_r stands for the rank of the considered run with regard to the criterion i. Thus, ranking values were defined to measure the performance of the model simulations with regard to

Table 4. Summary of Efficiency Criteria for the Three Calibration Variables

Efficiency Criteria	Calibration Period	Equation[a]
Nash-Sutcliffe of Q, R^2	9 Oct 07 to 2 Oct 08	$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Q_{i,\mathrm{obs}} - Q_{i,\mathrm{sim}}\right)^2}{\sum_{i=1}^{n}\left(Q_{i,\mathrm{obs}} - \overline{Q_{i,\mathrm{obs}}}\right)^2}$
Nash-Sutcliffe of $\log(Q)$, R^2_{log}	9 Oct 07 to 2 Oct 08	$R^2_{\log} = 1 - \frac{\sum_{i=1}^{n}\left[\log(Q_{i,\mathrm{obs}}) - \log(Q_{i,\mathrm{sim}})\right]^2}{\sum_{i=1}^{n}\left[\log(Q_{i,\mathrm{obs}}) - \overline{\log(Q_{i,\mathrm{obs}})}\right]^2}$
Root mean square error of mass balance, $RMSE_{acc}$	1 Oct 07 to 31 Mar 08	$\mathrm{RMSE} = \sqrt{\sum_{j=1}^{m}\left(\mathrm{MB}_{j,\mathrm{ref}} - \mathrm{MB}_{j,\mathrm{sim}}\right)^2}$
Root mean square error of mass balance, $RMSE_{abl}$	1 Apr 08 to 30 Sep 08	
Model efficiency in respect to glacier mass balance, MB_{eff}	1 Oct 07 to 30 Sep 08	$\mathrm{MB}_{\mathrm{eff}} = 1 - \frac{\mathrm{RMSE}}{\mathrm{MB}_{\mathrm{mean}}}$
Correctly predicted snow cover area, $CPSC_{year}$	9 Oct 07 to 2 Oct 08	$\mathrm{CPSC} = \frac{c_{\mathrm{corr}}}{c_{\mathrm{tot}} - c_{\mathrm{missing}}}$
Correctly predicted snow cover area, $CPSC_{summer}$	15 Jun 08 to 29 Aug 08	

[a] Q_{obs} stands for observed discharge; Q_{sim} stands for simulated discharge; and the index i stands for the time step; MB_{ref} and MB_{sim} stand for the reference and simulated glacier mass change; the index j stands for the altitude band; MB_{mean} stands for the mean absolute mass change of the glacier mass balance in all altitude bands for the years 1994 to 2008 and has the value of 2093 mm w eq; c_{corr} stands for number of correctly predicted cells; c_{tot} stands for the total of cells in the catchment; and $c_{missing}$ stands for the number of cells with no data due to cloud cover.

each of the observational data set, namely $P_r^{R^2}$ and $P_r^{R^2_{\log}}$ for the discharge, $P_r^{\mathrm{CPSC}_{\mathrm{year}}}$ and $P_r^{\mathrm{CPSC}_{\mathrm{summer}}}$ for the annual and summer snow cover, and $P_r^{\mathrm{RMSE}_{\mathrm{acc}}}$ and $P_r^{\mathrm{RMSE}_{\mathrm{abl}}}$ for glacier mass balance in the accumulation and in the ablation season, respectively. Normalizing for the number of runs, N, allows the ranking to be independent from the total number of runs used. This allows in turn to compute aggregated ranking values for each observed variable – P_r^Q for discharge, P_r^{SC} for snow cover and P_r^{MB} for glacier mass balance – by simple averaging of the relevant P_r^i terms. Likewise, aggregated ranking-values, P_r^{agg} can be defined for each combination of the efficiency criterion as

$$P_r^{\mathrm{agg}} = \frac{1}{m}\sum_{k=1}^{m}\left(P_r^i\right)_k, \tag{2}$$

where m is the number of considered efficiency criteria and P_r^i can indicate both a simple or an aggregated ranking. The performance of the model with respect to all of the observational data sets considered in this study can thus be measured by the overall ranking, P_r^{OA}, defined as follows:

$$P_r^{\mathrm{OA}} = \frac{1}{3}\left(P_r^Q + P_r^{\mathrm{SC}} + P_r^{\mathrm{MB}}\right). \tag{3}$$

[22] Likewise, aggregated overall ranking-values for the combination of two observational data sets can be computed by averaging the relevant ranking values. Thus, the parameters sets with the best overall performance are those with the highest P_r^{OA}. Subsequently, the aggregated ranking values for each data set were sorted in ascending order to determine the P_r^{OA} of the best runs selected according to single or combined observational data sets. For the sake of easier representation we normalized the overall ranking, P_r^{OAnorm} as

$$P_r^{\mathrm{OAnorm}} = P_r^{\mathrm{OA}}/P_{\max}^{\mathrm{OA}}, \tag{4}$$

where $P_{\max}^{\mathrm{OA}}$ is the best ranking value of all runs. The computation of P_r^{OAnorm} allows us to quantify the concurrent performance of all criteria considered, expressing the internal consistency of the model. Hence, a comparison of the mean overall performance of the best runs selected with different data sets allows us to assess the value of each datasets and their combinations for calibration purposes.

4. Results

4.1. Model Performance

[23] The performance of the model is first investigated with respect to the calibration period, the hydrologic year 2007–2008. This is done by selecting the ensemble of the best 100 simulations (out of the 10,000) with the highest P_r^{OAnorm}, corresponding to those runs which reveal the best performance with respect to discharge, snow cover, and mass balance. Figure 2 illustrates the overall ranking criterion, P_r^{OAnorm}, together with the performance of the normalized individual ranking criteria, P_r^i, in Figure 2a, and the comparison of the observed snow cover, glacier mass balance, and discharge with the simulated ones, in Figures 2b to 2d.

[24] P_r^{OAnorm} remains above 0.85 for all runs within the ensemble, revealing a good overall performance within the selected ensemble. Equally good are the individual efficiency criteria, the mean values of which are, respectively, $R^2 = 0.60$ and $R^2_{\log} = 0.85$ for the discharge, $\mathrm{CPSC}_{\mathrm{year}} = 0.90$ for the snow cover, and $\mathrm{MB}_{\mathrm{eff}} = 0.71$ for the glacier mass balance. However, as pointed out in Table 5, the individual efficiency criteria within the ensemble vary considerably: R^2 varies between 0.37 and 0.76, $\mathrm{RMSE}_{\mathrm{acc}}$ ranges between 224 and 904 mm w eq. (equivalent to about 75% of the total $\mathrm{MB}_{\mathrm{acc}}$), whereas $\mathrm{CPSC}_{\mathrm{year}}$ fluctuates more moderately between 0.88 and 0.91. The model performs generally well for snow cover extent and glacier mass balance, but performance in regard to observed discharge is for some runs low. Hence, overall model performance can be very good, while its performance in reproducing one single variable still fluctuates significantly. Nonetheless, the performance provided by the best parameters set with respect to discharge, $R^2 = 0.76$, snow cover, $\mathrm{CPSC}_{\mathrm{year}} = 0.91$, and mass balance, $\mathrm{MB}_{\mathrm{eff}} = 0.88$ is similar to results presented in studies performed in similar contexts [*Koboltschnig et al.*, 2008; *Michlmayr et al.*, 2008].

[25] More specifically, Figure 2b shows that during wintertime, when the entire catchment is covered by snow, the efficiency criterion CPSC is practically insensitive to changes in model parameters and is always higher than 0.98. During the transition seasons, autumn and spring, when the first snowfalls occur and snowmelt leads to increasing snow free areas, CPSC is more sensitive to the choice of model parameters and drops more frequently to lower values, leading to average monthly values in spring of ~0.75. Nevertheless, the annual values of $\mathrm{CPSC}_{\mathrm{year}}$ remain above 0.88 in all runs (Table 5). Similarly, Figure 2c illustrates that the mean of the ensemble simulations of mass balance is slightly overestimating accumulation and ablation during the two respective seasons. However, the reference mass balance is within 1 standard deviation (SD). Finally, the seasonal dynamics of the simulated discharge is predicted reasonably well by the 100 ensemble parameter set (Figure 2d). However, in spring and during heavy precipitation events the simulated discharge deviates from the observed one. This may be explained by the use of a constant temperature lapse rate, instead of seasonally varying temperature lapse rates measured in nearby valleys and presented in several studies in the European Alps [*Carrega*, 1995; *Rolland*, 2003; *Rüedinger*, 2010]. An implementation of a seasonally varying temperature lapse rate may improve model performance, but is not within the scope of this study. Furthermore, the meteorological data to drive the model is recorded outside the basin (Figure 1) and thus observed weather patterns might not always be representative for the entire Rhonegletscher basin.

[26] While the ensemble selected on the basis of good performance in all three available data sets (Q, SC, and MB) leads to good overall performance, ensembles selected regarding only one of the available data set yield better value of the efficiency criterion for that single variable, but show poor performance for the other variables. As shown in Table 5, the ensemble of 100 parameter selected on the basis of discharge data only ($R^2 = 0.76 - 0.84$), result in low values of $\mathrm{RMSE}_{\mathrm{acc}}$, $\mathrm{RMSE}_{\mathrm{abl}}$ and $\mathrm{CPSC}_{\mathrm{summer}}$. Likewise, the parameter sets that provide the highest performance in terms of glacier mass balance show poorer performance with respect to runoff ($R^2 = 0.04 - 0.78$) and snow cover

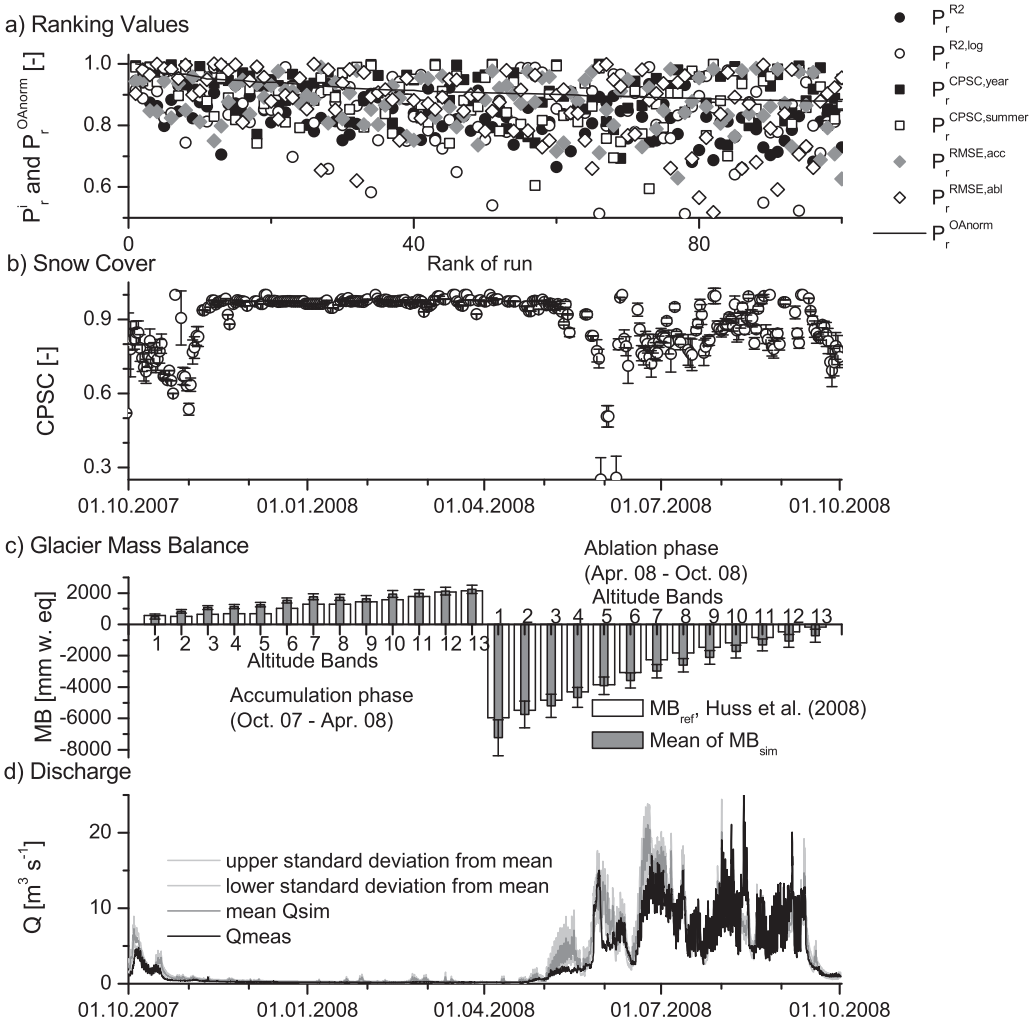

Figure 2. Model performance of the 100 MC simulations with the best overall model performance P_r^{OAnorm}. (a) Specific raking values, P_r^i, for all 6 ranking criteria and overall ranking, P_r^{OAnorm}, for the 100 best parameter sets. (b) The mean fraction of correctly predicted daily snow cover, CPSC. (c) The mean mass balance, MB, for 13 altitude bands (every 100 m elevation starting at 2100 m asl) are compared to the reference mass balance determined by *Huss et al.* [2008]. (d) Hydrographs of the 100 selected parameter sets. Error bars in Figures 2b and 2c illustrate the SD from the mean.

($CPSC_{summer}$ = 0.72 – 0.85), and those that best simulate the snow cover show poorer performance for discharge and glacier mass balance.

[27] This result points at the importance of developing a calibration procedure which allows identifying the parameter set which leads to a high internal consistency of the model. Figure 3 shows that the availability of multiple observational data sets and the use of an aggregated efficiency criteria, e.g., that defined by equation (2), can lead to the identification of the parameter set that, in a given basin context, yield the best overall model performance. Furthermore, comparing the mean overall performance, P^{OAnorm}, of model simulations selected by individual data sets or their combinations allows us to identify the value of

Table 5. Value of Efficiency Criteria of the 100 Best Parameter Sets With Respect to the Ranking Value Listed in the Header Line

Efficiency Criteria	Runs With Best P_r^{OAnorm}	Runs With Best P_r^Q	Runs With Best P_r^{SC}	Runs With Best P_r^{MB}
R^2	0.37–0.76	**0.76–0.84**[b]	–2–0.74	0.04–0.78
$R^2_{\log}$	0.73–0.91	**0.89–0.93**[b]	NA– 0.88[a]	NA– 0.91[a]
$RMSE_{acc}$	224–904	295–2693	232–1840	**239–488**[b]
$RMSE_{abl}$	261–2013	407–2509	355–5308	**265–918**[b]
$CPSC_{year}$	0.885–0.905	0.778–0.880	**0.901–0.907**[b]	0.857–0.905
$CPSC_{summer}$	0.791–0.857	0.483–0.764	**0.847–0.861**[b]	0.720–0.854

[a]For these parameter sets Q dropped for certain time periods to 0 and accordingly it was not possible to compute a $R^2_{\log}$.
[b]Bold fonts designate values of efficiency of parameters sets selected with the data set associated to the specific efficiency criteria of the data set.

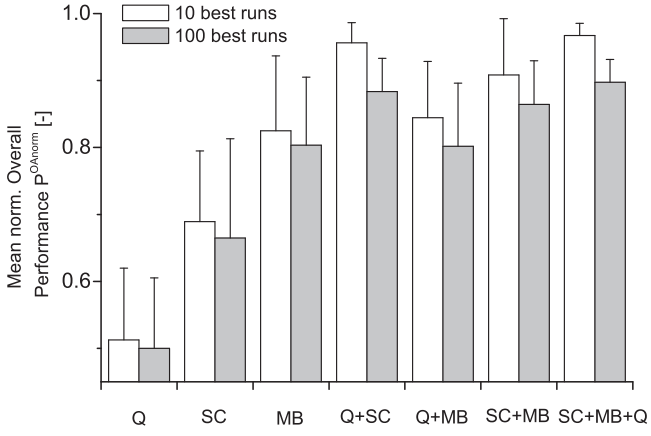

Figure 3. Mean overall performance, P^{OAnorm}, of the 10 and the 100 best runs selected with the observational data set indicated on the abscissa. The statistic is shown for the 10 and the 100 best runs in order to illustrate the consistency of the results, regardless of the number of runs within the selected ensemble. The error bars illustrate the SD from the mean.

observational data sets for model calibration. For the specific case study of the Rhonegletscher, Figure 3 suggests that the overall model performance is low when parameter sets are selected on the basis of discharge data only. Conversely, overall model performance is significantly higher when parameter sets are selected using glacier mass balance, which leads to a value of the overall efficiency criteria, P_r^{OAnorm}, of about 0.8. However, if discharge, snow cover, and glacier mass balance are simultaneously used to constrain the model, P_r^{OAnorm} increases significantly compared to constraining only to glacier mass balance, but less significantly when discharge and snow cover are the constraining data sets. In other words, discharge and snow cover data are in this particular case sufficient to identify parameter sets which lead to an internally consistent model.

4.2. Plausibility of Calibrated Parameters

[28] The robustness of the selected ensembles from the multivariable calibration was tested by assessing the plausibility of some calibrated parameters, which play a key role in the context of alpine catchments. Table 6 shows the basic criteria of each calibrated model parameter as obtained from the 100 simulations with the highest value of the efficiency criterion, P_r^{OAnorm}. Considering the key parameters controlling snow cover we could observe that their mean values agree quite well with the values obtained for the nearby Haut Glacier d'Arolla in a number of recent studies. For instance, the temperature gradient, T_{grad}, is similar to that estimated on observed data at the Haut Glacier d'Arolla by *Carenzo et al.* [2009] and at the Gornergletscher by *Kretz* [2007]. The mean values for the albedo factors, $p_1 = 0.87$ and $p_2 = 0.15$ (Table 6), are similar to those obtained by *Pellicciotti et al.* [2005] for the Haut Glacier d'Arolla. The mean values of the storage constants for the linear reservoirs of the snow- and ice-melt routing, K_{snow} and K_{ice}, are also comparable to those obtained in previous and more detailed studies [*Koch*, 2009]. Conversely, the snow- and ice-melt model parameters, TF and SRF, deviate from the values obtained by *Carenzo et al.* [2009] and show also a relatively high variability. Such deviation can be, however, explained considering that in this study the value of the threshold air temperature for melt, T_t, is also calibrated, whereas *Carenzo et al.* [2009] kept it fixed.

Table 6. Mean Values of Model Parameters for MC Runs With the Best Overall Performance

	10 Best Runs[b]		100 Best Runs[b]	
Model Parameter x	Mean Value, $\bar{x}$[a]	SD, σ	Mean Value, $\bar{x}$[a]	SD, σ
T_{t}	−1.01	0.57	−0.77	0.72
$T_{\mathrm{t,P}}$	−1.01	0.57	−0.77	0.72
T_{onset}	−1.01	0.57	−0.77	0.72
T_{grad}	0.0061	0.0007	0.0058	± 0.0008
T_{mod}	1.57	0.40	1.27	0.53
SRF	0.0040	0.0031	0.0047	0.0033
TF	0.247	0.058	0.224	0.083
p_1	0.869	0.078	0.871	0.088
p_2	0.148	0.031	0.150	0.027
α_{G}	0.229	0.114	0.216	0.111
α_{ground}	0.248	0.031	0.250	0.031
K_{snow}	59.3	23.5	55.7	26.4
K_{ice}	52.8	32.4	53.62	27.28
$P_{\mathrm{prec}}^{\mathrm{su}}$	0.014	0.003	0.015	0.004
$P_{\mathrm{prec}}^{\mathrm{wi}}$	0.075	0.014	0.083	0.020
d_1[c]	1.7 times BEK values	1.7 times BEK values	1.7 times BEK values	1.7 times BEK values
θ_{s1}[c]	1.06 times BEK values	1.06 times BEK values	1.05 times BEK values	1.05 times BEK values
K_{sh1}[c]	0.03 times initial values	0.03 times initial values	0.12 times initial values	0.12 times initial values
K_{sv1}[c]	0.03 times initial values	0.03 times initial values	0.12 times initial values	0.12 times initial values
K_{sh2}[c]	0.03 times initial values	0.03 times initial values	0.12 times initial values	0.12 times initial values
θ_{s2}	0.177	0.044	0.193	0.058
d_2[c]	1.51 times initial values	1.51 times initial values	1.66 times initial values	1.66 times initial values

[a] $\bar{x}$ denotes the average of the corresponding model parameter and σ illustrates the SD of the mean.
[b] The statistic is shown for the 10 and the 100 best runs in order to illustrate the consistency of the results, regardless of the number of runs within the selected ensemble.
[c] Soil parameters were varied while keeping the ratios of the different soil types constant. Accordingly, we present the mean factor used to adjust the soil parameters of each soil type, rather than presenting the mean value for each soil type.

[29] We further investigated the parameters controlling the snow and ice melt by plotting their values for each observational data set used in calibration. Figure 4 shows the box plots of five selected parameters, namely the temperature gradient, T_{grad}, the snow- and ice-melt parameters, TF and SRF, and the precipitation gradient in summer and in winter, respectively, P^{su}_{prec} and P^{wi}_{prec}. The boxes are limited by the upper and lower quartiles of the parameter distribution, while the thick line in the box shows the median. All of the parameters, show a high variability, regardless of the constraining observational data set, with the exception of P^{su}_{prec} and P^{wi}_{prec}, which are quite stable when constrained by glacier mass balance or by a combination of data sets. The two key parameters controlling the melt runoff, TF and SRF, seem to be best constrained by discharge observations, whereas the inclusion of snow cover and mass balance increases significantly the range of the estimated parameters. These two parameters are controlling short term fluctuations, and accordingly may be better represented by the discharge than by the snow cover or mass balance, which are mainly characterized by a seasonal dynamics. Similar controls on the other model parameters, though not shown here, seem to corroborate the conclusion that considering appropriate constraining data sets in the calibration phase leads to more plausible values of the parameters, the estimates of which are more physically meaningful than those obtained constraining only against the discharge.

4.3. Validation for a 14 year Period

[30] To better substantiate the conclusions out of the multivariable calibration we tested the predictive skills of the selected ensembles by simulating the period 1994–2007 and validating the model against both single observational data sets and their combinations. While discharge data and glacier mass balance were available for the entire time period, the model could be validated against snow cover data only after 2001. We performed model simulation for the 10 best parameter sets with respect to single or combinations of the observational data sets. This helped to keep computational time at a reasonable level without compromising the rigorousness of the assessment, since model parameters and the efficiency criteria computed for 10 and 100 best simulations are not significantly different, as shown in Figure 3 and Table 6.

[31] Figure 5 shows the time series of the efficiency criteria for the 14 year validation period as computed from simulations constrained to single data sets or to their combinations. The results substantially confirm those obtained for the calibration year. While mass balance allows to identify parameter sets leading to good simulation of discharge ($R^2 = 0.79$), very low values of the Nash-Sutcliffe criterion ($R^2 = 0.05$) are obtained by simulations with parameters sets constrained by snow cover images only. $CPSC_{year}$ and MB_{eff} are also poorly predicted by runs parameterized with constraining discharge data only. Thus, constraining the model to a single variable improves the model performance

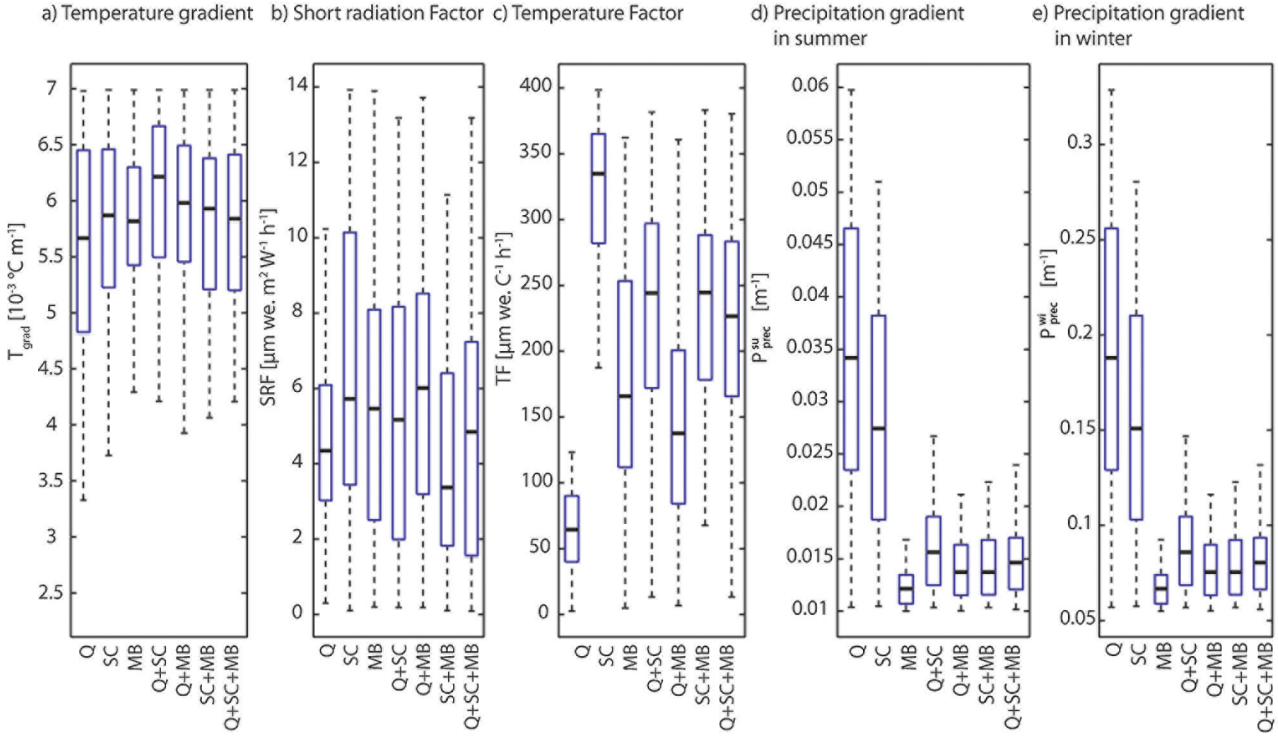

Figure 4. Box plots of selected model parameters of the best 100 MC runs during the calibration period selected in respect to the observational data set indicated on the abscissa. The boxes are limited by the upper and lower quartiles of the parameter distribution, while the thick line in the box shows the median. Whiskers extend from each end of the box to the most extreme values within 1.5 times the interquartile range from the ends of the box.

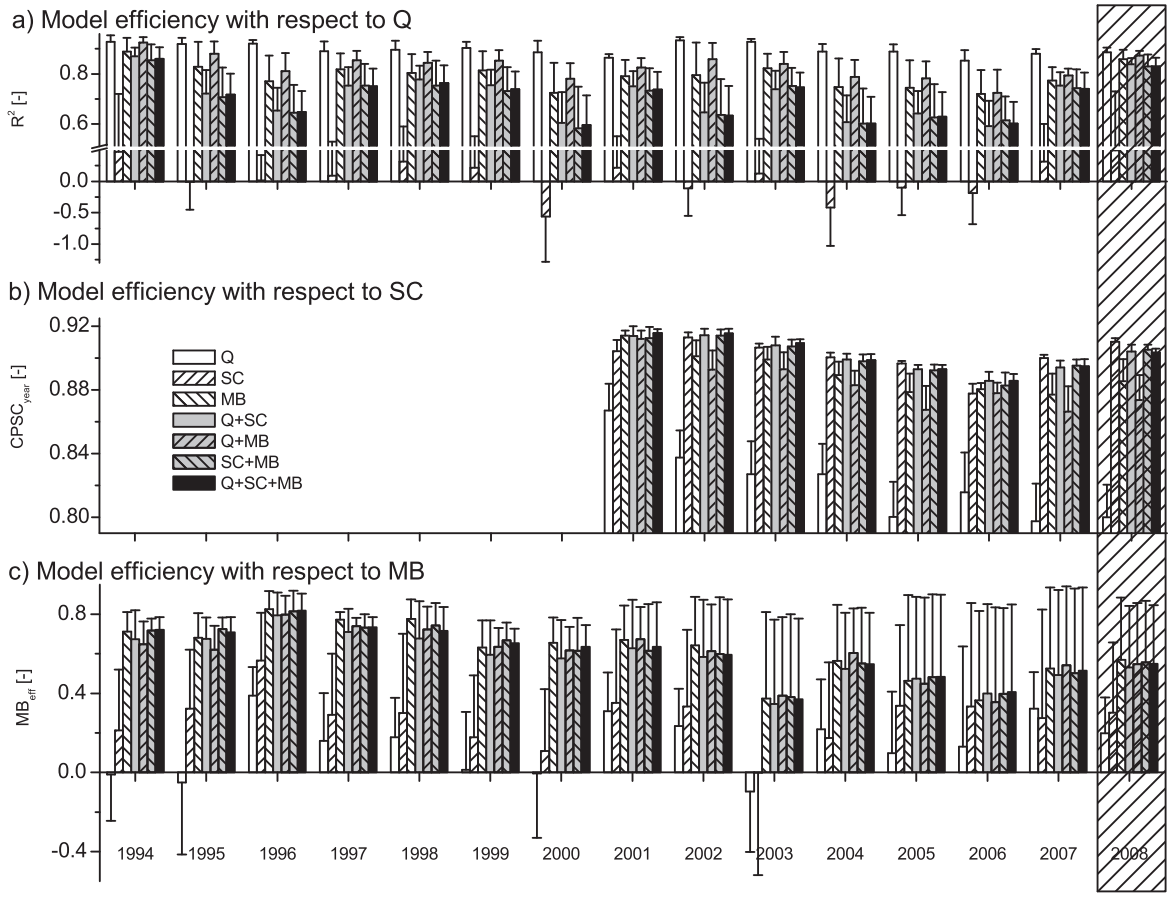

Figure 5. Mean model efficiency ((a) Nash-Sutcliffe value; (b) correctly computed snow cover extent; (c) mass balance efficiency) during the validation period (1994–2007) and calibration period (2008) of the 10 best MC runs (MODIS snow cover data are only available starting in 2001) selected according the observational data sets indicated in the legend. Striped area marks the calibration period (9 October 2007 to 2 October 2008). The error bars illustrate the SD from the mean.

with respect to that specific variable, but provides poor model performance for the remaining variables. The combinations of two data sets significantly improve the efficiency criteria of the third criterion in each individual year (Figure 5) and for the entire period (Figure 6). All parameter sets selected by concurrent constraining on two or more data sets provide values of the efficiency criteria, which are higher than 0.7 for R^2, 0.6 for MB_{eff} and 0.88 for $CPSC_{year}$.

[32] The somewhat limited variability of the mass balance efficiency criterion, MB_{eff}, over the 14 year validation period also indicates that the simplifying assumption of a constant glacier surface area had a limited impact on model results and does not affect the main objectives of this study. The multivariable calibration strategy also seems to lead to a model parameterization that allows the model to perform well during particularly extreme years, such as 2003, which was characterized by a heat wave that affected the entire European continent [*Schär et al.*, 2004], leading to a local increase of mean air temperatures of 1.1°C compared to the 14 year average. This extreme condition shows that a general decrease of the glacier mass balance efficiency criterion in 2003, regardless of the constraining data set, does not affect significantly the model performance with respect to the other simulated variables.

5. Discussion

[33] The main objective of this study was to evaluate the value of three observational data sets in (1) increasing model performance and (2) constraining model parameters by the means of Monte Carlo simulations. Nevertheless, the proposed Monte Carlo approach with a limited number of parameter sampling, does not explore likely the entire parameter space and does not represent a solution to the well-known equifinality problem [*Beven and Binley*, 1992; *Beven*, 2006]. It rather shows that, whenever available, multiple data sets should be used to aim at internally consistent calibration of physically based hydrological models.

[34] In an attempt to tackle the problem of equifinality we investigated the presence of interdependences among parameters. Some of these were indeed identified by a simple cross correlation analysis of the calibrated parameters for the ensemble with the highest P^{OAnorm} and showed to be consistent with results shown by *Carenzo et al.* [2009]. Statistically significant correlations ($P < 0.05$) in the 100 best runs were found between SRF and TF (temperature effects can be compensated by short wave radiation effects), P_{prec} and SRF (enhanced precipitation can be compensated with increased melt), T_{mod} and P_{prec} (increased

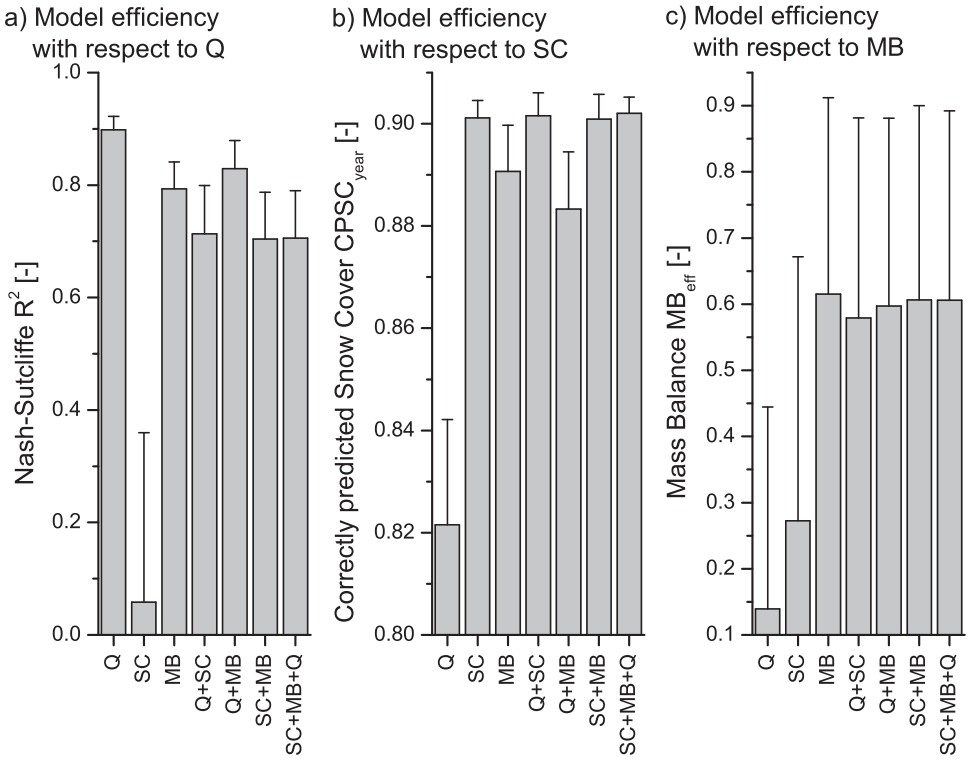

Figure 6. Model efficiency over the period 1994–2008 of the 10 best model runs selected using the observational data sets indicated in the legend. The error bars illustrate the SD from the mean. Note the different scale of the ordinates.

temperatures on the glacier can be compensated with enhanced precipitation). There is, however, no clear physically based dependence that would allow defining one as a function of the other, thus justifying removing some of them from the pool of calibrated parameters. Thus, as different model variables are sensitive to different parameters, we preferred to vary all relevant model parameters, which govern snow- and ice-melt, glacier mass balance, and discharge. A further advantage of varying all parameters is that the range of model performance reveals the uncertainty of the ensemble of best parameters sets.

[35] The most evident result is that using one single observational data set for parameter identification does not guarantee internal consistency, even in the case of physically based distributed models, because some processes may not be correctly simulated. In the specific context of Alpine basins, where glacier mass balance and snow cover are key variables, we could show that the combination of discharge and snow cover data sets has the highest constraining effects and leads to best overall model performance. This can be explained analyzing the information content of the three data sets. While hourly discharge records at the catchment outlet contain substantial information about the short-term temporal variability of the integrated response of a catchment, they do not provide evidence of the spatial variability of processes. Conversely, daily snow cover maps from satellite images represent well the spatial variability of snow cover extent also providing temporal information in the form of seasonal dynamics. The value of the latter depends on the considered area: in regions with frequent snow cover change the information content of SC is obviously more valuable than in regions with little variations in snow cover. Nevertheless, because the MODIS images only provide binary information—snow covered versus snow free—they do not contain any volumetric information about melt runoff generation. Our results demonstrate, however, that MODIS images in combination with discharge records can significantly improve model calibration. These results are consistent with previous works, which also demonstrated that snow cover images can improve model efficiency [*Immerzeel et al.*, 2009; *Parajka and Blöschl*, 2008], referring, however, to the calibration of conceptual or lumped hydrological models, where the need of constraining conceptual parameters is essential.

[36] The glacier mass balance reflects mainly spatial changes in precipitation during the accumulation phase, thus the integral of distributed solid precipitation. We observed three main effects in the calibration procedure. The first one was to constrain the precipitation gradient parameter, P_{prec}, as shown in Figure 4. The second effect was that the glacier mass balance information becomes redundant to improve overall model performance when discharge and snow cover data are used, as shown in Figure 3. This is not surprising given that spatial, volumetric and temporal information are included in discharge and snow cover data. Moreover, as MB measurements are only taken twice in the year, the information about short term variability is limited.

[37] For the specific case of the Rhonegletscher, mass balance data were available for every 100 m altitude band on the glacier containing a high degree of spatial information, as the time series was determined by means of a

model-based reconstruction. This suggests that data reconstructed by means of an independent model, representing adequately observed processes, can be used for calibration purposes. Finally, our results suggest that, in the context of Alpine regions with high degree of glaciation, glacier mass balance with a high spatial resolution may lead to a satisfactory model calibration of a physically based model. This is in agreement with recent studies about the value of glacier mass balance to calibrate hydrological models [*Koboltschnig et al.*, 2008; *Konz and Seibert*, 2010], which, however, are of conceptual nature.

[38] With respect to the effectiveness of multivariable calibration in constraining parameters, it is interesting to note that the most effective result was obtained for P_{prec}, which is essentially an empirical parameter, given the complexity of the space-time variability of precipitation in orographically complex regions. Conversely, all of the other model parameters varied considerably within the best 100 parameter sets. However, when combined observational data sets were used for parameter selection, some parameters, e.g., SRF and TF, seem to converge in the best 100 sets to a more restricted range of variability than that imposed a priori (see section 4.2). In order to reduce the variability, only more powerful constraining data sets such as distributed observational data sets of temperature and snow depth, can provide a more effective alternative.

6. Conclusion

[39] The fully distributed, physically based hydrological model TOPKAPI was used to investigate the predictive power of hourly discharge (Q), daily satellite snow cover images (SC) and biannual glacier mass balance (MB) and their combination to select consistent and reliable parameter sets from a uniform predefined range in the context of a Alpine basin. We could observe that model constrainment based on multiple data sets leads (1) to an increase of overall model performance as compared to estimation based on a single data set, (2) to parameter estimates partially converging to values that are comparable to physical plausible values, and, ultimately (3) to a set of parameters which are characterized by a high overall model performance, indicating a high internal consistency.

[40] The study showed that, despite the physically based nature of the model, parameter sets estimated from single observational data sets, regardless which, provide satisfactory simulations of the target variable used for selection, but are not sufficient to guarantee internally consistent simulations. The result seem to be particularly interesting, not only because it is often claimed that physically based models do not require, in principle, calibration, but also because the use of internally inconsistent models to simulate climate scenarios may lead to questionable predictions.

[41] We also observed that the type of information content (spatial, temporal, or volumetric) of datasets used for calibration should be complementary to each other in order to improve overall model performance. Data sets with different and complementary space, time and melt water volume information content (e.g., subdaily discharge provides volume and temporal information and daily snow cover provides spatial information) should be preferred to data sets, which supply similar information (e.g., glacier mass balance and discharge both provide volumetric information). In the context of the investigated Alpine case study, we could in fact show that the cumulative effect of snow cover satellite images and discharge data lead to similarly high model overall performance as if glacier mass balance were used as well. This opens new perspectives with regard to the application of complex physically based hydrological models in remote basins with limited data availability as increasingly available remotely sensed data sets can improve model performance in such areas.

[42] Alternatively, single datasets which contain spatial, temporal and volume information may also lead to high overall model performance. In the context of our specific basin, the particularly dense set of glacier mass balance observations led alone to a better model performance than discharge or snow cover images alone. This suggests that in Alpine basins with high degree of glaciation, glacier mass balance may also lead to satisfactory model calibration.

[43] While we could demonstrate the importance of multiple data sets to achieve an internally consistent calibrated model, the three presented data sets did not contain the necessary amount of spatial, volumetric and temporal information for an equifinality-free model calibration. We are presently investigating the issue of equifinality by assessing the added value of distributed and highly resolved ground measurements together with remotely sensed observations in a follow-up study.

[44] **Acknowledgments.** The present study is part of the ACQWA Project (Assessing climate impacts on the quantity and quality of water), funded within the seventh Framework Program of the European Union (grant agreement 212,250). We would like to thank the three anonymous reviewers who provided helpful comments to improve the manuscript. We finally thank A. Bauder (VAW, ETH Zurich) for the glacier mass balance data and P. Molnar (IfU, ETH Zurich) for many valuable discussions on statistical methods and calibration approaches.

References

Bandaragoda, C., D. G. Tarboton, and R. Woods (2004), Application of TOPNET in the distributed model intercomparison project, *J. Hydrol.*, *298*(1–4), 178–201, doi:10.1016/j.jhydrol.2004.03.038.

Beldring, S. (2002), Multi-criteria validation of a precipitation-runoff model, *J. Hydrol.*, *257*(1–4), 189–211.

Bergstrom, S., G. Lindstrom, and A. Pettersson (2002), Multi-variable parameter estimation to increase confidence in hydrological modelling, *Hydrol. Processes*, *16*(2), 413–421, doi:10.1002/hyp.332.

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. Processes*, *6*(3), 279–298.

Brock, B. W., I. C. Willis, and M. J. Sharp (2000), Measurement and parameterization of albedo variations at Haut Glacier d'Arolla, Switzerland, *J. Glaciol.*, *46*(155), 675–688.

Campo, L., F. Caparrini, and F. Castelli (2006), Use of multi-platform, multi-temporal remote-sensing data for calibration of a distributed hydrological model: An application in the Arno basin, Italy, *Hydrol. Processes*, *20*(13), 2693–2712, doi:10.1002/hyp.6061.

Cao, W. Z., W. B. Bowden, T. Davie, and A. Fenemor (2006), Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability, *Hydrol. Processes*, *20*(5), 1057–1073, doi:10.1002/hyp.5933.

Carenzo, M., F. Pellicciotti, S. Rimkus, and P. Burlando (2009), Assessing the transferability and robustness of an enhanced temperature-index glacier-melt model, *J. Glaciol.*, *55*(190), 258–274.

Carrega, P. (1995), A method for the reconstruction of mountain air temperatures with automatic cartographic applications, *Theor. Appl. Climatol.*, *52*(1–2), 69–84.

Ciarapica, L., and E. Todini (2002), TOPKAPI: A model for the representation of the rainfall-runoff process at different scales, *Hydrol. Processes*, *16*, 207–229, doi:10.1002/hyp.342.

Corripio, J. G. (2003), Vectorial algebra algorithms for calculating terrain parameters from DEMs and solar radiation modelling in mountainous terrain, *Int. J. Geogr. Inf. Sci.*, *17*(1), 1–23, doi:10.1080/13658810210157796.

Crow, W. T., and D. Ryu (2009), A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, *Hydrol. Earth Syst. Sci.*, *13*(1), 1–16.

De Bruin, H. A. R., and J. N. M. Stricker (2000), Evaporation of grass under non-restricted soil moisture conditions, *Hydrol. Sci. J.*, *45*(3), 391–406.

Eckhardt, K., and J. G. Arnold (2001), Automatic calibration of a distributed catchment model, *J. Hydrol.*, *251*(1–2), 103–109.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resour. Res.*, *45*, W06427, doi:10.1029/2008WR007255.

Frances, F., J. I. Velez, and J. J. Velez (2007), Split-parameter structure for the automatic calibration of distributed hydrological models, *J. Hydrol.*, *332*(1–2), 226–240, doi:10.1016/j.jhydrol.2006.06.032.

Franks, S. W., P. Gineste, K. J. Beven, and P. Merot (1998), On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, *Water Resour. Res.*, *34*(4), 787–797, doi:10.1029/97WR03041.

Grayson, R. B., and G. Blöschl (2000), Spatial modelling of catchment dynamics, in *Spatial Patterns in Catchment Hydrology: Observations and Modelling*, edited by R. B. Grayson and G. Blöschl, pp. 51–81, Cambridge University Press, Cambridge, U. K.

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.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763, doi:10.1029/97WR03495.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503, doi:10.1029/1999JD900154.

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.

Hock, R., and C. Noetzli (1997), Areal melt and discharge modelling of Storglaciaren, Sweden, in *Annals of Glaciology, 24*, edited by I. M. Whillans, pp. 211–216, Int. Glaciological. Soc., Cambridge, U. K.

Huss, M., A. Bauder, M. Funk, and R. Hock (2008), Determination of the seasonal mass balance of four alpine glaciers since 1865, *J. Geophys. Res.*, *113*(1), F01015, doi:10.1029/2007jf000803.

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.

Immerzeel, W. W., P. Droogers, S. M. de Jong, and M. F. P. Bierkens (2009), Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing, *Remote Sens. Environ.*, *113*(1), 40–49, doi:10.1016/j.rse.2008.08.010.

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.

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. Processes*, *23*(21), 3093–3109, doi:10.1002/hyp.7421.

Koboltschnig, G. R., W. Schoner, M. Zappa, C. Kroisleitner, and H. Holzmann (2008), Runoff modelling of the glacierized alpine Upper Salzach basin (Austria): Multi-criteria result validation, *Hydrol. Processes*, *22*(19), 3950–3964, doi:10.1002/hyp.7112.

Koch, N. (2009), Nonlinear analysis of melted water and related runoff in glacierised catchments, M.S. thesis, 111 pp., Swiss Fed. Inst. of Technol., Zurich, Switzerland.

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.

Konz, M., S. Uhlenbrook, L. Braun, A. Shrestha, and S. Demuth (2007), Implementation of a process-based catchment model in a poorly gauged, highly glacierized Himalayan headwater, *Hydrol. Earth Syst. Sci.*, *11*(4), 1323–1339.

Kretz, A. (2007), Modelling spatial and temporal variations in surface melt on Gornergletscher, M.S. thesis, 101 pp., Swiss Fed. Inst. of Technol., Zurich, Switzerland.

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.

Lindstrom, G. (1997), A simple automatic calibration routine for the HBV model, *Nord. Hydrol.*, *28*(3), 153–168.

Liu, Z. Y., M. L. Martina, and E. Todini (2005), Flood forecasting using a fully distributed model: Application of the TOPKAPI model to the Upper Xixian catchment, *Hydrol. Earth Syst. Sci.*, *9*(4), 347–364.

Makkink, G. F. (1957), Testing the Penman formula by means of lysimeters, *J. Inst. Water Eng.*, *11*, 277–288.

Michlmayr, G., M. Lehning, G. Koboltschnig, H. Holzmann, M. Zappa, R. Mott, and W. Schoner (2008), Application of the Alpine 3D model for glacier mass balance and glacier runoff studies at Goldbergkees, Austria, *Hydrol. Processes*, *22*(19), 3941–3949, doi:10.1002/hyp.7102.

Motovilov, Y. G., L. Gottschalk, K. Engeland, and A. Rodhe (1999), Validation of a distributed hydrological model against spatial observations, *Agric. For. Meteorol.*, *98*(9), 257–277.

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.

Normand, S. (2010), Effects of different meteorological input data on runoff prediction in the Tamor basin in Eastern Nepal, M.S. thesis, Swiss Fed. Instit. of Technol., Zurich, Switzerland.

Parajka, J., and G. Blöschl (2006), Validation of MODIS snow cover images over Austria, *Hydrol. Earth Syst. Sci.*, *10*(5), 679–689.

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.

Parajka, J., V. Naeimi, G. Blöschl, and J. Komma (2009), Matching ERS scatterometer based soil moisture patterns with simulations of a conceptual dual layer hydrologic model over Austria, *Hydrol. Earth Syst. Sci.*, *13*(2), 259–271.

Paul, F., H. Escher-Vetter, and H. Machguth (2009), Comparison of mass balances for Vernagtferner, Oetzal Alps, as obtained from direct measurements and distributed modeling, *Ann. Glaciol.*, *50*, 169–177, doi:10.3189/172756409787769582.

Pellicciotti, F., B. Brock, U. Strasser, P. Burlando, M. Funk, and J. Corripio (2005), An enhanced temperature-index glacier melt model including the shortwave radiation balance: Development and testing for Haut Glacier d'Arolla, Switzerland, *J. Glaciol.*, *51*(175), 573–587.

Pellicciotti, F., J. Helbing, A. Rivera, V. Favier, J. Corripio, J. Araos, J. E. Sicart, and M. Carenzo (2008), A study of the energy balance and melt regime on Juncal Norte Glacier, semi-arid Andes of central Chile, using melt models of different complexity, *Hydrol. Processes*, *22*(19), 3980–3997, doi:10.1002/hyp.7085.

Pellicciotti, F., T. Raschle, T. Huerlimann, M. Carenzo, and P. Burlando (2011), Transmission of solar radiation through clouds on melting glaciers: A comparison of parameterisations and their impact on melt modelling, *J. Glaciol.*, *57*(202), 367–381.

Ragettli, S. (2010), Modelling the runoff regime of the Aconcagua River Basin using a distributed hydrological model: Simulations of glacier and snow melt contributions to streamflow, M.S. thesis, 125 pp., Swiss Fed. Inst. of Technol., Lausanne, Switzerland.

Rolland, C. (2003), Spatial and seasonal variations of air temperature lapse rates in alpine regions, *J. Clim.*, *16*(7), 1032–1046.

Rüedinger, D. (2010), Charakterisierung von Temperatur- und Niederschlagszonen im Rhonetal, B.S. thesis, Swiss Fed. Inst. of Technol., Zurich, Switzerland.

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., and K. J. Beven (2009), Gauging the ungauged basin: How many discharge measurements are needed?, *Hydrol. Earth Syst. Sci.*, *13*(6), 883–892.

Sevruk, B. (2004), *Niederschalg als Wasserkreislaufelement*, 86 pp., Nitra, Slovakia.

Sevruk, B., and K. Mieglitz (2002), The effect of topography, season and weather situation on daily precipitation gradients in 60 Swiss valleys, *Water Sci. Technol.*, *45*(2), 41–48.

Sinclair, S., and G. G. S. Pegram (2010), A comparison of ASCAT and modelled soil moisture over South Africa, using TOPKAPI in land surface mode, *Hydrol. Earth Syst. Sci.*, *14*, 613–626.

Sorman, A. A., A. Sensoy, A. E. Tekeli, A. U. Sorman, and Z. Akyurek (2009), Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey, *Hydrol. Processes*, *23*(7), 1031–1040, doi:10.1002/hyp.7204.

Sorooshian, S., and V. Gupta (1995), Model calibration, in *Computer Models of Watershed Hydrology*, edited by V. P. Singh, pp. 23–68, Water Resour. Publ., Highlands Ranch.

Sudret, B. (2008), Global sensitivity analysis using polynomial chaos expansions, *Reliability Eng. Syst. Saf.*, *93*(7), 964–979, doi:10.1016/i.ress.2007.04.002.

Thornthwaite, C. W. (1948), An approach toward a rational classification of climate, *Geogr. Rev.*, *38*(1), 55–94.

Todini, E., and L. Ciarapica (2001), The TOPKAPI model, in *Mathematical Models of Large Watershed Hydrology*, edited by V. P. Singh, Water Resour. Publ., Littleton, Colo.

Uhlenbrook, S., and A. Sieber (2005), On the value of experimental data to reduce the prediction uncertainty of a process-oriented catchment model, *Environ. Modell. Softw.*, *20*(1), 19–32, doi:10.1016/j.envsoft.2003.12.006.

Wagner, W., N. E. C. Verhoest, R. Ludwig, and M. Tedesco (2009), Editorial "Remote sensing in hydrological sciences," *Hydrol. Earth Syst. Sci.*, *13*(6), 813–817.

P. Burlando, M. Konz, F. Pellicciotti, and S. Rimkus, Institute of Environmental Engineering, ETH Zurich, Wolfgang-Pauli-Str. 15, CH-8093 Zurich, Switzerland.

D. Finger, Institute of Geography, University of Bern, Hallerstr. 12, CH-3012 Bern, Switzerland. (david.finger@giub.unibe.ch)