Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., Güntner, A. (2014): The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia. - *Water Resources Research*, *50*, 3, p. 2002-2021

DOI: http://doi.org/10.1002/2013WR014382

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2013WR014382

Key Points:
- Calibration results in small trade-offs between discharge and snow cover
- Snow cover errors may be large when neglecting snow cover during calibration
- Value of increasing the number of snow cover scenes for calibration is studied

Correspondence to:
D. Duethmann,
doris.duethmann@gfz-potsdam.de

Citation:
Duethmann, D., J. Peters, T. Blume, S. Vorogushyn, and A. Güntner (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.

Received 4 JULY 2013
Accepted 7 FEB 2014
Accepted article online 12 FEB 2014
Published online 7 MAR 2014

The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia

Doris Duethmann[1], Juliane Peters[2], Theresa Blume[1], Sergiy Vorogushyn[1], and Andreas Güntner[1]

[1]GFZ German Research Centre for Geosciences, Section 5.4 Hydrology, Potsdam, Germany, [2]Institute for Cartography, Dresden University of Technology, Dresden, Germany

Abstract Including satellite-derived snow cover data for hydrologic model calibration can be a good way to improve model internal consistency. This study applied a multiobjective genetic algorithm to characterize the trade-off curve between model performance in terms of discharge and snow cover area (SCA). Using a Monte Carlo-based approach, we further investigated the additional information content of an increasing number of SCA scenes used in the calibration period. The study was performed in six snowmelt-dominated headwater catchments of the Karadarya Basin in Kyrgyzstan, Central Asia, using the hydrological model WASA and snow cover data from four melt seasons retrieved from AVHRR (Advanced Very High Resolution Radiometer). We generally found only small trade-offs between good simulations with respect to discharge and SCA, but good model performance with respect to discharge did not exclude low performance in terms of SCA. On average, the snow cover error in the validation period could be reduced by very few images in the calibration period. Increasing the number of images resulted in only small further improvements. However, using only a small number of images involves the risk that these particular images cause the selection of parameter sets which are not representative for the catchment. It is therefore advisable to use a larger number of images. In this study, it was necessary to include at least 10–16 images.

1. Introduction

Mountain water resources play an important role for the water supply of downstream areas [*Viviroli et al.*, 2011]. Particularly in dry regions with large populations, such as Central Asia, they are essential for irrigation, hydropower generation, and sustaining ecosystems that depend directly or indirectly on the river flow. Potential impacts of climate change on mountain water resources are thus of great concern for water management. Assessments of the hydrologic impacts are required for developing strategies to cope with changing conditions, which is often approached using hydrological models in combination with climate scenarios.

For such a task, the hydrological model needs to be able to extrapolate beyond the conditions during model calibration. This implicates that also internal processes, which have not undergone validation against observations, need to be simulated correctly [*Seibert*, 2000]. It has been recognized for a while that good discharge simulations at the catchment outlet cannot guarantee good internal functioning of the model [*Refsgaard*, 1997], as this may be an effect of error compensation [*Klemes*, 1986; *Seibert and McDonnell*, 2002]. Model consistency may be increased by including internal variables in the model calibration and evaluation procedure [*Güntner et al.*, 1999; *Seibert*, 2000]. Several studies have shown the utility of multivariable model calibration and validation, using variables like groundwater [*Juston et al.*, 2009; *Lamb et al.*, 1998; *Madsen*, 2000; *Seibert*, 2000], soil moisture [*Franks et al.*, 1998; *Parajka et al.*, 2009], streamflow salinity [*Mroczkowski et al.*, 1997], snow cover [*Parajka et al.*, 2007], or glacier mass balance [*Finger et al.*, 2011; *Konz and Seibert*, 2010; *Schaefli and Huss*, 2011; *Stahl et al.*, 2008].

In mountain catchments dominated by snowmelt runoff, the correct representation of snow processes is crucial. Snow cover patterns from remote sensing can therefore be a useful data source for constraining model parameters. While these data do not contain information on the snow water equivalent (SWE), they are spatially distributed, which makes them particularly useful for the evaluation of distributed models. They can thus be seen as complementary to the discharge time series, which are spatially integrated but give quantitative information on the water balance [*Finger et al.*, 2011]. Remotely sensed snow cover data

have been applied for hydrological modeling in different ways: as model forcing for snowmelt runoff models [e.g., *Li and Williams*, 2008], for model updating in data assimilation approaches [*Andreadis and Lettenmaier*, 2006; *De Lannoy et al*., 2012; *Liu et al*., 2013; *Rodell and Houser*, 2004; *Yatheendradas et al*., 2012; *Zaitchik and Rodell*, 2009], and for model calibration [*Corbari et al*., 2009; *Engeset et al*., 2003; *Finger et al*., 2011; *Koboltschnig et al*., 2008; *Parajka and Blöschl*, 2008; *Pellicciotti et al*., 2012; *Shrestha et al*., 2013; *Sorman et al*., 2009; *Udnaes et al*., 2007]. *Parajka and Blöschl* [2008] performed an extensive study over 148 catchments in Austria on the value of MODIS (Moderate-resolution Imaging Spectroradiometer) snow cover data for calibration of the HBV (Hydrologiska Byråns Vattenbalansavdelning) model. Their results are very encouraging, as they showed that, on average, the inclusion of MODIS snow cover during model calibration led to better snow cover and runoff simulations in the validation period. *Finger et al*. [2011] applied a Monte Carlo approach for the calibration of a grid-based model against discharge, satellite snow cover, and glacier mass balance. Their investigation of which data combinations are particularly useful for model conditioning showed a superior performance of the combination of discharge and snow cover data. However, studies explicitly showing the trade-off between snow cover and discharge performance are rare. *Parajka et al*. [2007] analyzed trade-off curves of model performance for discharge and snow cover area (SCA) derived from interpolated measurements of snow water equivalent (SWE). *Parajka and Blöschl* [2008] used a weighted sum approach to combine the objective functions for satellite-derived SCA and discharge into one criterion, but also analyzed the effects of varying the weights. In this study, we therefore explicitly analyze trade-offs between good model performance with respect to discharge and SCA. The results from the multiobjective optimization are also contrasted with results from Monte Carlo sampling and single-objective optimizations.

A further question concerns the amount of snow cover scenes required for model calibration. Whereas some remote sensing snow cover data, such as those from MODIS, can be obtained as ready product, others, like AVHRR (Advanced Very High Resolution Radiometer) or Landsat data, first need to be processed, as only the raw data are available. This can be very time consuming, e.g., due to the need for manual georeferencing and sensor calibration. It is therefore very valuable to know how the added information content of additional images declines with an increasing number of observation scenes. While for discharge data this question has been addressed by a number of studies [see, e.g., *Perrin et al*., 2007; *Juston et al*., 2009, and references therein], we could not find any information on how much remote sensing snow cover data should be used for model calibration. This study therefore investigates the value of increasing the number of snow cover scenes for the calibration of a hydrological model.

In the present study, a hydrological model is applied to headwater catchments of the Karadarya Basin in Kyrgyzstan, Central Asia. In these catchments, the runoff regime is strongly influenced by snowmelt runoff in spring and early summer and snow cover data are therefore expected to be very well suited for constraining the model and improving model consistency. Due to lower data availability after 1990, a time period before 1990 was selected for model calibration. MODIS snow cover data, which are only available from 2000, could therefore not be used and we instead resorted to AVHRR data.

We first analyze the trade-offs between the model performance with respect to discharge and SCA using a multiobjective genetic algorithm to identify pareto optimal solutions. Second, the study investigates within a Monte Carlo approach how the model performance in terms of snow cover prediction changes with an increasing number of SCA scenes in the calibration period. For the model evaluation against gridded satellite snow cover data, we suggest an approach which makes use of the information on the snow cover distribution with elevation and does not require threshold values for the comparison between simulated SWE and observed SCA.

2. Study Area

This study focuses on six headwater catchments of the Karadarya Basin located in Kyrgyzstan (Figure 1). Together with the Naryn River, the Karadarya forms the Syrdarya, which drains into the Aral Sea. The selected catchments are all upstream of the Andijan Reservoir and the influence of water management is generally assumed to be only marginal, except for the catchment Gulcha, where water is sometimes diverted into an irrigation channel. The catchment areas range from 170 to 3840 km^2 (Table 1). The catchments are characterized by high-elevation gradients and maximum elevations up to 4750 m (Figure 2).

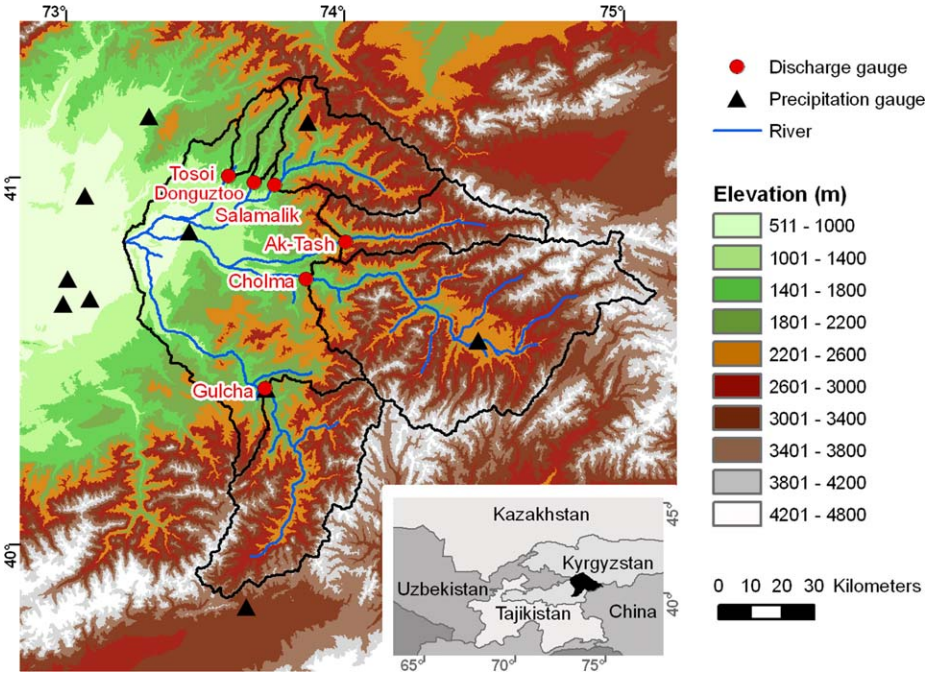

Figure 1. The six study catchments in the Karadarya Basin, including elevation, discharge gauges and precipitation gauges in the area. The inlet shows the location of the upper Karadarya Basin in Kyrgyzstan.

The predominant land cover is mostly grassland, and two catchments with lower elevations (Tosoi and Dongutzoo) also have a larger fraction of agricultural land. Forest, which often impedes remote sensing of snow cover [*Raleigh et al.*, 2013], only covers small fractions with a maximum value of 5% in Gulcha. A previous study analyzed the precipitation input to these catchments using hydrological modeling and observed discharge [*Duethmann et al.*, 2013]. While average annual precipitation at the gauges in vicinity to the study catchments ranges from 350 mm a^{-1} (measured at the stations in the lowland west of the study catchments) to 1050 mm a^{-1} (for the station in the northern mountain range), catchment average precipitation in the six study catchments is estimated to be between 700 and 1200 mm a^{-1}. Mean annual runoff over the period 1961–1990 ranges from $<$300 mm a^{-1} in the catchment Gulcha to nearly 800 mm a^{-1} in Ak-Tash (Table 1), resulting in annual runoff coefficients of approximately 0.4–0.6. The discharge regime is dominated by snowmelt with maximum discharges in spring and early summer.

3. Data and Methods

3.1. Extraction of Snow Cover From AVHRR Satellite Imagery

Snow cover data were extracted from images of the AVHRR instrument onboard the NOAA-9 and NOAA-11 satellites. The AVHRR sensor provides images with a resolution of 1.1 km at nadir. Scenes were selected

Table 1. Area, Glacier Coverage, Elevation Range, Mean Annual Runoff, and Runoff Coefficient of the Studied Subcatchments of the Karadarya Basin

			Elevation (m)				
	Area (km2)	Glacier Coverage (%)	Min.	Max.	Mean	Runoff[a] (mm a^{-1})	Runoff Coefficient[b]
Tosoi	216	0.0	1253	3165	2001	432	0.43
Donguztoo	166	0.0	1271	3502	1999	505	0.46
Salamalik	1180	0.5	1288	4381	2592	585	0.55
Ak-Tash	907	2.3	1728	4752	3121	778	0.62
Cholma	3840	1.9	1352	4753	3117	410	0.50
Gulcha	2010	0.7	1557	4623	3013	267	0.38

[a]Mean annual runoff over the period 1961–1990.
[b]Calculated for the period 1961–1990 using precipitation estimates by *Duethmann et al.* [2013].

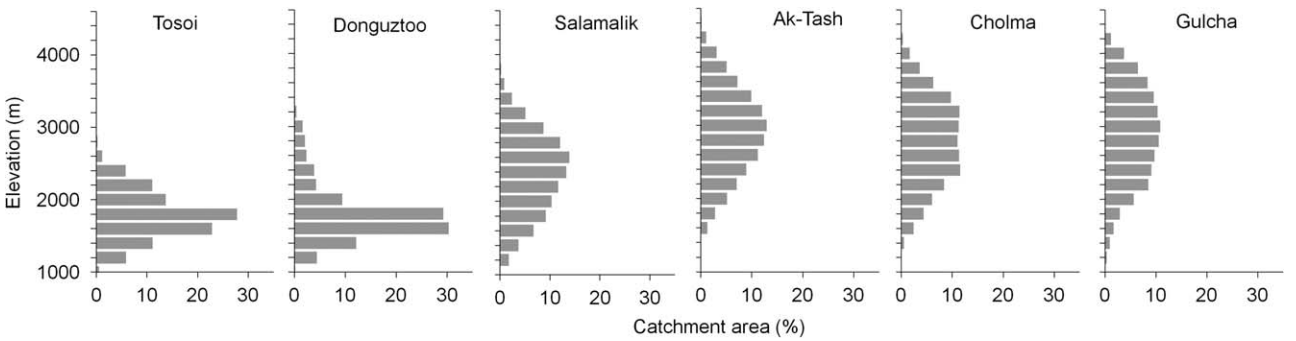

Figure 2. Distribution of the catchment area with altitude for the six study catchments (by 200 m elevation zones).

from the melt seasons (March–July) of the years 1986–1989, when both AVHRR data and suitable input data to drive the hydrological model were available. We focused on the melt season, as the parameters of the hydrological model were expected to be most sensitive to observed snow cover during this time. The time step between the images was approximately 9–14 days, but larger time steps were sometimes necessary in order to avoid high cloud coverage and poor data quality.

Before the images could be employed for snow cover delineation, careful calibration was required, including the calculation of albedo and radiances from the raw data as well as corrections for the sensor degradation and for nonlinearities in the measurements. In order to correct for the sensor degradation of channel 1 and 2, the calibration formulae of *Rao and Chen* [1995] were applied. Nonlinearities of the radiance measurements in channel 4 and 5 were adjusted using the corresponding error values from the NOAA Polar Orbiter Data User's Guide [*National Climate Data Center*, 1998]. Finally, radiances of channel 3–5 were converted to temperatures using the inverted Planck's radiation equation, and the Normalized Differential Vegetation Index (NDVI) was calculated from channels 1 and 2.

The data were then classified into snow, no snow, and clouds using a dichotomous multichannel classification scheme based on a model developed by *Voigt et al*. [1999]. Inputs to this model were differences in surface temperature, albedo, and NDVI, which is used for the extraction of the vegetation coverage. Initial values for the thresholds were taken from *Höppner and Prechtel* [2002]. On the basis of visual interpretation of the scenes, these thresholds were subsequently adapted to the study region and to the respective season with different values for spring and summer. The seasonal variation proved to be particularly important for the correct discrimination between snow and clouds.

3.2. Hydrological Model

The hydrological model WASA (Model of Water Availability in Semi-Arid Environments) [*Güntner*, 2002; *Güntner and Bronstert*, 2004] was originally developed for large semiarid basins, and later extended for erosion and sediment transport [*Mueller et al*., 2010] and for mountainous regions influenced by snow and glacier melt [*Duethmann et al*., 2013].

The model uses a semidistributed approach. The spatial discretization may either be based on hillslopes [*Francke et al*., 2008; *Güntner and Bronstert*, 2004], which also allows for lateral surface and subsurface redistribution along the hillslope, or on hydrological response units (HRUs), which enables faster simulations. This study applied the latter approach. HRUs were delineated based on 200 m elevation bands, resulting in 11–18 HRUs with a median size of 51 km^2 and a range of 0.01–452 km^2. Each HRU was associated with its dominant soil type, dominant land cover type, and its glacier fraction.

The model includes routines for snow accumulation, snow and glacier melt, interception, infiltration, percolation through a multilayer soil, and evapotranspiration. Runoff processes considered by the model are infiltration and saturation excess surface runoff, interflow and groundwater runoff. The model version used in this study separates subsurface flow into interflow and base flow based on a calibration parameter, and also considers that in a small fraction of the catchment (e.g., riparian areas, roads or rock areas connected to a stream) rainfall directly results into streamflow. The following section presents details of the snow module,

detailed descriptions of other model components can be found in *Güntner* [2002] and *Güntner and Bronstert* [2004].

For the calculation of snow accumulation, precipitation is separated into rainfall or snowfall based on a calibrated threshold temperature. Snowmelt is simulated using a temperature index approach [*Hock*, 2003], where the melt factor varies in a sinusoidal form between a minimum value at the winter solstice and a maximum value at the summer solstice. The increase of the melt factor from winter to summer reflects an increase in incoming solar radiation and a decrease of the snow albedo associated with aged snow [*Anderson*, 2006].

Due to wind drift, avalanches, and the spatial variability of precipitation, snow is rarely uniformly distributed after snowfall, and variability in the melt processes (for example, due to variations in shading) may further increase the variability of SWE within an elevation zone, which is however often neglected in hydrological models. Neglecting this variability of SWE within an elevation zone or HRU implicates that an elevation zone can only be snow covered or snow free, resulting in abrupt changes from completely snow-covered conditions to completely snow free at the end of a melt season. Furthermore, the comparison to remotely sensed snow cover is easier if also fractional snow cover areas are simulated. In the WASA model, the variability of SWE within an elevation zone is parameterized using a snow depletion curve, as for example described by *Liston* [2004]. The snow depletion curve in this case describes the fractional SCA as a function of the SWE divided by the maximum SWE at the end of the accumulation season (snow depletion curves may also be defined in other ways, for example, SCA as a function of time). During accumulation a spatially continuous snow cover is simulated; snowmelt then results in a gradually decreasing SCA. It has been observed that for a given catchment the spatial distribution of relative snow amounts is similar from year to year so that the shape of the snow depletion curve can be assumed constant in time [e.g., *Luce and Tarboton*, 2004]. The distribution of the observed SWE is often approximated by a lognormal distribution [*Donald et al*., 1995]. SCA can then be calculated analytically from the maximum simulated SWE at the end of the accumulation season and the cumulative melt depth [*Liston*, 2004]. This parameterization for the description of the fractional SCA has only one additional parameter, the coefficient of variation of the SWE distribution. It assumes that within each elevation zone there is always some part with a very thin snow cover so that with the beginning of snowmelt a snow-free area is created. For cases where this is not considered appropriate, one could introduce an additional parameter defining the SWE that must melt before any snow-free area starts to be exposed [*Donald et al*., 1995]. The situation that snowfall occurs during the melting phase needs to be considered separately. In this case, the model simulates a SCA of 100% and melt affects the whole area until the new snow is melted; the model then proceeds on the established snow depletion curve. If the sum of the remaining snow from the main accumulation period and the new snow exceed the previous maximum accumulation, the model considers this to be a new accumulation period and a new snow depletion curve is started. For HRUs containing a glacier, it is assumed that snow remains longest on the glacier so that glacier melt does not start until the snow covered area in that HRU is smaller than the glacier area.

This version of the model has 13 calibration parameters (Table 2). The parameter ranges were established based on literature values and previous experiences with the WASA model. The glacier melt parameter is

Table 2. Calibration Parameters Including Values for the Lower and Upper Bounds

Routine	Parameter	Unit	Lower Bound	Upper Bound
Snow and glacier melt	Min. snowmelt factor	mm $°C^{-1}$ d^{-1}	1	15
	Max. snowmelt factor	mm $°C^{-1}$ d^{-1}	1	15
	Glacier melt parameter	–	0	1
	Threshold melt temperature	°C	−2	2
	Coefficient of variation (*cv*)	–	0.001	1
Infiltration and percolation	*kf_corr_f*	–	0.01	100
	k_sat factor	–	0.01	100
Subsurface flow	*frac2gw*	–	0	1
	Interflow delay factor	days	10	100
	Groundwater delay factor	days	200	400
Generation of direct runoff from areas connected to the stream	*frac_riparian*	–	0	0.05
Spatial variability of saturated areas within a model unit	*sat_area_var*	–	0	0.3
Precipitation input	Precipitation bias factor	–	0.75	1.5

recalculated into a glacier melt factor with values ranging between the snowmelt factor and an upper boundary of 15 mm $^{\circ}C^{-1}$ d^{-1}.

The following input data were used to set up the WASA model for the Karadarya Basin: the SRTM (Shuttle Radar Topography Mission) digital elevation model [*Jarvis et al*., 2008] with a resolution of 90 m, the MODIS land cover product at a resolution of 500 m (MCD12Q1) [*Friedl et al*., 2002], mean monthly leaf area index (LAI) values from the 8 day MODIS LAI product for 2001–2008 at a resolution of 1 km (MOD15A2) [*Myneni et al*., 2002] aggregated by elevation zone, subcatchment, and land cover class, a soil map (1:1,500,000) digitized from the Kyrgyz Atlas [*Academy of Science of the Kyrgyz SSR*, 1987], and glacier areas delineated from a Landsat Multispectral Scanner (MSS) scene (resolution 79 m) in summer 1977.

A previous analysis evaluated various precipitation estimates for this area [*Duethmann et al*., 2013] and concluded that the precipitation product interpolated from gauge data using monthly precipitation fields derived by multilinear regression against elevation, longitude, and latitude was best suited for this area. This precipitation product ("MLR-all" in *Duethmann et al*. [2013]) was therefore also applied in the current study. The model further uses daily time series of solar radiation, temperature, temperature lapse rate and humidity. Due to the lack of observational data these data were derived from the ERA-40 reanalysis data [*Uppala et al*., 2005] downscaled to a resolution of 12 km using the regional climate model Weather Research and Forecasting Model (WRF) [*Skamarock et al*., 2008]. For details of the WRF simulations please refer to *Duethmann et al*. [2013]. The meteorological time series were applied to the hydrological model as area-mean values for each catchment.

Discharge data used for model calibration were provided by the Hydrometeorological Service of Kyrgyzstan. Water stage is recorded twice daily manually and continuously at the larger rivers under high flow conditions using a float driven recording sensor. Discharge is estimated about four times per month during low flow and eight times per month during high flow conditions using the velocity-area method with velocities derived from current meter measurements. Data at the gauge Gulcha are influenced by diversions into an irrigation channel upstream of this gauge. As timing and volume of these abstractions were not known, they could not be considered in the model.

3.3. Model Calibration

3.3.1. Objective Functions

For the calibration to discharge data, the following objective function was used:

$$objf\_q = 0.5 \times (NSE + LogNSE),$$

with

$$NSE = 1 - \frac{\sum_{t=1}^{T} (Q_{obs}(t) - Q_{sim}(t))^2}{\sum_{t=1}^{T} (Q_{obs}(t) - \overline{Q_{obs}})^2} \quad \text{and} \quad LogNSE = 1 - \frac{\sum_{t=1}^{T} (\log(Q_{obs}(t)) - \log(Q_{sim}(t)))^2}{\sum_{t=1}^{T} \left(\log(Q_{obs}(t)) - \overline{\log(Q_{obs})}\right)^2} \tag{1}$$

NSE is the Nash-Sutcliffe efficiency value between daily observed ($Q_{obs}(t)$) and simulated ($Q_{sim}(t)$) discharge and *LogNSE* is the Nash-Sutcliffe efficiency calculated on logarithmic flows. As the Nash-Sutcliffe efficiency is particularly responsive to errors in high discharge values, and the Nash-Sutcliffe efficiency for logarithmic flows is more sensitive to errors in low flows, an average of these two measures results in a more balanced evaluation of high and low flows. The maximum possible value of the objective function is 1, which would indicate perfect agreement between simulated and observed discharge.

Previous studies used different methods for evaluating model performance of semidistributed models with respect to raster-based remote sensing snow cover. Several studies simply compared catchment average simulated and observed SCA [*Engeset et al*., 2003; *Sorman et al*., 2009; *Udnaes et al*., 2007]. This approach however neglects the information of the SCA distribution with elevation, which is also contained in the snow cover data. The method suggested by *Parajka and Blöschl* [2008] takes account of the observed SCA in each elevation zone. As the hydrological model they applied simulated a uniform distribution of SWE in each elevation zone, they used an indirect approach for the comparison between simulated SWE and

observed SCA. They defined two kinds of errors—a snow underestimation error, if the model simulated no snow in an elevation zone, but the snow covered area was larger than a threshold, and a snow overestimation error, if the simulated SWE was above an SWE threshold, but the remote sensing data indicated no snow for this elevation zone. A disadvantage of this approach is the subjective choice of the thresholds, which can influence the results of the comparison. For the snow underestimation error it is for example not clear to what value the threshold for observed SCA should be set. A SCA threshold of, e.g., 0.05 means that the model has to simulate at least some snow (SWE >0) if the observed SCA is larger than 5%, otherwise this produces an underestimation error. By assuming a uniform SWE distribution this, however, also has the consequence that the model generates snowmelt from the entire elevation zone, while in reality snowmelt is only produced from the much smaller snow covered area. In the present study, this problem was avoided by introducing the parameterization for the fractional SCA in an elevation zone as described in section 3.2, which then allowed a direct comparison between simulated and observed snow cover fraction for each elevation zone. While introducing such a parameterization increased the number of parameters by one, it also allowed a more realistic representation of the fractional SCA, which could potentially also improve the runoff simulations.

For the comparison to simulated snow cover, the AVHRR snow cover data were summarized by catchment and elevation zone. An AVHRR cell (1.1 km $\times$ 1.1 km) was assigned to an elevation zone according to its median elevation, calculated from the 90 m SRTM digital elevation model. In a next step, the fractional snow and cloud cover by catchment and elevation zone were calculated. This approach thus particularly considers the large-scale heterogeneity at a scale larger than the AVHRR cell size. Observed snow cover values were only considered for model comparison if $>$20% of that elevation zone was cloud free and if there were more than three cloud-free cells. The objective function with respect to snow cover was then defined as root-mean-squared error (RMSE) between simulated and observed SCA:

$$objf\_sca = \sqrt{\left(\sum_{i=1}^{m} n(i)\right)^{-1} \cdot \sum_{i=1}^{m}\sum_{j=1}^{n(i)} \left(SCA_{obs}(i,j) - SCA_{sim}(i,j)\right)^2} \quad (2)$$

where $SCA_{obs}(i,j)$ and $SCA_{sim}(i,j)$ are the observed, respectively, simulated, SCA on observation day i in elevation zone j; m is the number of days with snow cover observations; and $n(i)$ the number of elevation zones with valid observations on day i. The best possible value of this objective function is 0, which would indicate perfect agreement between simulated and observed SCA.

Due to the formulation of the two objectives, multiobjective model calibration aims at identifying parameter sets which maximize *objf_q* and minimize *objf_sca*.

3.3.2. Multiobjective Calibration to Determine Pareto Fronts

As the optimum solutions for different objectives in general do not converge, multiobjective optimization aims at identifying a set of pareto optimal solutions instead of one single best solution. A solution is classified as pareto optimal (or nondominated) if there is no other solution which improves in one or more objectives without degrading at least one objective. Evolutionary algorithms are particularly suited to solve such multiobjective problems, as due to their population-based approach they can return a set of solutions within a single run [*Konak et al*., 2006]. For this study, we applied the Epsilon-Dominance Nondominated Sorted Genetic Algorithm II (ε-NSGAII) [*Kollat and Reed*, 2006]. This algorithm was selected, as in comparison studies which also included hydrological model calibration in their test problems, this algorithm was competitive or superior to other state-of-the-art multiobjective algorithms [*Kollat and Reed*, 2006; *Tang et al*., 2006]. ε-NSGAII is based on the NSGAII-algorithm [*Deb et al*., 2002], which uses a fast nondominated sorting algorithm and elitism. Elitism means that the nondominated solutions found so far are preserved and survive to the next generation. As an extension compared to the original NSGAII-algorithm, ε-NSGAII introduced the concept of ε-dominance, adaptive population sizing and automatic termination, reducing the number of algorithm parameters to be tuned. ε-dominance allows the user to specify the required precision in each objective. The user should set it to the difference in objective function values he or she considers to be relevant. The objective space is divided into multidimensional cells, with the dimension according to the number of objectives and the cell size in each dimension according to the ε-value in this objective. If there is more than one solution within a cell, only one solution—in case of minimization of all objectives, the

solution closest to the origin—is retained. This way the density of the final nondominated solutions is controlled. In a series of connected runs, the population size is adapted according to the number of archived nondominated solutions. A new run is started if there is no significant increase in the number or quality of the nondominated solutions over a selected number of generations. The overall search is terminated if there is no significant improvement between the archived solutions of two successive runs. For this study, we used an initial population size of 16 and a maximum run time of 40,000 generations. ε-values were set to 0.005 for *objf_q* and 0.001 for *objf_sca*. Other algorithm parameters were set to the values as suggested by *Kollat and Reed* [2006]. In order to reduce run time, a parallel version of the code [*Tang et al.*, 2007] was applied.

Snow cover data only were available for the period 1986–1989. As a 2 year period was regarded as being too short for model calibration with respect to discharge, and the model could not be run in the 1990s due to missing data, the model was first calibrated in simulation period A (1978–1981 and 1986–1987) and validated in simulation period B (1982–1985 and 1988–1989), and then calibration and validation period were swapped (calibration in simulation period B and validation in simulation period A). Thus, calibration and validation periods usually included 2 years of snow cover and 6 years of discharge data. Simulations were always performed for a continuous period. For example, for calibration in period A, simulations were performed for the period 1976–1987. From these simulations, *obj_q* was evaluated over 1978–1981 and 1986–1987, and *objf_sca* was evaluated over 1986–1987. Due to gaps in the discharge data (21 May to 31 December 1980 in Ak-Tash and 1989 in Salamalik) the calibration or validation periods were accordingly shorter in these two catchments. An additional 2 year period prior to the actual simulation period was used for model initialization.

3.3.3. Monte Carlo Parameter Calibration for Investigating the Value of an Increasing Number of Snow Cover Scenes

The additional information content of an increasing number of snow cover scenes was explored by calibrating the model using different subsets of snow cover images during the calibration period and evaluating the model performance against the snow cover images in the validation period. This resulted in a very large number of different calibration settings, as different possibilities of, e.g., selecting 5 snow cover images from 20 available images in the calibration period have to be considered. Optimizing the parameters for each subset of observations was therefore not a suitable approach. While for a small number of calibration settings multiobjective calibration usually results in better model performances with a lower number of simulations, Monte Carlo simulations can be more efficient if the number of different subsets of observations to which the model should be calibrated is very large. After once performing a large number of simulations, it is then possible to simply select well-performing simulations for a large number of different calibration settings without further computational cost. Assuming uniform distributions between the minimum and maximum bounds, 50,000 random parameter sets were generated using Latin hypercube sampling.

For each catchment, the RMSE of observed versus simulated SCA of each image was evaluated. Snow cover images for which the range of RMSE values over the 50,000 simulations was zero were not considered in the next step since they did not exert any constraining power. These were generally images where only elevation zones far below or above the snowline were cloud free and which were thus not challenging for the model. In the next step, the value of increasing the number of snow cover scenes was evaluated. Calibration and validation period were defined as described in section 3.3.2 and all simulations where *objf_q* in the calibration period was above 0.5 were selected. Now $n = 0, 1, \ldots, m$ images were randomly chosen from the calibration period, where m denotes the number of available images in the calibration period. In order to calculate the median, minimum, and maximum value of *objf_sca* during the validation period, the following procedure was applied (summarized in Figure 3). For each simulation, *objf_sca* with respect to the selected images in the calibration period was calculated and the best $s = 20$ parameter sets were selected (for $n = 0$, s parameter sets were chosen randomly). Selecting an ensemble of best performing parameter sets instead of selecting just one best parameter set is expected to result in a more robust behavior during the validation period. For the selected simulations, the model performance over the validation period was evaluated and averaged over the selected simulations. This was repeated $j = 10{,}000$ times in order to sample different combinations of selecting n from m images. Finally, the median, minimum, and maximum *objf_sca* of these 10,000 repetitions were calculated. The procedure depicted in Figure 3 was repeated for $n = 0, 1, \ldots, m$ images. It was also performed for the two different calibration periods (calibration in period A and validation in period B, and vice versa).

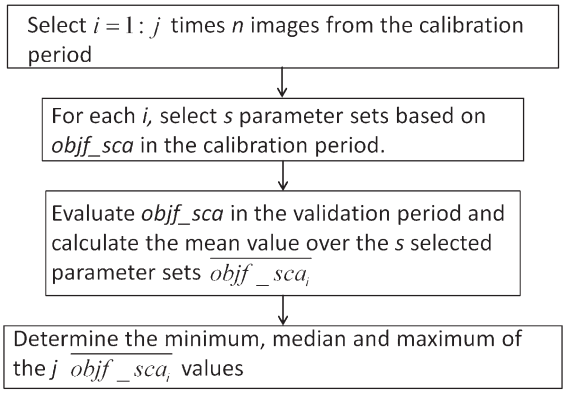

Figure 3. Procedure for calculating the minimum, median, and maximum snow cover error during the validation period, when using *n* snow cover images for model calibration.

4. Results and Discussion

4.1. Characterization of the Trade-Offs Between Good Performance for Discharge and Snow Cover Area

Using a multiobjective evolutionary algorithm, it was investigated how well discharge and SCA can be modeled simultaneously. Overall the model achieved good performances with objective function values for discharge between 0.62 and 0.93 (optimum = 1) and objective function values for SCA between 0.1 and 0.29 (optimum = 0). Plots of the pareto fronts show that the trade-offs between good simulations with respect to discharge and SCA are generally small in the studied catchments (Figure 4). In this figure, the red and blue dots indicate the solutions for calibration period A and B, respectively. Optimum solutions plot in the lower left corner. Only for the catchment Gulcha, a larger spread of solutions can be seen, for all other catchments, only up to six solutions were found. Solutions which differ by less than the specified ε-values of 0.005 for *objf_q* and 0.001 for *objf_sca* were not retained by the algorithm, avoiding overly precise trade-offs, which are not meaningful [*Kollat et al*., 2012]. Thus in most of the studied catchments a good model performance in terms of SCA did not preclude a good simulation of discharge. The model performances in the validation periods are close to the performances if the respective period was used for calibration (Figure 4), except for the catchment Gulcha, where the model seems to be less transferrable between different periods. The peculiarities at Gulcha with overall lower performances, a larger trade-off between discharge and snow cover performance, and lower transferability between the calibration and

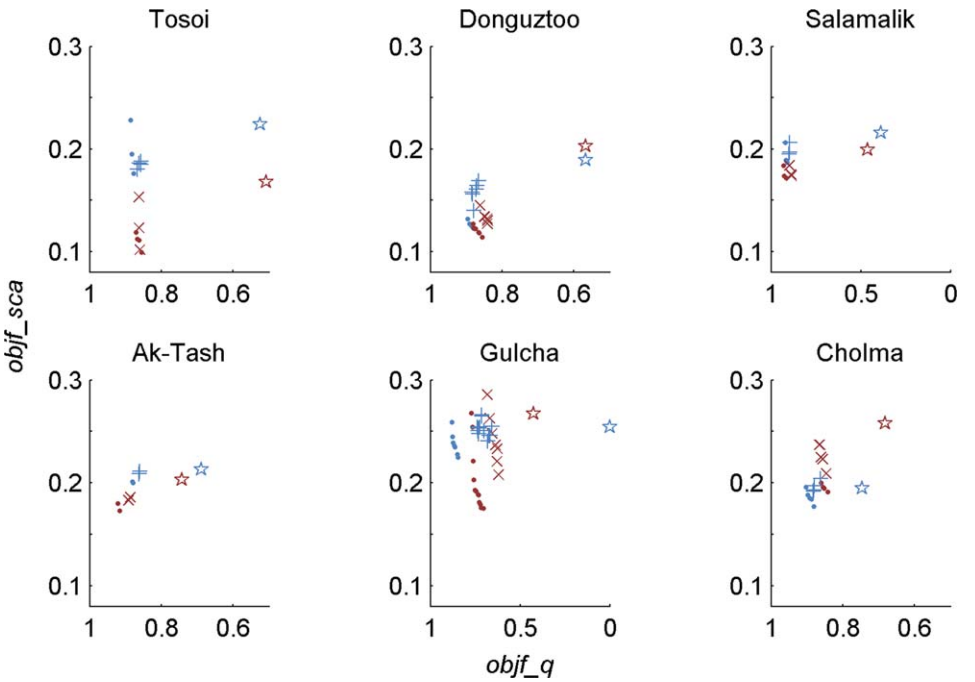

Figure 4. Trade-off curves of model performance against SCA (*objf_sca*) and discharge (*objf_q*) for the six study catchments. The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. Dots: calibration period; crosses: validation period; blue: simulation period A (1978–1981 and 1986–1987); red: simulation period B (1982–1985 and 1988–1989). For example, blue dots indicate the model performance in simulation period A of the solutions from the calibration in simulation period A, and blue crosses show the model performance in simulation period A of solutions from the calibration in simulation period B. Additionally, the stars show the performance of the uncalibrated model. Please note the different scales of the *x* axis.

validation period are likely to be caused by the unconsidered flow diversions into an irrigation channel upstream of this gauge.

We also compared the performance of the calibrated model to an uncalibrated model which simply uses mean parameter values between the upper and lower parameter bound of Table 2 (Figure 4). This shows that in all cases model performance was improved by calibration, and on average, *objf_q* improved by 0.33 and *objf_sca* improved by 0.03 (Figure 4).

The multiobjective optimization algorithm always outperformed the best solutions of the Monte Carlo simulations with randomly sampled parameters (Figure 5). The number of model evaluations used by the multiobjective optimization was between 4,168 and 12,472 with an average of 6,988. Thus, even though the

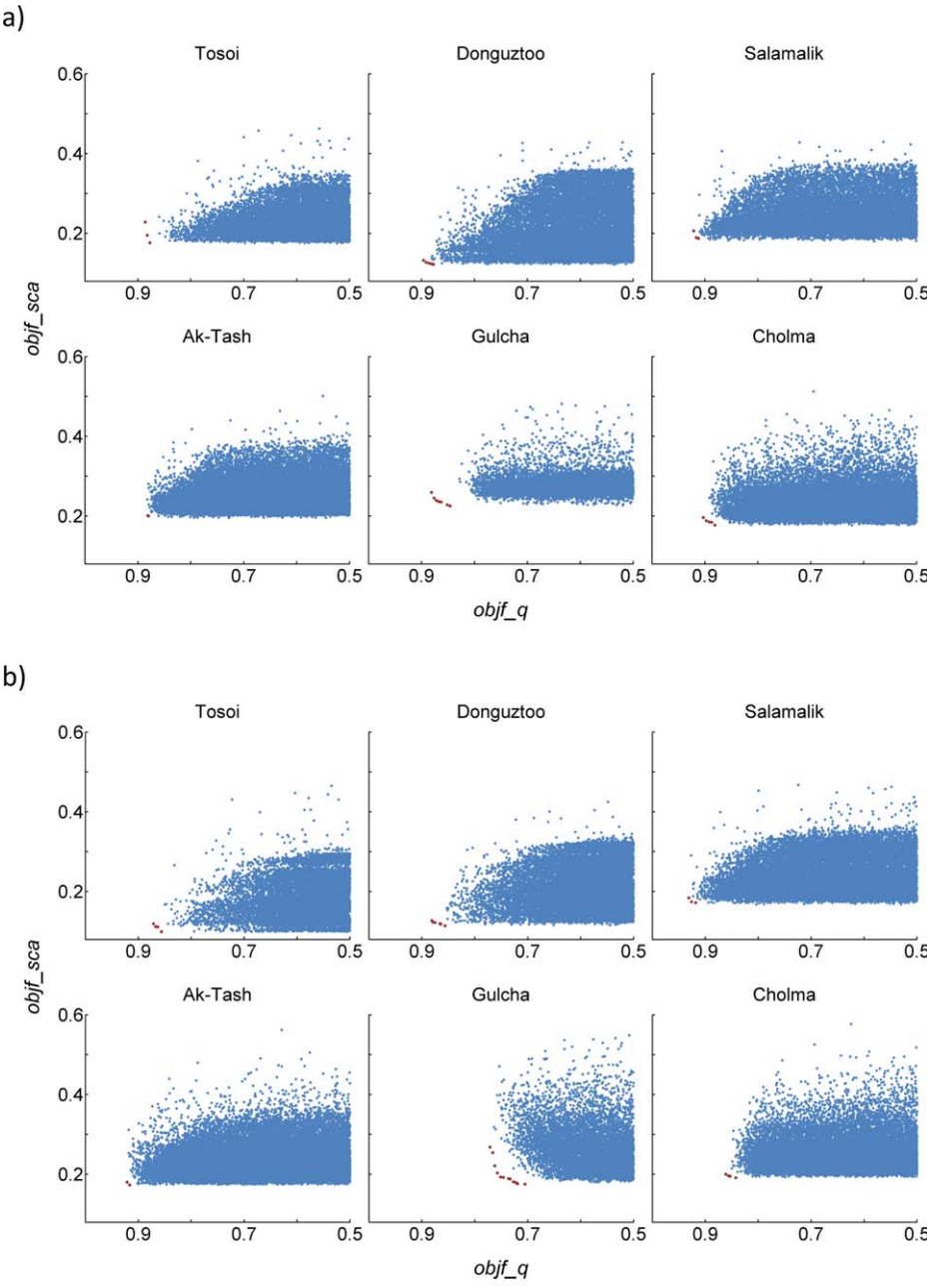

Figure 5. Scatterplots of the objective function with respect to SCA (*objf_sca*) against the objective function with respect to discharge (*objf_q*) showing both the solutions generated by Monte Carlo sampling (blue) and the solutions from the multiobjective optimization algorithm (red). The *x* axis is plotted in reverse order so that optimum solutions plot in the lower left corner. (a) Calibration in simulation period A (1978–1981 and 1986–1987) and (b) calibration in simulation period B (1982–1985 and 1988–1989).

average number of model evaluations in the multiobjective optimization was about seven times smaller than the number of Monte Carlo simulations applied in this study, better solutions were achieved through multiobjective optimization. The scatterplots in Figure 5 also show that it is possible to achieve good simulations with respect to discharge with a large snow cover error and vice versa. This was also confirmed by single-objective optimizations where the model was calibrated only against discharge or only against SCA. For the criterion which was included in the optimization (either discharge or snow cover area), they resulted in virtually the same model performance as the best performing solution with respect to this criterion from the multiobjective optimization, but only much lower performances were achieved for the neglected criterion. Namely, for the optimizations against SCA only, *objf_q* was on average 0.62 worse than the worst solution from the multiobjective optimization; and for the optimizations against discharge only, *objf_sca* degraded on average by 0.05. It is therefore necessary to consider both objectives for model conditioning.

4.2. Model Performance in Terms of Discharge and Snow Cover Area

In addition to the objective function values, which assess the model performance in an aggregated way, this section provides a more detailed analysis of the model performance with a particular focus on SCA prediction. Time series of simulated and observed fractional SCA can be compared for each elevation zone in each subcatchment (see examples in Figure 6). As the trade-offs between the solutions from the multiobjective optimization were only small, the figure depicts only one compromise solution from the multiobjective optimization indicated by the black line. In many cases, the observed decrease in snow cover is well represented by the model. Not only start and end point in time of snow cover depletion but also the fractional snow cover often compares well to the observed one. The model has some difficulties with snow events during the melt season. Possible reasons for this are uncertainties in the precipitation input, the precipitation phase (solid or liquid) and the model assumptions for snow events during the melt season. The interpolated precipitation is only based on relatively few stations so that the timing of precipitation events cannot always be representative for the entire catchment. Uncertainties in the precipitation phase are due to uncertainties in the temperature and lapse rate, as well as the fact that the temperature below which precipitation falls as snow is not a fixed value but may vary from event to event. If there are snow events during the melt season, the simulated SCA of the elevation zones with snowfall is increased to 100% (assuming that snowfall covers the entire elevation zone, see section 3.2) and only decreases again after this new snow has completely melted, though in reality this new snow might also disappear more gradually.

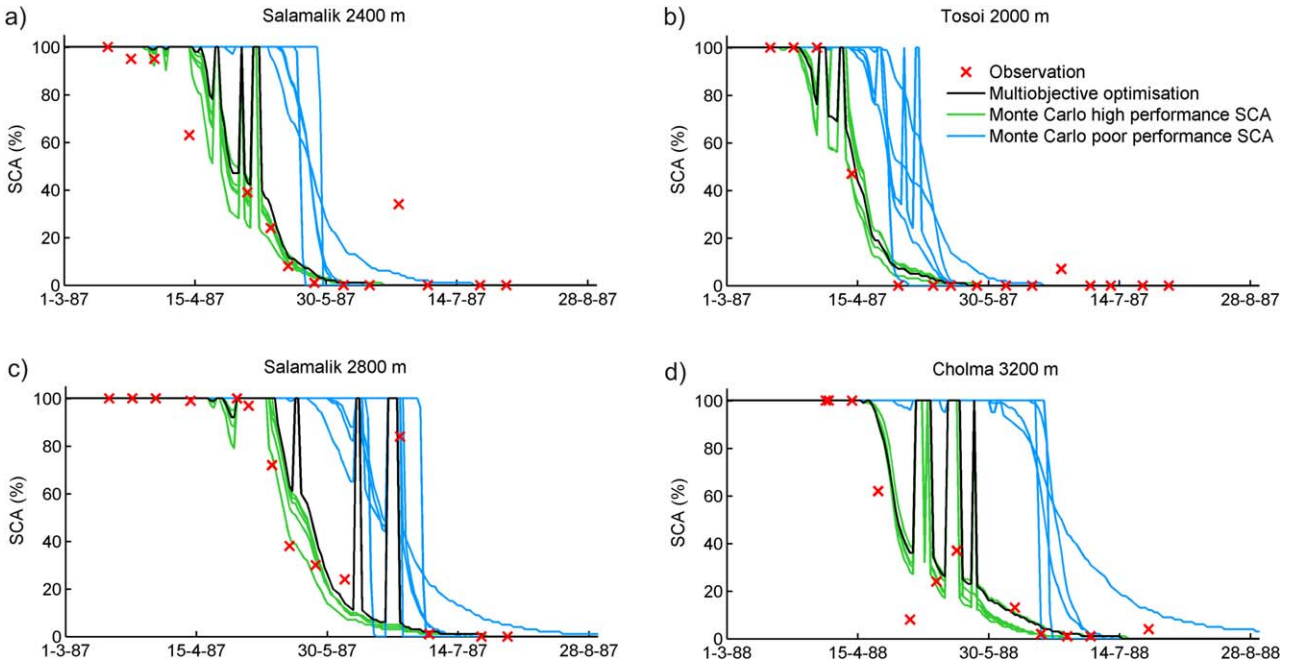

Figure 6. Time series of simulated versus observed SCA during the validation period, for selected catchments and elevation zones. The simulations show a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization, and solutions from the Monte Carlo sample with comparable performance in terms of discharge but contrasting performance in terms of SCA.

In order to visualize the difference between simulations with comparable performance in terms of *objf_q* but different performance in terms of *objf_sca*, Figure 6 also shows results from the Monte Carlo sample, selected as the five best and the five worst simulations in terms of *objf_sca* from all solutions with $objf\_q \geq objf\_q_{max} - 0.05$, where $objf\_q_{max}$ is the best Monte Carlo solution in terms of *objf_q* achieved for this catchment. The solutions with a poor performance considerably overestimate the snow cover. Despite a comparable performance with respect to discharge, large errors in the simulation of the snow cover can occur if snow cover observations are not considered during model calibration.

From a different perspective, Figure 7 shows simulated and observed SCA for all elevation zones of all study catchments as examples for three selected days. These plots generally show a consistent increase of the observed SCA with elevation with only little noise. A typical pattern, apparent on many dates, is that the elevation zone where for example around 50% of the area is snow covered is at a higher elevation in Cholma and Gulcha than in Tosoi and Donguztoo (Figures 7b and 7c). The faster decrease of the snow cover in Cholma and Gulcha may partly be explained by lower precipitation in these two catchments compared to Tosoi and Donguztoo. Differences in temperature, aspects or shading may furthermore explain the differences in snow cover decrease between the catchments. In accordance with the observations, the simulations also represent the gradual increase of SCA with elevation over several elevation bands. On some dates, the model shows a very abrupt increase of snow cover from 0% to 100% over only one or two elevation zones not reflected by the observations (e.g., 28 April 1986). This behavior can again be related to snow events during the melt season.

In order to summarize the behavior over all images, the mean bias by elevation zone averaged over all dates within the validation period was evaluated. The black dots in Figure 8 show the result for a compromise solution from the multiobjective optimization. This generally reveals only relatively small systematic biases with elevation. In the catchments with a larger elevation range, there is a tendency for overestimation at high elevations (e.g., Salamalik and Ak-Tash for both periods (Figures 8a and 8b) and Cholma and

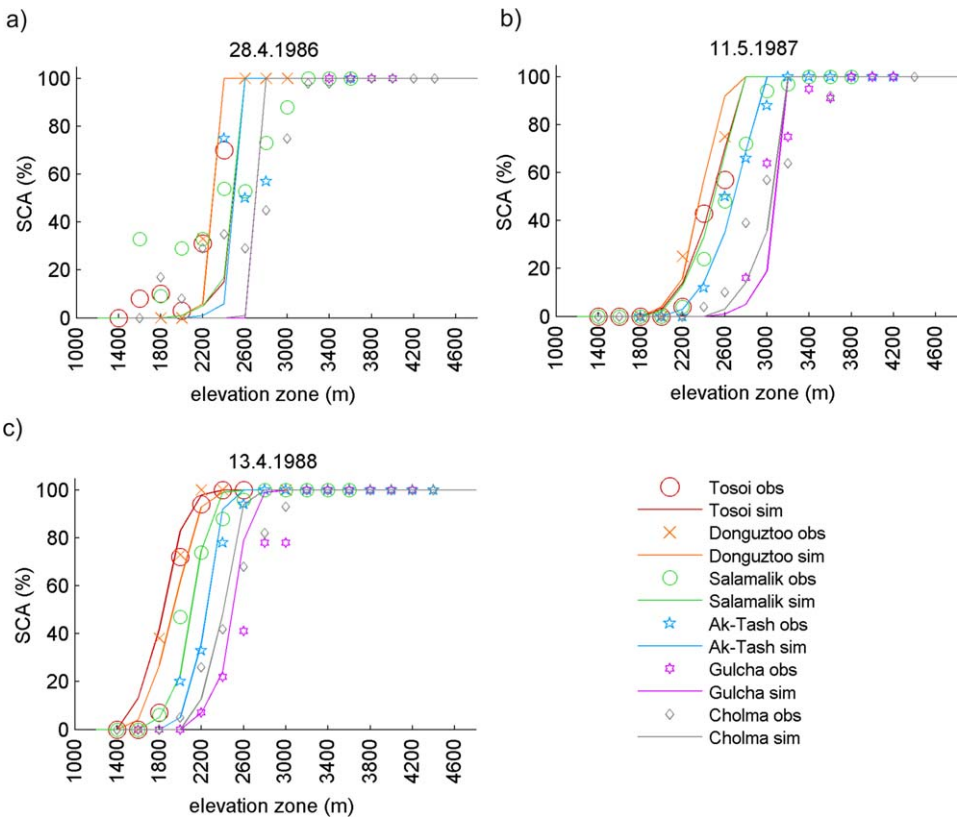

Figure 7. Simulated and observed SCA in the six study catchments, for the 28 April 1986, 11 May 1987, and 13 April 1988. Simulated snow cover is shown for the validation period and a selected compromise solution generated with the multiobjective optimization algorithm. The simulations also refer to discrete 200 m elevation zones but are drawn as continuous lines for better visibility.

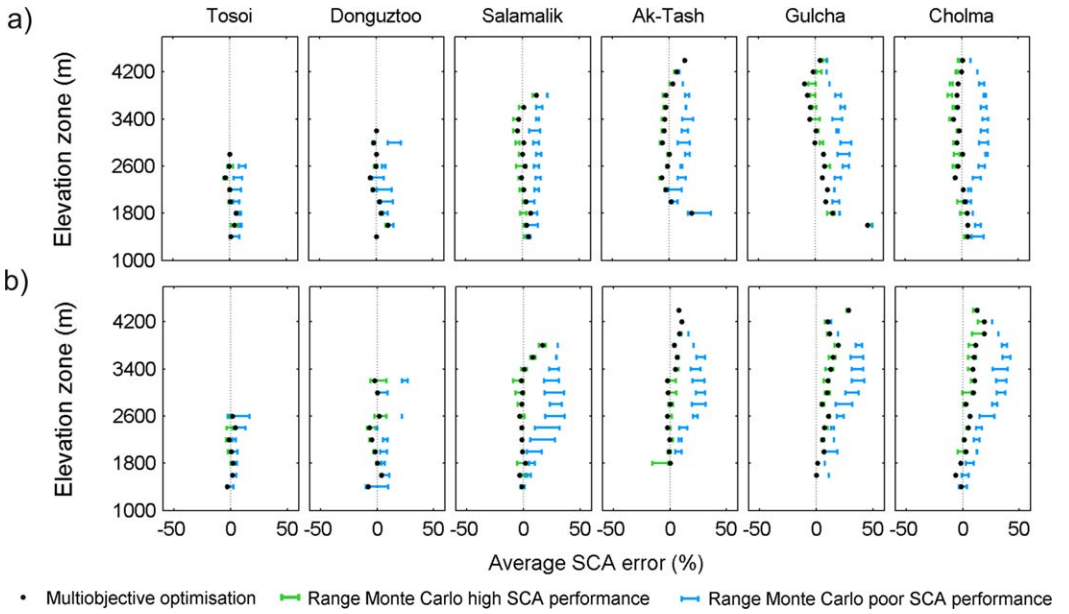

Figure 8. SCA bias by elevation zones averaged over the validation periods (top) 1986–1987 and (bottom) 1988–1989. The black dots show results for a compromise solution in terms of snow cover and discharge performance from the multiobjective optimization. Bars indicate ranges of values from selected Monte Carlo simulations with comparable performance in terms of discharge but contrasting performance in terms of SCA.

Gulcha for the model calibrated in period B (Figure 8b). Moreover, in the catchment Gulcha, SCA is overestimated at low elevations if the model is calibrated to simulation period A and validated in simulation period B (Figure 8a). In this case, the number of observations was very low; there are only four observations for the 1600 m elevation zone in the catchment Gulcha in the period 1986–1987 so that an overestimation of SCA at two of these dates caused a large average error. Figure 8 additionally also illustrates the bias by elevation zone for the selected sets of solutions from the Monte Carlo sample with comparable performance for discharge, but contrasting performance in terms of SCA. The simulations with a poor performance in terms of SCA generally show an overestimation of snow cover. This overestimation is however smaller in the catchments Tosoi and Donguztoo. In contrast to this, the selected solutions from the Monte Carlo sample with a good performance in terms of snow cover only have low bias values, comparable to the solutions from the multiobjective optimization.

As a result of the small trade-offs, the simulated hydrographs of the two extreme solutions from the multiobjective optimization which result in the best model performance for discharge and SCA, respectively, are most of the time very similar (results not shown). Therefore, Figure 9 only shows compromise solutions that are good both in terms of discharge and snow cover. The time series indicates that the degradation between calibration and validation period in Gulcha is to a large part the result of an underestimation of discharge in the years 1988–1989 of the model calibrated in simulation period A (1978–1981 and 1986–1987) and an overestimation in the years 1986–1987 of the model calibrated in simulation period B (1982–1985 and 1988–1989). These differences are likely a result of different abstraction volumes for irrigation during the two periods. Figure 9 also demonstrates some other deficiencies of the model, like too low variability of the simulated discharge in Salamalik, or a relatively strong overestimation of discharge of the base flow during the winter period 1987/1988 in Cholma. However, overall the model performs well, particularly considering the sparse data availability in this region.

As an example, Figure 10 also shows simulated discharge for solutions from the Monte Carlo sample, selected as the five best and five worst solutions with respect to *objf_sca* of all solutions with $objf\_q \geq objf\_q_{max}$—0.05, for 1 year in the catchment Ak-Tash. The shown example is typical also for other catchments and other years. The most striking difference between the two sets of simulations is that, compared to the solutions with poor SCA performance, the solutions with good SCA performance generally result in higher discharge at the beginning and lower discharge toward the end of the melting period. Thus, the simulations with a good performance with respect to snow cover have a higher tendency to

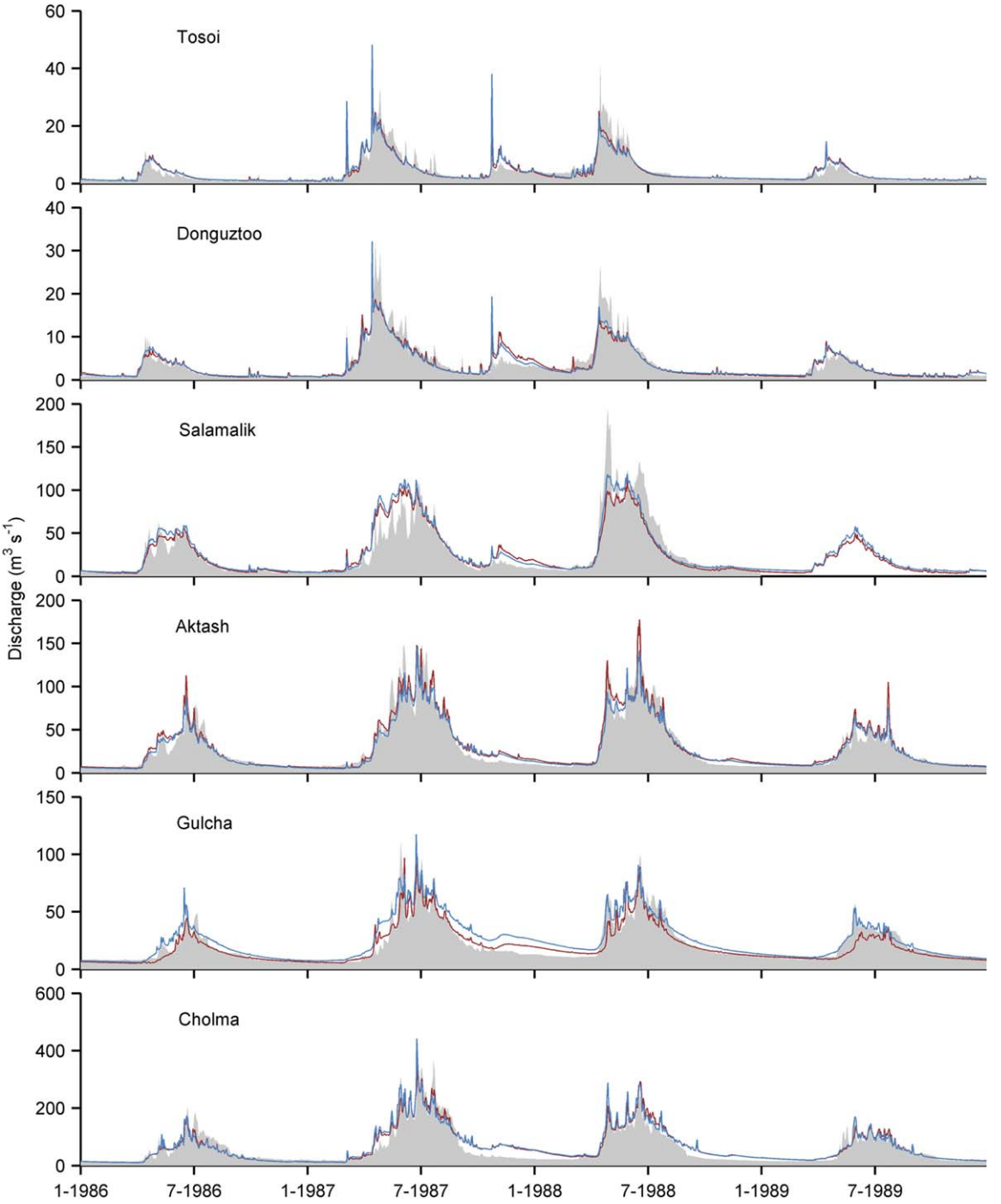

Figure 9. Time series of simulated versus observed discharge for a compromise solution (good performance for discharge and SCA) from the multiobjective optimization algorithm and the time period 1986–1989. Gray area: observed discharge; red line: simulated discharge for model calibration in simulation period A (1978–1981 and 1986–1987); blue line: simulated discharge for model calibration in simulation period B (1982–1985 and 1988–1989).

overestimate the observed discharge at the beginning of the melting season and to underestimate toward the end, while this is the other way round for the simulations with a high snow cover error. Overall, this results in a comparable performance with respect to discharge for the two sets of solutions.

4.3. Influence of the Two Objective Functions on Constraining Model Parameters

In order to illustrate the effect of the two different objectives on constraining the different model parameters, sets of solutions which perform well with respect to the two objective functions were selected from

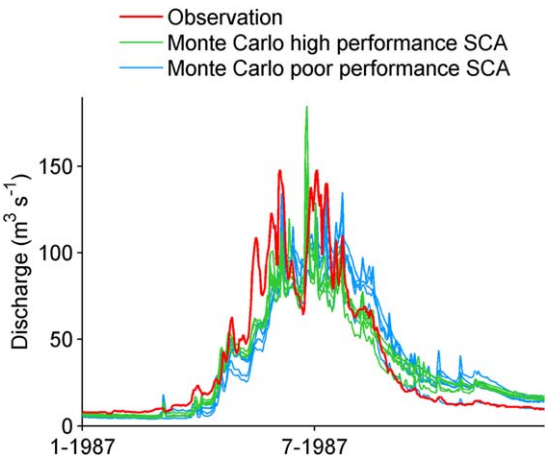

Figure 10. Observed discharge and simulated discharge, for the catchment Ak-Tash and the year 1987. Simulated discharge is shown for selected parameter sets from the Monte Carlo sample with overall comparable performance in terms of discharge, but good and poor performance with respect to SCA.

the Monte Carlo sample and evaluated with respect to their parameter distributions. For each catchment and for the two simulation periods, first the 5% best performing solutions in terms of *objf_q* were selected from the Monte Carlo sample. From these solutions, the 5% best performing solutions in terms of *objf_sca* were retained. Figure 11 shows cumulative density functions of the resulting parameter distributions. In these plots, uniform prior distribution would appear as straight lines. Note that the minimum and maximum melt factors were swapped when the minimum melt factor was larger than the maximum melt factor so that the prior distributions are not uniform. The two snowmelt factors and the threshold melt temperature are conditioned by both objectives. The coefficient of variation (*cv*), which determines the heterogeneity of the snow distribution, is largely controlled by *objf_sca*. Most other parameters, like frac2gw, the interflow delay factor, *sat_area_var*, and the precipitation bias factor are controlled by *objf_q*. Furthermore, some of the model parameters, in particular the glacier melt parameter and *kf_corr_f* hardly get constrained. If one selects all solutions with a comparable discharge performance ($objf\_q \geq objf\_q_{max}$—0.05) and then contrasts the five best and five worst simulations in terms of SCA (as also done for Figures 6, 8, and 10), the solutions with a good SCA performance are generally characterized by larger values for the parameter *cv*, higher melt factors, and in some cases (Cholma and Gulcha) lower threshold melt temperatures than the simulations with a poor SCA performance. Additionally, the solutions with low SCA performance have in some cases also lower values for the interflow delay factor, indicating that the too late snowmelt is compensated by faster subsurface transport.

Considering how the various parameters impact the model response, the model parameters are affected by the different objectives in a plausible way. A large variability in the parameter values is typical in many hydrological modeling applications, where often very different parameter sets may lead to similar performance of the objective function values [*Beven and Binley*, 1992]. The parameter distributions indicate that this problem can to some extent be alleviated when in addition to observed discharge also the performance in terms of SCA is considered. This is, for example, clearly demonstrated for the parameter *cv*. In other cases, adding the snow cover criterion may lead to a shift in the distribution, with possibly only small effects on further constraining the distribution. The main advantage of taking into account snow cover data should however be seen in improving model consistency.

4.4. Value of Increasing the Number of Snow Cover Observations During the Calibration Period

In order to determine how many snow cover images are necessary for successful model calibration, we investigated how the snow cover objective function in the validation period changed with the number of snow cover images in the calibration period. Generally, the median SCA error continually decreased with an increasing number of snow cover images in the calibration period (Figure 12). This decrease was strongest over the first 1–4 mages, while there was little further decrease when increasing the number of images from 10 to the maximum possible number of images in the calibration period (depending on catchment and period, 17–22 images).

However, a larger number of images was necessary to reduce the maximum SCA error. If the number of images is too low, the identified parameter sets may not be representative. A model calibrated to all images is likely to over- or underestimate individual observations, due to deficiencies of the model (including inputs and parameters) or errors in these observations. A small sample can in the worst case therefore be dominated by images where such a well-calibrated model would for example always overestimate the SCA. In such a case the SCA images would shift the model in the wrong direction. The maximum SCA error was

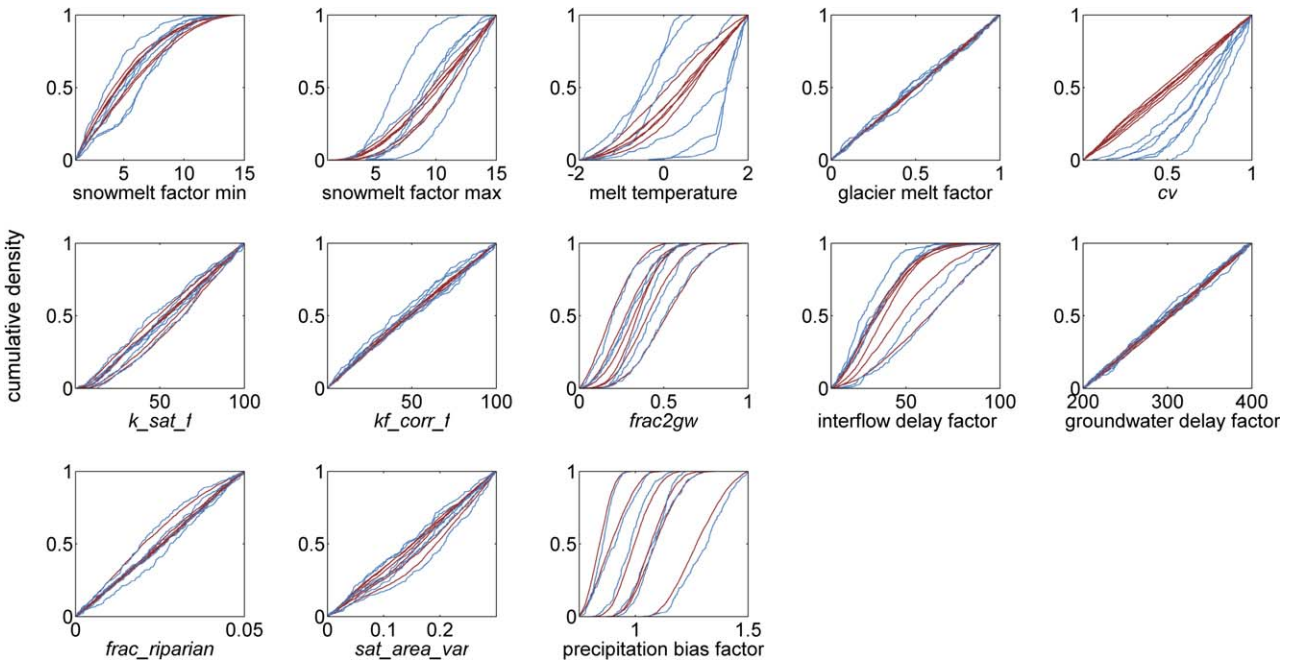

Figure 11. Cumulative distribution functions of parameter sets selected from the Monte Carlo sample as the 5% best performing solutions in terms of *objf_q* (red lines) and after further constraining these solutions selecting the 5% best performing solutions in terms of *objf_sca* (blue lines) for calibration to simulation period A. The six lines indicate the different catchments (for more details please refer to the text).

higher when using few images in the calibration period than when using no image and randomly selecting 20 parameter sets. The reason for this is that there is a very large number of possibilities to draw 20 simulations from the available simulations (all simulations where the objective function value with respect to discharge during the calibration was above 0.5) so that it is very unlikely to draw the 20 worst simulations by random chance. An image which indicates an overestimation of the model, although the model actually overall underestimates SCA, is thus much more effective in selecting simulations with poor performance with respect to SCA. In our study, in most catchments around 10 images were necessary to reduce the maximum error below the value for the case where no image was used.

In the catchments Tosoi and Donguztoo up to 16 images were necessary to reduce the maximum error below the level where no image was used for model calibration. A possible reason why a larger number of images was needed in these two catchments is the lower elevation range (Figure 2), resulting in a lower

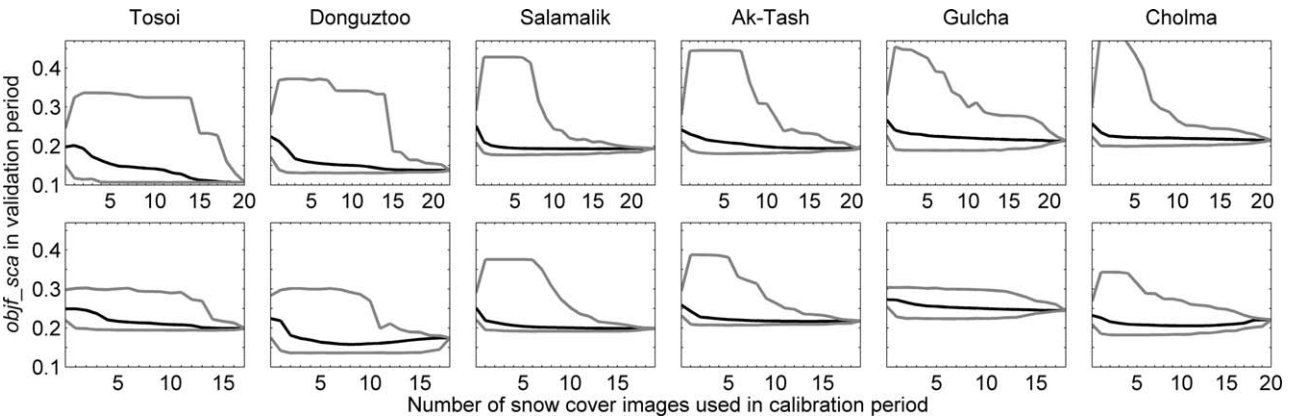

Figure 12. Value of the objective function with respect to SCA (*objf_sca*) in the validation period as a function of the number of snow cover images *n* used in the calibration period. The black line shows the median over 10,000 repetitions of selecting *n* images, the gray lines show the minimum and maximum, respectively. (top row) Results for calibration to simulation period A and validation in simulation period B, and (bottom row) results for the reversed case.

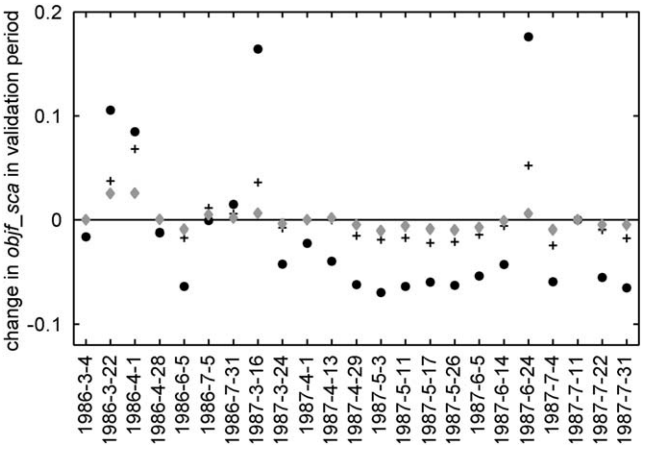

Figure 13. Median effect of individual images on *objf_sca* in the validation period when drawn as first (filled dots), second (crosses), or third (gray diamonds) image, for the catchment Salamalik with simulation period A for model calibration.

number of data points per image. The number of data points per images is determined by the number of elevation zones with maximum 80% cloud cover and more than three cloud-free cells (see section 3.3.1). The average number of data points per image in Tosoi and Donguztoo was 6.2 and 5.2, respectively; in contrast it was between 10.7 and 13.4 in the other catchments. Furthermore, many images in Tosoi and Donguztoo did not allow locating the transitional elevation zones between snow and snow-free areas, as they showed only snow-free or only snow-covered elevation zones. It is likely that images with a lower number of observations and images which do not allow locating the transitional elevation zones between snow and snow-free areas have a lower constraining power. Therefore, more images may be necessary to compensate the effect of images which tend to draw the model in the wrong direction.

Figure 13 illustrates the effect of individual images on *objf_sca* in the validation period when drawn as first, second, or third image, shown as the median change in *objf_sca* in the validation period. The example depicts the catchment Salamalik with simulation period A for model calibration. When using only one image for model calibration, most images result in an improvement of *objf_sca* in the validation period. In contrast, there are some images (22 March 1986, 1 April 1986, 16 March 1987, and 24 June 1987) which cause a large deterioration of *objf_sca* in the validation period. The effect of such images can be compensated when further images are added. Generally, the influence of individual images declines with an increasing number of images. While the median changes in *objf_sca* in the validation period, as shown in Figure 13, are mostly only small, the maximum positive or negative change can be much larger, since the influence of an image added as the second or third image of course also depends on which other image(s) are used for constraining the model. For example an image which has little constraining power when added as the first image, can in combination with particular other images still result in an improvement of the model performance in the validation period.

Snow cover scenes can only be useful for model calibration if there is variability in the simulated snow cover at that time. Therefore, the unconstrained model may be used for identifying suitable time periods in which snow cover data may be beneficial for model calibration. Figure 14 shows the range of simulated SCA of 1000 Monte Carlo simulations (generated using Latin hypercube sampling and the parameter ranges from Table 2) which have not been constrained by any observational data. Early in the melt season, data points at high elevations have no value in constraining the model, as there is no variability in the simulations. However, these snow cover scenes deliver useful observation points at low to medium elevations. In contrast, snow cover scenes in the late melt season deliver useful data points at medium to high elevations. Such relatively fast initial simulations, which can be performed a priori before snow cover data for model calibration is available, can be used for identifying variability hotspots in the simulations. Based on this, one can then select suitable time periods for which snow cover data should be processed.

5. Conclusions

This study evaluates the benefit of satellite-derived snow cover images for the calibration of a hydrological model in snow-dominated catchments in Central Asia. In most of the catchments, we found only small trade-offs between good simulations with respect to discharge and SCA. However, if the parameters were selected based on the discharge objective function only, this could lead to simulations with large snow cover errors. The fact that good discharge simulations were also achieved with large snow cover errors and

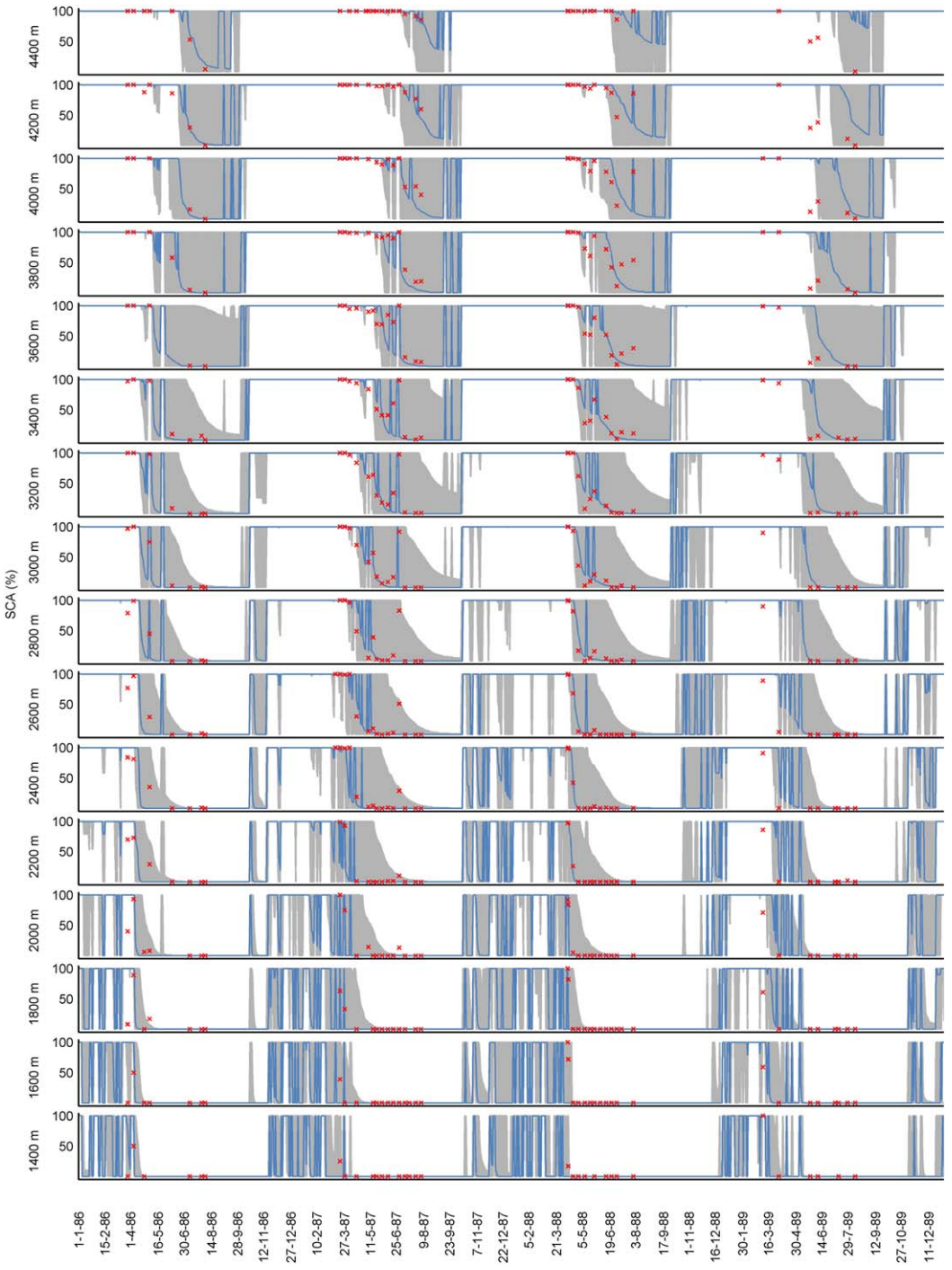

Figure 14. Simulated snow cover ranges of 1000 Monte Carlo simulations of each elevation zone over the time period 1986–1989 for the catchment Cholma (gray-shaded area). Red crosses show the snow cover observations. The blue line shows simulated snow cover for the validation period from a balanced solution of the multiobjective optimization.

thus for the wrong reasons demonstrates very clearly that SCA observations should be taken into account for model calibration in order to achieve higher internal consistency of the model. Using a parameterization for fractional SCA, allowed a direct comparison between simulated and observed SCA for each elevation zone without requiring any thresholds for the comparison of simulated SWE to observed SCA. This approach could also be advantageous for other semidistributed models, where the comparison to observed satellite-derived SCA is generally less straight-forward than in raster-based models.

While the average SCA error was already reduced with only few images, it is recommended to use a larger number of images (10–16 SCA scenes in this study) for model calibration in order to also reduce the maximum error in case of an unfavorable selection of satellite scenes. Our results should be seen as a starting point toward a more general understanding of the value of increasing the number of snow cover images for model calibration. As the six catchments investigated in this study are located in the same region, the results should at this point only be transferred to catchments with a similar physiographic setting and snowfall regime. Further studies should investigate how the results change with changes in the physiographic setting (e.g., catchments with a much lower elevation range) or snowfall regime (e.g., many snow events also during the melt season, or not one distinct snow season but rather a number of shorter snow events over the winter period). As a further step, it should also be explored whether there are characteristics of snow cover images (e.g., particular patterns of snow cover) which make them particularly useful for model calibration. It might for example be important that the set of images contains scenes with the snowline at low and at high elevations.

We recommend a wider application of SCA data for model calibration if snowmelt is an important runoff generation process in the catchment of interest. This study demonstrated that good discharge simulations could also be attained with large snow cover errors, if snow cover data were not taken into account for calibration. As satellite-derived SCA data are available globally, considering these data for calibration is a very good opportunity to improve hydrological modeling also in remote, data sparse areas. This is particularly important if the model is applied for an analysis of hydrological processes or for climate change scenarios.

Acknowledgments

This study was conducted within the projects CAWa (Water in Central Asia; http://www.cawa-project.net), funded by the German Federal Foreign Office (grant AA7090002), and SuMaRiO (Sustainable Management of River Oasises along the Tarim River; http://www.sumario.de/), funded by the German Federal Ministry of Education and Research (grant 01LL0918A). We thank Patrick Reed and Joshua Kollat from the Pennsylvania State University for providing the ε-NSGAII algorithm and Marina Köhler for her assistance with the computing cluster. The Central Asian Institute for Applied Geosciences (CAIAG) digitized soil and discharge data and the Hydrometeorological Service of Kyrgyzstan provided discharge and meteorological data. We would also like to thank the associate editor and two anonymous reviewers for their comments and suggestions which helped improving this manuscript.

References

Academy of Science of the Kyrgyz SSR (1987), *Atlas of the Kyrgyz Soviet Socialistic Republic, Natural Conditions and Resources*, vol. 1, State Agency for Cartogr. and Geod., Moscow.

Anderson, E. (2006), Snow accumulation and ablation model—SNOW-17, NOAA's National Weather Service Hydrology Laboratory NWSRFS user manual report, 61 pp, NOAA, Silver Spring, Md.

Andreadis, K. M., and D. P. Lettenmaier (2006), Assimilating remotely sensed snow observations into a macroscale hydrology model, *Adv. Water Resour.*, *29*(6), 872–886.

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

Corbari, C., G. Ravazzani, J. Martinelli, and M. Mancini (2009), Elevation based correction of snow coverage retrieved from satellite images to improve model calibration, *Hydrol. Earth Syst. Sci.*, *13*(5), 639–649.

De Lannoy, G. J. M., R. H. Reichle, K. R. Arsenault, P. R. Houser, S. Kumar, N. E. C. Verhoest, and V. R. N. Pauwels (2012), Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, *Water Resour. Res.*, *48*, W01522, doi:10.1029/2011WR010588.

Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan (2002), A fast and elitist multiobjective genetic algorithm: NSGA-II, *IEEE Trans. Evol. Comput.*, *6*(2), 182–197.

Donald, J. R., E. D. Soulis, N. Kouwen, and A. Pietroniro (1995), A land cover-based snow cover representation for distributed hydrological models, *Water Resour. Res.*, *31*(4), 995–1009.

Duethmann, D., J. Zimmer, A. Gafurov, A. Güntner, D. Kriegel, B. Merz, and S. Vorogushyn (2013), Evaluation of areal precipitation estimates based on downscaled reanalysis and station data by hydrological modelling, *Hydrol. Earth Syst. Sci.*, *17*, 2415–2434.

Engeset, R. V., H. C. Udnaes, T. Guneriussen, H. Koren, E. Malnes, R. Solberg, and E. Alfnes (2003), Improving runoff simulations using satellite-observed time-series of snow covered area, *Nord. Hydrol.*, *34*(4), 281–294.

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.

Francke, T., A. Güntner, G. Mamede, E. N. Müller, and A. Bronstert (2008), Automated catena-based discretization of landscapes for the derivation of hydrological modelling units, *Int. J. Geogr. Inf. Sci.*, *22*(2), 111–132.

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.

Friedl, M. A., et al. (2002), Global land cover mapping from MODIS: Algorithms and early results, *Remote Sens. Environ.*, *83*(1–2), 287–302.

Güntner, A. (2002), Large-scale hydrological modelling in the semi-arid North-East of Brazil, *PIK Rep. 77*, Potsdam Inst. for Clim. Impact Res., Potsdam, Germany.

Güntner, A., and A. Bronstert (2004), Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas, *J. Hydrol.*, *297*(1–4), 136–161.

Güntner, A., S. Uhlenbrook, J. Seibert, and C. Leibundgut (1999), Multi-criterial validation of TOPMODEL in a mountainous catchment, *Hydrol. Processes*, *13*(11), 1603–1620.

Hock, R. (2003), Temperature index melt modelling in mountain areas, *J. Hydrol.*, *282*(1–4), 104–115.

Höppner, E., and N. Prechtel (2002), Snow Cover Mapping with NOAA-AVHRR Images in the Scope of an Environmental GIS Project for the Russian Altai (South Siberia), in *Proceedings of the Commission on Mountain Cartography of the International Cartographic Association (ICA)*, Mt. Hood, Ore., 15-19, 26, May.

Jarvis, A., H. I. Reuter, A. Nelson, and E. Guevara (2008), *Holefilled Seamless SRTM Data*, Int. Cent. for Trop. Agric. (CIAT), Cali, Colombia.

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.

Klemes (1986), Operational testing of hydrological simulation models, *Hydrol. Sci. J.*, *31*(1), 13–24.

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.
Kollat, J. B., and P. M. Reed (2006), Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design, *Adv. Water Resour.*, *29*(6), 792–807.
Kollat, J. B., P. M. Reed, and T. Wagener (2012), When are multiobjective calibration trade-offs in hydrologic models meaningful?, *Water Resour. Res.*, *48*, W03520, doi:10.1029/2011WR011534.
Konak, A., D. W. Coit, and A. E. Smith (2006), Multi-objective optimization using genetic algoithms: A tutorial, *Reliab. Eng. Syst. Safety*, *91*, 992–1007.
Konz, M., and J. Seibert (2010), On the value of glacier mass balances for hydrological model calibration, *J. Hydrol.*, *385*(1–4), 238–246.
Lamb, R., K. Beven, and S. Myraboe (1998), Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model, *Adv. Water Resour.*, *22*(4), 305–318.
Li, X. G., and M. W. Williams (2008), Snowmelt runoff modelling in an arid mountain watershed, Tarim Basin, China, *Hydrol. Processes*, *22*(19), 3931–3940.
Liston, G. E. (2004), Representing subgrid snow cover heterogeneities in regional and global models, *J. Clim.*, *17*(6), 1381–1397.
Liu, Y. Q., C. D. Peters-Lidard, S. Kumar, J. L. Foster, M. Shaw, Y. D. Tian, and G. M. Fall (2013), Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska, *Adv. Water Resour.*, *54*, 208–227.
Luce, C. H., and D. G. Tarboton (2004), The application of depletion curves for parameterization of subgrid variability of snow, *Hydrol. Processes*, *18*(8), 1409–1422.
Madsen, H. (2000), Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, *J. Hydrol.*, *235*(3–4), 276–288.
Mroczkowski, M., G. P. Raper, and G. Kuczera (1997), The quest for more powerful validation of conceptual catchment models, *Water Resour. Res.*, *33*(10), 2325–2335.
Mueller, E., A. Güntner, T. Francke, and G. Mamede (2010), Modelling sediment export, retention and reservoir sedimentation in drylands with the WASA-SED model, *Geosci. Model Develop.*, *3*(1), 275–291.
Myneni, R. B., et al. (2002), Global products of vegetation leaf area and fraction absorbed PAR from year one of MODIS data, *Remote Sens. Environ.*, *83*(1–2), 214–231.
National Climate Data Center (1998), *NOAA Polar Orbiter Data User's Guide*, Suitland, Md.
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.
Parajka, J., R. Merz, and G. Blöschl (2007), Uncertainty and multiple objective calibration in regional water balance modelling: Case study in 320 Austrian catchments, *Hydrol. Processes*, *21*(4), 435–446.
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.
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.
Perrin, C., L. Oudin, V. Andreassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*(1), 131–151.
Raleigh, M. S., K. Rittger, C. E. Moore, B. Henn, J. A. Lutz, and J. D. Lundquist (2013), Ground-based testing of MODIS fractional snow cover in subalpine meadows and forests of the Sierra Nevada, *Remote Sens. Environ.*, *128*, 44–57.
Rao, C. R. N., and J. Chen (1995), Inter-satellite calibration linkages for the visible and near-infared channels of the advanced very high resolution radiometer on the NOAA-7, −9, and −11 spacecraft, *Int. J. Remote Sens.*, *16*(11), 1931–1942.
Refsgaard, J. C. (1997), Parameterisation, calibration and validation of distributed hydrological models, *J. Hydrol.*, *198*, 69–97.
Rodell, M., and P. R. Houser (2004), Updating a land surface model with MODIS-derived snow cover, *J. Hydrometeorol.*, *5*(6), 1064–1075.
Schaefli, B., and M. Huss (2011), Integrating point glacier mass balance observations into hydrologic model identification, *Hydrol. Earth Syst. Sci.*, *15*(4), 1227–1241.
Seibert, J. (2000), Multi-criteria calibration of a conceptual runoff model using a genetic algorithm, *Hydrol. Earth Syst. Sci.*, *4*(2), 215–224.
Seibert, J., and J. J. McDonnell (2002), On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration, *Water Resour. Res.*, *38*(11), 1241, doi:10.1029/2001WR000978.
Shrestha, M., L. Wang, T. Koike, H. Tsutsui, Y. Xue, and Y. Hirabayashi (2013), Correcting basin-scale snowfall in a mountainous basin using a distributed snowmelt model and remote sensing data, *Hydrol. Earth Syst. Sci. Discuss.*, *10*(9), 11,711–11,753.
Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, M. G. Duda, X. Huang, W. Wang, and J. G. Powers (2008), A description of the advanced research WRF Version 3, NCAR Tech. Note NCAR/TN–475+STR, 124 pp., National Center for Atmospheric Research, Boulder, Colo.
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.
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.
Tang, Y., P. Reed, and T. Wagener (2006), How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sci.*, *10*(2), 289–307.
Tang, Y., P. M. Reed, and J. B. Kollat (2007), Parallelization strategies for rapid and robust evolutionary multiobjective optimization in water resources applications, *Adv. Water Resour.*, *30*(3), 335–353.
Udnaes, H. C., E. Alfnes, and L. M. Andreassen (2007), Improving runoff modelling using satellite-derived snow covered area?, *Nord. Hydrol.*, *38*(1), 21–32.
Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*(612), 2961–3012.
Viviroli, D., et al. (2011), Climate change and mountain water resources: Overview and recommendations for research, management and policy, *Hydrol. Earth Syst. Sci.*, *15*(2), 471–504.
Voigt, S., M. Koch, and M. F. Baumgartner (1999), A multichannel threshold technique for NOAA AVHRR data to monitor the extent of snow cover in the Swiss Alps, in *Interactions Between the Cryosphere, Climate and Greenhouse Gases, Proc. IUGG 99 Symp. HS2*, IAHS Publ., Birmingham, U. K.
Yatheendradas, S., et al. (2012), Distributed assimilation of satellite-based snow extent for improving simulated streamflow in mountainous, dense forests: An example over the DMIP2 western basins, *Water Resour. Res.*, *48*, W09557, doi:10.1029/2011WR011347.
Zaitchik, B. F., and M. Rodell (2009), Forward-looking assimilation of MODIS-derived snow-covered area into a land surface model, *J. Hydrometeorol.*, *10*(1), 130–148.