**Hydrology and
Earth System
Sciences**

Integrating point glacier mass balance observations into hydrologic model identification

B. Schaefli[1,2] **and M. Huss**[3,4]

[1]Laboratory of Ecohydrology (ECHO), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[2]Water Resources Section, Delft University of Technology (TU Delft), Delft, The Netherlands
[3]Department of Geosciences, University of Fribourg, Fribourg, Switzerland
[4]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland

Abstract. The hydrology of high mountainous catchments is often predicted with conceptual precipitation-discharge models that simulate the snow accumulation and ablation behavior of a very complex environment using as only input temperature and precipitation. It is hereby often assumed that some glacier-wide annual balance estimates, in addition to observed discharge, are sufficient to reliably calibrate such a model. Based on observed data from Rhonegletscher (Switzerland), we show in this paper that information on the seasonal mass balance is a pre-requisite for model calibration. And we present a simple, but promising methodology to include point mass balance observations into a systematic calibration process.

The application of this methodology to the Rhonegletscher catchment illustrates that even small samples of point observations do contain extractable information for model calibration. The reproduction of these observed seasonal mass balance data requires, however, a model structure modification, in particular seasonal lapse rates and a separate snow accumulation and rainfall correction factor.

This paper shows that a simple conceptual model can be a valuable tool to project the behavior of a glacier catchment but only if there is enough seasonal information to constrain the parameters that directly affect the water mass balance. The presented multi-signal model identification framework and the simple method to calibrate a semi-lumped model on point observations has potential for application in other modeling contexts.

1 Introduction

Continuous simulation of discharge has become a standard tool for water resources management at the catchment-scale, e.g. to estimate design floods from long simulated time series (e.g. Hingray et al., 2010) or to predict climate and land use change impacts on hydrologic regimes (e.g. Horton et al., 2006). For this type of applications, conceptual models play an important role: they represent the dominant rainfall-runoff processes directly at the catchment-scale instead of resolving the small-scale physical processes (see Sivapalan et al., 2003). In high mountainous catchments, the focus of the present paper, such conceptual models represent e.g. the snow accumulation and ablation behavior of a very complex environment with a set of simple water balance equations using only precipitation and temperature as input (Hock, 2003; Klok et al., 2001; Schaefli et al., 2005; Stahl et al., 2008).

The predictive power and the reliable calibration of such conceptual models receives a continuously growing interest in hydrological research (Bardossy and Singh, 2008; Clark et al., 2010; Reusser et al., 2009; Vrugt et al., 2005; Yilmaz et al., 2008). In catchments with a significant glacier cover, this is particularly challenging since a precipitation-runoff model should not just "mimic" observed discharge but also reproduce snow accumulation and melt dynamics or the glacier mass change (e.g. Konz and Seibert, 2010). This reproduction of fluxes and storage changes is essential to correctly simulate the ongoing fast evolution of glaciers and related runoff dynamics (Moore and Demuth, 2001; Pellicciotti et al., 2010).

This has of course long been recognized and numerous authors have validated their simulations against observed

glacier mass balance (Braun and Renner, 1992; Koboltschnig et al., 2008; Konz et al., 2007). Schaefli et al. (2005) were among the first to include annual glacier-wide mass balance data into a rigorous multi-signal parameter estimation procedure. They constrained their daily discharge model on discharge and on the available three annual mass balance observations and showed that the resulting model reproduces reasonably well discharge and the altitudinal mass balance distribution. In absence of concomitant mass balance and discharge observations, Stahl et al. (2008) used reconstructed winter and summer mass balance data to assist step-wise model parameter selection and concluded that observed mass balances could greatly reduce the prediction uncertainty in glacier catchments. Konz and Seibert (2010) proposed a framework to investigate the question how useful such a limited amount of annual mass balance data in combination with a limited amount of discharge observations might be for the calibration of a glacio-hydrological model; they concluded that a single annual glacier-wide mass balance observation might contain useful information, especially in combination with carefully selected discharge observations.

One of the challenges in the combination of mass balance and discharge observations for model calibration is the integration of very different types of data containing spatially integrated information such as discharge or glacier mass balance, or point information such as snow height or ice ablation observations (Konz et al., 2007). We, therefore, developed a methodology that can make use of sparse point observations to calibrate and evaluate a semi-lumped precipitation-discharge simulation model for high mountainous catchments.

This work was triggered by the analysis of a case study in the Swiss Alps, the Rhonegletscher catchment, for which Schaefli et al. (2005) have presented a conceptual hydrological model (GSM-SOCONT) and Huss et al. (2008) a more complex glaciological model with higher spatial resolution. Both models simulate monthly, seasonal and annual glacier mass balances. However, the hydrological model GSM-SOCONT yields on average much less winter accumulation than the glaciological model and slightly more ablation during summer months. This striking difference, which would namely result in rather different predictions of glacier surface evolution, led us to a set of research questions: can a conceptual precipitation-runoff model calibrated (i) exclusively on runoff or (ii) on runoff and glacier-wide annual mass balance reproduce the seasonal glacier mass balances? or does a reliable calibration require to incorporate some in-situ measurements of snow accumulation and ablation processes? and if yes, how can such such point observations be incorporated?

Hereafter, we first present the case study (Sect. 2) and the used models (Sect. 3), before discussing in detail the method developed to merge different types of glacio-hydrological data into a precipitation-discharge model (Sect. 4). As illustrated and discussed in Sects. 5 and 6, this method provides valuable new insights into the use of multi-signal calibration
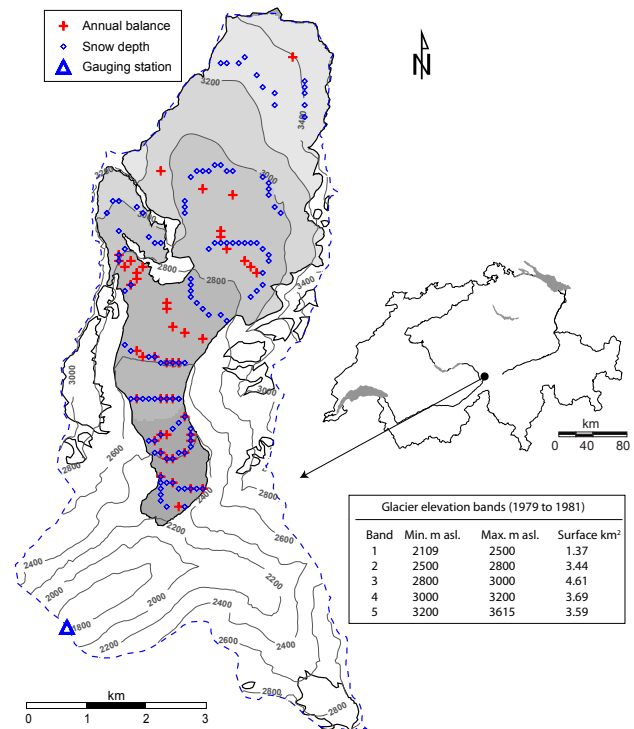
Fig. 1. Location of the case study in Switzerland and map of the catchment showing Rhonegletscher in gray shades and the measurement locations for 1979/1980 (glacier geometry of 1980).

for model development in general and into the question of how much information we need for discharge prediction in high Alpine environments.

2   Case study: Rhonegletscher catchment

The Rhonegletscher catchment at Gletsch has a size of 38.9 km² with a mean elevation of 2719 m a.s.l. (Fig. 1). The outlet is located at 1761 m a.s.l.; around 2.4% of the catchment area are covered by forest and another 19.0 % by other vegetation. Daily discharge measurements are available continuously since 1956.

Rhonegletscher is a medium-sized valley glacier and presently covers almost half of the catchment. Its size was 17.3 km² in 1980 and 15.9 km² in 2007 (Bauder et al., 2007). Several other small glaciers and ice patches are present in the catchment (around 2.3 km²), which are included in the total ice-covered area for discharge simulations. There are no meteorological measurement stations located within the catchment. The nearest stations are Grimsel (altitude 1980 m a.s.l., longitude 8°19′60″, latitude 46°34′18″), Ulrichen (1346 m a.s.l., 8°18′29″, 46°30′17″) and Oberwald (1375 m a.s.l., 8°20′48″, 46°32′03″). The station at Oberwald only measures precipitation and stopped recording in 1999. Grimsel is the closest station (distance to nearest

catchment point 1.5 km) but is located in a neighboring valley. Oberwald as well as Ulrichen lie downstream of the catchment outlet with a distance of 3.0 km and 7.4 km, respectively. The choice of the meteorological time series will be further discussed in Sect. 5.2.

The hypsometry of the catchment is derived from a digital elevation model with a resolution of 25 m (SwissTopo, 1995), the one of Rhonegletscher from photogrammetry that documents the changes in glacier surface elevation and area over the last decades. Based on this, Bauder et al. (2007) also determined the total glacier ice volume change over the corresponding time intervals.

A comprehensive data set of seasonal point mass balance observations is available from the direct glaciological measurements over five years, including a winter survey in April to May and a late summer field visit in September (Funk, 1985; Huss et al., 2008). Winter accumulation is measured using probings of the snow depth distributed all over the glacier and annual mass balance is determined at stakes drilled into the ice. Field data are available for the years 1979/1980 to 1981/1982 and for 2006/2007 and 2007/2008. For the individual surveys, between 10 and 50 ablation measurements and up to 800 snow soundings were performed.

3 Glacio-hydrological model

3.1 Original GSM-SOCONT

The hydrological model GSM-SOCONT uses a semi-lumped simulation approach. It computes snow accumulation and snow- and ice melt per elevation band (here 5 elevation bands for the glacier, Fig. 1, and 5 for the non-glacier part of the catchment). For each band, the reference precipitation and temperature time series are interpolated according to the difference between its mean elevation and the reference station altitude. For temperature, this interpolation uses a temperature lapse rate ρ ($°C\,m^{-1}$); for precipitation, it uses a proportional increase with altitude, γ_p ($\%\,m^{-1}$), of the amount observed at the reference meteorological station.

In the original formulation, the aggregation state of precipitation at the mean elevation of the band is determined based on a simple temperature threshold T_c ($°C$). To reduce the model parameter dependency on the spatial resolution (i.e. the number of elevation bands), we adopt here a slightly modified snowfall computation method. We compute the amount of snowfall $W(t,z)$ and of rainfall $V(t,z)$ at a given time t and an altitude z and integrate over the hypsometric curve of each elevation band.

The average snow melt per elevation band, is computed based on a direct linear relation with positive air temperature (Hock, 2003) using a simple degree-day approach:

$$M_s(t) = \begin{cases} a_s(T - T_m) & \text{if } T > T_m, \ H_s(t) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where a_s is the degree-day factor for snow melt ($mm\,d^{-1}\,°C^{-1}$), T_m is the threshold temperature for melting that is set to $0\,°C$ and $H_s(t)$ is the depth of the snow cover at time t in mm water equivalent. For glacier elevation bands, ice melt is assumed to occur only when $H_s(t) = 0$:

$$M(t)_i = \begin{cases} a_i(T - T_m) & \text{if } T > T_m, \ H_s(t) = 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where a_i is the degree-day factor for ice melt ($mm\,d^{-1}\,°C^{-1}$). The melt from glacier elevation bands is routed to the catchment outlet through two linear reservoirs, one for snow and one for ice, each having a linear outflow parameter k_s and k_i respectively. For elevation bands of the non-glacier part of the catchment, the melt water, together with rainfall, is routed through a nonlinear reservoir with one parameter (β) and a linear reservoir with two parameters ($\log(k)$, A) (for details see Schaefli et al., 2005).

The model does not contain a separate melt factor and routing parameter for firn (old snow that lasted more than a season). As we discuss in detail in a comment accompanying this paper (Schaefli and Huss, 2011), substantial areas of firn would only be exposed during extremely hot years and related model parameters would be difficult to identify based on mass balance and discharge data.

The above version of GSM-SOCONT has seven hydrological parameters to calibrate: the melt parameters (a_i, a_s), the rainfall and meltwater-runoff transformation parameters for the glacier part (k_i, k_s), and the non-glacier part ($\log(k)$, A, β). In addition, Schaefli et al. (2005) calibrated the meteorological parameter γ_p, whereas the parameter ρ was set to a fixed value ($-0.0065\,°C\,m^{-1}$).

3.2 Model modifications

As will become clear later in this paper, the model requires some structural modifications to reproduce the available discharge and glacier mass balance data. These step-wise modifications include first of all the use of seasonal parameters for the interpolation of temperature. This also reflects the underlying physical processes, since it is well known that temperature lapse rates show strong seasonal fluctuations (Blandford et al., 2008; Rolland, 2003), under the combined effect of dry and moist adiabatic lapse rates and their dependence on temperature (e.g. Miller and Thompson, 1970). The temperature decrease with altitude is on average lower during the cold season than during the warm season. This is also confirmed by an analysis of seasonal temperature variations at all meteorological stations within some tens of kilometers around Rhonegletscher, which showed that temperature lapse rates are strongest in June and reach a minimum from November to January.

The introduction of seasonal temperature lapse rates calls for a modification of the input interpolation to enhance parameter identifiability (further details in Sect. 5.3). We

replace the altitudinal precipitation correction factor with a solid input (snow) correction factor, γ_s (% m^{-1}), and a constant rainfall correction factor, γ_r (% m^{-1}).

The new interpolation of precipitation as a function of elevation reads as

$$P(t,z) = W(t,z) + V(t,z) = [1 + \gamma_s (z - z_{\text{ref}})] W(t, z_{\text{ref}})$$
$$+ [1 + \gamma_r (z - z_{\text{ref}})] V(t, z_{\text{ref}}) \tag{3}$$

where $W(z)$ and $V(z)$ are the snowfall and the rainfall at elevation z, and z_{ref} is the reference altitude of the meteorological station. Given that rainfall and snowfall occur under different meteorological situations, their average scaling with elevation can be assumed to be quite different and, accordingly, this separate altitudinal interpolation makes sense from a physical perspective.

A further modification is the introduction of a snow melt "delay" module to account for the delay of meltwater production at the beginning of the ablation season. From a physical point of view, this observed delay between the moment when air temperature rises above melt temperature and the moment when there is an effective water outflow from the snow pack, can be explained by: (i) the water retention capacity of the snow or (ii) the thermal state of the snow, where melt water is only released if the snowpack has been heated up to melt temperature. In temperature-index snow models, the thermal state of the snowpack cannot be quantified directly and, accordingly, there are very few examples in the literature where this concept is used (for an example, see Garçon, 1996). The concept of water retention capacity is, in exchange, often used in conceptual models (e.g. Kuchment and Gelfan, 1996; Seibert, 1999).

There is an important difference between the above two options for delaying snow melt: introducing a snow pack water retention capacity will only affect the discharge simulations; accounting for the thermal state of the snow pack, i.e. assigning some incoming energy to snowpack heating instead of melting, will have an effect on the discharge and on the mass balance computation. As tests during this study showed, accounting for the thermal state cannot improve the model performance, neither for daily discharge nor for mass balance simulations (details are available in the Supplement).

We, therefore, present here only the retention capacity module: instead of adding snow melt $M_s(t)$ directly to the linear routing reservoir as in the original GSM-SOCONT model, a meltwater delay is obtained by simulating separately the balance of the solid snow store $H_s(t)$ (mm) and of the liquid snow store $H_l(t)$ (mm):

$$\frac{dH_s(t)}{dt} = W(t) - M_s(t), \tag{4}$$

$$\frac{dH_l(t)}{dt} = M_s(t) - Q_s(t), \tag{5}$$

where $W(t)$ (mm d^{-1}) is the snowfall and $Q_s(t)$ (mm d^{-1}) is the water outflow from the snowpack computed as a function of the snowpack retention capacity η_s as follows:

$$Q_s(t) = \max[0, \tfrac{H_l(t) - \eta_s H_s(t)}{\Delta t}], \tag{6}$$

where Δt is the time step and $\eta_s H_s(t)$ represents the maximum water holding capacity of the snow pack at time t.

3.3 Reference mass balance simulation

To answer one of the main questions underlying this study, the mass balance simulations obtained with the glacio-hydrological model have to be compared to a reference data set. In absence of sufficiently long time series (the observed data cover only five non consecutive years), we use as a reference the monthly mass balance simulations of Huss et al. (2008), who used a distributed accumulation and melt model with a high temporal and spatial resolution.

Their model is driven by daily meteorological data and calculates the components of glacier surface mass balance for 25 m × 25 m grid cells. Thus, the spatial distribution of mass balance quantities is addressed in a more sophisticated way than in GSM-SOCONT.

Snow and ice melt is computed using a distributed temperature-index model (Hock, 1999), where the degree-day factors vary in space as a function of potential direct solar radiation to account for the effects of slope, aspect and shading. Instead of a simple altitude dependent precipitation gradient, this model uses a spatial snow accumulation pattern, derived from detailed winter accumulation measurements, to redistribute the precipitation input. Thus, effects of preferential snow deposition and wind-driven snow redistribution are taken into account. The amount of snowfall is estimated based on a linear transition from snowfall to rainfall between the thresholds of 0.5 °C and 2.5 °C. Measured precipitation is corrected using a multiplier to match the observed winter accumulation in the five seasons with directly observed field data.

Huss et al. (2008) combined this glacier mass balance model to a runoff routing scheme comparable to the one of GSM-SOCONT and calibrated it for the Rhonegletscher catchment. The parameters of the glaciological model were estimated based on the available observations of ice volume changes over subdecadal to multidecadal periods, while constraining the spatial mass balance variability with the seasonal in-situ field measurements. The glaciological parameters were, however, not explicitly calibrated on daily discharge; total annual discharge was only used to qualitatively assess the estimates of areal precipitation on the non-glacier part of the catchment.
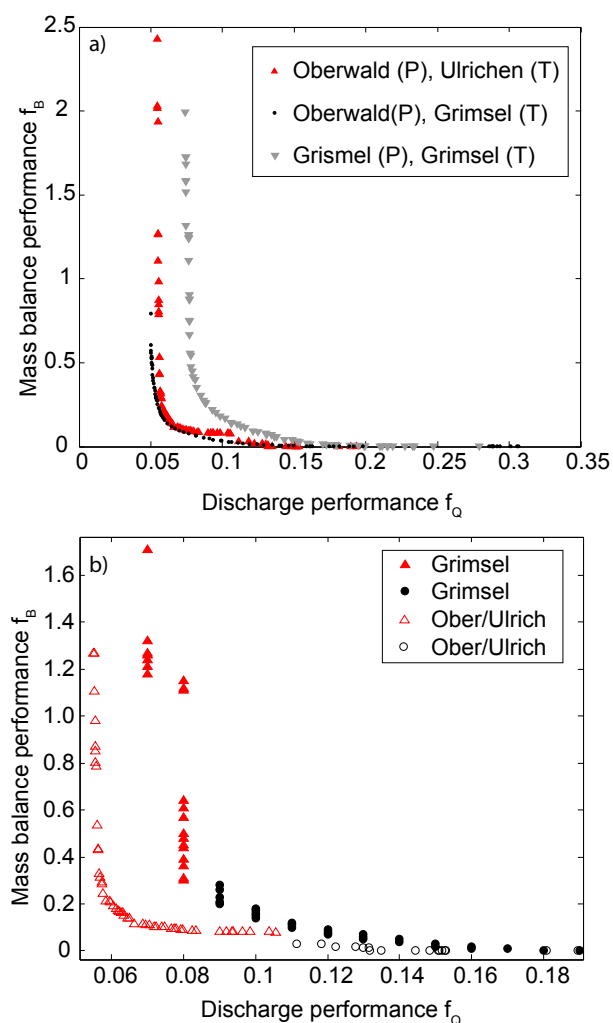
Fig. 2. Pareto-optimal frontiers (POF) illustrating the trade-off between model performance for the simulation of discharge ($f_Q(\theta)$) and of annual glacier-wide balance ($f_B(\theta)$) (the calibration presented in Schaefli et al., 2005, has $f_B = 0.57$ and $f_Q = 0.06$); **(a)** for different combinations of meteorological input stations for the original model; **(b)** for the model version with seasonal lapse rates (the color and marker coding is the same as in Fig. 4.

4 Model identification framework

4.1 Theoretical background

We propose a framework to extract as much information as possible from the available data sets. Rather than focusing on simply estimating the globally best parameter set, the method attempts to "inform the model" about potential data and model structural deficiencies while identifying the most likely ranges of parameter values.

We call these most likely parameter ranges posterior or updated ranges as in classical Bayesian parameter estimation (e.g. Kavetski et al., 2006), since they result from updating

the modeler's a priori knowledge about the parameter values with all available information about the system behavior.

If this updating includes several different types of data (as in the present case study), there are two possible approaches: we can either identify the posterior parameter range by confronting the model *sequentially* with different types of data, or by confronting it *simultaneously* through multi-objective optimization. The two different approaches yield different types of results and extract different information from the data.

4.1.1 Model improvement through multi-objective optimization

Multi-Objective Optimization (MOO) attempts to identify parameter sets that correspond to points on the so-called Pareto-optimal frontier (POF), the limit in the objective function space beyond which it is not possible to improve one objective function without worsening the others (Gupta et al., 1998; an example is shown in Fig. 2).

A real-world model with imperfect model structure and imperfect input data can generally not exactly reproduce any of the observed reference data sets. The trade-off between the different objective functions contains information about how well the model structure can explain the observed data. (This information is partly lost in sequential calibration where the result of each calibration step represents the joint information extracted from all previous steps).

It is not straightforward to use this objective function space trade-off for model development or for statistical interpretation of the results (see Reichert and Mieleitner, 2009). The trade-off between the objective functions is related to the non-dominated volume of the POF, i.e. the volume of the objective function space between the POF and the origin. Since each objective function has its own scale, no absolute meaning can be attached to this non-dominated volume. However, if the system to be optimized is modified (e.g. changed model structure) while keeping the objective function definitions unchanged, then the POF with the lower non-dominated volume has a lower trade-off. This property of the POF can be used to improve the model structure (for examples, see Fenicia et al., 2007, 2008; Schaefli et al., 2004) but also to select the most suitable model input data: a simulation resulting from model M_1 with input $\mathbf{Y}_1$ is judged to be better than a simulation from model M_1 with input $\mathbf{Y}_2$ if its POF has at least one better end-point or a smaller non-dominated volume.

It is noteworthy that in MOO, there is always the risk that the used optimization algorithm was simply not able to find points on the true POF. The above assessment can, thus, be misleading for problems difficult to optimize (e.g. a very high number of parameters). A general guideline is that increasing the degree of freedom of a model should always result in a lower trade-off, since the model is more flexible. If this is not the case, the results should be interpreted with care.

4.1.2 Sequential merging

If several different data sets are available, the model parameters can also be estimated in sequence following the classical Bayesian approach: after having estimated a first parameter distribution with a first data set, every additional data set can be used to update the parameter distribution. This approach is precluded for two cases: (i) if the different data sets are incompatible, i.e. if they contain inconsistent information (this could namely arise if the field data contain systematic measurement errors); or (ii) if the available data are sparse and not sufficient to reasonably formulate a statistical model (i.e. not enough data to develop and test the assumptions). Beven and Binley (1992) introduced the Generalized Likelihood Uncertainty Estimation (GLUE), partially to address the above problems. Later on, GLUE built the basis of a rejectionist framework (Freer et al., 2004), where acceptable parameter sets are selected among all a priori possible parameter sets by rejecting those that have too low model objective function performance or lie beyond some limits of acceptability (see, e.g. Winsemius et al., 2009).

While the formal Bayesian approach has been shown to be useful for model calibration on observed reference flux data (in particular discharge, e.g. Kavetski et al., 2006), its use with observed state variables is often precluded because these observations are sparse in time and in space (here we have for example data for only 5 time steps). The rejectionist framework, in exchange, can accommodate different types of data and performance criteria. For an example of sequential parameter rejection for glacio-hydrological modeling, see the work of Stahl et al. (2008).

4.2 Outline of proposed identification framework

We propose a step-wise model identification framework combining elements of both of the above approaches and classical statistical hypothesis testing. In a first step, we use MOO with the observed fluxes and state variables to improve the model structure based on all information that we can extract from the objective function trade-off.

In a second and third step, the parameters of the resulting model are sequentially updated to constrain the model parameter ranges from physically possible (a priori) ranges to ones that are plausible given the point reference data (winter and annual point balances), and to the final best parameter sets for discharge and glacier-wide annual balance simulation.

This sequential calibration combines a statistical hypothesis testing approach to extract the available information from the state observations and proceeds with classical model calibration minimizing the sum of squared residuals between an observed and a simulated times series.

As mentioned before, such a sequential parameter range updating approach only works if the different data sets do not contain inconsistent information. Here, this is ensured by the preliminary procedure of input data set selection and model structure updating.

4.3 Optimization algorithm

If nothing else is stated, parameter optimization is completed with the global optimization algorithm called Queueing Multi-Objective Optimiser (QMOO) developed by Leyland (2002). For an application of this optimizer to hydrology, see Schaefli et al. (2004, 2005). The algorithm has been designed to identify difficult-to-find optima and to solve far more complex problems than the ones presented in this paper, where only up to 14 parameters have to be identified. We, therefore, assume that all identified parameter sets correspond to the best identifiable solution for a given calibration run.

4.4 Model performance criteria

To assess the model performance with respect to seasonal glacier mass balance, the simulated seasonal balances (average per elevation band) have to be compared to the observed point balances. In the case of large sample sizes and assuming that the samples have been taken at representative locations within the elevation band, the straightforward method would be to estimate the spatial mean from the data and then minimize the residuals between this reference value and the simulated mean value.

Glaciological measurements consist typically of very small samples (i.e. 5 to 10 points per elevation band of a few hundred meters). Without any further source of information, it is highly questionable whether such a small sample contains enough information to estimate the spatial average.

Assuming that the samples have been taken at points that reflect the range of seasonal balances, we can, however, at least test whether the observed sample is *compatible* with the simulated average value for the elevation band. We use a two-sided sign test of the null hypothesis that the sample data come from an arbitrary continuous distribution with a given median m (Gibbons and Chakraborti, 2003). This is a simple non-parametric test where the test statistics is computed as follows: count the number of samples s_i that are lower than the median m, $s_i < m$ and the number of sample values $s_i > m$. The test statistics equals to the smaller of the two numbers. This test statistics is known to follow a binomial distribution for N trials (where N is the sample size) and a probability of success of each trial of 0.5. We can reject the null hypothesis at the 0.05 level that the sample comes from a continuous distribution with median m if the p-value (p) of the test statistics is smaller than 0.025. This test is robust for small sample sizes. We used the data from the winters 2006/2007 and 2007/2008, which have a high spatial resolution, to verify that the observed point balances have sufficiently symmetric distributions to not bias the statistical test

when we apply the above statistical test to the mean balance (the hydrological model only yields mean values) instead of the median.

The performance criterion for a band j and a year t_y becomes:

$$g(j, t, \theta) = \begin{cases} 0 \text{ if } p \gg 0.025 \\ 1 \text{ if } p < 0.025 \end{cases} \quad (7)$$

The performance criterion $G(\theta)$ considering all available years and elevation bands becomes:

$$G(\theta) = \frac{1}{J \cdot n} \sum_{j=1}^{J} \sum_{t=1}^{n} g(j, t, \theta). \quad (8)$$

It is important to point out that the above performance criterion G can be used to select acceptable parameter sets but not to find a single optimum parameter set. This simply follows from the binary nature of $g(j, t_y, \theta)$, which does not allow to further discriminate between acceptable sets.

For discharge, the performance criterion to be minimized is the complement to 1 of the classical Nash-Sutcliffe criterion (Nash and Sutcliffe, 1970):

$$f_Q(\theta) = \sum_{t=1}^{n} \left[\widetilde{Q}(t) - \widehat{Q}(t|\theta) \right]^2 \left\{ \sum_{t=1}^{n} \left[\widetilde{Q}(t) - \overline{Q} \right]^2 \right\}^{-1}, (9)$$

where $\widetilde{Q}(t)$ $(\mathrm{mm\,d^{-1}})$ is the observed discharge on day t, $\overline{Q}$ the corresponding mean value, $\widehat{Q}(t|\theta)$ is the simulated discharge given parameter set θ and n is the number of simulated time steps.

If time series of the glacier-wide annual mass balance are available, a similar performance criterion can be defined:

$$f_B(\theta) = \sum_{t=1}^{n_y} \left[\widetilde{B}(t_y) - \widehat{B}(t_y|\theta) \right]^2 \left\{ \sum_{t=1}^{n_y} \left[\widetilde{B}(t_y) \right]^2 \right\}^{-1}, \quad (10)$$

where $\widetilde{B}(t_y)$ $[(\mathrm{m\,y^{-1}})$ is the observed glacier-wide mass balance for the year t_y, $\widehat{B}(t_y|\theta)$ is the corresponding simulated value, n_y is the number of available yearly values.

4.5 Sequential updating and final calibration

After fixing all parameters to an initial guess, this sequential updating is completed as follows:

1. Select a state observation data set (e.g. point mass balance $\widetilde{b}[x, y, t|(x, y) \in j]$) and the corresponding model simulation $\widehat{b}[j, t|\theta]$.

2. Among all model parameters θ, select the subset θ_v that influences the simulation of $\widehat{b}[j, t|\theta]$.

3. Generate r random realizations of θ_v drawn in the prior (Table 2) or updated parameter ranges and compute the corresponding $\widehat{b}[j, t|\theta_v]$.

4. For each $\widehat{b}[j, t|\theta_v]$, test the null hypothesis H_0 that the observed sample $\widetilde{b}[x, y, t|(x, y) \in j]$ (point measurements) comes from a distribution having as median $\widehat{b}[j, t|\theta_v]$ against the alternative hypothesis that it does not (see Sect. 4.4).

5. Retain the parameter sets for which H_0 cannot be rejected for all available data points.

6. If available, repeat points 1 to 8 for additional state observation data sets.

7. Use the updated parameter ranges as prior parameter ranges to calibrate the model on the available time series (discharge, glacier-wide mass balance) (see Sect. 4.4).

5 Results

5.1 Reference data sets for parameter identification

We use the following reference data sets for parameter calibration:

– Daily discharge, the years 1974–1982 for calibration, 1983–1999 for validation (performance criterion f_Q);

– Observed winter point accumulation (direct measurements) for winters 1979/1980 to 1981/1982; for each year, we only use the 3 highest bands (to ensure that the melt parameters do not influence the model result). One sample (elevation band 4, winter 1981/1982) has less than 10 points and is excluded from the analysis. In total, we have $3 \times 3 - 1 = 8$ reference samples.

– Observed annual point balance (direct measurements) for the corresponding hydrological years. Only the two lowest elevation bands have more than 10 observations, resulting in 6 reference samples.

– Glacier-wide annual balance (performance criterion f_B): for the years 1979/1980 to 1981/1982, we have reference values estimated from direct measurements. For the validation of the hydrological model, we complete these reference values with the glacier-wide annual balances simulated with the glaciological model of Huss et al. (2008).

5.2 Selection of meteorological input series

There is the typical problem that the available meteorological time series have been observed at points located outside of the catchment. Assessing the information content of the available series for the modeling purpose is, thus, an essential modeling step. Figure 2a shows the trade-offs between the model performances for discharge simulation and for glacier-wide annual mass balance simulation using the
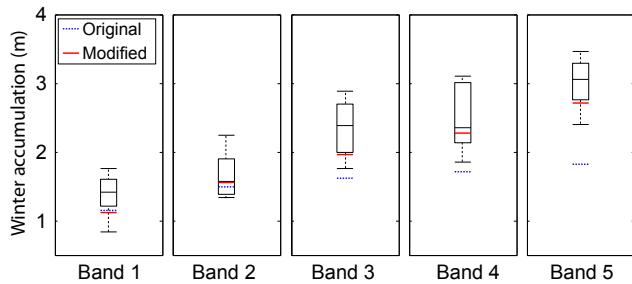
Fig. 3. Boxplots of point samples of winter accumulation for the winter 1979/80 per elevation band; the box represents the interquartile range and the median, the whiskers indicate the 80% range; the blue and red lines indicate the simulated mean values for the original GSM-SOCONT of Schaefli et al. (2005) and an acceptable solution with the modified model version.

Fig. 4. The points of the POF of Fig. 2b in the parameter space rather than in the objective function space.

different available input time series of precipitation and temperature. The 3 Pareto-optimal frontiers (POFs) show clearly that using the precipitation time series observed at Oberwald leads to much better discharge simulations and to a smaller trade-off between obtaining good results for discharge and for mass balance. Accordingly, we prefer this measurement station for this study (even if it stopped recording in 1999). Assessing in detail the differences in information content of the available meteorological series is beyond the scope of this paper.

5.3 Model structure modification

The GSM-SOCONT model (Schaefli et al., 2005) has 6 parameters that influence the glacier mass balance simulations: the ice and snow melt factors, a_i and a_s, the precipitation correction factor γ_p, the temperature lapse rate ρ and the threshold temperature for rainfall/snowfall separation T_c. In the original version, only the first 3 parameters are calibrated, the remaining were fixed, $\rho = 0.0065\,°\text{C}\,\text{m}^{-1}$ and $T_c = 0\,°\text{C}$. The original model, calibrated according to Schaefli et al. (2005) on discharge and glacier-wide annual balance for 3 hydrological years (1979/1980 to 1981/1982), cannot reproduce the seasonal mass balances. To illustrate this result, Fig. 3 shows boxplots of the point observations and the simulated winter accumulation per elevation band for one winter. The simulated mean value does not always lie between the observed 10% and 90% quantiles. This suggests that the inclusion of the winter balance observations adds additional information in the sense that it constrains the model differently than without using them.

An attempt of re-calibration of all of the above model parameters on the seasonal point balances, showed that GSM-SOCONT cannot reproduce both winter and annual point balances: for the winter balance, the smallest achievable value of $G_w(\theta)$ is 2 and for the annual balance $G_a(\theta) = 3$. This means that for the winter balance, there are least two
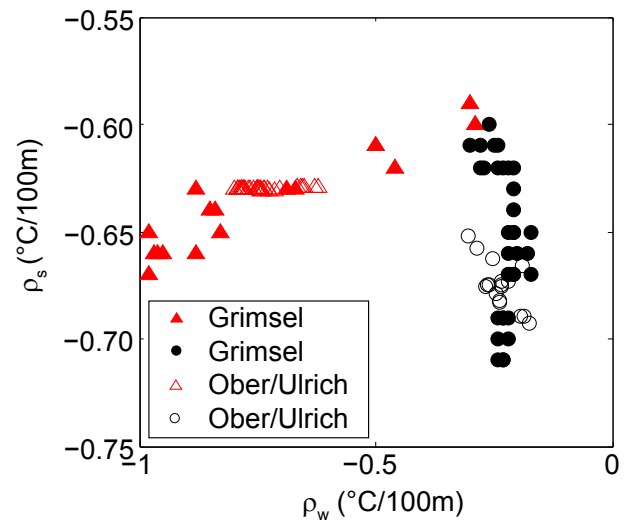
bands for which we can reject the hypothesis that the observed sample has the median value simulated by the model.

In a first step, we tried to use a linear transition between rainfall and snowfall (Rohrer et al., 1994) and to calibrate the threshold parameters between which there is a mixture of snowfall and rainfall. Even if this makes the model more flexible, it is still not possible to achieve $G_w(\theta) = 0$ and $G_a(\theta) = 0$ simultaneously.

We next tested the use of seasonally varying model parameters. We successively increased the number of seasonally varying parameters, starting with the introduction of the seasonal temperature lapse rates ρ_w and ρ_s (winter is defined as from 1 November to 30 April, which corresponds more or less to the dates of winter accumulation observation.) With this additional degree-of-freedom, the model can achieve $G_w(\theta) = 0$ and $G_a(\theta) = 0$, i.e. there are parameter sets that reproduce both the reference point data for winter accumulation and annual balance. This is not surprising: we increased the degree-of-freedom and provided the model with more flexibility to reproduce observed data. It can now, for example, simulate snow melt during the winter season.

This modification also reduces the performance trade-off between glacier-wide annual balance and discharge simulation (Fig. 2b). It is noteworthy that the performance with the Grimsel input station is still much worse, confirming our initial choice not to consider this station. The new model version has, however, still an important shortcoming: considering all good solutions (i.e. parameter sets) on the POF of Figure 2b, the seasonal lapse rates show a distinct dependence on each other for good discharge parameter sets versus good mass balance parameter sets (see Fig. 4). This implies that this model version needs different pairs of temperature lapse rates for mass balance simulation than for discharge simulation.
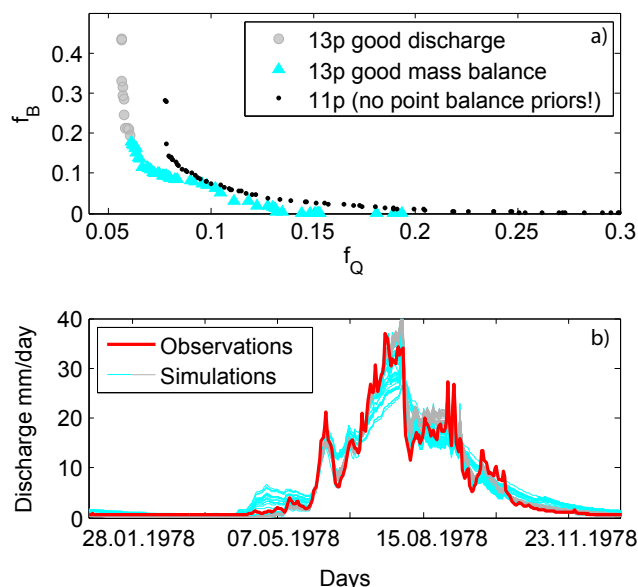
Fig. 5. (a) Trade-off between discharge performance criterion and mass balance performance criterion for two intermediate versions of the modified GSM-SOCONT; 11p stands for 11 parameters (for parameters see Table 1); note: the figure does not show the tail end of the POF for the 11p model, the best performance for f_Q is 0.06 as for the other model versions. **(b)** discharge simulations corresponding to all parameter sets of the 13 parameter model of **(a)** using the same color coding (cyan/grey).

The seasonal lapse rates have, for evident reasons, a strong interaction with the precipitation input correction factor. A solution to overcome the above structural shortcoming is the introduction of seasonal input correction factors. As tests showed, using different altitudinal gradients leads to complex parameter interactions with the rainfall/snowfall separation thresholds, which is solved by the separate interpolation of snowfall and rainfall (see Sect. 3.2).

The new model structure has the nice property that we can fix the rainfall/snowfall separation thresholds (to 0 °C and 2 °C) and still obtain parameter sets that reproduce the reference winter accumulation and annual point balance. This result holds independent of the chosen meteorological input series (but, of course, for each series, we obtain different parameter sets).

If we optimize this new model structure on annual glacier-wide mass balance and discharge, there remains an important trade-off between achieving good mass balance and good discharge simulations (see Fig. 5a, model 11p). Introducing a separate summer snow melt factor a_{ss}, as in many similar studies, to account, e.g., for albedo differences (Stahl et al., 2008), partly reduces this trade-off; but good mass balance simulations lead to far too much winter discharge. Adding a separate summer snow accumulation correction factor considerably reduces the trade-off (see Fig. 5a, model 13p). However, there is still too much discharge in early spring for
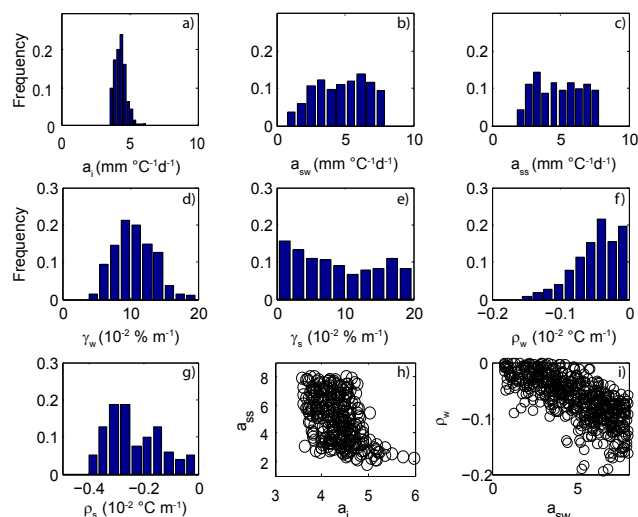
Fig. 6. (a)–(g) Distribution of behavioral parameter sets for the seasonal point mass balances, **(h)** and **(i)** scatter plots of the parameters showing the strongest linear correlations.

all good mass balance simulations (see Fig. 5b). This problem is addressed by introducing a water retention capacity η_s for the snow layer (see Sect. 3.2).

The *final improved model structure*, used in the rest of this paper, has 14 parameters to calibrate:

- Fixed rainfall/snowfall separation thresholds (0 °C and 2 °C) and linear transition between them.

- Winter and summer temperature lapse rates (ρ_w, ρ_s).

- Ice melt factor a_i, constant throughout the year.

- Winter and summer snow melt factors (a_{sw}, a_{ss}),

- Winter and summer snow accumulation correction factors (γ_w, γ_s).

- Summer rainfall correction factor γ_r.

- Six water-runoff transfer parameters, A, $\log(k)$, β, k_i, k_s, η_s.

5.4 Sequential calibration

The default parameter values to initialize the sequential calibration are given in Table 1. The updating starts with the parameters that influence winter accumulation, i.e. the winter accumulation correction factor γ_w, the winter snow melt factor a_{sw} and the temperature lapse rate ρ_w. During this period, the other parameters have no influence on the model output. Figure 6b, d and f show the distribution of the parameter sets for which we cannot reject the null hypothesis that the observed samples have the simulated median value for all

Table 1. Calibrated and default GSM-SOCONT parameter values; the default values are from (Schaefli et al., 2005). The model version is identified with the number of parameters (8p stands for the original version with 8 parameters, > 8p stands for all model versions with more than 8 parameters). For the final model structure with 14 parameters, the best parameter sets under f_Q and under f_B are given.

Name	Unit	Default	Model	f_Q	f_B
a_i	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	11.5	> 8p	6.3	3.5
a_{sw}	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	6.6	12p	5.8	5.5
a_{ss}	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	6.6	12p	3.1	2.6
a_s	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	6.6	8p, 11p	–	–
k_i	d^{-1}	4.7	> 8p	3.5	–
k_s	d^{-1}	5.2	> 8p	5.9	–
A	mm	2147	> 8p	1038	–
$\log(k)$	$\log(s^{-1})$	−9.9	> 8p	−7.8	–
β	$m^{4/3}\,s^{-1}$	301	> 8p	20 232	–
η_s	–	–	14p	0.38	–
γ_p	$\%\,(100\,m)^{-1}$	3.1	8p	–	–
ρ	${}^{\circ}C\,m^{-1}$	−0.0065	8p	–	–
ρ_w	${}^{\circ}C\,m^{-1}$	−0.0065	> 11p	−0.0013	−0.0013
ρ_s	${}^{\circ}C\,m^{-1}$	−0.0065	> 11p	−0.0041	−0.0041
γ_r	$\%\,(100\,m)^{-1}$	3.1	> 11p	−2.5	–
γ_w	$\%\,(100\,m)^{-1}$	3.1	> 11p	7.5	6.7
γ_s	$\%\,(100\,m)^{-1}$	3.1	> 13p	6.4	0.02

retained elevation bands and winters (a total of 100 000 parameter sets was randomly drawn in the initial priors of Table 2). The temperature lapse rates are very small and they show a certain correlation with the snow melt factors (linear correlation of −0.6, see Fig. 6i). This result is to be expected. Melt factors are known to be lapse-rate specific (Shea et al., 2009) since, to melt a given quantity of snow, the melt factor has to be higher if the temperature is lower (i.e. of the temperature lapse rate is more negative). For the ice melt factor, this relationship is obscured by the fact that ice melt only occurs if the glacier is snow-free. The other parameters do not show any pair-wise dependence structure.

We reduce the prior for a_{sw}, γ_w and ρ_w to the above identified ranges and draw another 100 000 parameter sets including the summer melt factor, a_{ss}, the summer temperature lapse rate, ρ_s, the summer accumulation correction factor γ_s and the ice melt factor a_i. The distributions of the parameter sets for which the null hypothesis cannot be rejected for all elevation bands and years are shown in Fig. 6 and summarized in Table 2. The summer snow melt factor a_{ss} and the ice melt factor a_i show a very strong dependence on each other (see Fig. 6h). This dependence shows the typical shape reported by previous studies (see Schaefli et al., 2005, Fig. 4 or Hock et al., 1999, Fig. 4).

Figure 7a shows the simulated annual glacier-wide balance corresponding to one of the acceptable parameter sets. This simulation is strikingly different from the reference simulation of Huss et al. (2008). This illustrates that the point mass balances contain some information but not enough to

Table 2. Prior and updated parameter ranges for the generation of random parameter sets; if there is no updated range, this implies that the point mass balance data does not contain information for this parameter; the lower limit for ρ_w and ρ_s is the average lapse rate in the troposphere (Miller and Thompson, 1970).

Name	Unit	Prior	Updated	Ref. data
a_i	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	(2.0, 12.0)	(3.5, 6.5)	Annual
a_{sw}	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	(0.5, 8.0)	(1.5, 8.0)	Winter
a_{ss}	$mm\,d^{-1}\,{}^{\circ}C^{-1}$	(0.5, 8.0)	(1.5, 8.0)	Annual
ρ_w	${}^{\circ}C\,m^{-1}$	(−0.0065, 0)	(−0.0016, 0)	Winter
ρ_s	${}^{\circ}C\,m^{-1}$	(−0.0065, 0)	(−0.0042, 0)	Annual
γ_w	$\%\,(100\,m)^{-1}$	(0.0, 20.0)	(4.5, 18.5)	Winter
γ_s	$\%\,(100\,m)^{-1}$	(0.0, 20.0)	(0, 20.0)	Annual

constrain the model. This is partly due to the fact that for annual point balance, only the lowest elevation bands have sufficiently large samples.

We complete a final parameter calibration minimizing f_Q and f_B, limiting the search space to the parameter ranges that are acceptable given the point balance observations. Given that neither discharge nor glacier-wide annual balance contain direct information on the temperature lapse rates, we fix their value to a previously identified possible parameter couple (see Table 1). Even under the final updated model structure, there remains a trade-off between f_Q and f_B, corresponding to a considerable range of possible system
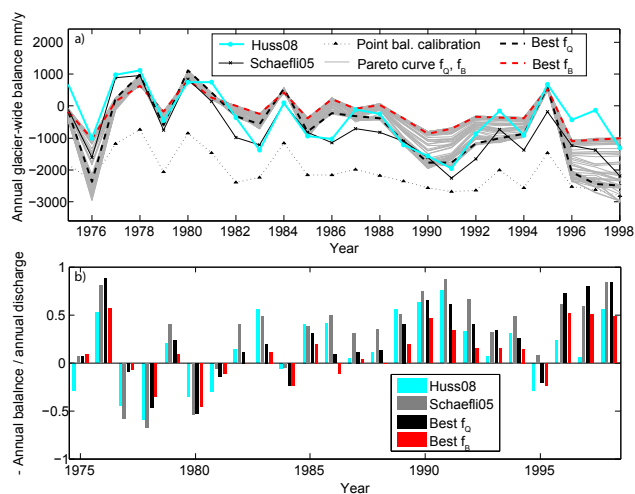
Fig. 7. (a) Annual glacier-wide simulations of Huss et al. (2008), Schaefli et al. (2005) and of all Pareto-optimal parameter sets (f_Q and f_B) for the final model structure; shown is also another simulation ($a_i = 3.9$, $a_{sw} = 7.6$, $a_{ss} = 7.3$, $\gamma_w = 0.12$, $\gamma_s = 0.08$ $\rho_w = -0.0009$, $\rho_s = -0.0033$, units see Table 2) that is acceptable considering only the point mass balance observations; **(b)** ratio between glacier storage change and annual discharge (the simulation of Huss et al. (2008) is divided by the observed annual discharge).

representations; this is illustrated in Figures 7 and 8 that show the glacier-wide mass balance simulations and the discharge simulations for all parameter sets on the Pareto-optimal frontier. The best performing parameter sets under f_Q and f_B are given in Table 1. The performance of the best f_Q simulation in terms of mimicking the observed discharge is comparable to the calibration presented by Schaefli et al. (2005), with a Nash value of 0.92 for the calibration period and 0.90 for the entire period 1976–1999. Compared to the observed mean interannual discharge as benchmark (see Schaefli and Gupta, 2007), this corresponds to a benchmark efficiency of 0.46 for the entire period. This is lower than for the original model (0.55) but still acceptable in light of the fact that the model has been calibrated on the relatively cold years 1976–1982. The relative bias of the best simulation is -1% for the calibration period and 3% for the entire period (the relative bias between a simulated variable x and a reference variable y is defined as $(x - y) \, y^{-1}$).

The rainfall/melt water-runoff transformation parameters of the non-glacier part have considerably different values than the original GSM-SOCONT, which is due to the modification of the transfer module with the introduction of η_s.

Finally, we would like to report an important detail finding: in previous work, we used a warm-up period of two years for model calibration, assuming that this is sufficient to build-up the initial snowpack. Given the new model formulation, including the snowpack retention capacity, a longer warm-up period might be required if the simulation starts during a cold period where the seasonal snowpack does not
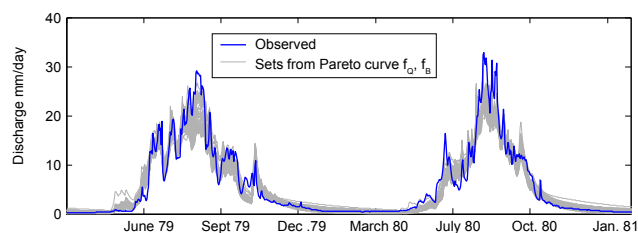
Fig. 8. Observed discharge and simulations corresponding to all Pareto-optimal parameter sets (f_Q and f_B) for the final model structure for two years of the calibration period.

melt away for the two or three highest glacier elevation bands and the highest non-glacier elevation band. This effect is particularly pronounced if, as in the present case, the model initialization period is relatively dry and the build-up of the "correct" initial snowpack takes more than a winter. We, therefore, used an initialization period of four years for the final model calibration.

6 Discussion

6.1 Multi-objective calibration for model improvement

The presented case study revealed the importance of considering reference data on flux and on state variables for model development. First of all, for the choice of appropriate input series: considering discharge performance or mass balance performance individually, none of the available measurement stations would appear to contain more valuable information. Considering the trade-off between the two objective functions reveals that there is a considerable difference in information content. We believe that this type of analysis has great potential for hydrological modeling where assessing the (dis-)information content of input time series is still one of the big open questions (Beven and Westerberg, 2011).

During the model modification process, it is again the trade-off between the reproduction of these two model variables that provides clear indications about the value of adding a parameter or modifying a submodule. The multi-signal approach allows the detection of a subtle model improvement through the introduction of the snow layer retention capacity η_s: adding this parameter does not lead to a better discharge simulation, i.e. we cannot find a parameter set under the new model structure (including η_s) that outperforms the best discharge simulation under the old model structure. But the best mass balance parameter sets do also a fairly good job for discharge simulation and vice-versa.

Our results also underline the value of analyzing the parameter sets associated with the simulation trade-offs, i.e. of analyzing the parameter sets corresponding to points on the Pareto-optimal frontier (Fig. 4): the strikingly different relationship between winter and summer lapse rate

parameters for good discharge simulation versus good mass balance simulation was a very strong hint for a model structural deficiency. Again, this type of model diagnostics could be very valuable in other hydrological contexts.

6.2 Reproduction of water balance terms

Figure 7 shows the simulated glacier-wide annual mass balance for all parameter sets on the POF and the simulation corresponding to the original model structure calibrated on f_Q and on f_B. Just as the original model, the new model strongly overestimates the negative cumulative mass balance at the end of the simulation period leading to a strong bias. The original model shows a mean bias compared to the simulation of Huss et al. (2008) of 294%. In the new model, this is reduced to 101% for the best simulation under f_Q and equals −33% for the best simulation under f_B.

The variability of the monthly glacier-wide balance of the reference simulation is also better reproduced by the new simulation. This can be measured by the coefficient of variation, which corresponds to the standard deviation of a variable x divided by the mean of x. The coefficient of variation of the reference simulation is −14.9, the new model gives −7.1 under f_Q and −13.3 under f_B whereas the original model had a far too low value (−3.9).

The most important difference between the original GSM-SOCONT and the new model version becomes visible if we consider the individual water balance terms (Table 3). The total area-averaged precipitation is much higher for the new model than for the original model, evaporation is lower. These differences in the mass balance terms can result in big differences in the relative contribution of glacier storage change to total annual runoff (Fig. 7b) in individual years. This relative contribution is not only a proxy for climate sensitivity but also a key variable for water management and in particular hydropower production (Schaefli et al., 2007).

The question which of the two water balance representations is closer to the "truth" is extremely difficult to answer. There are estimated reference values on a 2 km by 2 km grid for evaporation and precipitation (see Table 3), but given the resolution and the complexity of the topography, these are just rough estimates for such a small catchment. A rather strong hint is, however, the following: for the years 1974–1990 (to which Table 3 refers to), the reference simulation of Huss et al. (2008) corresponds to a mean glacier storage change of −260 mm w.e. (water equivalent) per year relative to the ice-covered area (around 17 km^2). Relative to the entire catchment area (38.9 km^2), this equals −110 mm w.e. a^{-1}, a value that is very close to the storage change simulated with the hydrological model for the same period (−140 mm w.e. a^{-1}), whereas the original model gives a value of −250 mm w.e. a^{-1}.

Neither the original nor the new model can reproduce the annual glacier-wide mass balance simulations towards the end of the simulation period. As discussed by Huss et al.

Table 3. Water balance terms for the original GSM-SOCONT (orig.), the new final model (1974–1990) calibrated on f_Q and the values extracted from the hydrological atlas of Switzerland, which provides estimated average values of precipitation P (mm a^{-1}) on a 2 km by 2 km grid (1971–1990, Schwarb et al., 2001) and of actual evaporation E (mm a^{-1}) (1973–1992, Menzel et al., 1999) on a 1 km by 1 km grid; $\triangle S/\triangle t = P - Q - E$ is the storage variation (mm a^{-1}); for the last column, this value is put in brackets since it is computed from data sources referring to different periods; snowfall is indicated relative to total precipitation.

Name	Orig.	New	Atlas
E	180	150	280
P	1940	2670	1950
P glacier	2050	3040	
P non-glacier	1830	2430	
P (glac.)/P (non-glac.)	89%	80%	
$\triangle S/\triangle t$	−250	−140	(−470)
Discharge Q	2010	2660	2250
Snowfall glac.	85%	82%	
Snowfall non-glac.	70%	71%	

(2009), this period corresponds to a period of intense melting where traditional temperature-index models might become unreliable. Huss et al. (2009) found, in fact, that for this period, the degree-day factors are consistently smaller than their long-term mean value. The question how to account for such anomalies in temperature-index types hydrological models, especially for projections into the future, remains open to date.

6.3 Information content of glacio-hydrological data

Our results suggest that even small samples of point mass balance observations contain valuable information to calibrate four additional hydro-meteorological model parameters with respect to the original GSM-SOCONT. The obtained trade-off for the reproduction of winter balance versus annual balance with the original model provides evidence that both data sets contain complementary information. The fact that this trade-off can be removed through simple model modifications suggests that the point data sets are not inconsistent.

Such observations are often deemed to be too sparse to be useful to calibrate a simple semi-lumped hydrological model. Our results, however, demonstrate that the data are indeed useful to identify a model structure that can reproduce the seasonal mass balance dynamics and furthermore, it constrains the plausible ranges of the parameters. This conclusion is at least partly unexpected: in fact, it is often assumed that winter accumulation data do not contain any information on temperature lapse rates arguing that in glacier catchments, the low winter temperatures lead to permanent snow fall independent of the temperature lapse rate.

The updated plausible parameter ranges contain other hints that support the value of the proposed method to extract information from the mass balance samples: (i) the temperature lapse rates are less strong in winter than in summer as expected for theoretical reasons and (ii) the parameters do not show unexpected pair-wise dependencies. The dependence of a_{ss} and a_i shows the typical shape obtained in classical calibration on time series (Fig. 6).

Our case study shows that model calibration with annual glacier-wide mass balance data in addition to discharge, as done in similar previous studies (Konz and Seibert, 2010; Schaefli et al., 2005), might lead to wrong intra-annual mass balance dynamics, whereas using only point mass balance observations might lead to wrong glacier-wide mass balance simulations.

From a more general perspective, reliable mass balance simulations can only be obtained if there is data-based information or a priori knowledge either on incoming and outgoing fluxes or on one of the two and the storage change. Seasonal glacier mass balance observations can provide such direct information on all terms. Another option is the use of data about the snow covered area (Koboltschnig et al., 2008). Discharge data alone, however, will only yield information on outgoing fluxes, which is, furthermore, obscured by the hydrological processes in the ice-free part.

6.4 Transferability of the results

The conclusion that using only discharge and annual glacier-wide data might lead to an erroneous simulation of sub-annual mass balance terms and related fluxes is partly related to the simple model structure. A more complex, and, thus, more flexible model might well give a more plausible pattern of sub-annual mass balance dynamics even if it is only calibrated on discharge and glacier-wide annual balance – either by chance or because the modeler had sufficient a priori knowledge. Distinguishing between the two cases is not a trivial task but a step-wise model modification as the one proposed here certainly contributes to avoid spurious results due to overparameterization (see also Clark et al., 2008).

The above main conclusion is, nevertheless, transferable to many other similar simulation settings in climates where there is a distinct accumulation and ablation season. This applies to most glacier catchments in the world (see Kaser et al., 2010) but namely not to glacier catchments in the tropical Andes, where ablation also occurs during the accumulation season.

Most other conclusions are at least partly dependent on the simulation time step, the model complexity and the available input data. For example, the estimation of seasonal temperature lapse rates would not be necessary in presence of a temporally distributed temperature input field. This need could also have been masked if we had used an extended temperature-index approach where degree-day melt factors are corrected to account for the time-varying potential radiation (Hock, 1999).

7 Conclusions

Getting the water balance right is a basic requirement for any hydrologic model – yet, in many environments, this is very difficult to achieve. In the case of glacier catchments, it is often assumed that some glacier-wide annual mass balance estimations, in addition to observed discharge, are good enough to obtain reliable estimates of the different water balance terms and to develop prediction models. In this paper, we present evidence that information on the seasonal mass balance is a pre-requisite to reliably calibrate a hydrological model and we demonstrate the value of a simple, but promising method to use small samples of point observations for this calibration.

The observed seasonal mass balance data could only be reproduced through a model structure modification. The identified new model structure shows some interesting features; it has seasonal lapse rates and a separate snow accumulation and rainfall correction factor. These last two parameters are required for the model to be sufficiently flexible to compensate differently for missing processes during summer and winter (e.g. raingauge undercatch, redistribution through wind etc.). In summary, we can say that a simple conceptual model, as the one presented here, can be a valuable tool to project the behavior of a glacier catchment but only if we have enough (seasonal) information to constrain the parameters that directly affect the mass balance. Since in most catchments around the world, virtually no observed ground-based data are available, future research could focus on the question of how to extract such information from remotely-sensed data. It would also be interesting to assess how the information content of seasonal mass balance data varies across hydro-climatological regions.

Glacier hydrology deals with systems where the main water storage term, the glaciers, can be directly observed – an invaluable advantage over other ecosystems. We showed in this paper possible ways of taking advantage of this direct data on storage and flux to select appropriate input time series and to improve the hydrologic model structure. We believe that such catchments offer ideal test cases for further developments in this direction. Since, in addition, these catchments are undergoing rapid change, they can also be used to assess the ability of hydrologic models to predict change.

Supplementary material related to this article is available online at: http://www.hydrol-earth-syst-sci.net/15/1227/2011/ hess-15-1227-2011-supplement.pdf.

References

Bauder, A., Funk, M., and Huss, M.: Ice-volume changes of selected glaciers in the Swiss Alps since the end of the 19th century, Ann. Glaciol., 46, 145–149, doi:10.3189/172756407782871701, 2007.

Bárdossy, A. and Singh, S. K.: Robust estimation of hydrological model parameters, Hydrol. Earth Syst. Sci., 12, 1273–1283, doi:10.5194/hess-12-1273-2008, 2008.

Beven, K. J. and Binley, A.: The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process., 6, 279–298, doi:10.1002/hyp.3360060305, 1992.

Beven, K. and Westerberg, I.: On red herrings and real herrings: disinformation and information in hydrological inference, Hydrol. Proc., in press, doi:10.1002/hyp.7963, 2011.

Blandford, T. R., Humes, K. S., Harshburger, B. J., Moore, B. C., Walden, V. P., and Ye, H. C.: Seasonal and synoptic variations in near-surface air temperature lapse rates in a mountainous basin, J. Appl. Meteorol. Climatol., 47, 249–261, doi:10.1175/2007jamc1565.1, 2008.

Braun, L. N. and Renner, C. B.: Application of a conceptual runoff model in different physiographic regions of Switzerland, Hydrol. Sci. J., 37, 217–231, 1992.

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00b02, doi:10.1029/2007wr006735, 2008.

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D., and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrol. Proc.,15, 523–543, doi:10.1002/hyp.7902, 2010.

Fenicia, F., Solomatine, D. P., Savenije, H. H. G., and Matgen, P.: Soft combination of local models in a multi-objective framework, Hydrol. Earth Syst. Sci., 11, 1797–1809, doi:10.5194/hess-11-1797-2007, 2007.

Fenicia, F., Savenije, H. H. G., Matgen, P., and Pfister, L.: Understanding catchment behavior through step-wise model concept improvement, Water Resour. Res., 44, W01402, doi:10.1029/2006WR005563, 2008.

Freer, J. E., McMillan, H., McDonnell, J. J., and Beven, K. J.: Constraining dynamic TOPMODEL responses for imprecise water table information using fuzzy rule based performance measures, J. Hydrol., 291, 254–277, doi:10.1016/j.jhydrol.2003.12.037, 2004.

Funk, M.: Räumliche Verteilung der Massenbilanz auf dem Rhonegletscher und ihre Beziehung zu Klimaelementen, Ph.D. thesis, Eidgenössische Technische Hochschule Zürich, nebis, 1985.

Gibbons, J. and Chakraborti, S.: Nonparametric Statistical Inference, Marcel Dekker, New York, 2003.

Garçon, R.: Prévision opérationnelle des apports de la Durance à Serre-Ponçon à l'aide du modèle MORDOR, Bilan de l'anne 1994–1995, Houille Blanche, 5, 71–76, 1996.

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

Hingray, B., Schaefli, B., Mezghani, A., and Hamdi, Y.: Signature-based model calibration for hydrologic prediction in mesoscale Alpine catchments, Hydrolog. Sci. J., 55, 1002–1016, doi:10.1080/02626667.2010.505572, 2010.

Hock, R.: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, J. Glaciol., 45, 101–111, 1999.

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282, 104–115, doi:10.1016/S0022-1694(03)00257-9, 2003.

Horton, P., Schaefli, B., Hingray, B., Mezghani, A., and Musy, A.: Assessment of climate change impacts on Alpine discharge regimes with climate model uncertainty, Hydrol. Process., 20, 2091–2109, doi:10.1002/hyp.6197, 2006.

Huss, M., Bauder, A., Funk, M., and Hock, R.: Determination of the seasonal mass balance of four Alpine glaciers since 1865, J. Geophys. Res., 113, F01015, doi:10.1029/2007JF000803, 2008.

Huss, M., Funk, M., and Ohmura, A.: Strong Alpine glacier melt in the 1940s due to enhanced solar radiation, Geophys. Res. Lett., 36, L23501, doi:10.1029/2009GL040789, 2009.

Kaser, G., Grosshauser, M., and Marzeion, B.: Contribution potential of glaciers to water availability in different climate regimes, Proc. Nat. Acad. Sci. USA, 107, 20223–20227, doi:10.1073/pnas.1008162107, 2010.

Kavetski, D., Kuczera, G., and Franks, S. W.: Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory, Water Resour. Res., 42, W03407, doi:10.1029/2005WR004368, 2006.

Klok, E., Jasper, K., Roelofsma, K., Gurtz, J., and Badoux, A.: Distributed hydrological modelling of a heavily glaciated Alpine river basin, Hydrolog. Sci. J., 46, 553–570, doi:10.1080/02626660109492850, 2001.

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

Konz, M., Uhlenbrook, S., Braun, L., Shrestha, A., and Demuth, S.: Implementation of a process-based catchment model in a poorly gauged, highly glacierized Himalayan headwater, Hydrol. Earth Sys. Sci., 11, 1323–1339, doi:10.5194/hess-11-1323-2007, 2007.

Konz, M. and Seibert, J.: On the value of glacier mass balances for hydrological model calibration, J. Hydrol., 385, 238–246, doi:10.1016/j.jhydrol.2010.02.025, 2010.

Kuchment, L. S. and Gelfan, A. N.: The determination of the snowmelt rate and the meltwater outflow from a snowpack for

modelling river runoff generation, J. Hydrol., 179, 23–36, 1996.

Leyland, G. B.: Multi-Objective Optimisation Applied to Industrial Energy Problems, Ph.D. thesis, available at: http://library.epfl.ch/theses, Ecole Polytechnique Fédérale de Lausanne, Switzerland, 2002.

Miller, A. and Thompson, J. C.: Elements of meteorology, Merill Publishing, Columbus, Ohio, p. 402, 1970.

Menzel, L., Lang, H., and Rohmann, M.: Mean annual actual evaporation 1973–1992, in: Hydrological Atlas of Switzerland, Plate 4.1, Service Hydrologique et Géologique National, Berne, 1999.

Moore, R. D. and Demuth, M. N.: Mass balance and streamflow variability at Place Glacier, Canada, in relation to recent climate fluctuations, Hydrol. Proc., 15, 3473–3486, doi:10.1002/hyp.1030, 2001.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models, Part I, a discussion of principles, J. Hydrol., 10, 282–290, 1970.

Pellicciotti, F., Bauder, A., and Parola, M.: Effect of glaciers on streamflow trends in the Swiss Alps, Water Resour. Res., 46, W10522, doi:10.1029/2009WR009039, 2010.

Reichert, P. and Mieleitner, J.: Analyzing input and structural uncertainty of nonlinear dynamic models with stochastic, time-dependent parameters, Water Resour. Res., 45, W10402, doi:10.1029/2009WR007814, 2009.

Reusser, D. E., Blume, T., Schaefli, B., and Zehe, E.: Analysing the temporal dynamics of model performance for hydrological models, Hydrol. Earth Syst. Sci., 13, 999–1018, doi:10.5194/hess-13-999-2009, 2009.

Rohrer, M. B., Braun, L. N., and Lang, H.: Long-term records of snow cover water equivalent in the Swiss Alps, 1. Analysis, Nord. Hydrol., 25, 53–64, 1994.

Rolland, C.: Spatial and seasonal variations of air temperature lapse rates in Alpine regions, J. Clim., 16, 1032–1046, doi:10.1175/1520-0442(2003)016<1032%3ASASVOA>2.0.CO%3B2, 2003.

Schaefli, B. and Gupta, H.: Do Nash values have value?, Hydrol. Proc., 21(15), 2075–2080, doi:10.1002/hyp.6825, 2007.

Schaefli, B., Hingray, B., and Musy, A.: Improved calibration of hydrological models: use of a multi-objective evolutionary algorithm for parameter and model structure uncertainty estimation, in: Hydrology: Science and Practice for the 21st Century, edited by: Webb, B., British Hydrological Society, London, 362–371, 2004.

Schaefli, B. and Huss, M.: Interactive comment on "Simulation of high mountainous discharge: how much information do we need?" by B. Schaefli and M. Huss Hydrol. Earth Syst. Sci. Discuss., 7, C4779–C4785, 2011

Schaefli, B., Hingray, B., Niggli, M., and Musy, A.: A conceptual glacio-hydrological model for high mountainous catchments, Hydrol. Earth Syst. Sci., 9, 95–109, doi:10.5194/hess-9-95-2005, 2005.

Schaefli, B., Hingray, B., and Musy, A.: Climate change and hydropower production in the Swiss Alps: quantification of potential impacts and related modelling uncertainties, Hydrol. Earth Syst. Sci., 11, 1191–1205, doi:10.5194/hess-11-1191-2007, 2007.

Schwarb, M., Daly, C., Frei, C., and Schär, C.: Mean annual precipitation throughout the European Alps, 1971–1990, in: Hydrological Atlas of Switzerland, Plate 2.6, Service Hydrologique et Géologique National, Berne, 2001.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agricul. Forest Meteorol., 98–99, 279–293, 1999.

Shea, J. M., Moore, R. D., and Stahl, K.: Derivation of melt factors from glacier mass-balance records in western Canada, J. Glaciol., 55, 123–130, doi:10.3189/002214309788608886, 2009.

Sivapalan, M., Blöschl, G., Zhang, L., and Vertessy, R.: Downward approach to hydrological prediction, Hydrol. Process., 17, 2101–2111, doi:10.1002/hyp.1425, 2003.

Stahl, K., Moore, R., Shea, J., Hutchinson, D., and Cannon, A.: Coupled modelling of glacier and streamflow response to future climate scenarios, Water Resour. Res., 44, W02422, doi:10.1029/2007WR005956, 2008.

SwissTopo: Digital height model of Switzerland – DHM25, 1995.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, Water Resour. Res., 41, W01017, doi:10.1029/2004WR003059, 2005.

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

Yilmaz, K. K., Gupta, H. V., and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44, W09417, doi:10.1029/2007wr006716, 2008.