Merits and Limits of SWAT-GL: Application in Contrasting Glaciated Catchments

Timo Schaffhauser[1], Florentin Hofmeister[1,2], Gabriele Chiogna[1], Fabian Merk[1], Ye Tuo[1], Julian Machnitzke[1], Lucas Alcamo[1], Jingshui Huang[1], and Markus Disse[1]

[1]School of Engineering and Design, Technical University of Munich, Munich, Germany
[2]Bavarian Academy of Sciences and Humanities, Munich, Germany

Correspondence: Timo Schaffhauser (t.schaffhauser@tum.de)

Abstract. The recently released SWAT-GL aims to overcome multiple limitations of the traditional hydrological model SWAT (Soil Water Assessment Tool) in glaciated mountainous catchments. SWAT-GL intends to increase the applicability of SWAT in these catchments and to reduce misapplication when glaciers have a significant role in the catchment hydrology. It thereby relies on a mass balance module, based on a degree-day approach similar to SWAT's snow melt module, extended by a glacier evolution component which is based on the delta-h (Δh) parameterization. The latter one is a mass conserving approach which enables the spatial distribution of ice thickness changes and thus dynamic glacier retreat. However, the extended SWAT version was not yet comprehensively benchmarked. Hence, our paper aims to benchmark SWAT-GL with four different benchmark glaciers which are part of the USGS (United States Geological Survey) Benchmark Glacier Project. The benchmarking considers a comprehensive evaluation procedure, where the routine is optimized on glacier mass balance and hypsometry as well as snow cover. Snow cover is included to consider snow-glacier feedbacks appropriately. Besides, a sensitivity analysis using Elementary Effects (or Method of Morris) is performed to give a detailed picture on the importance of the introduced glacier processes, as well as the relevance of the interactions with the already-existing snow routine. We intentionally did not include discharge in the optimization procedure to fully demonstrate the capabilities of SWAT-GL in terms of glacier and snow processes. Results demonstrate that SWAT-GL is able to ~~perform reasonably well in~~ represent the characteristics of contrasting glaciated catchments, which underlines SWAT-GL's applicability and transferability. We could further show its strong (non-linear) interactions with the existing snow routine suggesting a simultaneous calibration of the snow components. While snow and glacier processes were adequately represented in the catchments, discharge was not necessarily represented sufficiently when excluded in the optimization procedure. However, SWAT-GL has been shown to be easily capable of reproducing discharge when used in a stand-alone optimization, although this may come at the expense of model consistency.

1 Introduction

We recently submitted a paper that introduces a new glacier routine to SWAT to overcome current limitations in its applicability, especially in glaciated catchments (Schaffhauser et al., 2024). The work is built on previous efforts of multiple groups which intend to address common constraints of conceptual and physically-based hydrological models in glacier-dominated catchments.

Examples include the work of Seibert et al. (2018) or Li et al. (2015) for HBV (Hydrologiska Byråns Vattenbalansavdelning), Wortmann et al. (2016) for SWIM (Soil and Water Integrated Model) or Shannon et al. (2022) for the DECIPHeR model (Dynamic fluxEs and ConnectIvity for Predictions of HydRology). However, to our knowledge many hydrological models do not include glacier routines by default and glacier-focused extensions are often only available to the developing groups, although trends are clearly towards publishing model code and making it openly accessible. Despite these improvements, applications in glaciated basins remain challenging due to missing (or very simple) glacier representations, whereby modelers might rely on external couplings of glaciological and hydrological models (Adnan et al., 2019; Wu et al., 2015; Stoll et al., 2020; Du et al., 2022; Naz et al., 2014; Wiersma et al., 2022; Chen et al., 2017). As it is commonly known, glacio-hydrological models applied to small and highly glacierized catchments often have no or a rather rough representation of additional hydrological components (e.g., evapotranspiration) (Hassan et al., 2021; Ali et al., 2017; Pradhananga et al., 2014), potentially leading to an integration problem, at least when the model domain is extended (Tiel et al., 2020; Wortmann et al., 2016). A further problem is that it is not always clear in the existing literature whether, for example, a glacier routine is coupled with or integrated into an hydrological model, if this hydrological model by default does not take glacier processes into account (e.g., SWAT, ~~VIC (Variable Infiltration Capacity)~~VIC—Variable Infiltration Capacity—, HBV) and is used to simulate glacio-hydrological processes. The terms *integrated* and *coupled* seem to be used interchangeably, thus impairing reproducibility. From our perspective, integration should suggests a model expansion, while coupling suggests the use of an additional model. Besides, distinctions are not always easy and clear especially what coupling exactly means. In our understanding, coupling usually refers to a chained approach, where external information, e.g. from a snow module, are used as input for a hydrological model without an exchange of information of the two.

Depending on the research question and data available, several glacier routines with different complexity are available for simulating glacier mass balances and melt contribution in hydrological models. An unlimited ice storage that generates melt water based on a calibrated degree-day factor represents the simplest empirical routine (Naz et al., 2014). However, this routine cannot consider glacier evolution, such as glacier retreat. Conceptual routines, such as the volume-area scaling (VA scaling) (Bahr et al., 2015) or the Δh-parameterization (Huss et al., 2010), simulate the spatial dynamic of a glacier as a simplified function of glacier extent, thickness, and elevation range (Tiel et al., 2020). The largest limitation of these methods is the lack of an actual representation of the ice flow dynamic, which can be simulated with full physics-based algorithms (Zekollari et al., 2022). Since ice flow modules require several input data and the definition of distinctive boundary conditions, such as bedrock roughness, the application is usually beyond the scope of common water balance simulations in glaciated catchments. In recent years, the coupling of water balance simulation models with global glacier models has proven to be a valuable method for predicting the hydrological response of catchments in mountainous regions under a changing climate (Pesci et al., 2023).

Past SWAT-specific efforts to improve capabilities in glacier-fed catchments other than those from Schaffhauser et al. (2024) for example refer to the work of Luo et al. (2013), who implemented a ~~volume-area scaling (VA scaling)~~ VA scaling for glacier evolution along with a degree-day-based mass balance module. The modified SWAT version was further applied in several studies, mostly focusing on China (Gan et al., 2015; Luo et al., 2018; Ma et al., 2015; Shafeeque et al., 2019; Wang et al., 2018) and recently integrated in SWAT+ (Yang et al., 2022). However, to the best of our knowledge in none of the publications the model code was made publicly available. Besides, Ji et al. (2019) for instance implemented an ice melt routine based on a degree-day approach but did not account for glacier evolution explicitly. Unfortunately, none of the approaches was included in any of the official SWAT revisions. SWAT-GL was developed to tackle these issues and to provide a freely available and user-friendly SWAT version for glaciated catchments (Schaffhauser et al., 2024). Besides, the chosen approach, namely the Δh-parameterization from Huss et al. (2010) has proven to be a robust method to simulate glacier evolution in glaciated catchments (Huss and Hock, 2015). ~~Mass balance simulations are similar to most glacio-hydrological models performed with a simple temperature-index approach (Tiel et al., 2020). The empirical Δh-parameterization is called annually to translate the cumulative mass balance change to a change in glacier geometry. The concept assumes that lower elevated areas closer to the glacier terminus receive stronger ablation than higher elevated ones (Huss et al., 2008, 2010). Therefore, glaciers are divided into different elevation sections (ES) for the application. In addition to its spatial distributed applicability the method is mass-conserving and can be applied with glacier outlines and glacier thickness data only (Li et al., 2015; Seibert et al., 2018).~~ However, no comprehensive evaluation of SWAT-GL has been conducted to date. As glaciers and high-mountainous catchments are usually rather data-scarce (Tuo et al., 2016), testing the performance of glacio-hydrological models for long observed time series of good quality is challenging. Moreover, in many cases the available variables to calibrate and validate the model are limited to discharge only, with these gauges often located much further downstream and not close to the glacier. Evaluating the glaciological routines of glacio-hydrological models by discharge alone, representing a superposition of multiple signals, might be problematic, as it might not reflect the signatures which would be visible in the glaciological components. In other words, a satisfactory representation of discharge used to evaluate catchment glaciology (or other processes), albeit often done, might be inadequate. Nevertheless, a sound evaluation of newly introduced schemes in glacio-hydrological models (e.g., glacier components into a hydrological model) should be desired and aimed for. If a mass balance module is implemented together with an evolution module (as in SWAT-GL), in the best case both are evaluated individually and complemented by discharge assessments.

The USGS Benchmark Glacier Project (O'Neel et al., 2019) is a promising attempt to overcome current limitations in data accessibility and modelling efforts of high-mountainous and glaciated basins. The project involves five glaciers, four where long-term measurements are available and one for which the project expanded more recently. The glaciers, namely Gulkana, Wolverine, Lemon Creek, South Cascade and Sperry Glaciers, are located across the Northern United States and thus characterized by various climate regimes (O'Neel et al., 2014; O'Neel et al., 2019). As each glacier is situated within the catchment of a close-by discharge gauge, they are well-suited for glacio-hydrological studies. Long-term hydrological, meteorological, glaciological as well as geodetic measurement are available for each glacier, which range back to the 1950s in terms of mass balance and glacier area observations. O'Neel et al. (2019) found that mass loss is not only present from the beginning of the measurements but has actually increased for four of the five glaciers since the 1990s. A trend which is likely to continue under global temperature projections (Tebaldi et al., 2021).

This paper thus aims to ~~benchmark~~ evaluate the recently developed SWAT-GL with respect to its glaciological components. Thus, a sensitivity analysis (SA) using the method of Morris (Morris, 1991), or Elementary Effects (EE), is conducted for screening and ranking of the new input factors under different conditions. Besides, the model is evaluated against long-term glacier mass balance and glacier area-altitude (hypsometry) measurements. Due to the feedbacks between the snow and glacier routine, the model evaluation also considers the performance to simulate snow cover. Discharge is used to cross-validate SWAT-GL under the hypothesis whether a well-performing snow and glacier routine in alpine catchments is sufficient to reproduce discharge in ~~this kind of environments. The benchmarking~~ these environments. Lastly, a comparison of the snow and glacier optimized SWAT-GL model with a discharge-only optimized SWAT-GL model is performed to provide an upper benchmark and to see how glacier representations are affected. The evaluation is performed for four highly glaciated catchments across the US based on the USGS Benchmark Glacier Project (O'Neel et al., 2019).

2 Materials & Methods

In the following we will briefly introduce the USGS Benchmark Glacier Project, the chosen datasets for the benchmarking, the study area as well as the benchmarking and SA approach.

2.1 Datasets & USGS Benchmark Glacier Project

The study is based on the USGS Benchmark Glacier Project (O'Neel et al., 2019) which provides data for five long-term monitoring glaciers across the Northern United States (Mcneil et al., 2016; Baker et al., 2018). The five sites are distributed over Alaska, Washington and Montana and thus represent coastal as well as inland locations. Long-term meteorological, geodetic and glaciological measurements starting from the 1950s or 1960s onward are available for four of the glaciers. For the relatively new Sperry Glacier in the program only short time series (from 2005 on) are available: therefore it was excluded in this study. In the following we only refer to the Gulkana (GG), Wolverine (WG), South Cascade (SCG) and Lemon Creek (LCG) glaciers. Seasonal mass balance estimates are derived from geodetically calibrated, conventional glacier-wide mass balance observations (Mcneil et al., 2016). The project combines measurements with homogeneous data processing methods to allow for inter-glacier comparisons. An overview of the acquisition years of the geodetic surveys can be found in O'Neel et al. (2019). Glaciological field visits of each glacier take place every spring and fall. Summarizing, the following glaciological variables were used from the USGS Glacier Benchmark Project, total annual glacier area (km^2), annual net mass balance change (m w.e.), annual glacier hypsometry (km^2 at a specific elevation range) (see Table 1). Glacier hypsometries hereby represent the area-altitude distribution of the glacier.

Continuous daily meteorological time series (precipitation & temperature) are directly available on-site for the GG & WG. However, although on-site measurements are also available for the LCG & SCG, the time series are rather short and show a relatively high amount of missing values. For reasons of comparability we follow the approach of O'Neel et al. (2019) and use the closest representative station, which is Juneau Airport (LCG) and Diablo Dam (SCG), respectively. The latter two are also part of the official data of the USGS Benchmark Glacier Project (Baker et al., 2018).

However, as SWAT-GL needs minimum daily (Tmin) and maximum daily temperature (Tmax) which was not continuously available in the project for GG (starting 1995) and WG (starting 1997), a regression model was established to produce con-

Table 1. Overview of datasets used. P represents precipitation (mm), T temperature (°C), Q discharge (m^3/s) and SC snow cover (%). Glaciological data is a merged representation of annual net mass balance change (B_{gl} in m w.e.), total annual glacier area (A_{gl} in km^2) and annual glacier hypsometry (H_{gl} in km^2 at a specific elevation range). *var.* indicates that measurements stem from various locations or refer to the whole glacier. The elevation in the glaciological dataset section refers to the total glacier elevation range.

Glacier Basin	Site	Variable	Time Step	Lat	Lon	Elevation [masl]	Temporal Coverage	Missing [%]
Meteorological								
GG	On-site	P	Daily	63.26	-145.41	1,480	1964-2022	9
		T						12
WG	On-site	P	Daily	60.39	-148.94	990	1964-2022	12
		T						14
LCG	Juneau Airport	P	Daily	58.35	-134.56	6	1936-2022	<1
		T						<1
SCG	Diablo Dam	P	Daily	48.71	-121.14	272	1914-2022	<1
		T						<1
Hydrological								
GG	Phelan Creek	Q	Daily	63.24	-145.47	1,127	1966-2023	19
WG	Wolverine Creek	Q	Daily	60.37	-148.9	366	1964-2023	57
LCG	Lemon Creek	Q	Daily	58.39	-134.42	204	1951-2023	41
SCG	SF Cascade	Q	Daily	48.37	-121.07	1,613	1957-1993	28
Snow								
All Basins	Basin mean	SC	Monthly	-	-	-	2002-2023	0
Glaciological								
GG	On-site	B_{gl}	Annual	*var.*	*var.*	1,185-	1966-2022	0
		A_{gl}				2,420	1965-2022	0
		H_{gl}						0
WG	On-site	B_{gl}	Annual	*var.*	*var.*	466-	1966-2022	0
		A_{gl}				1,653	1965-2022	0
		H_{gl}						0
LCG	On-site	B_{gl}	Annual	*var.*	*var.*	543-	1953-2022	0
		A_{gl}				1,550	1946-2022	0
		H_{gl}						0
SCG	On-site	B_{gl}	Annual	*var.*	*var.*	1,619-	1959-2022	0
		A_{gl}				2,439	1950-2022	0
		H_{gl}						0

tinuous daily Tmin and Tmax time series. In detail, daily mean temperature was used as predictor of either Tmax or Tmin in the period where all three variables were available. Subsequently, the regression model was used to predict Tmax and Tmin backwards for the periods before 1995 (GG) and 1997 (WG), respectively. Data gaps of up to three days were linearly interpolated and longer gaps were regressed using daily data from the closest meteorological station of each glacier. The approach is similar to O'Neel et al. (2019), with the only difference that they used monthly regression for longer gaps. We also investigated the potential of ERA5-Land (Copernicus Climate Change Service, 2019) data for gap-filling which was inadequate especially due to a significant precipitation excess throughout the year compared to the station data (not shown in this study).

Hydrological data was obtained from the USGS National Water Information System (U.S. Geological Survey, 1994). We used the closest available gauge for each glacier to determine the total basin area. In detail, the discharge data of the Phelan Creek (representing the GG basin), Wolverine Creek (WG basin), Lemon Creek (LCG basin) and the SF Cascade (SCG Basin) was used. Details about the meteorological and hydrological sites and time series are found in Table 1.

Snow cover (SC) data was derived from the MOD10A1 & MYD10A1 V061 NDSI (Normalized Difference Snow Index) products (500 m resolution). The NDSI is based on optical sensors from MODIS (Moderate Resolution Imaging Spectroradiometer) and is calculated as the difference between the reflection in the green spectrum (GREEN) and the shortwave infrared (SWIR) divided by the sum of the two (Dozier, 1989).

$$I_{NDSI} = \frac{B_{GREEN} - B_{SWIR}}{B_{GREEN} + B_{SWIR}} \tag{1}$$

where I_{NDSI} is the NDSI and B_{GREEN} and B_{SWIR} are the green and SWIR bands, respectively. For the classification of snow or no-snow pixels a NDSI threshold of 0.4 was used, where values above 0.4 (Hofmeister et al., 2022) indicate snow pixels and smaller values are classified as snow-free. Daily fractional SC (%) on the basin and subbasin scale was then calculated as the average of snow-covered pixels within each basin. Subsequently monthly aggregates were produced. MODIS NDSI data was available from 2002 up to now. A full overview of all datasets is given in ~~Tab.~~ Table 1.

Auxiliary datasets used were elevation data from the Shuttle Radar Topography Mission (SRTM) (NASA JPL, 2013), the Randolph Glacier Inventory V6 (RGI) (RGI Consortium, 2017) as well as ice thickness estimates from Farinotti et al. (2019).

2.2 Study Area

The location of the USGS Benchmark Glaciers combined with the location of the corresponding hydrological and meteorological stations used for each glacier is shown in Fig. 1. Note that the representative meteorological stations which were used to force the SWAT-GL models of the SCG and LCG are not shown as they are situated outside of the basin boundaries (see O'Neel et al. (2019) for more details). Besides, the map contains the basin boundaries for each glacier which were used as model domain. The total glacier area in each catchment is slightly higher than the individual glacier area of each glacier, as the basins can include several adjacent glaciers. However, the main glacier fraction can be accounted to the four benchmark glaciers in each basin.

The basins have an area of 28.4 km^2 (GG, 64% glaciated 2009), 23.9 km^2 (WG, 69% glaciated 2006), 29.3 km^2 (LCG, 50% glaciated 2005) and 5.9 km^2 (SCG, 58% glaciated 1958). Basin-wide glacier fractions were determined using Randolph Glacier Inventory data (RGI Consortium, 2017). Each glacier hereby represents a distinct climate regime, where the most northward located GG is characterized by a continental (high-latitude) climate (O’Neel et al., 2014; O'Neel et al., 2019). WG, in contrast,

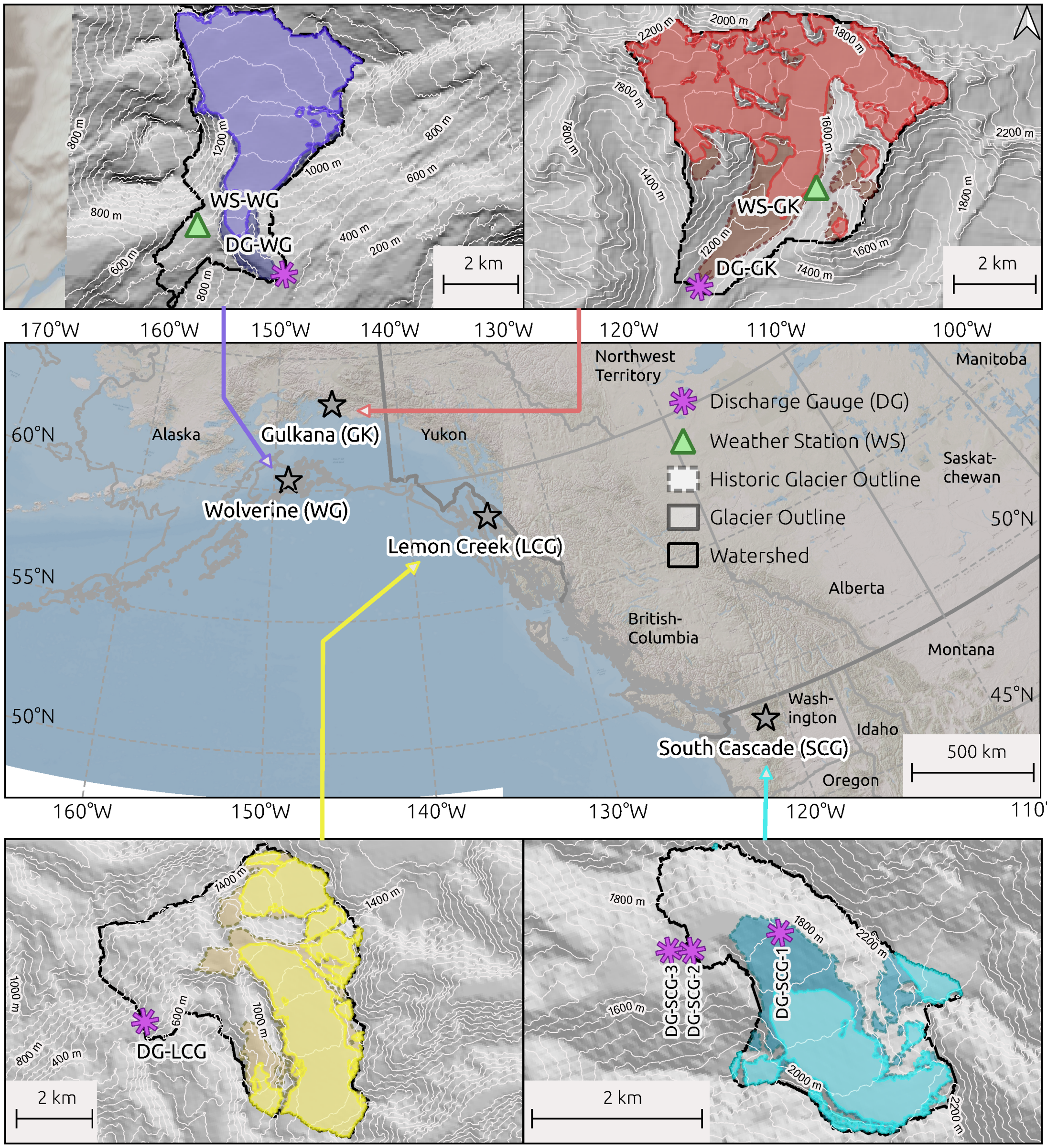

Figure 1. Overview of the four USGS Benchmark Glaciers used in this study. Note that the SCG & LCG meteorological stations that were used are remote stations which is why they are not visible in the map. The transparent outline refers to a historical date, the filled outline to a recent date. The dates are: 1957/2021 GG, 1950/2020 WG, 1948/2021 LCG, 1958/2021 SCG.

is characterized by a maritime (high-latitude) climate regime (O'Neel et al., 2014; O'Neel et al., 2019). LCG represents another high-latitude maritime glacier, while SCG represents a mid-latitude maritime glacier (O'Neel et al., 2019; Horlings, 2016). All glaciers are retreating, where the SCG shows the strongest relative recession with a glacier area loss of more than 40% (1.3 km^2) since 1950. GG lost around 18% (3.3 km^2) of its area since 1965, LCG decreased by 16% (3.3 km^2) from 1946 up to now and WG receded around 12% (2 km^2) compared to 1965. Glacier recession magnitudes reveal a gradient from North to South (O'Neel et al., 2019). Mass balance rates are provided in Fig. 3, which show an increasing negative (statistically significant) trend at all sites (O'Neel et al., 2019). According to O'Neel et al. (2019) total uncertainty, consisting of a geodetic and glaciological component, in the mass balance estimates is around 0.2 m w.e. a^{-1} except for GG where it it higher with 0.4 m w.e. a^{-1}.

The following mean climate characteristics of each basin were evaluated based on the meteorological stations listed in ~~Tab.~~ Table 1 and the period 1971-2000. The continental GG with an annual average precipitation of 1,480 mm and a peak in August/ September is significantly drier than its maritime counterparts. SCG with an annual average of 1,970 mm and WG with an annual mean of 2,375 mm show the highest precipitation values among the four. While WG has its precipitation peak roughly in September/ October it also shows consistent high precipitation during the Winter months and a dry period in summer. SCG is also characterized by a summer low, followed by an increase in precipitation during Autumn and ending in a strong late Autumn and Winter peak (November - January). The annual precipitation totals of the LCG are around 1,480 mm with a less pronounced peak in Autumn. The glacier generally has a lower gradient between wet and dry period, which makes precipitation more evenly distributed throughout the year. In terms of temperatures, all glaciers reach their maximum in either July or August and their yearly minimum in January. The mean climates are illustrated in Figure 2. However, it should be noted that no lapse rates have been applied for the climate classification, the meteorological station used for LCG and SCG come with an elevation difference of more than 500-1,500 m (LCG) and 1,300-2,100 m (SCG). Lapse rates (temperature and precipitation) were later calibrated through the optimization procedure (see section 3.3).

All gauges belong to intermittent streams, which can fall dry during the winter months (Fig. 3). While the corresponding streams of GG and WG had almost no flow in the available time series (see ~~Tab.~~ Table 1) from December to April/May, the streams of SCG and LCG carried water sporadically during these months. Annual average flows are 2.4 m^3/s (GG), 2.7 m^3/s (WG), 5.5 m^3/s (LCG), 28.1 m^3/s (SCG) evaluated for the period 2002-2022 for all glaciers except SCG, where the years from 1972-1992 had to be chosen (due to a lack of observations). Inter-annual variability is highest at GG (Coefficient of ~~Variation (CV) of~~ Variation—CV—of 0.21) and lowest at SCG (CV of 0.13). Except ~~in the SC~~ for the SCG basin, we can see a tendency of a slight shift in the flow period towards an earlier onset of the melt season ~~.~~ (Fig. 3).

2.3 SWAT-GL

The recently developed SWAT-GL (Schaffhauser et al., 2024) is a modified version of the traditional hydrological model SWAT (Arnold et al., 1998), which includes glacier dynamics based on the Δh approach developed by Huss et al. (2010). The empirical Δh-parameterization is called annually to translate the cumulative mass balance change to a change in glacier geometry. The concept assumes that lower elevated areas closer to the glacier terminus receive stronger ablation than higher elevated ones (Huss et al., 2008, 2010). Therefore, glaciers are divided into different elevation sections (ES) for the application. In addition to its spatial distributed applicability the method is mass-conserving and can be applied with glacier outlines and glacier thickness data only (Li et al., 2015; Seibert et al., 2018).

It basically consists of two modules, a mass balance and a glacier evolution module. Mass balance estimations are based on a degree-day approach, similar to the already existing snow routine of SWAT. Glacier evolution is implemented by means of the Δh approach (Huss et al., 2008, 2010). For detailed technical explanations we refer to Schaffhauser et al. (2024), as we only provide a short summary of the main points here.

In general, the mass balance is formulated as:

$$W_t = W_{t-1} - M_t \cdot (1 - \beta_f) - S_t + C_t \tag{2}$$

with W_t being the water equivalent of ice [mm $H_2O \cdot d^{-1}$] at day t, M_t represents the melt rate [mm $H_2O \cdot d^{-1}$], S_t represents the sublimation rate [mm $H_2O \cdot d^{-1}$], C_t refers to the accumulation rate [mm $H_2O \cdot d^{-1}$] and β_f is an adjustable refreezing factor of ice during melt periods. M_t is calculated analogously to snow melt in the standard SWAT using a distinct melt factor.

The physically-based Δh-parameterization is able to simulate spatially distributed glacier retreat. The core of the approach is that glaciers are discretized in elevation sections (ES), where each has an inherent storage and receives distinct ice thickness changes. The ES are normalized for the glacier elevation range and a characteristic (normalized) ice thickness change is assigned to each zone, according to:

$$E_{norm,i} = \frac{E_{max} - E_i}{E_{max} - E_{min}} \tag{3}$$

~~with~~ where $E_{norm,i}$ is the normalized elevation of ES i [-], E_{max} and E_{min} refer to the maximum and minimum glacier elevation [m], and E_i is the actual elevation of ES i [m].

Lower altitudes hereby receive stronger ablation than higher ones. The characteristic ice thickness change for each normalized elevation varies with glacier size. One of three parameterizations is used separately for each glacier, which are thus classified as small (<5 km^2), medium (5-20 km^2) or large (>20 km^2). The empirical relationship is illustrated in Fig. 4. It is important to note that the Δh-parameterization is called annually (at the end of a glaciological year) to redistribute (lumped) annual mass balance changes over the individual ES of a subbasin to simulate glacier retreat. The normalized ice thickness change formulas of Fig. 4 follows the general form

$$\Delta h_i = (E_{norm,i} + a)^y + b \cdot (E_{norm,i} + a) + c \tag{4}$$

where a, b, c, y are coefficients which vary for glacier size and Δh_i represents the normalized ice thickness change for an Elevation E_i. We use the parameters based on Huss et al. (2010). Theoretically the parameters could be derived specifically for any glacier if the required data is available (e.g. two DEMs at different dates). The dimensionless ice thickness change Δh_i is rescaled using a scaling factor f_s [m] to receive the change in meters for every glaciological year (5).

$$f_s = \frac{V_a}{\sum_{i=1}^{n} A_i \cdot \Delta h_i} \tag{5}$$

with V_a referring to the annual glacier volume change expressed in water equivalent [m^3], that is calculated by multiplication of annual EW values (see Eq. 2) and the subbasin area. A_i is the area of ES i with n being the total number of ES. Annual ice thickness changes are then calculated via

$$h_{i,1} = h_{i,0} + f_s \cdot \Delta h_i \tag{6}$$

where $h_{i,1}$ is the updated ice thickness [m water equivalent] after each glaciological year of ES i, $h_{i,0}$ is the ice thickness [m water equivalent] in ES i before the application of Δh parameterization. If $h_{i,1}$ is ≤ 0 the ES is assumed ice-free causing an update of the glacier extent.

SWAT-GL classifies glaciers on the subbasin scale, meaning that a simplified assumption is used where all glaciated areas within a subbasin are considered as one glacier object. The implementation of the glacier routine in SWAT introduces five new parameters which control glacier melt (and refreezing) and accumulation. Besides, one new output file containing annual glacier mass balance information (for each glaciological year) as well as two new input files which require some preprocessing are introduced. The two input files refer to the parameterization on the HRU scale as well as the glacier initialization with respect to hypsometry, ice thickness and volume. The source code of SWAT-GL together with an example is freely accessible via GitLab https://gitlab.com/lshm1/swat-g.

2.4 Sensitivity Analysis

In order to provide a comprehensive picture of SWAT-GL we also performed a global sensitivity analysis using the Method of Morris or Elementary Effects (EEs) (Morris, 1991; Saltelli et al., 2008). The method which is based on a multiple-starts perturbation approach and thus belongs to the one at a time (OAT) methods is able to determine approximate sensitivities at a relatively low computational cost (Saltelli et al., 2008; Pianosi et al., 2016). A lower computational cost hereby refers to the fact that the total number of simulations required is significantly reduced compared to other methods, such as for the Sobol method. The EEs test has thus been established as a robust method for screening and ranking of the input factors (Pianosi et al., 2016).

For sampling we used the radial design, where r sample points from a Latin hypercube sampling serve as well-spread starting points in the input space (Campolongo et al., 2011). The total sample size N follows the form $r(M+1)$, with M being the number of input factors. For r a value of 500 was chosen, resulting in 7,500 model evaluations ($M = 14$) (Sarrazin et al., 2016). A value of r=500 can be considered as sufficient for screening and ranking purposes, for example Vanuytrecht et al. (2014) has shown that a stable ranking was achieved using only 25 trajectories, whereby the relatively small numbers of simulations necessary for ranking was further confirmed by Nossent et al. (2011). The basic idea of the radial design is that from a starting point, one factor M is varied while keeping all others fixed. This results in M steps (varying each factor once) that are performed for each sampling point (r). In general, r EEs are calculated per input factor which are then averaged to provide a global sensitivity metric μ_i for each input factor i. The calculation itself is based on finite differences. To account for non-monotonic effects in the model, μ_i^* is used based on the absolute values of the EEs (Campolongo et al., 2011). The formulation is:

$$\mu_i^* = \frac{\sum_{i=1}^{r} |\mathrm{EE}_i|}{r} \tag{7}$$

An EE of an input factor i can be calculated as follows:

$$\mathrm{EE}_i = \frac{Y(X_1, X_2, ..., X_{i-1}, X_i + \Delta, ..., X_M) - Y(X_1, X_2, ..., X_M)}{\Delta} \tag{8}$$

with $X = (X_1, X_2, ..., X_M)$ being the individual values of the factors and Δ the step (or perturbation). We also calculated the standard deviation σ_i of the EEs as a proxy of the interaction of input factor i with the other factors. It ~~further~~ also describes whether the model output is linearly or non-linearly affected ~~from~~ by an input factor (Garcia Sanchez et al., 2014; Merchán-Rivera et al., 2022). To avoid confusion with other sigmas used in this work, we define sigma of an EE as $\sigma_{EE,i}$. The ratio of ~~$\sigma_{EE,i}/\mu^*$~~ $\sigma_{EE,i}/\mu_i^*$ also provides insights ~~on~~ into whether the effect of a factor is monotonous or almost monotonous as explained in Garcia Sanchez et al. (2014); Merchán-Rivera et al. (2022). This is especially important to derive information whether the interaction of snow and glacier processes are represented (adequately). Due to the different scales of the input factors, the standardization ~~from~~ of Sin and Gernaey (2009) was applied. The SA is used to support and complement the benchmarking procedure to i) determine the importance of the newly introduced parameters under different conditions, ii) rank the parameters to get insights into how the most influential parameters differ between the catchments and iii) identify whether the glacier routine interacts appropriately with the snow routine. For the SA the combined calibration period from 2002-2015 and validation period from 2016-2022 was used.

2.5 ~~Calibration and Validation~~ Model Evaluation Procedure

The core of the study is to evaluate the capabilities of SWAT-GL to simulate multiple glaciological components in glaciated catchments. This not only involves the assessment of multiple performance criteria in the four glaciated catchments, but is also complemented by a SA in order to demonstrate the need of a robust glacier representation. In addition, the ~~benchmarking~~ evaluation is also used to ~~evaluate~~ identify potential model weaknesses of SWAT-GL that ~~need to be~~ should be addressed in future developments of the model. Besides, a comparison of a snow and glacier optimized SWAT-GL model with a discharge only optimized SWAT-GL model is performed.

2.5.1 Multi-objective Optimization

Although hydrological models traditionally focus on the simulation of discharge, which is also one of the goals of SWAT-GL, ~~we will evaluate~~ the glacier routine ~~mainly~~ will be mainly assessed in terms of representing glacier and snow processes. Being aware of the relevance of discharge, the variable will be presented for cross-validation purposes ~~throughout the paper~~ under the hypothesis that an adequate (at least monthly to annual) discharge representation in heavily glacier and snow-dominated catchments can be achieved by a reasonable representation of the snow and glacier components~~alone~~. In detail, we calibrated and validated each of the four models based on snow cover (monthly), glacier mass balance variations (annual) and glacier hypsometries (annual) ~~. For an adequate representation of the snow and glacier routine, monthly snow cover estimates were~~

~~considered to be appropriate. As mentioned, the effects on discharge~~ using an automated multi-objective optimization. As stated before, discharge effects (daily to annual) are provided ~~in addition~~ alongside to evaluate model consistency ~~. Thus, only parameters affecting the newly introduced glacier routine as well as~~ of SWAT-GL. Therefore, given the snow and glacier focused optimization, only parameters associated with SWAT's ~~snow processes, due to its strong interactions with the glacier component,~~ original snow routine and the newly introduced glacier routine of SWAT-GL were used (see ~~Tab.~~ Table 2). As a result 14 parameters were considered. A model run comprised the full available glaciological time series of each catchment, due to differences in the availability of snow cover and the glaciological components. Thus, across glaciers calibration phases can differ. For example, the glaciers were initialized for the starting year of the mass balance time series (GG 1966, WG 1966, LCG 1953, SCG 1959), while MODIS SC was available from 2002 onward. SC was therefore calibrated from 2002-2015 and validated for the period (2016-2021). For glacier mass balances and glacier hypsometries two calibration phases were used (in contrast to snow cover), one which was the same period as for snow cover (2002-2015) and one at the beginning of the model run (GG 1971-1985, WG 1971-1985, SCG 1962-1976, LCG 1953-1967). Discrepancies in the first calibration and validation phase across glaciers stem from different starting dates when measurements started. It was aimed to make use of the full glaciological time series to account for transient behavior in the models. The remainder of the time series was then used as validation phase (one period matching the one of snow cover (2016-2021) and one covering the remaining time series at the end of each glaciers first calibration phase up to 2001). The second validation phase was used in order to make full use of the available information and to assess SWAT-GL over a long time scale. The summary of the temporal settings is given in Table 3. Note that the periods for discharge were intended to match those of the other variables, data restrictions however did not allow for a perfect match.

~~We used a~~ We used an automatic Multi-Objective ~~Automatic~~ optimization (MOO) procedure, where each of the three prescribed variables referred to one objective (SC, MB, hypsometry). The optimization was performed using the widely used evolutionary NSGA-II (Nondominated Sorting Genetic Algorithm) algorithm (Deb et al., 2002). Based on nondomination sorting and the introduction of a crowding distance operator to favor solutions which are less-crowded (high crowding distance), NSGA-II iteratively finds solutions which are uniformly spread at the Pareto front. The population size of a generation was 100 and the maximum number of generation was set to 100. We used ~~Simulated Binary Cross Over~~ simulated binary cross over with a cross over probability of 0.9 and ~~Polynomial Mutation~~ polynomial mutation with a mutation probability of 0.3.

For ~~all variables a normalized form of the~~ glacier and snow variables a Normalized Root-Mean-Square Error (NRMSE), based on the standard deviation of the observations of each variable, was consistently used as objective function (OF) (Eq. 10). The NRMSE increases comparability between the individual OFs and allows to minimize the residuals between observed and simulated values of all variables. For ~~SC~~ snow cover we excluded winter months in the optimization procedure to put more weight on the months where snow cover is dynamic as snow cover is usually 100% from December to at least April. The effect is more pronounced in basins with less snow cover dynamics and therefore a relatively high minimum summer ~~SC~~snow cover. The simulation of a permanent snow cover then leads to good OF values. We will further discuss this issue during the paper. The standard form of the RMSE can be defined as follows:

Table 2. Parameters and their relative ranges used for the benchmarking of SWAT-GL.

Parameter	Description	Minimum	Maximum
SFTMP	Snowfall temperature [°C]	0	4.5
SMTMP	Snowmelt temperature [°C]	0	4.5
SMFMX	Melt factor for snow on June 21 [mm H_2O/(°C·day)]	0.1	7
SMFMN	Melt factor for snow on December 21 [mm H_2O/(°C·day)]	0.1	7
TIMP	Snow temperature lag factor [-]	0	0.5
SNOCOVMX	Snow water equivalent threshold where 100% snow cover occur [mm]	2	75
SNO50COV	Fraction of SNOCOVMX at which 50% snow cover occur [-]	0.1	0.9
TLAPS	Temperature Lapse Rate [°C/km]	-9	-5
PLAPS	Precipitation Lapse Rate [mm/km]	550	1800
GLMLTMP	Threshold temperature for glacier melt [°C]	0	4.5
GLMFMX	Melt factor for ice on June 21 [mm H_2O/(°C·day)]	3.5	13
GLMFMN	Melt factor for ice on December 21 [mm H_2O/(°C·day)]	3.5	10
β_f / f_{frze}	Refreezing factor of glacier melt [-]	0.001	0.01
f_{accu}	Conversion factor of snow to ice [-]	0.1	0.6

$$\mathrm{RMSE}_x = \sqrt{\frac{\sum_{t=1}^{n}(O_{x,t} - S_{x,t})^2}{n}} \tag{9}$$

with $O_{x,t}$ being observed and $S_{x,t}$ simulated components of variable x, which is either snow cover, glacier mass balance or hypsometry, t refers to the time step (monthly or annual depending on the variable) and n represents the number of available data points. The standardization follows the form:

$$\mathrm{NRMSE}_x = \frac{\mathrm{RMSE}_x}{\sigma_x} \tag{10}$$

where σ_x is the standard deviation of the observations of each variable. However, as glacier hypsometries provide areal time series for multiple glacier elevations, the individual RMSE of each elevation was calculated and then averaged to obtain one RMSE value which was standardized in a last step using the standard deviation of observed total glacier area. This gives a more equal weight to all elevations to get rid of solutions where individual elevations might have a high non-standardized error. In other words, if we would use an average of the NRMSE of all individual elevations, those with small observed standard deviations (e.g., higher-elevated and less dynamic ones) could lead to an excessive degradation of the overall OF. The PBIAS is used to show the difference between simulated and observed cumulative mass balance at the end of the simulation period and can be formulated as:

$$\mathrm{PBIAS} = 100 \cdot \frac{\sum_{t=1}^{n}(S_t - O_t)}{\sum_{t=1}^{n} O_i} \tag{11}$$

with the same declarations as above.

Discharge, used for cross-validation purposes, is based on the Kling Gupta Efficiency (KGE), which is calculated as:

$$\mathrm{KGE} = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \tag{12}$$

where r refers to the Pearson Correlation, α to the error in flow variability and β to the bias term with:

$$\alpha = \frac{\sigma_s}{\sigma_o}$$

$$\beta = \frac{\mu_s}{\mu_o}$$

with σ and μ referring to the standard deviation and mean of the simulation and observation, respectively.

2.5.2 Preprocessing & Glacier Initialization

SWAT-GL needs distributed glacier thickness and glacier area information as input for each ES and subbasin. However, this data is usually not easily available for various years. In most cases only globally and openly available datasets such as glacier areas from the RGI (RGI Consortium, 2017) and ice thickness from Farinotti et al. (2019) or Millan et al. (2022), representing a fixed point in time, are available to use. Alternatively, if geodetic information is available at different times glacier thickness can also be directly inferred for each of them. As the USGS Benchmark glacier project provides geodetic data for several years we have chosen the best available DEM closest to the mass balance observation start of each glacier. Best hereby refers to a full coverage of the glacier basins along with a minimum of missing values. The thickness was then estimated using the GlabTob2 model (Linsbauer, A et al., 2009; Linsbauer et al., 2012; Frey et al., 2014). Glacier outlines were also available at several times in the Benchmark project and the year closest to the chosen DEM was selected to initialize SWAT-GL. If a mismatch between DEM and glacier outline acquisition year and mass balance observation start was present, we corrected the initial glacier volume by adding the mass loss or gain that happened since the observation start to the DEM acquisition year. E.g., if mass balance measurements started 1966 and the best DEM and glacier outline was available from 1977, the cumulative mass balance estimates until 1977 were added to the initial volume. The volume was then distributed to the individual ESs while maintaining the original volume fractions of the bands in the total volume. ES sections were defined with a spacing of 100 m.

2.5.3 Statistical Tests

Results for all variables ~~will particularly~~ focus on the last generation of the optimization and the best simulations of each variable. Apart from the described methodology, we tested SWAT-GL's ability to reproduce observed mass balance nonstationarities, inter-annual variability and the monotonic relationship between simulated and observed mas balance. For this purpose, the full time series was used, homogeneity was tested with the Wilcoxon-Rank Sum Test (WRS) (Wilcoxon, 1945), the ~~Pettit~~ Pettitt Test (Pettitt, 1979) and trends were detected using a modified Mann-Kendall version from Hamed and Ramachandra Rao (1998) which considers autocorrelation. For the WRS the first calibration and the last available validation periods were used.

The statistical tests are applied to give an indication whether SWAT-GL is generally able to deal with potential nonstationarities in the catchments, as the model is applied for relatively long simulation periods.

2.5.4 Single-Objective Optimization

Lastly, for demonstration purposes ~~we performed~~ a Single-Objective Optimization (SOO) was conducted using Differential Evolution (DE) (Storn and Price, 1995), with the adaptations described in Dawar and Ludwig (2014) for the WG and two variables, namely mass balance and discharge. Results for mass balance optimized and discharge optimized models were compared with the results stemming from the MOO introduced before (covering MB, SC and hypsometry). The results exemplify ~~what potential users could expect~~ potential upper benchmarks for pure discharge or mass balance calibrations, ~~which~~ where especially the first is often (albeit questionable) the case in hydrological studies. Thus, results demonstrate impacts of SOO models on general model consistency and the missing trade-offs. Discharge was evaluated based on the KGE and mass balance again on the NRMSE. The time periods are equal to those in 3.

~~Parameters and their relative ranges used for the benchmarking of SWAT-GL.~~

Table 3. Overview of the calibration and validation phases. Note: As discharge was used for cross-validation purposes only, it thus has two validation phases rather than a calibration and validation phase. For the SCG no discharge data was available in the 2000s leading to only one validation period. The asterisk for the WG indicates that validation period II is rather poor due to mainly missing values. For snow cover only one calibration and validation phase was used due to the relatively short temporal coverage of the product, while for glacier variables two calibration and validation periods were used. A minus indicates that a specific second calibration or validation period was not used for this variable.

Glacier	Mass Balance		Snow Cover		Hypsometry		*Discharge*	
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Validation I	Validation II
Gulkana	1971 - 1985	1986 - 2001	2002 - 2015	2016 - 2021	1971 - 1985	1986 - 2001	1971 - 1978	1990 - 2001
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021
Lemon Creek	1953 - 1967	1968 - 2001	2002 - 2015	2016 - 2021	1953 - 1967	1968 - 2001	1953 - 1967	1968 - 1973
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021
South Cascade	1962 - 1976	1977 - 2001	2002 - 2015	2016 - 2021	1962 - 1976	1977 - 2001	1962 - 1976	1977 - 1992
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	-	-
Wolverine	1971 - 1985	1986 - 2001	2002 - 2015	2016 - 2021	1971 - 1985	1986 - 2001	1971 - 1981	1986 - 2001*
	2002 - 2015	2016 - 2021	-	-	2002 - 2015	2016 - 2021	2002 - 2015	2016 - 2021

3 Results

In the following, the results of the ~~SA and optimization procedure of~~ Morris SA and SWAT-GL ~~are presented~~MOO optimization are presented, complemented by results providing details on the final parameterizations derived from the MOO optimization, results indicating SWAT-GL's representation of inhomogeneities and variability as well as SWAT-GL's representation of discharge.

First, the results of the SA are shown ~~followed by the sections related to the optimization procedure. The results of the optimization focus on the simulations of the last generation (with N=100 simulations/evaluations)unless otherwise stated. Table 4 provides a summary of the performance metrics for all glaciers based on the last generation of the optimization. Discharge is discussed separately from the results of~~ (Sect. 3.1), followed by an overview of the final parameter sets across all catchments (Sect. 3.2). Afterwards, the results of SWAT-GL's MOO are illustrated, which represent the main purpose of the paper (Sect. 3.3). The MOO results are supplemented by a short statistical analysis focusing on whether SWAT-GL captures inhomogeneity and variability appropriately (Sect. 3.4), as well as results how the MOO (excluding discharge) is capable of simulating discharge (Sect. 3.6). Lastly, results of two SOO models that were optimized using discharge or mass balance, respectively, are compared to the MOO results of simulating mass balance, discharge and snow cover.

It is important to highlight that discharge was excluded from the MOO procedure, as the primary objective of the study is to evaluate SWAT-GL's ~~glacier and snow performance as it was only used for cross-validation purposes.~~ ability to represent snow and glacier processes. However, given that all catchments are strongly driven by snow and glacier processes, discharge simulations from the MOO are discussed and further compared to a SOO model that solely considers discharge and, therefore, can be seen as an upper benchmark for the discharge representation in these catchments by SWAT-GL.

~~SA results based on the EE method for all four catchments and 14 parameters. The slopes (σ_{EE}/μ^*) of the different lines that classify parameter effects on the model outputs as linear, monotonous, almost monotonous and non-linear are: 0.1 (dotted line), 0.25 (dashed line), 0.5 (solid line).~~

3.1 SWAT-GL's Glacier and Snow Parameter Sensitivity

The SA is based on the EEs method. Our results for all catchments are presented as scatterplot between μ^*, the mean sensitivity of a factor (parameter), and σ as proxy for the interactions of a factor (~~Figure~~ Fig. 5).

A common pattern that all catchments share, albeit to varying extents, is their spread around the 1:1 line that differentiates between non-linear and almost-monotonous effects (Garcia Sanchez et al., 2014). Moreover, it is shown that in general the more sensitive parameters (larger μ^*) tend to have higher interactions as well as stronger potential non-linear model responses. It is also shown that the model response of all catchments strongly depends on GLMFMX (blue hexagon) that controls the maximum value of the degree-day factor of ice (and thus the amount of glacier melt that can occur at a specific day of the year). In terms of factor ranking~~GLMFMX is either~~, GLMFMX is the most or the second-most influential factor. It is the most important parameter in the WG & GG basins, where the respective meteorological stations are located directly at the glacier. However, GLMFMX is substituted by the temperature lapse rate (TLAPS, grey circle) at the SCG & LCG, where the respective meteorological stations are located outside of the catchment and at a significant lower elevation than the glaciers. Due to the difference in altitude (which is part of the precipitation correction formulation) between station and elevation band centers it is inherent that the lapse rates become more important. Due to the temperature dominance of both, snow and glacier processes, the sensitivity of the precipitation lapse rate (PLAPS, grey diamond) is less pronounced and strongest at the high-elevated SCG. Among the four most important factors is the threshold temperature of glacier melt (GLMLTMP~~)~~, blue upward triangle), which controls the onset of melt and has an effect on the timing of melt events as well as the amount of melt. The temperature lapse rate and glacier melt temperature can favor similar conditions or act contradictory (decrease of melt temperature favors earlier melt onset and small or no temperature lapse rate as well). In general, SWAT-GL is strongly temperature-dominated in all catchments.

However, the relevance of precipitation in the SCG basin might be a special characteristic (with regard to the PLAPS ranking in the other basins). ~~Besides, (with an exception~~ In addition, (except for the SCG ~~) the SNOCOVMX parameter~~ and LCG) SNOCOVMX (red pentagon) is ranked among the four most sensitive parameters~~in the catchments~~. The parameter determines a threshold of snow water equivalent (SWE) that ~~is required to cause~~ corresponds to a 100% coverage of snow. As glacier melt can only occur when the glacier is snow-free, the parameter directly affects glacier melt, which explains its relevance in the catchments.

Overall, the ranking shows that across all basins, GLMFMX, TLAPS, GLMLTMP, SNOCOVMX and PLAPS are found to be most often among the most influential parameters.

With respect to potential ~~interaction~~ interactions and non-linear model responses, the accumulation factor (f_accu, blue square) that is responsible for the snow metamorphism (or turnover from snow to ice) takes a dominant role at the WG, GG and SCG. This is plausible as it couples snow and glacier processes by transforming a specific fraction of snow lying on the glacier to ice and thus affecting both storages. Although it is not among the most sensitive factors, it can have a high significance in certain situations due to its possible interactions. Although the most influential parameters receive high σ_{EE} values, they do not necessarily fall in the non-linear area (GG, WG). ~~However, all models show generally~~ The strong non-linearity visible for the SCG & LCG might be caused by the fact that the underlying measurement stations are located outside of the catchments at relatively low elevations. All models generally show a non-linear or monotonous behavior and are potentially characterized by interactions rather than a linear relationship.

~~Furthermore, we can identify at least~~ Additionally, we identify around 6 to 8 parameters that are less or non-influential~~parameters~~, which would reduce the ~~dimension~~ dimensionality of the parameter space ~~to a 6 or 8 dimensional problem in the different models, respectively~~for the respective models. However, given the strong interactions of the parameters, the parameter space was kept constant, including the entire space.

3.2 Inter-Basin Comparison of Optimized Glacier Parameters

3.3 ~~Inter-Basin Comparison of Optimized Glacier Parameters~~

A comparison between the values of the final parameter sets of all catchments is shown in Fig. 6. As the main purpose is to evaluate the glacier routine introduced in SWAT-GL, the comparison is limited to the five glacier parameters. Results are presented for the ~~RMSE~~ NRMSE values of the annual glacier mass balance only. The parameter values of the GG are relatively well-spaced in the parameter space with an exception for the GLMFMX parameter, which controls the maximum amount of glacier melt. The parameter tends to cluster at its lower boundary for GG, LCG and SCG. The lower bound of GLMFMX is associated with a reduction of strong negative mass balance rates, which might lead to an overestimated ablation. ~~Analogously to the~~ The parameter is among the most interactive ones, as shown before, indicating that parameter clustering does not necessarily result in a distinct objective function distribution. As Fig. 6 only provides a snapshot for the NRMSE of MB, the pattern of the objective function space for SC for example could look differently. Analogously to GG, the final parameterizations of the LCG are generally well-spread. An exception exhibits ~~here~~ the glacier melt temperature (GLMLTMP) and GLMFMX that both show clustered values. In contrast to the maximum melt factor, the glacier melt threshold temperature groups at its (relatively high) upper bound for the LCG, WG and SCG. The two patterns indicate the necessity of high melt rates, which should not occur too early. The final parameter distribution of the WG is more narrow compared to the other catchments. It is shown that especially for the accumulation rate (f_{accu}), small values are desired to avoid large positive mass balance simulations.

Looking at the best parameters for different objectives, it is shown that the best SC simulations can deviate significantly from the best MB simulations (in both the difference in the NRMSE of the MB and the parameter values). Large deviations in terms of the NRMSE of the MB can be seen for the SCG and WG (large vertical difference of blue cross and x). With respect to the best final parameters of these two variables, large differences can be observed for GLMLTMP, GLMFMX of the WG (large horizontal distance of blue cross and x), GLMFMN of the SCG and LCG. Moreover, it already becomes apparent that the best discharge simulations do not necessarily coincide with the best simulations of glacier mass balance or snow cover, despite the strong dependency of discharge on snow and glacier melt, which is further discussed in the following chapters.

Table A1 provides a detailed overview of the ranges and median values for each glacier parameter and each catchment at the end of the optimization.

3.3 Evaluating SWAT-GL's Representation of Glacier & Snow Processes

~~Performance of SWAT-GL for all variables and glaciers with respect to the best simulation of the last generation of the optimization procedure. Note: Discharge was not calibrated but is only shown for cross-validation purposes. Discharge thus has two validation phases following the periods assigned to the glacier mass balance evaluation of each glacier (see Section 3.3). Cum. B_{gl} refers to the mismatch between observed and simulated cumulative mass balance at the end of the time series and is therefore not attributed to any of the calibration or validation periods. Negative values of Cum. B_{gl} indicate that the model is underestimating mass balance losses.~~

The performance of the optimization procedure is shown in Table 4. Statistical results for discharge are presented alongside the other variables for demonstration purposes, although it was not part of the optimization procedure. Results for discharge comprise two validation periods which were chosen analogously to the ones of glacier mass balance, if data was available.

With respect to glacier mass balance estimates, lowest NRMSE values were found for the WG model (both, calibration and validation), followed by the SCG model. In contrast, the WG model shows the worst performance w.r.t. hypsometries, for which best performance was reached in the GG basin. Concerning snow cover, the LCG and SCG model achieve the best performance metrics in both the calibration and validation period (0.43 and 0.44 for LC model, 0.35 and 0.36 for SC model). ~~No significant degradation was~~ Slight variations were observed between the calibration and validation phases for any of the objective functions of all variables included in the optimization procedure. It should be noted that a direct inter-comparison of the absolute NRMSE values between the catchments is difficult, as the standard deviation of the observations used for the

Table 4. Performance of SWAT-GL for all variables and glaciers with respect to the best simulation of the last generation of the optimization procedure. Numbers in brackets belong to the initial results based on a latin hypercube sampling that serve as starting values for the optimization to provide an indication on the optimization performance. Note: Discharge was not calibrated but is only shown for cross-validation purposes. Discharge thus has two validation phases following the periods assigned to the glacier mass balance evaluation of each glacier (see Sect. 3.3). Cum. B_{gl} refers to the mismatch between observed and simulated cumulative mass balance at the end of the time series and is therefore not attributed to any of the calibration or validation periods. Negative values of Cum. B_{gl} indicate that the model is underestimating mass balance losses. Optimal values for the NRMSE and PBIAS are 0 and 1 for the KGE, respectively.

Glacier	NRMSE [-]						KGE		PBIAS [%]
	Mass Balance		Snow Cover		Hypsometry		Discharge		Cum. B_{gl}
	Calibration	Validation	Calibration	Validation	Calibration	Validation	Validation 1	Validation 2	Complete
Gulkana	0.76 (0.80)	0.68	1.08 (1.36)	0.98	0.34 (0.39)	0.28	0.62 (0.71)	0.62	-21.68
Lemon Creek	0.74 (0.86)	0.70	0.43 (0.71)	0.44	0.70 (0.70)	0.75	0.29 (0.32)	0.19	-1.60
South Cascade	0.67 (0.90)	0.69	0.35 (0.46)	0.36	1.13 (1.13)	0.90	0.82 (0.81	0.59)	-55.17
Wolverine	0.51 (0.58)	0.56	0.87 (1.40)	0.99	2.29 (2.30)	1.79	0.64 (0.78)	0.64	-11.16

normalization has a dominant effect on the values. For example, the standard deviation of the mass balance observations of the WG is a factor of 1.6 - 9.4 higher than that of the other glaciers (calibration period). The same applies for the snow cover results of the SCG, where observed snow cover standard deviations are 1.3 - 2.4 times above those of the other glaciers.

A clearer picture emerges from the graphical representation of the optimization results in ~~Figure~~ Fig. 7. The discrepancy at the end of the simulation periods of the cumulative mass balance ranges from -1.6% to -55.17%. ~~However, the~~ The relatively large outlier of -55.17% arises from the SCG, where the mass balance loss stagnates in the 1990s. The abrupt change in the mass balance is also indirectly reflected in the cross-validated results of discharge. While the KGE in the beginning of the simulation is around 0.82, a significant drop to 0.59 can be observed in the second evaluation period. Problems in the SCG model become especially apparent when focusing on the lower bound of the model range (grey shading in Fig. 7 of cumulative mass balance). Here, we can notice a very poor representation of the inter-annual signal as the simulations show two very abrupt drops and long periods of stagnation. It is likely that the glacier has retreated to altitudes, which are not subject to temperature-induced melting. The wide range visible in the cumulative mass balance ~~can~~ could be caused by only a few solutions and does not allow for conclusions about the real distribution of the final simulations. We thus show in Fig. A1 the individual cumulative mass balance representations of all optimized solutions together with the distribution of the cumulative mass balance at the end of the simulation period (all 100 values of the last year of simulation for each glacier). It is found ~~,~~ that particularly the upper bound of the GG and WG, as well as the lower bound of the LCG and SCG are caused by a small subset of solutions. The LCG with an almost perfect fit at the end of the simulation period, however, is overestimating ablation ~~over~~ for a large part of the simulation period. Overall, the models are underestimating ablation rates after the 2000's (with an exception for the WG). This becomes even more evident by looking at the annual mass balance rates (last column Fig. 7). All models perform well in simulating monthly snow cover, in both, the calibration and validation phase. The spread of the models (grey shading) is relatively large and includes simulations with almost no snow cover in summer. The WG consists of a period of positive mass balance in the 1980s which is likely causing the upward tendency of the simulation range (the share of simulations with a positive cumulative mass balance until the end of the simulation).

Summarizing, ~~we assess~~ the objective function values indicate SWAT-GL ~~'s results in general as satisfactory to very good,~~ being generally capable of representing the glacier dynamics with the exception of ~~SCG, which shows sharply declining accuracy of the results~~ the SCG that shows a sharply declining drop in performance over the course of the simulation period. However, a benchmark model would be necessary for an absolute interpretation of the values.

~~Simulation results of the last generation of the optimization procedure for the cumulative mass balance (Column 1), monthly snow cover (Column 2) and annual mass balance (Column 3). Each row corresponds to one glacier. Blue represents the best evaluation of the last generation for the respective variable, black refers to observations and grey shadings indicate the range of all evaluations part of the last generation. The dashed lines with the Cal. annotation indicates the individual calibration phases of each study area. The remainder of each time series was used for validation.~~

3.4 SWAT-GL's Ability to Capture Mass Balance Inhomogeneity and Variability

~~As~~ As SWAT-GL was tested for relatively long simulation periods~~were used,we evaluate whether SWAT-GL is capable~~ ,its capabilities to capture potential inhomogeneities ~~which are~~ present in the mass balance observations are evaluated. Furthermore, it was investigated how ~~basic summary statistics of mass balance simulations, such as~~ mass balance variability, e.g. represented by the Coefficient of Variation (CV) as proxy for inter-annual variability, ~~differ from observations~~are represented by SWAT-GL. Inhomogeneities were ~~not only determined by~~ determined by a trend detection based on ~~the~~ a modified Mann-Kendall ~~, but also~~ and additionally using the Pettitt and Wilcoxon-Rank Sum Test (WRS). The Null Hypothesis of both methods is that the time series contain no change. All results are provided in Table 5. Results of the MK, WRS and Pettitt tests are provided as integers, 0 or 1, indicating whether the null hypothesis was rejected (1) or not (0).

Table 5. Summary of statistical results for the simulated and observed mass balance time series over the whole simulation period. The summary table consists of the Spearman Correlation (ρ), the modified Mann-Kendall after Hamed and Ramachandra Rao (1998) considering autocorrelation (MMK_H), the Sen's Slope estimator, the Coefficient of Variation (CV) as well as the Pettitt and Wilcoxon-Rank Sum (WRS) Test. Results of the MK, WRS and Pettitt tests are provided as integers where 1 indicates the null hypothesis was rejected and 0 that is was accepted.

Glacier	Mass Balance										
	ρ	MK_H		Sen's Slope		CV		Pettitt		WRS	
	-	Sim	Obs	Sim	Obs	Sim	Obs	Sim	Obs	Sim	Obs
Gulkana	0.67	0	1	-5.142e+4	-2.479e+5	1	1.30	0	0	0	1
Lemon Creek	0.62	0	1	-5.846e+4	-2.107e+5	0.88	1.21	0	0	0	0
South Cascade	0.79	1	0	3.281e+4	-1.113e+4	2.75	1.72	1	0	0	0
Wolverine	0.83	1	1	-5.311e+5	-3.931e+5	2.52	2.72	1	1	0	0

While the Spearman Correlation suggests a satisfying to good agreement in the monotonic relationship between simulated and observed mass balance in all catchments, the models failed to represent the trend detection results of the observations. Based on the modified Mann-Kendall test that takes serial correlation into account, SWAT-GL was only in one catchment able to correctly classify the trend statistic. The Sen's slope estimator, used to represent trend magnitudes, differed especially for the outlier SCG model where the sign was mismatched. In the WG and LCG models the simulated trend component was underestimated, while it was overestimated for the WG. The CV, with generally large values mostly above 1, was underestimated in two as well as overestimated in two cases. While it was general in an acceptable range, the SCG again exhibited an outlier. The annual mass balance time series, which were trend-corrected in the presence of a trend, have been further tested on inhomogeneities based on the non-parametric Pettitt and Wilcoxon-Rank Sum Tests.

In summary, the simulations agreed relatively well with the observations at the 0.05 significance level. However, SWAT-GL was not capturing the shift in median detected in the observed time series of the ~~WG (~~GG (based on the WRS) and ~~further~~ additionally rejected the Null Hypothesis of the Pettitt Test ~~in case of~~ for the SCG. However, it ~~must be noted that the poor simulations of the SCG affect the meaningfulness of the~~ can be observed that poor SCG simulations affect the significance of test results.

3.5 Cross-Validation of Discharge

Discharge in all catchments was cross-validated on the daily scale, assuming that a reasonable fit is achievable, when glacier and snow-related processes are well represented in the heavily glaciated catchments. The performance, evaluated based on the KGE, can be found in Table 4. The temporal coverage of the validation periods of each catchment is found in ~~Tab.~~ Table 3. First, we see that the GG and WG models show no difference in the KGE values of the respective calibration and validation periods. Second, a significant drop in quality for the SCG model (from calibration to validation) is found. Lastly, the LCG model acts as a strong outlier with KGE values <0.3. In contrast, the other three catchments almost entirely show KGE values >0.6, results often considered satisfactory in hydrological studies according to the classifications of Moriasi et al. (2007); N. Moriasi et al. (2015) . Interestingly, the worst-performing glacier during the glacier-based optimization (SCG) exhibits the best overall discharge performance. ~~However, its~~ Its good quality in the first validation period stems mainly from an overestimated ablation that reduces the underestimation of available water for discharge compared to the other glaciers. Given the large glacier influence in the catchment, the good performance w.r.t. discharge is likely caused for wrong reasons. This stresses the fact that models that are evaluated only for discharge are questionable for comprehensive hydrological investigations. The other glaciers are characterized by a high PBIAS towards the observations (simulations have less water than observations). ~~Besides, the~~ The second validation phase of the SCG model, in contrast to the other glaciers, does not cover the 2000s, which are associated with even higher instationarities. Covering the 2000s would likely further reduce the performance metric. ~~The large PBIAS is most significant~~ PBIAS is larger in the LCG model, where streamflow is underestimated by 45% ~~or~~ and 51% in the calibration and validation phase, respectively.

~~This behavior is emphasized when looking at Fig. 8, which illustrates mean annual discharge,~~ In addition to the daily performance of Table 4, we further evaluate mean annual flows as illustrated in Fig. 8 together with two separate periods of simulated mean daily discharge (averaged discharge for each day of the year over the indicated periods). For the WG, LCG and GG models a distinct underestimation of annual flows is present in the simulations. When ~~we center (correct the time series by its mean) the annual discharge time series (see Fig A2), we find~~ annual flows are centered (mean corrected), a good monotonic relationship for the WG model ~~.~~ is observed (see Fig. A2). For the SCG, the deviation of observed and simulated annual discharge increases over time. This further indicates that a temporal coverage of the 2000s of the SCG model would further degrade its results. The annual dynamics of the LCG and GG are only partially met and particularly the larger positive anomalies of the LCG are not captured by the model. Results suggest that inter-annual discharge variability is slightly lower in the model than observed. In general, simulated annual flows of SCG show a decreasing tendency with time, which could be caused by the strong recession initiated at an early stage of the simulation period as shown before (~~See~~Sect. 3.3). This could then cause a pronounced underestimation of glacier melt contribution especially at end of the simulations as the actually contributing elevations disappeared already.

Evaluating mean daily discharge of the different periods further stresses the substantial undercatch of flow in the simulations (Fig. 8 all glaciers). The periods were chosen so that they lie from each other to highlight potential model deficiencies in the representation of nonstationarities. The LCG and GG models do not show any significant change in the amount of discharge between the early and late period. This points to relatively stable glacier conditions. In contrast, flows in the SCG basin essentially decrease in the later period, which is in line with the aforementioned results. A similar, albeit not as pronounced, pattern is found for the WG model. The model of the SCG shows the largest range of simulated flows over the year (~~grey~~ gray shadings). The share of simulations with a positive mass balance for the WG (see ~~Section~~ Sect. 3.3) is likely causing the very low simulation bound of discharge in the basin, with almost no flow until August.

3.6 Comparison of Single-Objective and Multi-Objective Optimization Results of Discharge and Mass Balance

To illustrate the capabilities of SWAT-GL in terms of discharge and mass balance simulations, two SOOs were conducted. One for each of the two variables. The SOO of discharge was ~~based on KGEand~~ evaluated using the KGE, while that of mass balance ~~on NRMSE to be consistent~~ relied on the NRMSE, ensuring consistency with the MOO counterpart ~~.~~ presented earlier. Results of the two SOO models that do not consider any trade-offs in process representations are compared with the MOO model results introduced and shown in Sect. 3.3 to 3.5 (using SC, MB and hypsometry). The test case was conducted for the WG only and should replicate a typical hydrological ~~modelling~~ modeling case where people rather use discharge or mass balance only than multiple objectives. It ~~should be emphasized here~~ is important to clarify that we are not ~~saying that the approach is desirableor good practice, but thatdespite known shortcomings of single-objective~~ endorsing this approach as desirable, rather, we highlight that, despite well-known shortcomings of SOO studies, the approach ~~is still~~ remains a common practice. For the SOO of the mass balance~~the parameter choice~~ , parameter choices and ranges are similar to Table 2. ~~However, as we did not consider any lags for discharge in the MOO we added SWAT's SURLAG~~ For the SOO of discharge the following adaption in terms of parameter choices was included: To account for lags in discharge as well as the general infiltration behavior, SURLAG (surface runoff lag) and CN2 ~~parameter for~~ (Curve Number for moisture conditions II) parameters were introduced for the SOO of discharge ~~to maximize~~ in contrast to the MOO parameter space of Table 2. The introduction aims to improve SWAT-GL's capabilities in the representation of streamflow.

Fig. 9 illustrates the results. The first line of plots (a ~~) and b)~~and b, respectively) ~~refers to the SOO model of~~ focus on discharge and the second to the one of MB (c ~~)~~and d)~~)~~. The first column shows the SOO results for the variable which was optimized (a ~~)~~and c) ~~)~~and the second column for the variable that was not used in the respective SOO (MB in b~~)~~, Q in d)~~)~~. The SOO results clearly demonstrate the sharp increase in KGE values when discharge is used directly as objective, compared to KGE values from the MOO. The former best KGE of 0.64 (~~Tab.~~ Table 4) is substituted by a relatively high upper bound of 0.92. The results are reached after 40 generations already and lead to a median KGE shift from 0.5 to 0.9. Besides, the median KGE of the MOO represents the lower bound of the SOO. While the simulated mean annual flow of the MOO procedure showed a significant bias compared to the observed flow (see Fig. 8), the SOO results in PBIAS values of -0.78% (not shown) and is capable to bring reasonable amounts of water in the system. The SOO of discharge shows a substantial degradation in representing annual mass balance compared to the MOO (~~b))~~shown in b). The best solution of the SOO of discharge exhibits has a NRMSE value which is 0.37 above the MOO counterpart (or 78%).

When looking at annual mass balances, although we see a much better representation already after 20 generations, the "best" performing simulations smaller or equal to the 10% percentile are similar between MOO and SOO. This ~~even goes so far that the~~ extends to the point where the minimum NRMSE of 0.51 achieved in the MOO ~~, is not beaten~~ is not surpassed by the SOO procedure even after convergence. This ~~indicates the appropriateness of the MOO using~~ highlights the suitability of an MOO with SWAT-GL ~~in~~ for complex glaciated and snow-fed catchments. However, the MOO shows slightly poorer results for discharge compared to those resulting from the SOO for mass balance (d)). Above the 20th percentile the performance of the SOO has on average a KGE approximately 0.05 above the one from the MOO. Below the 20th percentile the performance of the MOO is largely degraded. When looking at a variable that was not included in any of the SOO models but in the MOO, such as SC, the superiority of the MOO approach becomes visible. Similar to Fig. 9), the SC results are illustrated in Fig. A5) suggesting a clear dominance of the MOO model over the two SOO models.

It was shown that the MB-based SOO was superior in representing both, MB and discharge, while the discharge-based SOO only demonstrated higher capabilities in representing discharge compared to the MOO. However, interpretation should be done with caution, as for example both SOO models show degraded results in representing further variables as shown in Fig. A5. Furthermore, the comparability, especially for discharge, is constrained due to the fact it was not considered in the MOO at all.

~~Cumulative distribution function (CDF) plot of the objective functions for the single-objective optimization (SOO) and multi-objective optimization (MOO) results of daily discharge and annual mass balance for the Wolverine Glacier. a) and b) compare the MOO results with the SOO model optimized for discharge. c) and d) compare the MOO results with the SOO model optimized for mass balance. a) and c) show the results for the optimized variable of the SOO, b) and d) show how the corresponding non-optimized variable (MB for b) and Q for d)) perform. The CDF of the MOO refer to the last generation (Generation = 100), while the selected SOO results refer to generations at a relatively early stage of the optimization close to the point where convergence is reached. In detail, the selected generation of the discharge optimized model refers to generation 40, and generation 20 for the mass balance optimized model. Black indicates the CDF from the MOO, red the one from the~~

~~discharge SOO model and blue the mass balance SOO model. The subscripts Q and MB indicate discharge and mass balance, respectively. The sample size N of the CDFs is 100 which refers to the general sample size in the optimization. Results focus on the calibration period of the WG.~~

4 Discussion & Outlook

As a profound evaluation of SWAT-GL's performance in different glaciated catchments was missing so far, the intention was to contribute to close this gap with our work. Note that further information ~~especially regarding~~ with respect to technical details of SWAT-GL and future plans about model improvements are found in Schaffhauser et al. (2024).

4.1 Glacier Parameterizations & Process Representation

We used the Method of Morris to identify (screen) and rank glacier and snow parameters in the four basins (Song et al., 2015; Pianosi et al., 2016; Sarrazin et al., 2016).

In general, it was shown that lapse rates together with parameters controlling the maximum degree-day factor for ice (and thus glacier melt) are among the most sensitive parameters in all catchments. The strong temperature-dependence ~~inherent in~~ of SWAT-GL is further emphasized ~~through the relevance~~ by the high sensitivity of the threshold temperature ~~that triggers glacier melt~~ when glacier melt occurs. An important role plays the SWE threshold that determines when a (sub-)basin is fully snow-covered. The parameter links the old snow routine and the newly integrated glacier routine, as glacier melt can only occur under snow-free conditions on the glacier. SWAT's snow cover received growing attention in multi-objective calibration studies that try to improve model consistency (Tuo et al., 2018; Grusson et al., 2015). The fraction of snow cover is directly affecting the amount of daily snow melt (lower fractions reduce the amount of snow melt) and indirectly glacier melt. As any degree of snow cover can be achieved with any SWE, there is also the risk to accomplish good snow cover results with implausible amounts of snow. This circumstance is, contradictory to its importance, rarely discussed in the literature. SWE measurements ~~are, although seldom available, of tremendous importance as they can be used for plausibility checks of snowamounts. Reanalysis~~ can therefore add significant benefits as they provide valuable insights on actual amounts of snow. However, resolutions of reanalysis or remote sensing-based SWE ~~product resolutions as well as a relative large spread~~ products, along with the significant variability observed in product comparison studies~~are still an obstacle. Further, SWE would probably~~, remain a challenge. SWE observations would be more suitable ~~to draw conclusions on~~ for drawing conclusions about precipitation inputs compared to ~~snow cover only~~relying solely on snow cover. We want to further emphasize that the intermediate sensitivity of precipitation lapse rates might be misleading. The objectives chosen might not be valid to allow for precipitation-related conclusions as none of the three variables is based on absolute volumes (of snow or ice). For example, a separate consideration of summer ablation and winter accumulation would provide a more realistic picture of system in- and outputs.

As SWAT-GL is still in its early stages, the SA was conducted for diagnostic purposes, involving screening and ranking among different catchments. This was done independently of the optimization purpose. In future applications, the dimensions of the parameter space should be reduced accordingly. The SA for example suggested a reduction of the parameter space (by 6-

8 dimensions) in the different catchments of the study. The derived glacier parameter sets after optimization, are relatively well-spread in the parameter space. However, in the demonstration catchments, the maximum melt factor tends to group at its defined lower bound. This indicates a potential reduction of the lower bound for an even better representation of glacier mass balance. ~~Physically our~~ The chosen values could be reduced, however, as SWAT-GL internally makes a plausibility check between the estimated snow and ice degree-day factors, a further reduction might make internal corrections of the degree-day factor more likely. ~~Besides~~In addition, further reducing the lower bound of the parameter might exacerbate the strong underestimation of flow. The high values of the glacier melt temperature imply that the models seem to compensate for other temperature-dependent processes as the model seems to try to delay glacier melt. This indicates that glaciers are, despite the good SC representation, snow free and exposed to melting too early. A lag factor similar to the temperature lag factor of snow already present in SWAT could also give further control in the timing of glacier melt. However, our study has shown that the snow lag factor is not very sensitive, although its lower bound was chosen ~~in a way~~ to avoid abrupt and extreme snow melt events. We further propose that alternative solutions concerning the lag of ice and snow melt might be explored and evaluated in order to decouple them more clearly from the lag factor related to effective precipitation.

Overall, the standard deviation of the Elementary Effects indicate that glacier and snow processes behave strongly non-linear and exhibit potential interacting effects, which we see as a further indication of SWAT-GL's suitability. The moderate interaction ability for SNOCOVMX is considered to be unusual, as it links snow and glacier processes which would suggest higher interaction and/or non-linearity. Future work might put attention on Time-Varying Sensitivity Analysis (TVSA), such as DYNIA or also using EEs, to obtain further insights in parameter dominance at different scales and periods over time (Chiogna et al., 2024). Especially in the context of climate impact assessment, insights of a potential loss in model skill due to a reduction in the dominance of (historically working) parameter sets in non-stationary systems become crucial (Wagener, 2022).

4.2 SWAT-GL's Performance in Representing Glaciated Catchments

The optimization procedure using NSGA-II for snow cover, glacier hypsometry and glacier mass balance worked well for the highly glaciated catchments with the exception of the SCG. For the SCG an abrupt change in the mass balance estimates in the middle of the simulation period causes implausible results. However, snow cover estimates were very good in all catchments and we highly recommend to use snow cover as an objective function for an adequate representation of mountain hydrology when using SWAT-GL. ~~Especially due to the fact that~~ This is particularly true as MODIS (or other) snow cover data is relatively easy to access and ~~readily available, which is not the case for measurements of~~ widely available, unlike glacier mass balance ~~estimates~~measurements, which are more challenging to obtain. In data-scarce regions, predominantly the typical setting of high-mountainous areas, snow cover in combination with downstream measured discharge might often be the only sources of data for calibration and validation.

While annual net mass balance was well represented, it was noticed that glacier melt tends to start too early leading to an extended overlapping period where snow and glacier melt contribute equally to runoff generation. The mass balance estimates are better represented for the bigger glaciers or catchments respectively. However, SWAT-GL was introduced to provide a simple but efficient approach to represent glacier dynamics on multiple scales. ~~It is assumed that most application will generally~~

~~be beyond the scale of the example of the~~ While SWAT applications often focus on larger scales compared to the relatively small SCG, ~~which might also be one reason for the relatively bad performance . Moreover, small-scale processes such as snow redistribution are equalized on larger scales (e.g., mesoscale) and thus less dominant.~~ more evaluation is necessary on the reasons for the bad performance and whether it might be linked to the small spatial scale. Something which is especially important to investigate given the overall trend of shrinking glaciers across the world.

As hypsometry measurements were available, they were used in the optimization process. In future work total glacier area might be a suitable alternative to the individual hypsometry time series. As the Δh-parameterization assumes upper parts of a glacier to be more stable we conclude that the approach might fail to represent the dynamics of the upper elevation sections at the SCG and LCG while it seems more appropriate for the WG and GG (see Fig. A3). Using total area changes could therefore improve the representation of the overall MOO, as it would circumvent the attempt to reproduce a pattern SWAT-GL is structurally not able to. Similarly, if individual hypsometry time series are used one might consider to put less weight on the upper parts of the glacier.

We have shown that, by using discharge as a single-objective, as done for the WG, the performance could be substantially ameliorated (KGE >0.9). Using mass balance in the SOO (again WG) we have shown that the best solutions of the MOO were comparable to the ones resulting from the SOO with respect to the achieved NRMSE values. The statistical results of the mass balance estimates significantly dropped using discharge for SOO and could not compete with the MOO mass balance results. In contrast, the discharge performance as a result from the mass balance SOO was better than the discharge representation of the MOO. Unlike mass balance, discharge was not part of the MOO objectives and partly constraints the interpretability. Nevertheless, the difficulty to achieve model consistency in highly glaciated and mountainous catchments became particularly visible in the LC catchment. The studied glaciers generally have a high contribution to streamflow, as for example found in O'Neel et al. (2014) (for GG and WG). Since we consistently overestimate ablation for the LCG, it is initially contradictory that we obtain streamflow underestimations of up to about 50% (ablation in SWAT-GL mainly refers to glacier melt). Potential reasons can be manifold. A wrong representation of snow amounts and distribution, despite a good snow cover fit, or a simple underestimation of liquid precipitation (or a combination of the two) might be potential reasons. Moreover, similarly to what is described in O'Neel et al. (2014) the model could also underestimate summer ablation and winter accumulation which govern the mass balance, which would again be related to precipitation. However, it seems that precipitation input might be too low since the LCG meteorological data stems from a remote valley station and the model relies on optimized lapse rates. Precipitation estimates in mountainous catchments can generally be considered as a complex field coming with large uncertainties that are then reflected in our hydrological models (Evin et al., 2024). Therefore future modelling work could also try to not only use net mass balance but (if available) make further use of seasonal mass balance (winter accumulation & summer ablation) derivations in the calibration strategy (Schaefli and Huss, 2011). The winter mass balances could also be used to additionally validate the precipitation inputs from the stations and to adjust the precipitation laps rates.

In general, it became evident that SWAT-GL has great capabilities to be applied in glaciated catchments, also for longer, non-stationary time scales. It is assumed that the simple degree-day approach integrated in the mass balance module alone could cause significant improvements in glaciated catchments. The simplicity of the approach also leads to high transferability with manageable effort.

4.3 Glacier Initialization

A further very sensitive factor that affects the simulation results, in particular under long simulation periods which are likely subject to persistent climate change effects, is the initialization of the glacier mass. Datasets such as from Farinotti et al. (2019) or Millan et al. (2022) provide valuable information on glacier thickness and thus mass initialization. But attention should be paid when simulation periods start decades ago before the considered thickness estimates. A comparison of the mass differences between the Farinotti estimates and our own calculations based on GlabTop-2 for the earliest possible DEM and outline (see ~~Section~~ Sect. 2.5) reveal substantial differences. The magnitudes are between -11% to +19%. The contradictory signs are produced as the Randolph Glacier Inventory outlines for the different glaciers stem from different acquisition years which are sometimes earlier than the outlines used for GlabTop-2. This emphasizes the importance of the initialization assumptions. In greater detail, temperature conditions back then might not be suitable to trigger glacier melt in an appropriate magnitude as lower glacier bounds are simply located too high. Basically, there would be a mismatch between the link of glacier elevation and runoff generation. This becomes evident when examining areal losses as fraction of initial area of each glacier over the simulation period (see Appendix A4), where fractional area losses range from more than 10% (WG) to more than 40% (SCG). The Δh approach implemented in SWAT-GL does not consider glacier flow and does allow for glacier area growth (not be confused with accumulation in ice water equivalent of a specific ES) in its current version, which is relevant especially for long simulation periods with phases of ~~grwoth~~growth.

5 Conclusions

The recently extended version of SWAT, called SWAT-GL, was tested in representing the hydrology of four highly glaciated basins. The new SWAT-GL, which makes use of a physically-based glacier evolution routine has proven to provide robust hydrological simulations of catchments that are characterized by nivo-glacial processes. It thus serves its purpose and adds a valuable contribution to the hydrological modelling community, and in particular, the SWAT community.

We have also identified traditional model consistency issues prevalent in hydrological modeling and demonstrated their significance, even when multiple glacier and snow processes are included in the calibration procedure. Although SWAT-GL substantially improves model consistency, such problems should receive more attention. While we could show SWAT-GL's applicability even under long transient conditions, constraints remain and require further efforts to address. This is particularly true for climate impact studies, where simulation periods can exceed 100 years. In such studies, we advocate a minimum requirement that assesses the suitability of model components for climate impact statements to avoid flawed conclusions. Transient conditions, for example, could significantly affect degree-day factors, making initial choices inappropriate. A topic that is rarely addressed and discussed.

We identified parameter clustering at the edges of the initial parameter ranges, which indicate solutions that could impair physical plausibility. Moreover, contradictory patterns in the representation of snow and glacier processes (and discharge) were found. For example, a good representation of snow and glacier processes partly resulted in an unsatisfying representation of streamflow. We demonstrated that an adequate to good snow cover simulation does not necessarily lead to an accurate representation of glacier components. These basic insights, although partly recognized, go beyond SWAT-GL applications and are of general importance for the modeling community. The sensitivity analysis of SWAT-GL emerged a strong temperature-dependence of the model. This underpinned the importance and role of lapse rate parameterizations, also as a major source of uncertainty, in high mountain catchments. In addition, it was shown that the parameter space of glacier and snow-related parameters could be significantly reduced across all basins, suggesting potential applicability to other study areas.

Even though SWAT-GL was tested in catchments that are unprecedented in terms of data availability, the authors see no restrictions in its transferability to areas with poor data. Global datasets of ice thickness estimates and glacier outline set a suitable baseline to apply SWAT-GL. Although relatively small glaciated catchments were employed, the approach can be scaled up without imposing any substantial additional computational demand or physical limitation on the approach.

In conclusion, the most significant merit we discovered with SWAT-GL was its ability to adequately represent glacier processes in contrasting catchments. This encourages its further use in modeling glaciated and high-mountain catchments. However, there are technical limitations, including the requirement for introducing supplementary concepts to enhance the model's flexibility, along with structural limitations.

Appendix A

A1

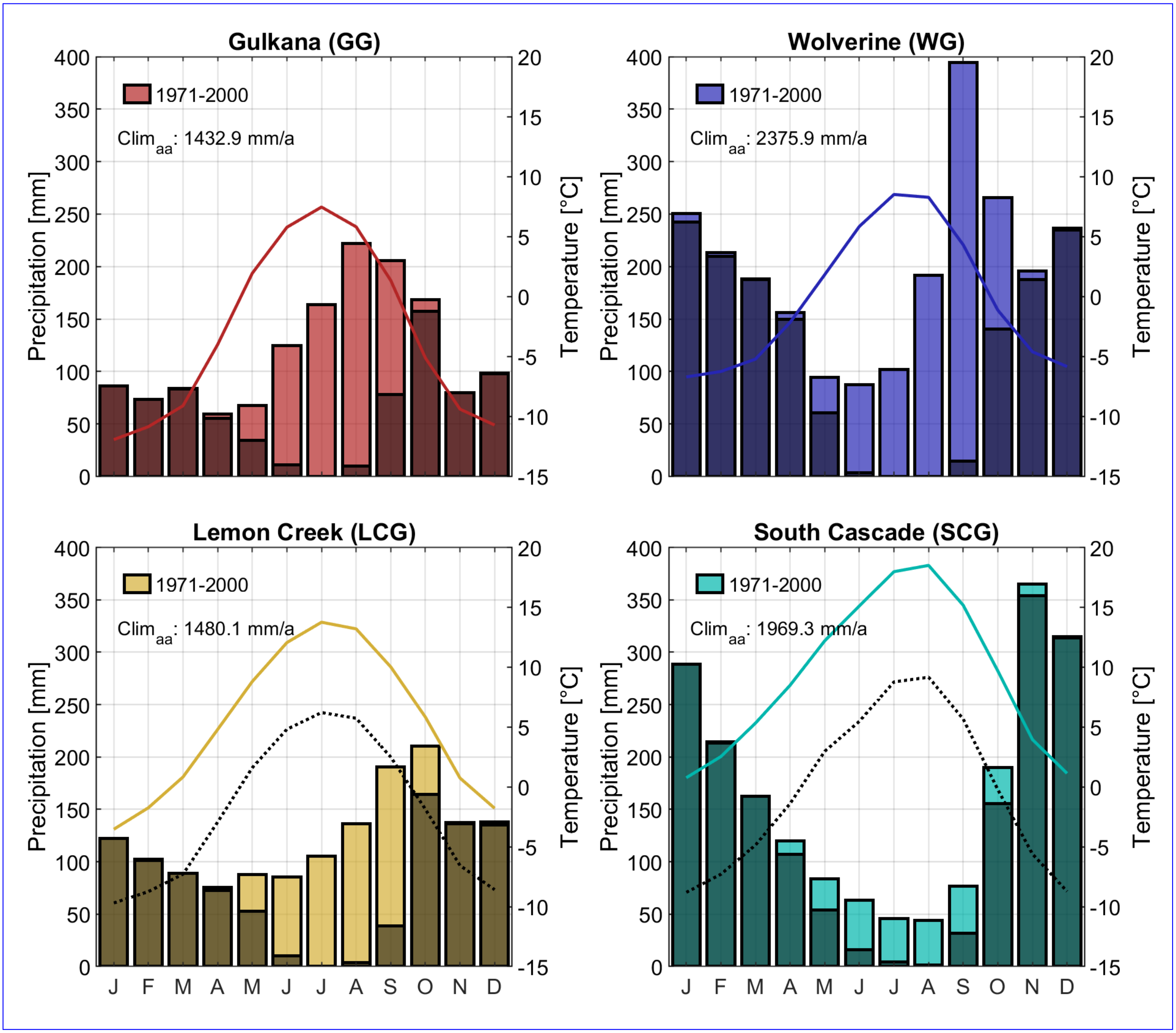

Figure 2. Overview of the mean climate regimes of the four glaciers according to the stations of ~~Tab.~~ Table 1 and the reference period 1971-2000. Lines refer to monthly mean temperatures and bars to mean monthly precipitation sums. ~~It has to be noted that no lapse rates were applied, what causes~~ As the ~~high monthly mean temperatures of~~ LCG and SCG ~~as~~ stations are remote and lower-elevated ~~stations were used~~ (20 and 40 km apart ~~, respectively~~and at 6 m and 270 m), black dotted lines representing temperature based on a -6.5 °C/km were provided additionally. The letters from J to D correspond to the months January to December. Clim$_{aa}$ refers to the annual average precipitation for the indicated periods. Black overlapping bars represent a solid precipitation proxy assuming that snowfall occurs when temperature is below 1 °C.

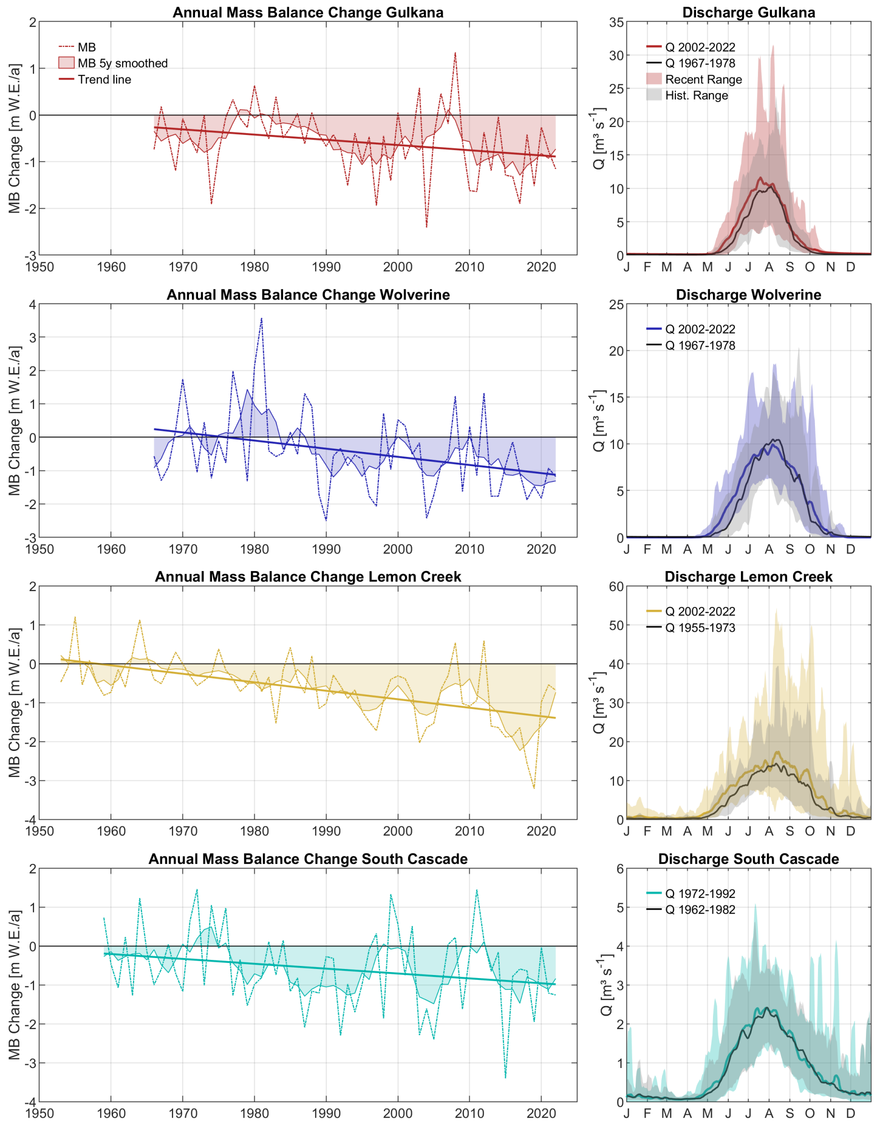

Figure 3. Overview of the annual mass balance rates of all glaciers merged with the mean daily discharge of two periods for each glacier. The recent periods refer to 2002-2022 (GG, WG, LCG) and 1972-1992 (LCG) and the older periods refer to 1967-1978 (GG, WG), 1955-1973 (LCG) and 1962-1982 (SCG) as indicated in the day of the year plots.

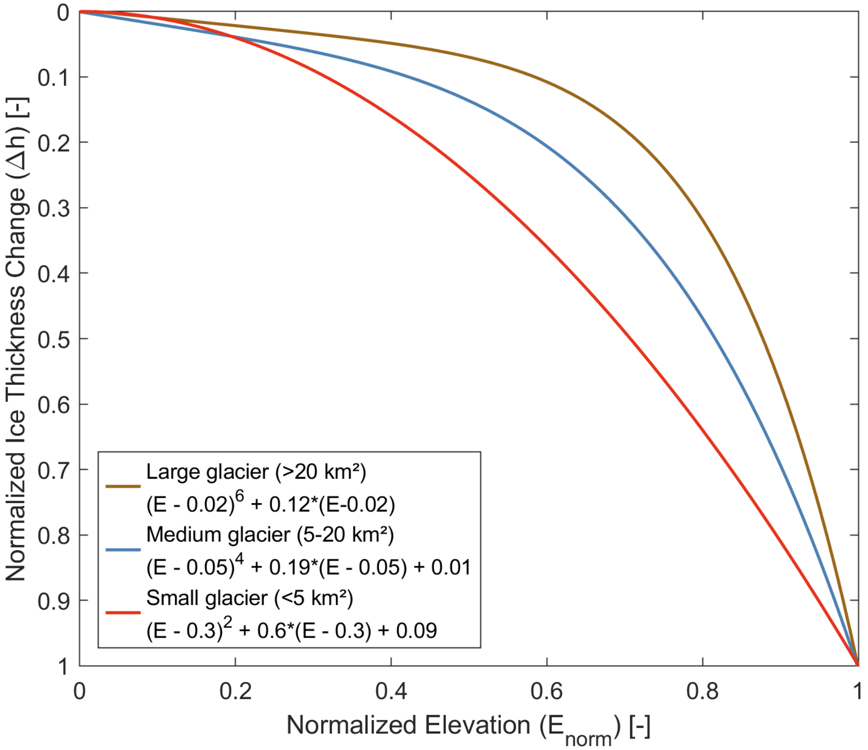

Figure 4. Empirical relationship of normalized glacier elevation and the normalized ice thickness change based on Huss et al. (2010).

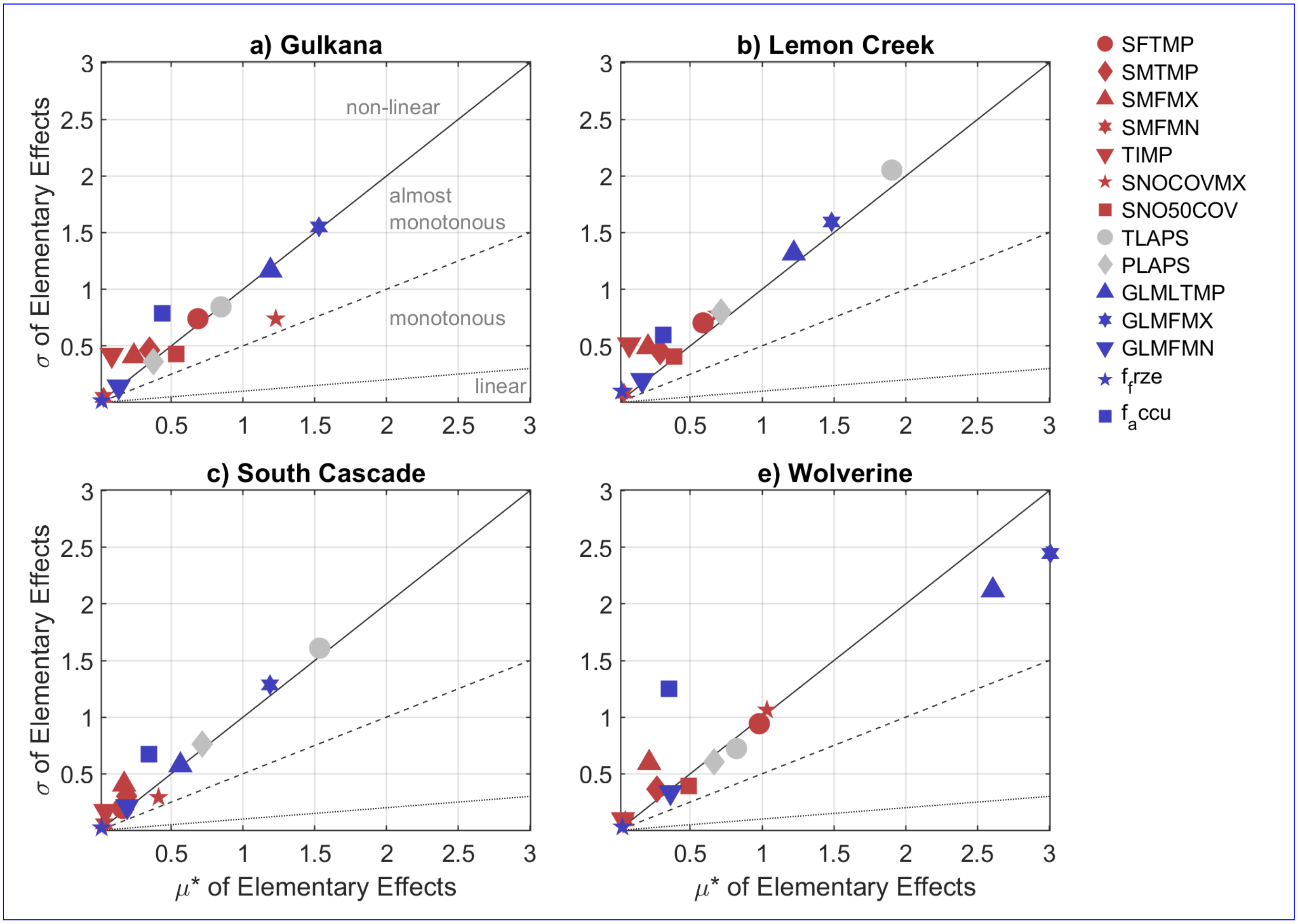

Figure 5. SA results based on the EE method for all four catchments and 14 parameters. The slopes (σ_{EE}/μ^*) of the different lines that classify parameter effects on the model outputs as linear, monotonous, almost monotonous and non-linear are: 0.1 (dotted line), 0.5 (dashed line), 1 (solid line). Red signs refer to snow parameters, gray indicate lapse rates and blue glacier parameters.

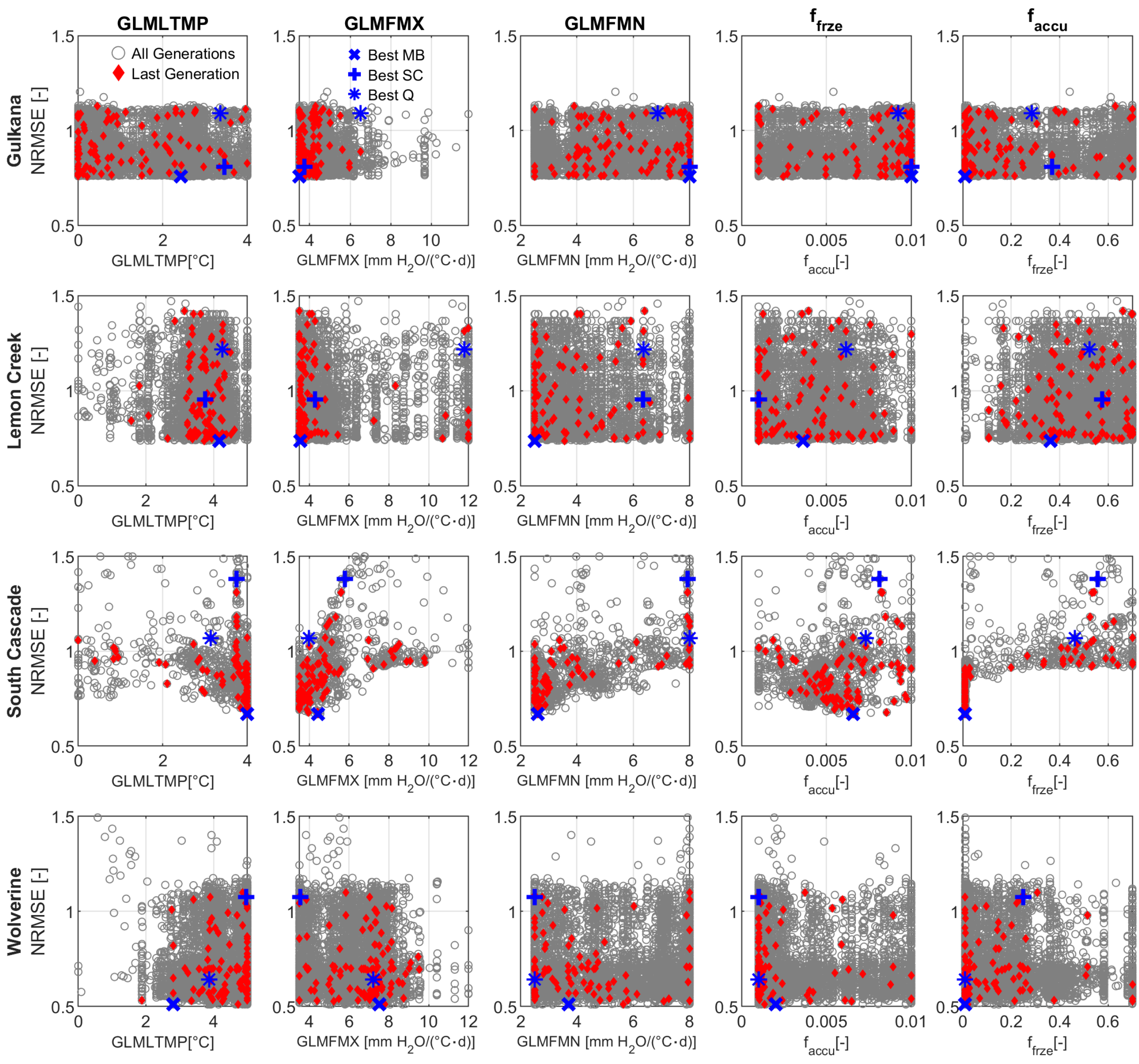

Figure 6. Parameter space illustrated for all glacier-related parameters for all generations (grey) and the final generation (red). The shown NRMSE values refer to the results of annual glacier mass balance simulations. The three blue symbols refer to the individual best solutions of mass balance, snow cover and (cross-validated) discharge in the last generation. Best hereby means the either the lowest NRMSE (for SC or MB) or the highest KGE (for discharge) within the last generation.

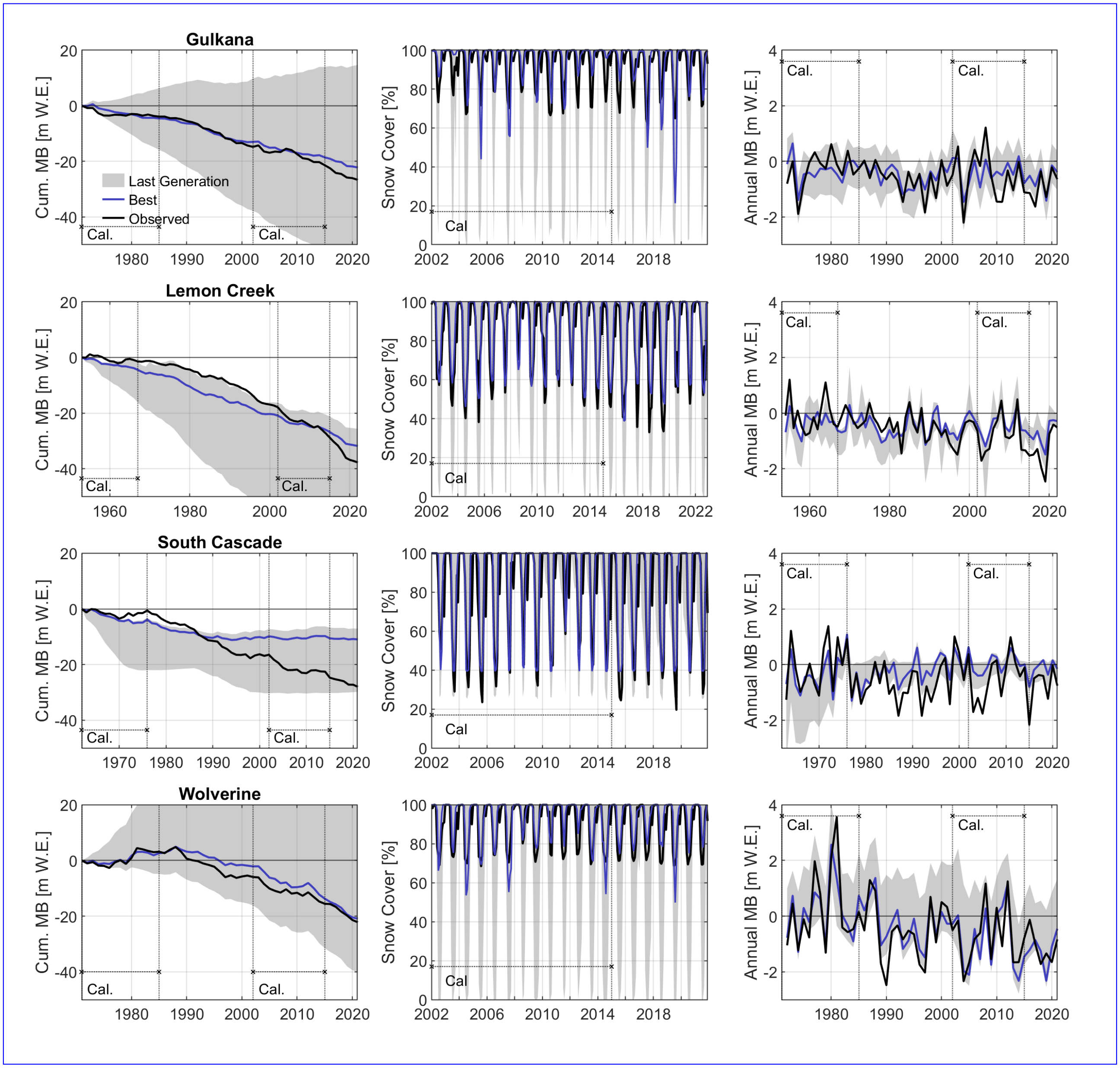

Figure 7. Simulation results of the last generation of the optimization procedure for the cumulative mass balance (Column 1), monthly snow cover (Column 2) and annual mass balance (Column 3). Each row corresponds to one glacier. Blue represents the best evaluation of the last generation for the respective variable, black refers to observations and grey shadings indicate the range of all evaluations part of the last generation. The dashed lines with the Cal. annotation indicates the individual calibration phases of each study area. The remainder of each time series was used for validation.

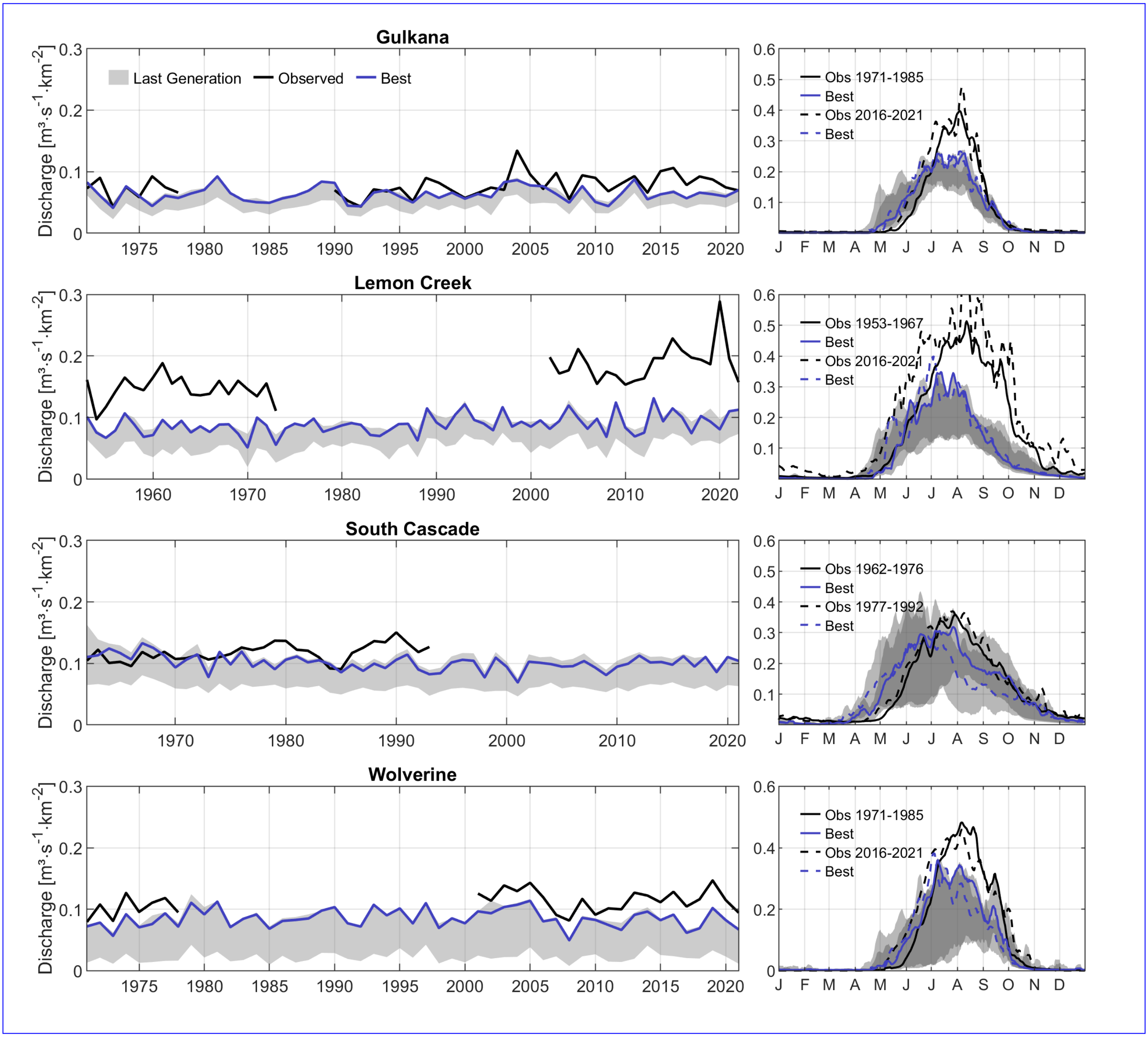

Figure 8. Simulation results for cross-validation of discharge in all four catchments based on mean annual flows (Column 1) and hydrological regimes (Column 2) at the end of the optimization. Hydrological regimes are represented by the mean of daily flows for each day of the year (from day 1 to 366) for the earliest available slice of validation period 1 (solid black line represents observation; solid blue line the best simulation; see also Table 3) and the latest available period of validation period 2 (dashed black line represents observation; dashed blue line the best simulation). Blue lines (simulation) cover the same period as their black counterparts.

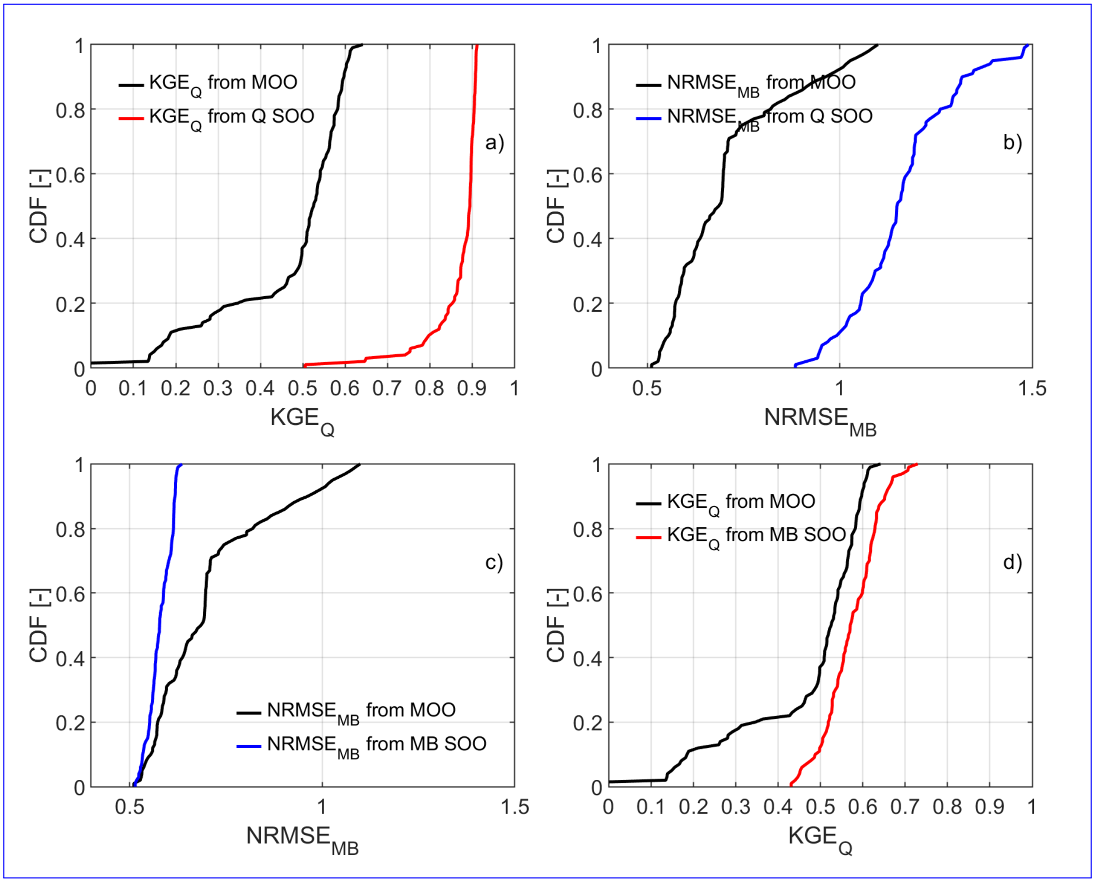

Figure 9. ~~Simulation results~~ Cumulative distribution function (CDF) plot of the objective functions for the ~~cross-validation~~ single-objective optimization (SOO) and multi-objective optimization (MOO) results of daily discharge ~~in all four catchments for mean~~ and annual ~~flows~~ mass balance for the Wolverine Glacier. a) and b) compare the MOO results with the SOO model optimized for discharge. c) and d) compare the MOO results with the SOO model optimized for mass balance. a) and c) show the results for the optimized variable of the SOO, b) and d) show how the corresponding non-optimized variable (~~Column 1~~MB for b) and ~~average flows~~ Q for d)) perform. The CDF of ~~each day of~~ the ~~year~~ MOO refer to the last generation (~~Column 2~~Generation = 100)~~at~~, while the ~~end~~ selected SOO results refer to generations at a relatively early stage of the optimization close to the point where convergence is reached. ~~Mean daily flows over~~ In detail, the ~~year (from day~~ selected generation of the discharge optimized model refers to ~~366) are averaged~~ generation 40, and generation 20 for the ~~earliest available slice of validation period 1 (solid black line represents observation; solid~~ mass balance optimized model. Black indicates the CDF from the MOO, red the one from the discharge SOO model and blue ~~line~~ the ~~best simulation) (see Tab~~mass balance SOO model. ~~3)~~ The subscripts Q and ~~the latest available slice~~ MB indicate discharge and mass balance, respectively. The sample size N of ~~validation period 2 (dashed black line represents observation; dashed blue line~~ the ~~best simulation)~~CDFs is 100 which refers to the general sample size in the optimization. ~~Blue lines (simulation) cover~~ Results focus on the ~~same~~ calibration period ~~as indicated for~~ of the ~~corresponding black lines~~WG.

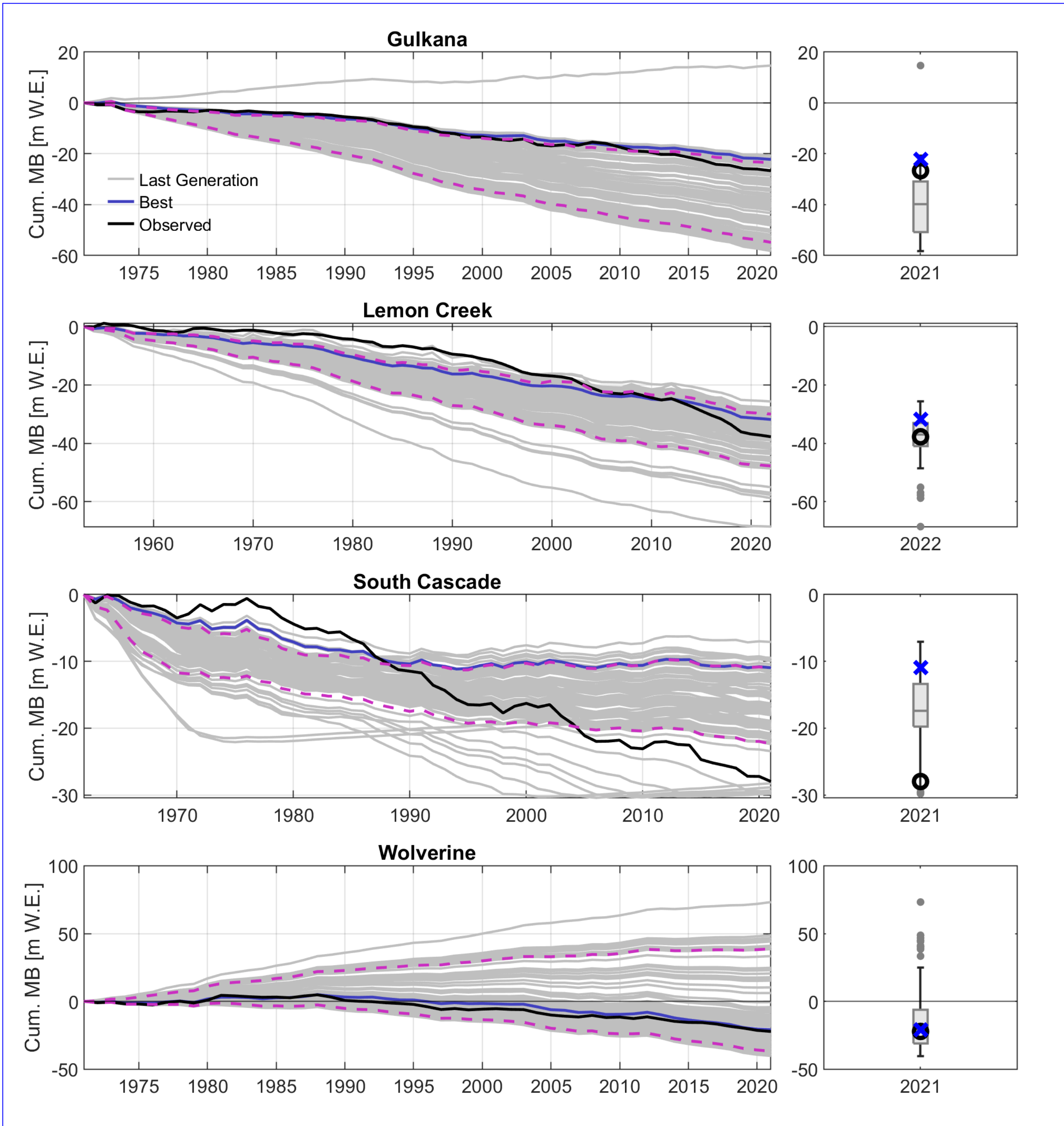

Figure A1. Similar to cumulative mass balance plot of Fig. 7 but providing all individual solutions of the final generation instead of the range. The boxplots provide an estimate of the distribution of the cumulative mass balance at the end of simulation period. For example the boxplot of GG consists of the 100 individual cumulative mass balance estimates of year 2021. Blue crosses indicate the results of the best annual mass balance representation in the optimization and black circle indicate the final observed value. Dashed purple lines refer to the 10th and 90th percentiles.

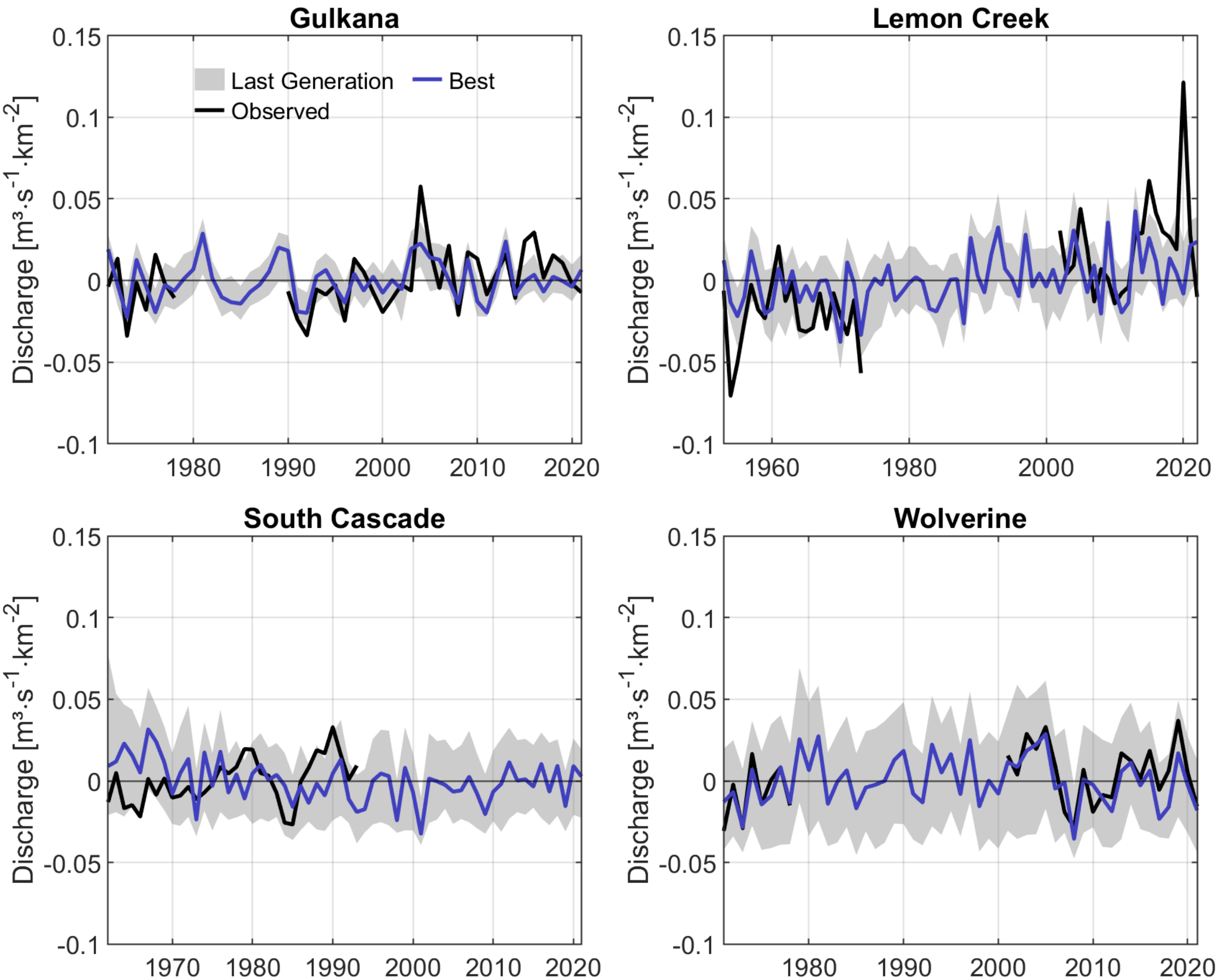

Figure A2. Simulation results for centered specific mean annual discharge.

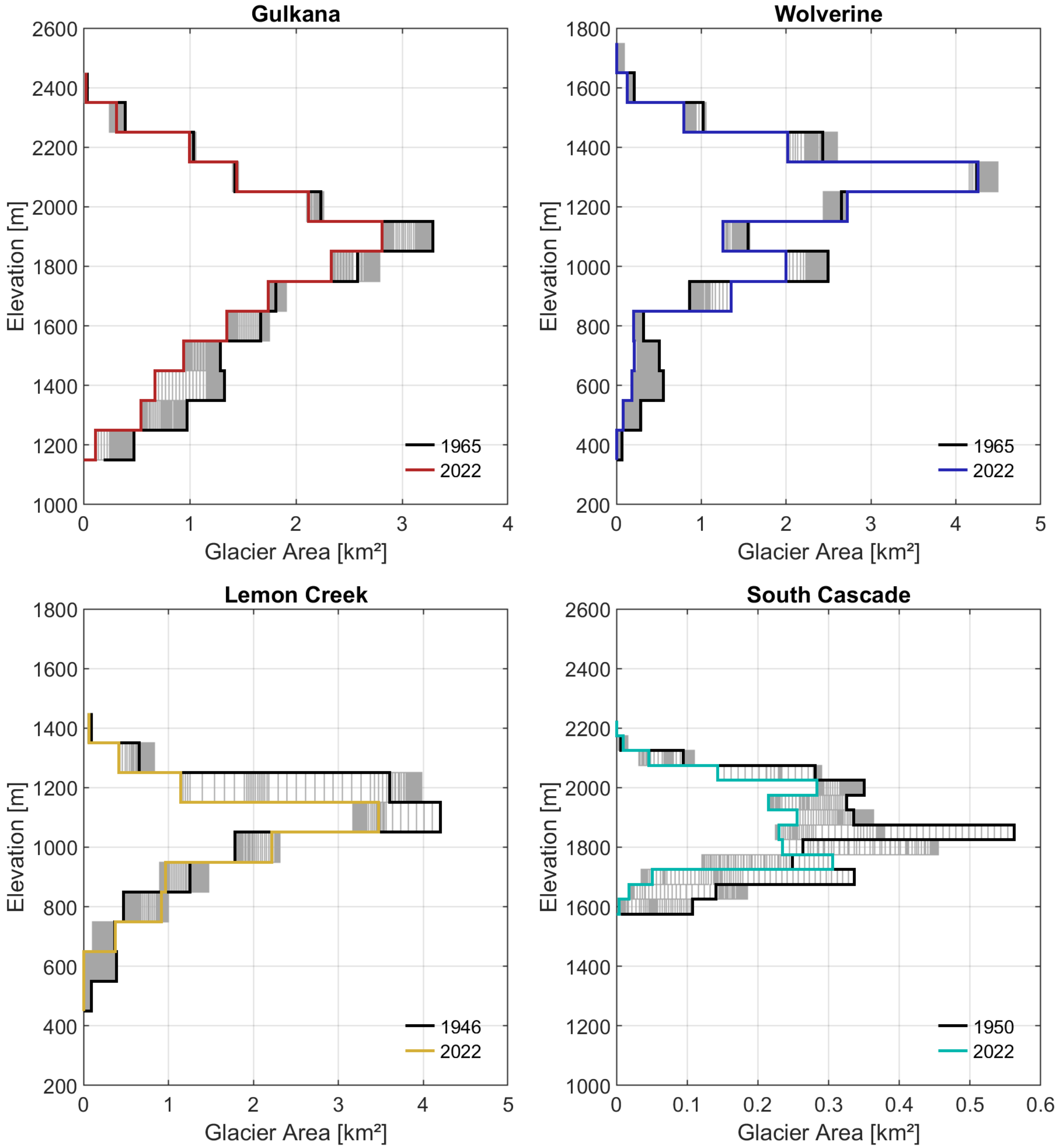

Figure A3. Annual glacier hypsometry observations for all four glaciers. Grey indicates each year where data was available while black represents the first available year and the individual colors the last available year of measurements.

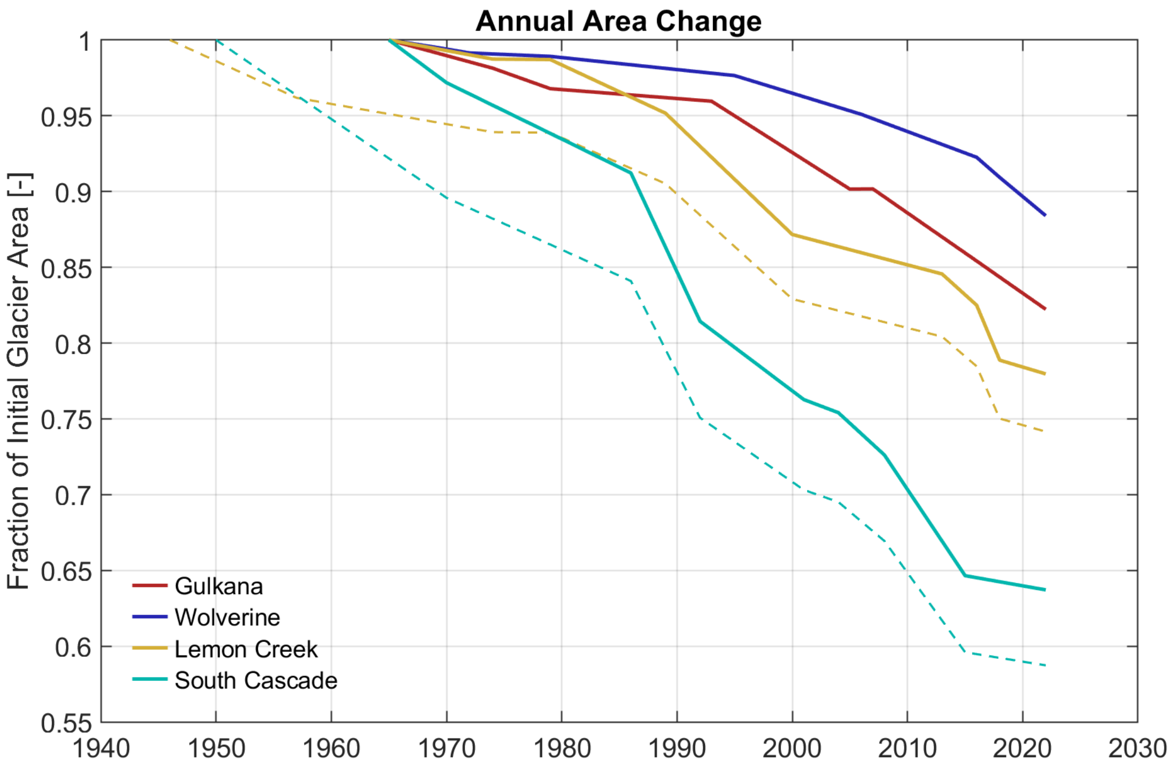

Figure A4. Observed annual glacier area observations of all glaciers expressed as fraction of initial area. Solid lines represent the earliest date of overlap between all glaciers and dashed lines represent the dates at which the mass balance measurement start in case it deviates from the starting point of the solid counterpart (only the case for SCG and LCG).

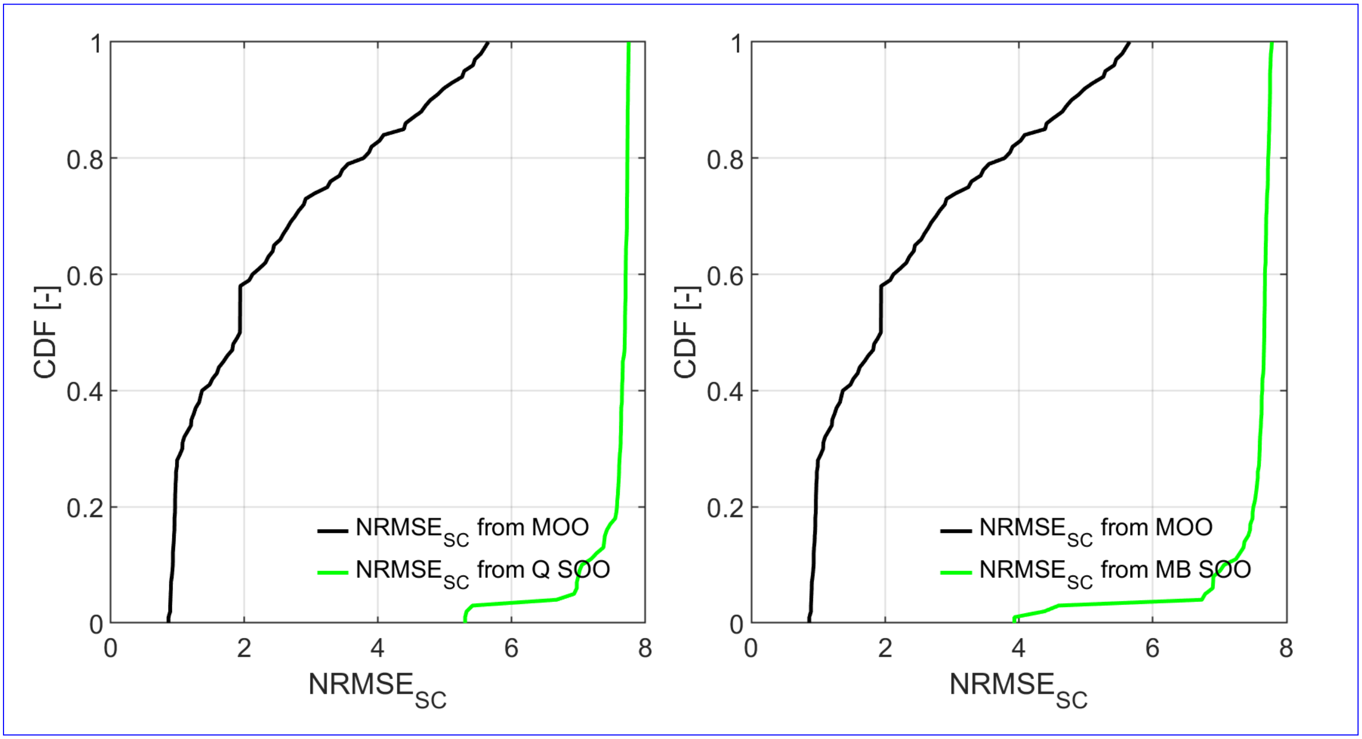

Figure A5. ~~Similar to cumulative mass balance plot of~~ Like Fig. ~~7~~ 9 but ~~providing all individual solutions of the final generation instead of the range. The boxplots provide an estimate of the distribution of the cumulative mass balance at the end of simulation period. In detail,~~ for ~~example at the GG the boxplot consists of the 100 individual cumulative mass balance estimates of year 2021. Blue crosses indicate the results of the best annual mass balance representation in the optimization and black circle indicate the final observed value~~snow cover.

Table A1. Parameter ranges and median values of all glaciers for the last generation of the optimization. Note: only the glacier parameters are shown here, which is the main novelty and purpose of the article.

Glacier	GLMLTMP			GLMFMX			GLMFMN			f_{frze}			f_{accu}		
	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median	Min	Max	Median
GG	0	4.00	0.70	3.50	6.51	4.05	2.50	8.00	5.76	0.001	0.01	0.0092	0.01	0.70	0.09
LCG	1.56	4.62	3.75	3.50	12.00	3.70	2.50	8.00	2.95	0.001	0.01	0.0033	0.01	0.70	0.45
SCG	0	4.00	3.92	3.50	12.00	4.40	2.50	8.00	2.74	0.001	0.01	0.0059	0.01	0.70	0.01
WG	1.88	5.00	4.49	3.50	9.53	7.13	2.50	8.00	3.30	0.001	0.01	0.001	0.01	0.70	0.03

Author contributions. TS, FM, FH, GC, MD, YT, LA, JM manuscript review and editing; TS, FM, MD study design; TS, FM, JM data acquisition; TS, FM, FH, JM data analysis and modelling; TS, FM, FH, GC, MD, YT, LA, JM model evaluation. TS manuscript draft.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. We gratefully acknowledge support from the United States Geological Survey (USGS) Benchmark Glacier Project for providing the comprehensive datasets. This work was supported by the EU Horizon ~~2020~~ Europe project WE-ACT (ID: 101083481). The authors also want to thank Scott Boyce from USGS for assisting the work.

References

Adnan, M., Kang, S., Zhang, G., Saifullah, M., Anjum, M. N., and Ali, A. F.: Simulation and Analysis of the Water Balance of the Nam Co Lake Using SWAT Model, Water, 11, 1383, https://doi.org/10.3390/w11071383, 2019.

Ali, S. H., Bano, I., Kayastha, R. B., and Shrestha, A.: COMPARATIVE ASSESSMENT OF RUNOFF AND ITS COMPONENTS IN TWO CATCHMENTS OF UPPER INDUS BASIN BY USING A SEMI DISTRIBUTED GLACIO-HYDROLOGICAL MODEL, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W7, 1487–1494, https://doi.org/10.5194/isprs-archives-xlii-2-w7-1487-2017, 2017.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., and Williams, J. R.: LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT, Journal of the American Water Resources Association, 34, 73–89, https://doi.org/10.1111/j.1752-1688.1998.tb05961.x, 1998.

Bahr, D. B., Pfeffer, W. T., and Kaser, G.: A review of volume-area scaling of glaciers, Reviews of Geophysics, 53, 95–140, https://doi.org/10.1002/2014rg000470, 2015.

Baker, E. H., Mcneil, C. J., Sass, L., Peitzsch, E. H., Whorton, E. N., Florentine, C. E., Clark, A. M., Miller, Z. S., Fagre, D. B., and O'Neel, S. R.: USGS Benchmark Glacier Mass Balance and Project Data, https://doi.org/10.5066/F7BG2N8R, 2018.

Campolongo, F., Saltelli, A., and Cariboni, J.: From screening to quantitative sensitivity analysis. A unified approach, Computer Physics Communications, 182, 978–988, https://doi.org/10.1016/j.cpc.2010.12.039, 2011.

Chen, Y., Li, W., Fang, G., and Li, Z.: Review article: Hydrological modeling in glacierized catchments of central Asia – status and challenges, Hydrology and Earth System Sciences, 21, 669–684, https://doi.org/10.5194/hess-21-669-2017, 2017.

Chiogna, G., Marcolini, G., Engel, M., and Wohlmuth, B.: Sensitivity analysis in the wavelet domain: a comparison study, Stochastic Environmental Research and Risk Assessment, https://doi.org/10.1007/s00477-023-02654-3, 2024.

Copernicus Climate Change Service: ERA5-Land hourly data from 2001 to present, https://doi.org/10.24381/CDS.E2161BAC, 2019.

Dawar, D. and Ludwig, S.: Differential evolution with dither and annealed scale factor, in: 2014 IEEE Symposium on Differential Evolution (SDE), IEEE, https://doi.org/10.1109/sde.2014.7031528, 2014.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6, 182–197, https://doi.org/10.1109/4235.996017, 2002.

Dozier, J.: Spectral signature of alpine snow cover from the landsat thematic mapper, Remote Sensing of Environment, 28, 9–22, https://doi.org/https://doi.org/10.1016/0034-4257(89)90101-6, 1989.

Du, X., Silwal, G., and Faramarzi, M.: Investigating the impacts of glacier melt on stream temperature in a cold-region watershed: Coupling a glacier melt model with a hydrological model, Journal of Hydrology, 605, 127 303, https://doi.org/10.1016/j.jhydrol.2021.127303, 2022.

Evin, G., Le Lay, M., Fouchier, C., Penot, D., Colleoni, F., Mas, A., Garambois, P.-A., and Laurantin, O.: Evaluation of hydrological models on small mountainous catchments: impact of the meteorological forcings, Hydrology and Earth System Sciences, 28, 261–281, https://doi.org/10.5194/hess-28-261-2024, 2024.

Farinotti, D., Huss, M., Fürst, J. J., Landmann, J., Machguth, H., Maussion, F., and Pandit, A.: A consensus estimate for the ice thickness distribution of all glaciers on Earth, Nature Geoscience, 12, 168–173, https://doi.org/10.1038/s41561-019-0300-3, 2019.

Frey, H., Machguth, H., Huss, M., Huggel, C., Bajracharya, S., Bolch, T., Kulkarni, A., Linsbauer, A., Salzmann, N., and Stoffel, M.: Estimating the volume of glaciers in the Himalayan–Karakoram region using different methods, The Cryosphere, 8, 2313–2333, https://doi.org/10.5194/tc-8-2313-2014, 2014.

Gan, R., Luo, Y., Zuo, Q., and Sun, L.: Effects of projected climate change on the glacier and runoff generation in the Naryn River Basin, Central Asia, Journal of Hydrology, 523, 240–251, https://doi.org/10.1016/j.jhydrol.2015.01.057, 2015.

Garcia Sanchez, D., Lacarrière, B., Musy, M., and Bourges, B.: Application of sensitivity analysis in building energy simulations: Combining first- and second-order elementary effects methods, Energy and Buildings, 68, 741–750, https://doi.org/10.1016/j.enbuild.2012.08.048, 2014.

Grusson, Y., Sun, X., Gascoin, S., Sauvage, S., Raghavan, S., Anctil, F., and Sáchez-Pérez, J.-M.: Assessing the capability of the SWAT model to simulate snow, snow melt and streamflow dynamics over an alpine watershed, Journal of Hydrology, 531, 574–588, https://doi.org/10.1016/j.jhydrol.2015.10.070, 2015.

Hamed, K. H. and Ramachandra Rao, A.: A modified Mann-Kendall trend test for autocorrelated data, Journal of Hydrology, 204, 182–196, https://doi.org/10.1016/s0022-1694(97)00125-x, 1998.

Hassan, J., qing Chen, X., Kayastha, R. B., and Nie, Y.: Multi-model assessment of glacio-hydrological changes in central Karakoram, Pakistan, Journal of Mountain Science, 18, 1995–2011, https://doi.org/10.1007/s11629-021-6748-9, 2021.

Hofmeister, F., Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., and Chiogna, G.: Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment, Journal of Hydrology X, 15, 100 123, https://doi.org/10.1016/j.hydroa.2022.100123, 2022.

Horlings, A.: A Numerical Modeling Investigation on Calving and the Recession of South Cascade Glacier, Tech. rep., https://doi.org/10.15760/honors.307, 2016.

Huss, M. and Hock, R.: A new model for global glacier change and sea-level rise, Frontiers in Earth Science, 3, https://doi.org/10.3389/feart.2015.00054, 2015.

Huss, M., Farinotti, D., Bauder, A., and Funk, M.: Modelling runoff from highly glacierized alpine drainage basins in a changing climate, Hydrological Processes, 22, 3888–3902, https://doi.org/10.1002/hyp.7055, 2008.

Huss, M., Jouvet, G., Farinotti, D., and Bauder, A.: Future high-mountain hydrology: a new parameterization of glacier retreat, Hydrology and Earth System Sciences, 14, 815–829, https://doi.org/10.5194/hess-14-815-2010, 2010.

Ji, H., Fang, G., Yang, J., and Chen, Y.: Multi-Objective Calibration of a Distributed Hydrological Model in a Highly Glacierized Watershed in Central Asia, Water, 11, 554, https://doi.org/10.3390/w11030554, 2019.

Li, H., Beldring, S., Xu, C.-Y., Huss, M., Melvold, K., and Jain, S. K.: Integrating a glacier retreat model into a hydrological model – Case studies of three glacierised catchments in Norway and Himalayan region, Journal of Hydrology, 527, 656–667, https://doi.org/10.1016/j.jhydrol.2015.05.017, 2015.

Linsbauer, A., Paul, F., and Haeberli, W.: Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach, Journal of Geophysical Research: Earth Surface, 117, n/a–n/a, https://doi.org/10.1029/2011jf002313, 2012.

Linsbauer, A, Paul, F, Hoelzle, M, Frey, H, and Haeberli, W: The Swiss Alps without glaciers – a GIS-based modelling approach for reconstruction of glacier beds, https://doi.org/10.5167/UZH-27834, 2009.

Luo, Y., Arnold, J., Liu, S., Wang, X., and Chen, X.: Inclusion of glacier processes for distributed hydrological modeling at basin scale with application to a watershed in Tianshan Mountains, northwest China, Journal of Hydrology, 477, 72–85, https://doi.org/10.1016/j.jhydrol.2012.11.005, 2013.

Luo, Y., Wang, X., Piao, S., Sun, L., Ciais, P., Zhang, Y., Ma, C., Gan, R., and He, C.: Contrasting streamflow regimes induced by melting glaciers across the Tien Shan – Pamir – North Karakoram, Scientific Reports, 8, https://doi.org/10.1038/s41598-018-34829-2, 2018.

Ma, C., Sun, L., Liu, S., Shao, M., and Luo, Y.: Impact of climate change on the streamflow in the glacierized Chu River Basin, Central Asia, Journal of Arid Land, 7, 501–513, https://doi.org/10.1007/s40333-015-0041-0, 2015.

Mcneil, C. J., Sass, L., Florentine, C., Baker, E. H., Peitzsch, E. H., Whorton, E. N., Miller, Z., Fagre, D. B., Clark, A. M., and O'Neel, S. R.: Glacier-Wide Mass Balance and Compiled Data Inputs: USGS Benchmark Glaciers, https://doi.org/10.5066/F7HD7SRF, 2016.

Merchán-Rivera, P., Geist, A., Disse, M., Huang, J., and Chiogna, G.: A Bayesian framework to assess and create risk maps of groundwater flooding, Journal of Hydrology, 610, 127 797, https://doi.org/10.1016/j.jhydrol.2022.127797, 2022.

Millan, R., Mouginot, J., Rabatel, A., and Morlighem, M.: Ice velocity and thickness of the world's glaciers, Nature Geoscience, 15, 124–129, https://doi.org/10.1038/s41561-021-00885-z, 2022.

Moriasi, D. N., Arnold, J. G., Liew, M. W. V., Bingner, R. L., Harmel, R. D., and Veith, T. L.: Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations, Transactions of the ASABE, 50, 885–900, https://doi.org/10.13031/2013.23153, 2007.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33, 161–174, https://doi.org/10.1080/00401706.1991.10484804, 1991.

N. Moriasi, D., W. Gitau, M., Pai, N., and Daggupati, P.: Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria, Transactions of the ASABE, 58, 1763–1785, https://elibrary.asabe.org/abstract.asp?aid=46548&t=3, 2015.

NASA JPL: NASA Shuttle Radar Topography Mission Global 1 arc second, https://doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003, 2013.

Naz, B. S., Frans, C. D., Clarke, G. K. C., Burns, P., and Lettenmaier, D. P.: Modeling the effect of glacier recession on streamflow response using a coupled glacio-hydrological model, Hydrology and Earth System Sciences, 18, 787–802, https://doi.org/10.5194/hess-18-787-2014, 2014.

Nossent, J., Elsen, P., and Bauwens, W.: Sobol' sensitivity analysis of a complex environmental model, Environmental Modelling & Software, 26, 1515–1525, https://doi.org/https://doi.org/10.1016/j.envsoft.2011.08.010, 2011.

O'Neel, S., Hood, E., Arendt, A., and Sass, L.: Assessing streamflow sensitivity to variations in glacier mass balance, Climatic Change, 123, 329–341, https://doi.org/10.1007/s10584-013-1042-7, 2014.

O'Neel, S., McNeil, C., Sass, L. C., Florentine, C., Baker, E. H., Peitzsch, E., McGrath, D., Fountain, A. G., and Fagre, D.: Reanalysis of the US Geological Survey Benchmark Glaciers: long-term insight into climate forcing of glacier mass balance, Journal of Glaciology, 65, 850–866, https://doi.org/10.1017/jog.2019.66, 2019.

Pesci, M. H., Schulte Overberg, P., Bosshard, T., and Förster, K.: From global glacier modeling to catchment hydrology: bridging the gap with the WaSiM-OGGM coupling scheme, Frontiers in Water, 5, https://doi.org/10.3389/frwa.2023.1296344, 2023.

Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, Applied Statistics, 28, 126, https://doi.org/10.2307/2346729, 1979.

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environmental Modelling & Software, 79, 214–232, https://doi.org/10.1016/j.envsoft.2016.02.008, 2016.

Pradhananga, N. S., Kayastha, R. B., Bhattarai, B. C., Adhikari, T. R., Pradhan, S. C., Devkota, L. P., Shrestha, A. B., and Mool, P. K.: Estimation of discharge from Langtang River basin, Rasuwa, Nepal, using a glacio-hydrological model, Annals of Glaciology, 55, 223–230, https://doi.org/10.3189/2014aog66a123, 2014.

RGI Consortium: Randolph Glacier Inventory - A Dataset of Global Glacier Outlines, Version 6, https://doi.org/10.7265/4M1F-GD79, 2017.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

Sarrazin, F., Pianosi, F., and Wagener, T.: Global Sensitivity Analysis of environmental models: Convergence and validation, Environmental Modelling & Software, 79, 135–152, https://doi.org/10.1016/j.envsoft.2016.02.005, 2016.

Schaefli, B. and Huss, M.: Integrating point glacier mass balance observations into hydrologic model identification, Hydrology and Earth System Sciences, 15, 1227–1241, https://doi.org/10.5194/hess-15-1227-2011, 2011.

Schaffhauser, T., Tuo, Y., Hofmeister, F., Chiogna, G., Huang, J., Merk, F., and Disse, M.: SWAT-GL: A new glacier routine for the hydrological model SWAT, JAWRA Journal of the American Water Resources Association, https://doi.org/10.1111/1752-1688.13199, 2024.

Seibert, J., Vis, M. J. P., Kohn, I., Weiler, M., and Stahl, K.: Technical note: Representing glacier geometry changes in a semi-distributed hydrological model, Hydrology and Earth System Sciences, 22, 2211–2224, https://doi.org/10.5194/hess-22-2211-2018, 2018.

Shafeeque, M., Luo, Y., Wang, X., and Sun, L.: Altitudinal Distribution of Meltwater and Its Effects on Glacio-Hydrology in Glacierized Catchments, Central Asia, JAWRA Journal of the American Water Resources Association, 56, 30–52, https://doi.org/10.1111/1752-1688.12805, 2019.

Shannon, S. R., Payne, A., Freer, J., Coxon, G., Kauzlaric, M., Kriegel, D., and Harrison, S.: A snow and glacier hydrological model for large catchments – case study for the Naryn River, Central Asia, https://doi.org/10.5194/hess-2022-51, 2022.

Sin, G. and Gernaey, K. V.: Improving the Morris method for sensitivity analysis by scaling the elementary effects, in: Computer Aided Chemical Engineering, pp. 925–930, Elsevier, https://doi.org/10.1016/s1570-7946(09)70154-3, 2009.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, Journal of Hydrology, 523, 739–757, https://doi.org/10.1016/j.jhydrol.2015.02.013, 2015.

Stoll, E., Hanzer, F., Oesterle, F., Nemec, J., Schöber, J., Huttenlau, M., and Förster, K.: What Can We Learn from Comparing Glacio-Hydrological Models?, Atmosphere, 11, 981, https://doi.org/10.3390/atmos11090981, 2020.

Storn, R. and Price, K.: Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces, Journal of Global Optimization, 23, 1995.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K. B., Hurtt, G., Kriegler, E., Lamarque, J.-F., Meehl, G., Moss, R., Bauer, S. E., Boucher, O., Brovkin, V., Byun, Y.-H., Dix, M., Gualdi, S., Guo, H., John, J. G., Kharin, S., Kim, Y., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, 2021.

Tiel, M., Stahl, K., Freudiger, D., and Seibert, J.: Glacio-hydrological model calibration and evaluation, WIREs Water, 7, https://doi.org/10.1002/wat2.1483, 2020.

Tuo, Y., Duan, Z., Disse, M., and Chiogna, G.: Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy), Science of The Total Environment, 573, 66–82, https://doi.org/10.1016/j.scitotenv.2016.08.034, 2016.

Tuo, Y., Marcolini, G., Disse, M., and Chiogna, G.: A multi-objective approach to improve SWAT model calibration in alpine catchments, Journal of Hydrology, 559, 347–360, https://doi.org/10.1016/j.jhydrol.2018.02.055, 2018.

U.S. Geological Survey: USGS Water Data for the Nation, https://doi.org/10.5066/F7P55KJN, 1994.

Vanuytrecht, E., Raes, D., and Willems, P.: Global sensitivity analysis of yield output from the water productivity model, Environmental Modelling & Software, 51, 323–332, https://doi.org/https://doi.org/10.1016/j.envsoft.2013.10.017, 2014.

Wagener, T.: On the Evaluation of Climate Change Impact Models for Adaptation Decisions, p. 33–40, Springer International Publishing, https://doi.org/10.1007/978-3-030-86211-4_5, 2022.

Wang, X., Zhang, Y., Luo, Y., Sun, L., and Shafeeque, M.: Combined use of volume-area and volume-length scaling relationships in glacio-hydrological simulation, Hydrology Research, 49, 1753–1772, https://doi.org/10.2166/nh.2018.137, 2018.

Wiersma, P., Aerts, J., Zekollari, H., Hrachowitz, M., Drost, N., Huss, M., Sutanudjaja, E. H., and Hut, R.: Coupling a global glacier model to a global hydrological model prevents underestimation of glacier runoff, Hydrology and Earth System Sciences, 26, 5971–5986, https://doi.org/10.5194/hess-26-5971-2022, 2022.

Wilcoxon, F.: Individual Comparisons by Ranking Methods, Biometrics Bulletin, 1, 80, https://doi.org/10.2307/3001968, 1945.

Wortmann, M., Bolch, T., Krysanova, V., and Buda, S.: Bridging glacier and river catchment scales: an efficient representation of glacier dynamics in a hydrological model, https://doi.org/10.5194/hess-2016-272, 2016.

Wu, F., Zhan, J., Wang, Z., and Zhang, Q.: Streamflow variation due to glacier melting and climate change in upstream Heihe River Basin, Northwest China, Physics and Chemistry of the Earth, Parts A/B/C, 79-82, 11–19, https://doi.org/10.1016/j.pce.2014.08.002, 2015.

Yang, C., Xu, M., Fu, C., Kang, S., and Luo, Y.: The Coupling of Glacier Melt Module in SWAT+ Model Based on Multi-Source Remote Sensing Data: A Case Study in the Upper Yarkant River Basin, Remote Sensing, 14, 6080, https://doi.org/10.3390/rs14236080, 2022.

Zekollari, H., Huss, M., Farinotti, D., and Lhermitte, S.: Ice-Dynamical Glacier Evolution Modeling—A Review, Reviews of Geophysics, 60, https://doi.org/10.1029/2021rg000754, 2022.