Snow accumulation over glaciers in the Alps, Scandinavia, Central Asia and Western Canada (1981-2021) inferred from climate reanalyses and machine learning

Matteo Guidicelli[1], Matthias Huss[1,2,3], Marco Gabella[4], and Nadine Salzmann[5,6]

[1]Department of Geosciences, University of Fribourg, Fribourg, Switzerland
[2]Laboratory of Hydraulics, Hydrology and Glaciology (VAW), ETH Zurich, Zurich, Switzerland
[3]Swiss Federal Institute for Forest, Snow and Landscape Research (WSL), Birmensdorf, Switzerland
[4]Federal Office of Meteorology and Climatology MeteoSwiss, Locarno-Monti, Switzerland
[5]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[6]Climate Change, Extremes and Natural Hazards in Alpine Regions Research Center CERC, Davos, Switzerland

Correspondence: Matteo Guidicelli (matteo.guidicelli@unifr.ch)

Abstract. Reanalysis products for remote high-mountain regions provide estimates of snow precipitation. However, this data is inherently uncertain and assessing a potential bias is difficult due to the typically very low quantity and quality of available in-situ observations, limiting their reliability to evaluate long-term effects of climate change. In this study, we make use of a hardly used source of direct snow precipitation observations at high elevations. We use the winter mass balance data of 95 glaciers distributed over the Alps, Western Canada, Central Asia and Scandinavia, and compare them with the total precipitation from the ERA-5 and the MERRA-2 reanalysis products during the snow accumulation seasons from 1981 until today. We propose a machine learning model to adjust the precipitation of reanalysis products to the elevation of the glaciers, and consequently to derive snow water equivalent (SWE) estimates over glaciers without ground observations and to fill spatial and temporal observational data gaps. We use a gradient boosting regressor (GBR), which combines several meteorological variables from the reanalyses (e.g. air temperature, relative humidity) with topographical parameters. These GBR-derived estimates are evaluated against the winter mass balance data using on the one hand independent glaciers (site-independent GBR) and on the other hand independent accumulation seasons (season-independent GBR). Both approaches resulted in reduced biases and increased correlation between the precipitation of the original reanalyses and the winter mass balance data of the glaciers. Generally, the GBR models have also shown a good representation of the spatial (vertical elevation intervals) and temporal (years) variability of the winter mass balance on individual glaciers. The SWE estimates that our GBR models are able to provide can thus be used to significantly improve the calibration of glaciological and hydrological models in observation-scarce regions.

1 Introduction

Climate change considerably alters the high-mountain cryosphere (e.g. Beniston, 2012; Vorkauf et al., 2021; Marty, 2008; Beniston et al., 2018; Bormann et al., 2018). Vanishing glaciers and changes in the seasonal snow regime are leading to reduced water storage capacity in worldwide high-mountain regions, impacting adjacent lowlands far away (e.g. Viviroli et al.,

2007; Immerzeel et al., 2020). Cryospheric hazards such as slope failures and glacier lake outburst floods (e.g. Gobiet et al., 2014; Rasul and Molden, 2019) are other impacts of climate change in mountain regions, which are typically felt by mountain society (e.g. Adger et al., 2003; Hock et al.). However, the elevation dependency of precipitation trends is unclear: precipitation trends in station observations are often inconsistent with no systematic changes with elevation, while precipitation increases in gridded datasets are weaker at higher elevations (e.g. Pepin et al., 2022). It is thus crucial to improve our understanding of the local climate-cryosphere interaction in order to implement appropriate adaptation strategies (e.g. Stone et al., 2013; Salzmann et al., 2014; Huss et al., 2017; Barandun et al., 2020). However, at the local scale and particularly at very high altitudes, snow (and precipitation) in-situ observations are typically very scarce, spatially not optimally distributed, with low temporal resolution or time series are short or with important gaps because of technical challenges, difficult accessibility and thus complicated and lavish maintenance (e.g. Beniston et al., 2012; Tapiador et al., 2012). This is an important limitation for studies focusing on the long-term effects of climate change on the snowpack, which require snow accumulation data covering decadal periods (e.g. Seiz et al., 2010). Thus, the further development of techniques to spatially and/or temporally transfer the available observational series between sites and/or filling data gaps, is critical and urgently needed (e.g. Salzmann et al., 2014).

Understanding the precipitation regime and snow accumulation dynamics at highest elevations of mountain ranges is crucial, as – despite their relatively limited areal extent – these regions act as water towers, releasing vital amounts of water in the form of glacier and snow melt during the dry season, i.e. when it is most urgently needed (e.g. Immerzeel et al., 2020). Due to the scarcity and inherent problems of conventional precipitation gauges in high-mountain environments dominated by snow and strong winds (e.g. Sevruk et al., 2009), snow water equivalent (SWE) at very high elevations is mostly only measured as the cumulative snow accumulation on glaciers. Measurements of SWE on glaciers are typically used for the determination of winter mass balance (Cogley et al., 2011), an important variable in international glacier monitoring (e.g. Zemp et al., 2013). The main process of snow accumulation is the total precipitation received by the glacier during the accumulation season. Since melting is often negligible during this time period, SWE on glaciers represents a reliable measure of local winter precipitation and was thus used for a comparison with precipitation products in different studies (e.g. Gugerli et al., 2020; Guidicelli et al., 2021). However, other processes such as deposition of hoar, freezing rain or snow drift caused by winds and avalanching can also influence the accumulation (Dadic et al., 2010; Gascoin et al., 2013).

Currently, the correct partitioning of mass fluxes on glaciers puts limits on the usefulness of many large-scale remote sensing data sets: whereas, the volume change of glaciers over time can be inferred with remote sensing observations (e.g. Hugonnet et al., 2021), the unambiguous calibration of glaciological and hydrological models requires the two components of glacier mass balance (accumulation and melt), which cannot be constrained with remote observations. The improvement of snow accumulation estimates in observation-scarce regions is thus highly relevant.

Worldwide spatio-temporally continuous information on precipitation, snow depth and SWE is also provided by climate reanalyses that merge physical laws with the assimilated satellite and ground observations (e.g. Hersbach et al., 2020; Gelaro et al., 2017). However, the performance of reanalysis results can vary greatly depending on the region and the elevation range of interest (Sun et al., 2018). Large biases in reanalysis precipitation are particularly observed in high-mountain regions (e.g. Liu and Margulis, 2019; Zandler et al., 2019). The scarcity of observations available for assimilation and the coarse resolution of

such models limit their accuracy in areas of complex topography and their suitability for studies at a local scale (e.g. Salzmann and Mearns, 2012, (for snow)).

As a result, downscaling of precipitation estimates of reanalyses is necessary to represent the local conditions in high-mountain regions. Different statistical and dynamical downscaling methods exist (cf. Maraun et al., 2010), which has also been employed and evaluated over glacierized regions (e.g. Mölg and Kaser, 2011). For instance, Liston and Elder (2006) developed a quasi–physically based, meteorological model to produce high-resolution (30 m to 1 km horizontal grid) atmospheric forcings for several variables, where the precipitation adjustment is a nonlinear function of the elevation difference between the grid and the point of interest. The same equation was used by Gupta and Tarboton (2016), who proposed an approach to downscale the MERRA (Rienecker et al., 2011) variables. They obtained a Nash-Sutcliffe efficiency greater than 0.70 for downscaled monthly precipitation at 173 SNOTEL (Snow Telemetry) sites. Fiddes and Gruber (2014) adapted this method for the Swiss Alps by including a climatological parameter based on the Alpine precipitation data set provided by the Climatic Research Unit (gridded monthly precipitation totals at 10 arc-min resolution over the Alps, for the period 1800–2003). Their product allowed improving the purely lapse-rate-based approach of Liston and Elder (2006), obtaining a correlation coefficient of 0.6 (versus 0.5) against the annual precipitation observed at 40 ANETZ (MeteoSwiss automatic meteorological network) stations. Recently, machine learning methods have demonstrated their high performance to statistically downscale reanalyses (and global climate models) estimates of precipitation and other meteorological variables, from sub-daily and daily (e.g. Serifi et al., 2021; Wang et al., 2021) to monthly and seasonal (e.g. Sachindra et al., 2018; Najafi et al., 2011; Sun and Tang, 2020) resolution. However, downscaling methods for snow (and precipitation) are rarely assessed at very high elevations, mainly due to the scarcity of ground observations. Consequently, long-term effects of climate change on the snowpack at very high elevations are not well understood yet (e.g. Seiz et al., 2010).

In this study, we thus aim at providing improved observation-independent SWE estimates at highest elevations of different mountain ranges across the Earth. In order to achieve this goal, we develop and evaluate a machine learning approach based on gradient boosting regressor (GBR) models (see Friedman, 2001) to adjust the total precipitation of reanalysis dataset (ERA-5 and MERRA-2) over the accumulation season. The GBR model is then used to derive SWE estimates over glaciers in the Alps, Scandinavia, Central Asia and Western Canada. Data on snow accumulation distribution at the end of the accumulation season covering a period of up to 41 years from 95 glaciers were used to train our approach. More specifically, the GBR models aim at allowing the spatio-temporal transferability of the learned information over the 95 glaciers to other glaciers with no ground observations and/or filling gaps of observational series.

2 Study sites and data

The study was conducted on 95 glaciers located in the Alps, Scandinavia, Central Asia and Western Canada (Fig. 1), where the longest time series and the highest density of winter glacier mass balance data are available. In the following, we describe the different data sources used in the study.

2.1 Reanalysis data

We used data from ERA-5 and MERRA-2 reanalyses since they are among the currently most used reanalysis products with the highest spatial resolution, covering the longest time period in all the regions of our study.

2.1.1 ERA-5

ERA-5 is the fifth generation of the European Centre for Medium-Range Weather Forecasts atmospheric reanalyses of the global climate (see Hersbach et al., 2020, for more information). In this study, we used several variables from the ERA5 hourly data on single levels from 1979 to present (Hersbach et al., 2018b), and the ERA5 hourly data on pressure levels from 1979 to present (Hersbach et al., 2018a), all with a spatial resolution of 0.25° x 0.25° (∼30 km). All variables were resampled on a daily timescale before usage. The list of variables selected for the analysis is reported in Table B1. The ERA-5 precipitation variable used in the study is "tp" (total precipitation) from the ERA-5 single levels.

2.1.2 MERRA-2

MERRA-2 is the second version of the Modern-Era Retrospective Analysis for Research and Applications (see Gelaro et al., 2017, for more information). In this study, we used several variables from the MERRA-2 Land Surface Diagnostics (Global Modeling and Assimilation Office (GMAO), 2015b), the MERRA-2 Single-Level Diagnostics (Global Modeling and Assimilation Office (GMAO), 2015c) and the MERRA-2 Analyzed Meteorological Fields (Global Modeling and Assimilation Office (GMAO), 2015a). All variables have a spatial resolution of 0.5° x 0.625° (∼50 km), and we resampled them on a daily timescale before usage. The list of the selected variables is reported in Table B2. The MERRA-2 precipitation variable used in the study is "PRECTOTLAND" (total precipitation) from the MERRA-2 Land Surface Diagnostics.

2.2 Winter mass balance data

The World Glaicer Monitoring Service (WGMS) compiles and publishes standardized observations on changes in mass, volume, length and area of glaciers collected by national monitoring programmes and local observers around the world (glacier fluctuations (see Zemp et al., 2021, for more details)).

We used the winter mass balance data separated per elevation intervals (EE-MASS-BALANCE data sheet in WGMS, 2020) and we refer to them as B_w in this study. Point observations are also available (EEE-MASS-BALANCE POINT data sheet in WGMS, 2020) but are not used in this study because of the smaller amount of glaciers with complete information reported (observation dates, elevation, coordinates). We only considered the B_w data where the elevation interval is indicated in the WGMS database. The glacier area related to each elevation interval was also used to weight the B_w data. In addition, we considered the average slope and aspect of the glaciers by using the information provided in the Randolph Glacier Inventory version 6 (RGI Consortium, 2017).

The winter mass balance is the result of the balance between the gain of snow which accumulates over the glacier, as well as refreezing of liquid water within the snowpack, and the loss caused by melting and sublimation over the accumulation season.

Other processes such as snow drift caused by winds can also influence the accumulation. However, the amount of snow melt is typically minor compared to the snow accumulation and is thus neglected in the comparison with the precipitation totals performed in this study. Snow accumulation is expressed in SWE (e.g. Østrem and Brugman, 1991), which is calculated by multiplying the measured snow depth with the respective bulk density of the snowpack. The snow depth is typically measured with a snow probe or ground-penetrating radar, while the snow density is usually measured in snow pits or by coring and is subsequently extrapolated to all observations on a glacier. The WGMS database only provides information on SWE but does not generally allow tracing whether density was directly measured or not.

The B_w data used in this study correspond to the mean winter balance for the glacier area contained in the respective elevation interval. Various spatial extrapolation techniques were applied by the national observers to infer elevation band average snow accumulation from the (sparse) point observations, which can be challenging due to important local-scale variability in snow depth (e.g. Dadic et al., 2010; Helfricht et al., 2014; Sold et al., 2016). Unfortunately, the WGMS database does not allow tracing the methods used, hence, resulting in an uncertainty that is difficult to be estimated. Often, no direct snow depth and density observations are available at the most extreme elevations of the glaciers because of high surface slopes and difficult accessibility. The employed techniques in the framework of the Swiss national program GLAMOS (Glacier Monitoring Switzerland, providing the data for the Swiss sites to the WGMS) are described in Huss et al. (2021). The impact of the inter- and extrapolation of direct SWE measurements acquired on glaciers to obtain B_w data used in this study on our results is discussed in Section 5.2.5.

The starting date of the accumulation season is not precisely known and is often determined with a stratigraphic system (since the date of the minimum surface in the previous summer) (e.g. Mayo et al., 1972; Cogley et al., 2011). The date of the minimum surface varies between the years and also across the glacier, in fact, the snow accumulation starts typically later at lower elevations than at higher elevations (Huss et al., 2009). However, in this study we used a unique starting date for the entire glacier according to the information provided in the WGMS database. The end of the season is determined by the day of the snow survey that is indicated in the WGMS database. In this study, we cumulated precipitation amounts over the accumulation season. The impact of the date considered as beginning of the accumulation season on our results is discussed in Section 5.2.5.

3 Methods

First, we derived total or average of all variables provided by the reanalyses for the entire accumulation season. Subsequently, a machine learning model to adjust the total precipitation (see Sec. 2.1.1 and 2.1.2) of the reanalyses over glaciers for the accumulation season was developed to derive SWE estimates. We use a GBR (gradient boosting regressor), which makes use of several meteorological variables (original and downscaled) and topographical parameters as input variables (predictors). In principle, a different adjustment factor of precipitation might be needed depending on the precipitation phase. However, as we only adjust the total precipitation occurring during the accumulation season, the adjustment factors used here represent the "average" adjustment factor of all precipitation events. Moreover, the snowfall variable was used as a predictor in order

to enable the GBR model to learn that a different "average" adjustment factor must be applied depending on the fraction of snowfall and total precipitation (i.e. depending on the main precipitation phase during the accumulation season).

The applied methods to downscale other meteorological variables used by the GBR model are described below.

3.1 Downscaling temperature and relative humidity

In addition to the original variables, the GBR requires some downscaled variables of the reanalyses as predictors at the glacier elevation, including air temperature, dew point temperature and relative humidity (for MERRA-2 and ERA-5), vertical velocity of air motion (for ERA-5 only) and specific humidity (for MERRA-2 only). The downscaling procedure was applied at a daily resolution using a linear interpolation between the values of the two closest pressure levels to the elevation of the B_w data of the glaciers. This downscaling approach is illustrated in the Supplementary material (Fig. S1).

If information regarding the relative humidity was not directly provided by the reanalyses, we applied approaches presented by Liston and Elder (2006) and Gupta and Tarboton (2016) to derive it. The applied equations are described in the Appendix (Section A).

3.2 Downscaling precipitation

In order to derive downscaled precipitation estimates over the glaciers, we built a machine learning model and we applied a pre-existing lapse-rate-based approach that we considered as a benchmark. The approaches are described below and the results obtained with the two methods are compared afterward.

3.2.1 Benchmark

Liston and Elder (2006) proposed a lapse-rate-based approach to downscale reanalysis' precipitation by accounting for the elevation difference between the point of interest and the grid of the reanalysis. Whereas they applied the approach to the MERRA reanalysis, we applied the same approach to MERRA-2 and ERA-5 reanalysis data and used the resulting adjusted precipitation as a benchmark:

$$P_{\mathrm{adj}} = P_{\mathrm{reanalysis}} \frac{1 + \kappa(H_{\mathrm{point}} - H_{\mathrm{reanalysis}})}{1 - \kappa(H_{\mathrm{point}} - H_{\mathrm{reanalysis}})}, \tag{1}$$

where $P_{\mathrm{reanalysis}}$ is the precipitation of the reanalysis, $H_{\mathrm{reanalysis}}$ is the elevation of the reanalysis' grid, P_{adj} is the adjusted precipitation at the altitude of the point of interest (H_{point}) and κ is a monthly adjustment factor (cf. table 1 of Liston and Elder, 2006). In our study, we used an average factor $\kappa = 0.3214$, corresponding to the average between October and April. The precipitation adjusted with this approach on Findelgletscher is illustrated in Fig. S1.

3.2.2 GBR model (Gradient Boosting Regressor)

In order to represent a potential non-monotonic increase of snow accumulation (and precipitation) with elevation and to provide different adjustments of the original reanalysis' precipitation depending on the region and the site, we built more complex

models based on machine learning. All the models are built with the open source "scikit-learn" library for machine learning in Python (cf. Pedregosa et al., 2011). Specifically, we built a series of GBR models, each one consisting of an ensemble of weak learning models (estimators) represented by regression trees. In our case, the goal of the GBR models is to predict the logarithmic adjustment factors of the reanalysis' precipitation with respect to the B_w on glaciers (Eq. 2). In a GBR, the trees are built sequentially, and the subsequent trees learn from the errors of the previous trees, minimizing the residuals between their predictions and the reference values.

$$F_{\mathrm{dB,ref}} = 10 \log_{10} \frac{B_{\mathrm{w}}}{P_{\mathrm{reanalysis,tot}}}, \tag{2}$$

where $P_{\mathrm{reanalysis,tot}}$ is the total precipitation of the reanalysis over the accumulation season. The GBR models aim at minimizing the cost function defined in Equation 3, corresponding to the mean squared error between the predicted and reference logarithmic adjustment factors.

$$MSE_{\mathrm{dB^2}} = \frac{1}{n} \sum_{i=1}^{n} (F_{\mathrm{dB,pred},i} - F_{\mathrm{dB,ref},i})^2 \tag{3}$$

Different hyperparameters characterize a GBR. In this study, we applied a grid search to optimize the number of estimators (number of additive trees), the maximum depth that each tree can reach, the minimum number of samples required to be at a leaf node of a tree, and the learning rate. A 10-fold cross-validation was applied with different combinations of hyperparameters. The hyperparameters that were able to minimize the mean squared error of the validation data were chosen. Finally, the GBR model with the chosen hyperparameters was tested on independent data.

The validation and the test data were defined differently depending on the goal of the GBR model. For both reanalysis products (ERA-5 and MERRA-2), we built two different GBR models with two different goals and two different cross-validation and test schemes. The first GBR model is site-independent and aims at "extrapolating" the B_w data in time and space (over glaciers with no B_w data). Thus, groups of data (folds) in the 10-fold cross-validation contain data of different glaciers and the site-independent GBR model with the chosen hyperparameters was tested on an independent glacier. This process was repeated for each glacier, which was used to test the GBR-model defined with the data of the other glaciers (see Fig. 2). The second GBR model is season-independent and aims at "extrapolating" the B_w data in time only (filling data gaps over glaciers with discontinuous records of B_w). For these cases, groups of data in the 10-fold cross-validation contain data of different years but different groups can contain data of different years of the same glacier. Finally, the season-independent GBR model with the chosen hyperparameters was tested for an independent year of a given glacier. This process was repeated for each year and each glacier.

The average optimal hyperparameters for all the studied glaciers are reported in Tab. 1. The resulting site-independent model is more generalized (since no information regarding the glacier where the model is validated and tested was provided), while the season-independent model is more detailed and can split into individual sub-models adapted to a small number of samples (since it can exploit the B_w data of the tested glacier).

All the variables presented in Section 2.1 and listed in Section B of the Appendix were considered by the GBR as predictors (separately for ERA-5 and MERRA-2). In addition, we derived and used the differences between the downscaled variables

(cf. Section 3.1) and the estimates at the grid of the reanalysis. Some variables were not only averaged considering all days in the accumulation season, but also considering only the days with a relevant amount of precipitation, here arbitrarily set to 5 mm. The GBRs also use the latitude and longitude of the glacier (unique value for the entire glacier), as well as the aspect and slope of the glacier (unique value for the entire glacier). A summary of the predictors used by the GBRs is reported in Table B3. Furthermore, the B_w data were weighted by considering the area of the glacier related to the respective elevation interval. Larger glaciers (and elevation intervals related to larger areas) thus receive more weight in the training of our models than smaller glaciers (and elevation intervals related to smaller areas).

3.3 Evaluation metrics for the models' estimates

3.3.1 Adjustment factors

In order to evaluate the bias between the B_w data and the estimates of the models (original reanalyses, benchmark or GBR), we computed the adjustment factor f (dimensionless) as:

$$f = \frac{B_\mathrm{w}}{E_\mathrm{model}}, \tag{4}$$

where E_model is the estimate of a model. The adjustment factor F_dB is expressed in decibels and is used to derive supplementary evaluation metrics:

$$F_\mathrm{dB} = 10 \log_{10} \frac{B_\mathrm{w}}{E_\mathrm{model}} \tag{5}$$

3.3.2 Glacier-wide means

When deriving a glacier-wide factor (or glacier-wide B_w) for a single accumulation season, we computed a weighted mean of the area contained in the individual elevation intervals. These seasonal glacier-wide values were used to derive the Pearson's correlations (r), the root mean squared error (RMSE), and the fraction of standard error (FSE), between the glacier-wide B_w and the model estimate. The FSE corresponds to the RMSE divided by the glacier-wide B_w.

3.3.3 Regional metrics

In order to further validate the performance of the GBR models, we derived the glacier-wide F_dB described by Equation 5 for every accumulation season and every glacier with B_w data ($F_{\mathrm{dB,\,mean}}$). We thus analyzed the four investigated regions separately by deriving a mean factor per region, as:

$$F_{\mathrm{dB,\,region}} = \frac{\sum_{g=1}^{n} \sum_{s=1}^{m_g} F_{\mathrm{dB,mean},g,s}}{n \sum_{g=1}^{n} m_g}, \tag{6}$$

where n is the number of glaciers and m_g is the number of accumulation seasons with B_w data for the glacier g.

4 Results

In the following, we first present the analysis of the predictors' importance in the GBR models (Sec. 4.1), followed by the main results of our study, i.e. the performance of the GBR models over glaciers in the Alps, Scandinavia, Central Asia and Western Canada (Sec. 4.2).

4.1 Importance of predictors in the GBR models

In order to understand the importance of the predictors used by the GBR models (i.e. those not related to the elevation of the glaciers and their elevation difference with the reanalysis' grid), we evaluated the changes in terms of overall GBR model performance when suppressing groups of predictors. For both ERA-5 and MERRA-2 site-independent GBR models, the smallest RMSE results when using all predictors (Fig. 3a and b). The RMSE particularly increases when suppressing the MERRA-2 single level and pressure levels variables from the predictors. In turn, for both ERA-5 and MERRA-2 season-independent GBR models, the smallest RMSE results when suppressing the single level and pressure levels variables from the predictors (Fig. 3c and d). The RMSE increases most when suppressing the year, the topographical parameters and the glaciers coordinates simultaneously as predictors.

However, skipping reanalysis variables from the set of predictors leads to higher errors for some individual glaciers, especially in the representation of the temporal variability of the snow accumulation. In fact, excluding the reanalysis variables, the year is the only predictor that could be used to predict a different adjustment factor depending on the accumulation season (all the other predictors are constant in time). Therefore, and to allow a fairer comparison between site-independent and season-independent GBRs, in all our following analyses we always include all predictors.

In order to infer the importance of the predictors for the individual study regions, we built an individual GBR for each region. We furthermore performed a principal component analysis (PCA) considering the ten predictors most frequently used by the GBRs for each region. In the Alps, lower factors between B_w and ERA-5 precipitation result at lower latitudes, and the glaciers affected by 100 m westerly winds (negative u component of the wind speed) have generally higher factors than those affected by easterly winds (Supplementary Fig. S2a). In Scandinavia, we notice a cluster of five glaciers with smaller ERA-5 factors and higher downscaled temperatures during precipitation events (Fig. S2c). In Central Asia, the glaciers' aspect is the predictor that most clearly discriminates between high and low factors between B_w and both ERA-5 and MERRA-2 precipitation. Glaciers with $\sim$North-facing slopes show smaller ERA-5 factors and $\sim$East-facing slopes higher MERRA-2 factors (Fig. S2e and f). In Western Canada, lower ERA-5 factors correlate with larger precipitation amounts and lower elevation of the glaciers, while MERRA-2 factors are clearly lower at higher latitudes, which are characterized by stronger southerly winds at 850 hPa (Fig. S2g and f).

4.2 Performance of the GBR models

Overall, the GBR models indicate better agreement in terms of bias, spatial and temporal correlation with the B_w data than the original reanalyses and the benchmark for the majority of the studied glaciers. In the following we report in detail on the comparison between the B_w data, the precipitation estimates of the reanalyses and the GBR models' estimates.

4.2.1 Glacier-wide reanalysis' bias adjustment

Figure 4 shows the comparison between all glacier-wide B_w values and the models' estimates. MERRA-2 precipitation underestimates B_w more importantly than ERA-5 precipitation in all regions (Fig. 4a and b), with an overall RMSE of 946 mm (mean absolute error (MAE) of 749 mm) against 793 mm (611 mm) of ERA-5. Excluding the Alps, the correlation between the B_w data and the ERA-5 precipitation is always higher than the correlation with the MERRA-2 precipitation. The adjusted estimates obtained with the site-independent and the season-independent GBRs allowed us to consistently reduce (increase) the bias (Pearson correlation (r)) between the precipitation of the original reanalyses and B_w (overall RMSE of 433 mm, MAE of 326 mm, r of 0.86 for the MERRA-2 site-independent GBR; RMSE of 410 mm, MAE of 307 mm, r of 0.87 for the ERA-5 site-independent GBR; RMSE of 293 mm, MAE of 211 mm, r of 0.94 for the MERRA-2 season-independent GBR; RMSE of 275 mm, MAE of 200 mm, r of 0.94 for the ERA-5 season-independent GBR). These results demonstrate the need of an adjustment of reanalyses data to reproduce snow accumulation on glaciers, which are, otherwise, largely underestimated in all four regions involved in this study.

In order to make an in depth analysis of the model performance, we also derived a glacier-wide factor between the B_w data and reanalysis-based models' estimates (Eq. 4) for each accumulation season and for each site separately (Fig. 5). By comparing Fig. 5b and 5c, it is clear that, in Central Asia, the factors for adjusting the MERRA-2 reanalysis' precipitation are much larger than the factors for the ERA-5 precipitation. The benchmark method overestimates the B_w for many glaciers in the Alps (both ERA-5 and MERRA-2) and several glaciers in Central Asia (ERA-5). The site-independent and especially, the season-independent GBRs are better scaled with respect to the B_w data than the original reanalyses and the benchmarks. In general, the variability of the factors for each glacier is strongly affected by the number of available accumulation seasons with B_w data (Fig. 5a). A lower variability is usually observed for glaciers with a small number of seasons with B_w data.

Figure 6 shows the mean regional factor between the B_w data and the models' estimates as a function of the accumulation seasons from 1981 to 2019. It indicates that the original reanalyses clearly underestimate snow accumulation on glaciers, except for ERA-5 in Central Asia, where, as a consequence, the benchmark overestimates B_w. However, temporal variations in the mean regional bias are also affected by the considered set of glaciers that fluctuates over the analyzed years. In the Alps, we observe increasing biases of the original reanalyses in recent years, where a much larger number of glaciers is available. In Scandinavia, the bias of MERRA-2 and ERA-5 is similar and all the models are generally not able to remove it completely. In, Central Asia, there is a tendency for all models to yield lower adjustment factors before the 2000s than afterwards. However, this has to be interpreted with care, because only one glacier was considered between 2002 and 2014. The continuity of the available B_w data in Western Canada is too limited to analyze temporal changes in the adjustment factors.

In order to evaluate the robustness of the GBR models to reduce the glacier-wide bias of the reanalysis, we performed a temporal and spatial validation of their predictions (Fig. 7). The performance of the season-independent models improves when using more accumulation seasons in the training data (Fig. 7a, c, e and g). Training the models with more than 20 seasons, however, does not seem to further improve performance consistently. The performance of the site-independent models is constant because they are never trained with B_w data of the tested glacier. When no B_w data of the tested glacier is used to train the season-independent models (as for the site-independent models), their performance is worse than the site-independent models, confirming the importance of a specific optimization scheme depending on the goal of the model.

As also expected, the performance of the site-independent models decreases when data of neighbouring glaciers are excluded from the training (Figs. 7b, d, f and h). The highest impact is on the performance of the MERRA-2 site-independent GBR in Central Asia. Overall, the bias of the site-independent GBR models remains comparable to the bias obtained with the benchmark method even when excluding all other glaciers located within 1000 km from the training. For the season-independent models, we always kept the B_w data of the tested glacier in the training data, and only excluded the other glaciers. This explains why the season-independent models perform better and are less sensitive to the removal of neighbouring glaciers from the training process.

The good performance of the GBRs in terms of bias suggests that they can be used for SWE estimates over glaciers where no ground observations are available (site-independent GBRs) and for filling data gaps of the recorded observations (season-independent GBRs). However, the performance generally decreases when the glacier is not in proximity to the glaciers used to train the GBR models. Furthermore, we assume that the resulting performance strongly depends on the characteristics of the glacier with respect to the glaciers used in the training. The results indicate moreover that the season-independent GBRs outperform the site-independent GBRs to reduce the bias against B_w data, especially in regions with a limited number of glaciers with snow accumulation data. In conclusion, filling data gaps is much simpler than estimating SWE on glaciers with no observations.

4.2.2 Spatial snow accumulation variability on individual glaciers

In order to evaluate the ability of the GBR models to reproduce the spatial variability of the snow accumulation over individual glaciers, we compared the vertical profiles of B_w to the estimates of the models. For Rhonegletscher for instance (Alps, Fig. 8a), both site-independent and season-independent GBRs are able to represent the shape of the vertical profile of the B_w, which is characterized by an increasing B_w until 3350 m a.s.l. and a more stable/decreasing B_w in the upper part of the glacier. This vertical profile cannot be reproduced by using the benchmark approach, where, by definition, the precipitation is monotonically increasing with the elevation. Decreasing B_w is also clearly indicated in the upper part of Abramov glacier (Central Asia, Fig. 8b) in 1992. As suggested by the point observations reported, this is certainly the result of extrapolating to elevation ranges not or only poorly covered with data. However, this has a limited influence on the GBR models than the lower part of the glacier, as it received higher weights because of the larger areas (see Sec. 2.2). The total precipitation estimated by the original reanalyses was well reproduced compared with the B_w. The site-independent GBRs are not able to adjust the precipitation by consistently reducing the bias with B_w. On the other hand, the season-independent GBRs are able to better fit the altitudinal distribution of

B_w. In this case, we observe that the maximum B_w coincides with the maximum downscaled MERRA-2 relative humidity. In the case of Storglaciären (Scandinavia, Fig. 8c), B_w is underestimated by the benchmark, while the GBR models (the season-independent especially) are able to represent the steep increase of snow accumulation over the glacier. In the case of Sykora glacier (Western Canada, Fig. 8d), all GBR models show a good agreement with B_w data. By comparing the coefficients of variation, it is clear that the season-independent GBRs are able to reproduce better the amplitude of the spatial variability of the B_w than the site-independent GBRs (see Supplementary Tab. S1). Furthermore, the correlations demonstrate that the GBRs outperform the benchmark method to reproduce the B_w of almost all glaciers of this study (Tab. S1).

4.2.3 Temporal snow accumulation variability on individual glaciers

In general, the GBR models show a better performance in reproducing the relative changes of B_w among individual years for the same glacier than the original reanalysis (see Tab. 2). The correlation between the GBR models' estimates and the B_w over the years is often much higher than for the original reanalysis. The level of significance of the correlation between the original ERA-5 or/and MERRA-2 improves when the GBR models are applied, especially for Goldbergkees, Graasubreen, and Ts. Tuyuksuyskiy glacier. However, in some cases we still have low correlations, indicating that the models are not suitable to represent the temporal variability. Furthermore, although the season-independent GBRs are the best models to reduce the bias, the relative changes among years are sometimes better explained by the site-independent GBRs. In fact, the number of years of data of the tested glacier used to train the season-independent GBRs does not seem to impact his performance in term of temporal correlation with the B_w data (see Supplementary Fig. S3). The number of years with available B_w data is typically much smaller in Western Canada and Central Asia than in the Alps and Scandinavia (see Fig. 5a), therefore, we could not robustly evaluate the ability of the models to represent the temporal variability of the B_w data for these regions.

5 Discussion

The GBR models developed, evaluated and presented in this study showed a better overall agreement in terms of bias, spatial and temporal correlation with the B_w data than the original reanalyses and the benchmark (lapse-rate-based approach described in Sec. 3.2.1) for the majority of the studied glaciers in the Alps, Scandinavia, Central Asia and Western Canada. In the following, we provide a comprehensive discussion of the approach and the results.

5.1 Advantages and disadvantages of gradient boosting regressors

5.1.1 Differences with lapse-rate-based approaches

With the exception of some specific sites, our GBR models outperformed the benchmark method (lapse-rate-based approach (Sec. 3.2.1)) in the Alps, Scandinavia, Central Asia and Western Canada regarding the reduction of the bias against glacier-wide B_w data (Figs. 5 and 6). This suggests that complex models such as our GBRs are needed to adjust reanalysis to different glaciers sites, which can be characterized by different topographical and climatic conditions, and where the performance of

reanalysis' estimates can vary greatly depending on the region (e.g. Sun et al., 2018). In fact, (independent) B_w data was used to train our GBR models, allowing the GBRs to learn specific characteristics of actual snow accumulation on glaciers, and to transfer them to unknown sites (site-independent GBRs) and unknown seasons (season-independent GBRs).

The GBR models also outperform the benchmark to reproduce the spatial variability of the snow accumulation on individual glaciers. B_w data indicate decreasing SWE in the uppermost sections of many glaciers which may be attributed to preferential snow deposition redistribution processes, caused by the interplay between snow, wind and the generally steep topography (e.g. Sold et al., 2016; Gerber et al., 2019). The ability of GBR models to model non-linear relationships allows a better representation of the vertical profiles of snow accumulation than the benchmark method (Fig. 8, Tab. S1). In fact, the observed spatial variability of B_w could not be reproduced with the benchmark method, which by definition cannot represent decreasing values with the elevation (cf. Eq. 1).

Both, the GBR models and the benchmark do not require direct ground observations to be applied. However, the performance of the GBR models is influenced by the amount of data used to train the models and strongly depends on the characteristics of the glacier with respect to the glaciers used to train the models.

5.1.2 Differences with other machine learning algorithms

We have chosen a tree-based algorithm because of its higher readability in terms of the predictors' usage compared to other machine learning methods (e.g. Huysmans et al., 2011; Freitas, 2014). A disadvantage of tree-based algorithms, however, could be that this approach does not predict continuous values. Yet here, we aim at predicting an adjustment factor depending on a classification based on the used predictors, which is exactly the purpose of a tree-based algorithm. The choice of a gradient boosting instead of other tree-based algorithms (e.g. random forest (Breiman, 2001)) is motivated by the fact that gradient boosting is a gradient descent algorithm, where each additional tree tries to reduce the bias (which is the main goal of our study) rather than the variance of the predictions.

5.2 Impact of the data used to train the GBR models

5.2.1 Site-independent and season-independent GBRs

The lower generalization of the season-independent GBRs compared with the site-independent GBRs allows the splitting into individual sub-models adapted to a small number of samples (see Tab. 1). This enables to exploit the B_w data of the tested glacier by creating a specific sub-model, but can result in an overfit of the training data. In contrary, the higher generalization of the site-independent GBRs allows learning on overall relationships between the used predictors and the reference adjustment factors (Eq. 2).

The used training data and the selected hyperparameters also have a direct influence on the predictors needed by the GBR models to reduce the cost function (Eq. 3). In fact, the use of reanalysis variables (from single level and pressure levels) as predictors, caused an increase of the overall RMSE of both ERA-5 and MERRA-2 season-independent GBRs against the B_w data of all glaciers of the study (Fig. 3c and d). However, despite the high correlation of the downscaled reanalysis

variables (cf. Section 3.1) with the elevation of the glaciers, their inclusion in the set of predictors for the training of the site-independent GBRs reduced the overall RMSE (Fig. 3a and b). This difference can be explained by the combined effect of using data of the tested glacier in the training of the season-independent GBRs, and defining a small minimum number of samples required to create a leaf node of the GBR. In fact, the season-independent GBR can theoretically exploit the coordinates to split into individual sub-models adapted to individual glaciers. Therefore, the season-independent GBRs can learn the adjustment factors observed in the other accumulation seasons of the tested glacier and predict a similar adjustment factor for the tested accumulation season, with no need of learning overall relationships between the reanalysis predictors and the reference adjustment factors (Eq. 2).

5.2.2 Spatial and temporal transferability of the GBR models

The GBR models were trained with almost 100 glaciers distributed over the four regions on three continents. Between the regions we observed different robustness and performances. The performance of the GBR models tends to decrease when removing B_w data of neighbouring glaciers from the training process (Fig. 7b, d, f and h). Neighbouring glaciers were removed from the training as a function of the distance (range) from the tested glacier. Our results suggest that more available glaciers with B_w data would probably greatly improve the performance in Central Asia and Western Canada, where our dataset is limited in terms of number of monitored glaciers and the horizontal spacing between different sites is considerable. In the Alps, the network of monitored glaciers is much denser. Thus, more glaciers are excluded from the training for shorter distances than in other regions, impacting the performance of the site-independent GBR models. When the range of excluded neighbouring glaciers is extended to 1000 km, a strongly reduced number of glaciers of the same region is still used in the training, meaning that the models are almost exclusively trained with the glaciers of the other regions (the site-independent GBR becomes almost a region-independent GBR). The climate conditions and the complexity of the weather processes can be very different among the four investigated regions (and even within the individual regions). A region-independent model is thus not expected to provide accurate results. In Scandinavia, a linear precipitation gradient with elevation is more appropriate than in the more complex topography of the Alps and Central Asia (e.g. Rasmussen and Andreassen, 2005). Thus, the site-independent GBR models are only performing slightly better than the benchmark when the full set of the other glaciers is used in the training, indicating that a simpler lapse-rate-based approach might be preferable. However, considering the four regions, the bias of the region-independent GBR models remains comparable to the bias obtained with the benchmark method, which is independent from any ground observation.

The performance of the season-independent GBR models improved consistently when including in the training only a few other seasons of B_w data related to the tested glacier (Fig. 7a, c, e and g). This thus demonstrates the uniqueness of the snow accumulation distribution over each glacier that cannot be easily reproduced by using the relations learned at other glaciers. However, this indicates that the snow accumulation distribution, and its relation with precipitation, is similar in different years (see also e.g. Grünewald et al., 2013; Sold et al., 2016). For our application, there would be added benefit from B_w data on additional glaciers rather than on additional seasons.

5.2.3 Representation of the temporal variability of the snow accumulation

All GBRs aimed at minimizing the MSE between the predicted and reference logarithmic adjustment factors (Eq. 3). The improvement of the temporal correlation between the original reanalysis and the B_w data is thus a consequence of bias-adjusted estimates over accumulation seasons rather than a primary goal of the GBRs. A sensitivity test reported in the Supplementary material (Fig. S3) suggests that the season-independent GBRs are not very sensitive to the number of years of data of the tested glacier used for training. Their performance is comparable to the site-independent GBRs (Tab. 2). Furthermore, only in a few cases the site-independent GBRs show a performance inferior to the original reanalysis or the benchmark method (e.g. Ts. Tuyuksuyskiy glacier). These promising results suggest that our new estimates could also be used to derive SWE trends with generally higher accuracy than the original reanalyses, thus potentially providing insights on the relation between climate change and both snow accumulation and precipitation at the highest elevations of mountain ranges, where virtually no direct precipitation records are available. Still, the limited number of glaciers with abundant B_w data coverage available over sufficient number of years do not allow us to perform a complete application of this approach.

5.2.4 Impact of the chosen reanalyses on the GBR models

At a regional scale, the total precipitation estimated in the accumulation season by the original MERRA-2 has shown larger biases than the original ERA-5 when compared to B_w on glaciers. The coarser spatial resolution of MERRA-2 is certainly a factor causing larger biases in complex high-mountain areas (e.g. Zandler et al., 2019; Chen et al., 2021). In fact, a coarse resolution directly implies that mountains are more strongly smoothed. The absolute elevation of a grid cell is thus lower for a coarse resolution and the estimated precipitation also refers to the lower elevation of the grid cell.

The performance of the original ERA-5 and MERRA-2 has a direct impact on the GBR models. However, the GBR models were able to compensate for such differences in the bias. In fact, the biases of the ERA-5 and MERRA-2 GBR models are much closer to each other than the biases of the original reanalyses (see Fig. 4a, b, c and d). The differences between the performance of our GBR models are also caused by the different predictors that have been used. For instance, we considered all the topographical predictors describing the reanalysis's subgrid complexity of both reanalysis products and ERA-5 is providing more descriptors than MERRA-2 (see Tab. B1 and B2).

5.2.5 Influence of the winter mass balance data accuracy on the GBR models

Our study strongly relies on observed snow accumulation data on glaciers. However, various problems are related to the direct measurements of snow accumulation on glaciers thus leading to uncertainties in the observations (e.g. Zemp et al., 2013; Sold et al., 2016; O'Neel et al., 2019; Huss et al., 2021). Most importantly, snow accumulation measured at individual points needs to be extrapolated in space to obtain B_w data used in our analysis. At the highest elevations of glaciers with a typically difficult accessibility for manual observations, results are often purely based on extrapolation techniques (e.g. Østrem and Brugman, 1966; Cogley et al., 2011; Huss et al., 2021). Given that the WGMS database does not generally report how this was achieved and how many actual observations were available in a given elevation interval, it is difficult to assess the integrative

uncertainty in the B_w data used. In order to illustrate the importance of the extrapolated B_w data used in this study, we more closely inspected point winter snow observations for 12 Swiss sites and three years (2016-2018) based on a dataset with higher resolution and full documentation (GLAMOS, 2021).

Figure 9a indicates that a lower number of manual observations was typically performed at the lowest and the highest elevations of the glaciers. In some elevation bands, even no manual observations are available and B_w data refer to an extrapolation. However, a much larger number of manual observations is typically performed in the elevation intervals corresponding to the largest areas of the glaciers (Fig. 9a and c). As indicated by Fig. 9e, considerable uncertainties might exist in the analyzed vertical profiles of B_w. However, the weighting function dependent on the area of the intervals used in the training of the GBR models assigned more importance to the B_w data in such observation-rich areas. Furthermore, the main results of the study relate to glacier-wide B_w data, which for most glaciers is very close to the glacier-wide mean of the manual observations as also indicated in Fig. 9e (only in three out of 34 cases the ratio of glacier-wide mass balance to the average of all individual observations is larger than 1.10 and in no case the ratio is lower than 0.90).

Another source of uncertainty that is difficult to assess is the starting date of the accumulation season. We considered the same starting date for all elevation intervals, even though it varies over the glacier's elevation range. The accumulation of snow starts later at low elevations and earlier at high elevations. Therefore, the different elevations also collect different precipitation totals (as the periods differ). The impact on the study of the date considered as beginning of the accumulation season has been evaluated with a sensitivity test (Fig. 9). For the same 12 Swiss glaciers and three years as above we rely on the more detailed data set of point winter mass balance data that documents start dates of measured cumulative snow precipitation of the winter season for each location individually. Start dates have been inferred based on a distributed glaciological modelling approach driven by daily local weather data (Huss et al., 2021). The total precipitation of ERA-5 and MERRA-2 were derived over these varying starting dates and were compared with the total precipitation obtained with non-varying, average starting dates (Fig. 9b, d and f). Figure 9b indicates that at high (low) elevations, the accumulation season can start up to 20 days before (after) the unique date that we considered for all elevation intervals. These differences may translate in different amounts of total precipitation. In extreme cases, the total MERRA-2 (Fig. 9d) or ERA-5 (Fig. 9f) precipitation that would be obtained with varying dates would be almost twice (or half) of the total precipitation that we considered. However, the impact on the main results presented above is limited because these large differences are typically observed at the highest/lowest elevations of the glaciers, where the glacier area is minor and thus, a lower weight is assigned to the B_w data in the training of the GBR models. Moreover, the main results of the study are based on glacier-wide values and for both MERRA-2 and ERA-5 we observe very small differences in terms of glacier-wide precipitation totals for the majority of the Swiss glaciers. Only in six out of 34 cases the glacier-wide ratio is smaller (larger) than 0.97 (1.03), and only in two cases it exceeds 1.10.

This analysis suggests that the used B_w data and the considered start dates can lead to relevant uncertainties in the analysis of vertical profiles. However, this does not generally have a relevant impact on our conclusions, which are mainly based on glacier-wide values.

6 Conclusions

In this study, we developed and evaluated a machine learning approach based on gradient boosting regressor models to adjust the total precipitation of reanalysis datasets (ERA-5 and MERRA-2) over the accumulation season on glaciers. The high performance achieved with our approach allowed us to use it to derive observation-independent SWE estimates over glaciers in the Alps, Scandinavia, Central Asia and Western Canada. Data on snow accumulation distribution at the end of the accumulation season covering a period of up to 41 years from 95 glaciers (Zemp et al., 2021) were used to train our approach.

The most important variables that were automatically selected by our GBR models were those related to the elevation difference between the glacier surface and the terrain model underlying the reanalyses. The latitude and longitude of the studied sites were also frequently used in order to discriminate between regions that are characterized by different climate conditions and weather systems, allowing the GBR models to be split into individual sub-models adapted to specific sub-regions.

In general, the total precipitation of the reanalyses largely underestimates observed snow accumulation on glaciers. The largest (relative) regional underestimation is observed in Central Asia for MERRA-2 and in Scandinavia for ERA-5 (Fig. 4). The GBR models allowed reducing these biases. In Central Asia and Western Canada, the correlation between the original reanalyses' estimates and the snow accumulation on the analyzed glaciers has considerably increased with the season-independent GBRs only. With the exception of some specific glaciers, our GBR models outperformed the benchmark method (lapse-rate-based approach) in the Alps, Scandinavia, Central Asia and Western Canada by reducing the bias of the original reanalysis against the B_w data (Fig. 5). This suggests that complex models such as our GBRs are needed to adjust reanalysis data to different glaciers located in different topographical settings and climatic conditions, and to overcome the varying performance of reanalysis data for different region of the world.

Our results furthermore indicate that the season-independent GBRs outperform the site-independent GBRs to reduce the bias, which consequently makes filling temporal data gaps much simpler than estimating SWE on glaciers where no observations are available. Thus, denser network of ground-based SWE measurements and/or improved remote sensing observations, are of great importance to further develop methods that allow spatio-temporal transferability of the observed snow and/or precipitation in high-mountain areas.

The GBR models, compared to the original reanalyses, have moreover shown improved performance in reproducing temporal changes (over years) of snow accumulation for the majority of the analyzed glaciers. Generally, our GBR models would allow deriving more accurate SWE trends than the original reanalyses, thus potentially providing insights on the relation between climate change and snow accumulation over glaciers.

We finally demonstrated that machine learning models (with robust cross-validation schemes) can be powerful instruments to adjust precipitation estimates over glaciers. The new information that our approach is able to deliver can significantly improve the calibration of glaciological and hydrological models in different regions of the world, in particular for regions where the quantity and quality of observations is very limited and the spatial resolution and performance of reanalysis products is too low.

Appendix A: Equations used to derive the relative humidity

The relative humidity is not directly provided by all the reanalysis products, therefore we derived it by applying a similar approach to Liston and Elder (2006) and Gupta and Tarboton (2016), which is presented hereafter.

A1 ERA-5

The relative humidity is not directly provided at the grid level, therefore, we combined the 2 m temperature ($t2m$) and dew point temperature ($d2m$) as follows:

$$r2m^* = \frac{a * exp(\frac{b*d2m}{c+d2m})}{a * exp(\frac{b*t2m}{c+t2m})} \tag{A1}$$

where $r2m^*$ is the computed 2 m relative humidity and for ice/snow, $a = 611.21\,Pa$, $b = 22.452$ and $c = 272.55\,°C$.

A2 MERRA-2

MERRA-2 is not providing the relative humidity at the grid and at the pressure levels, furthermore, the dew point temperature is not provided at the pressure levels either, therefore, we combined the specific humidity and the pressure in order to derive them (at the grid and at the pressure levels). For ice/snow, $a = 611.21\,Pa$, $b = 22.452$ and $c = 272.55\,°C$.

Vapour pressure:

$$e^* = \frac{QV * P}{0.622 + QV} \tag{A2}$$

where the specific humidity $QV = QV10M$ for the grid and $QV = QV_{\text{levels}}$ for the pressure levels, the pressure $P = PS$ for the grid and $P = P_{\text{levels}}$ for the pressure levels. The vapour pressure e^* was named $e10M^*$ for the grid and e^*_{levels} for the pressure levels.

Dew point temperature:

$$Td^* = 273.15 + \frac{c * \ln(\frac{e^*}{a})}{b - \ln(\frac{e^*}{a})} \tag{A3}$$

The dew point temperature Td^* was named $Td10M^*$ for the grid and Td^*_{levels} for the pressure levels.

Relative humidity:

$$RH^* = \frac{a * exp(\frac{b*Td^*}{c+Td^*})}{a * exp(\frac{b*T}{c+T})} \tag{A4}$$

where the temperature $T = T10M$ for the grid and $T = T_{\text{levels}}$ for the pressure levels. The relative humidity RH^* was named $RH10M^*$ for the grid and RH^*_{levels} for the pressure levels.

Appendix B: Derivation and list of the variables used in the models

In Table B1 and Table B2, we report the complete list of variables selected from the reanalyses products. In Table B3 we provide a summary of all the variables used by the GBR models.

Author contributions. MGu conducted the analysis and wrote the manuscript with inputs from all co-authors. MH derived the snow accumulation data for the Swiss glaciers analyzed in Section 2.2 and 5.2.5 and contributed to the related discussion. MGa and NS contributed to the design of the research and helped with continuous discussions concerning the results. All co-authors contributed to the final form of the manuscript.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. The study is part of the High-SPA 200021_178963 project, which is funded by the Swiss National Science Foundation (SNSF). We would like to acknowledge the WGMS and all the groups providing freely available in-situ observations on glaciers. The authors would also like to acknowledge NASA for providing the freely available MERRA-2 products (Global Modeling and Assimilation Office (GMAO), 2015a, b, c). Hersbach et al. (2018a, b) were downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store. The results contain modified Copernicus Climate Change Service information 2020. Neither the European Commission nor ECMWF is responsible for any use that may be made of the Copernicus information or data it contains.

References

Adger, W. N., Huq, S., Brown, K., Conway, D., and M., H.: Adaptation to climate change in the developing world, Progress in Development Studies, 3 (3), 179–195, https://doi.org/10.1191/1464993403ps060oa, 2003.

Barandun, M., Fiddes, J., Scherler, M., Mathys, T., Saks, T., Petrakov, D., and Hoelzle, M.: The state and future of the cryosphere in Central Asia, Water Security, 11, 100 072, https://doi.org/10.1016/j.wasec.2020.100072, 2020.

Beniston, M.: Is snow in the Alps receding or disappearing?, Wiley Interdisciplin. Rev. Clim. Change, 3, 349–358, https://doi.org/10.1002/wcc.179, 2012.

Beniston, M., Stoffel, M., Harding, R., Kernan, M., Ludwing, R., Moors, E., Samuels, P., and Tockner, K.: Obstacles to data access for research related to climate and water: Implications for science and EU policy-making, Environmental Science and Policy, 17, 41–48, https://doi.org/10.1016/j.envsci.2011.12.002, 2012.

Beniston, M., Farinotti, D., Stoffel, M., Andreassen, L. M., Coppola, E., Eckert, N., Fantini, A., Giacona, F., Hauck, C., Huss, M., Huwald, H., Lehning, M., López-Moreno, J.-I., Magnusson, J., Marty, C., Morán-Tejéda, E., Morin, S., Naaim, M., Provenzale, A., Rabatel, A., Six, D., Stötter, J., Strasser, U., Terzago, S., and Vincent, C.: The European mountain cryosphere: a review of its current state, trends, and future challenges, The Cryosphere, 12, 759–794, https://doi.org/10.5194/tc-12-759-2018, 2018.

Bormann, K., Brown, R., Derksen, C., and Painter, T.: Estimating snow-cover trends from space, Nature Clim Change, 8, 924–928, https://doi.org/10.1038/s41558-018-0318-3, 2018.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Chen, Y., Sharma, S. andX Zhou, X., Yang, K., Li, X., Niu, X., Hu, X., and Khadka, N.: Spatial performance of multiple reanalysis precipitation datasets on the southern slope of central Himalaya, Atmospheric Research, 250, 105 365, https://doi.org/10.1016/j.atmosres.2020.105365, 2021.

Cogley, J., Hock, R., Rasmussen, L., Arendt, A., Bauder, A., Braithwaite, R., Jansson, P., Kaser, G., Möller, M., Nicholson, L., and Zemp, M.: Glossary of Glacier Mass Balance and Related Terms. IHP-VII Technical Documents in Hydrology, 86, 2011.

Dadic, R., Mott, R., Lehning, M., and Burlando, P.: Wind influence on snow depth distribution and accumulation over glaciers, Journal of Geophysical Research F: Earth Surface, 115, F01 012 (8 pp.), https://doi.org/10.1029/2009JF001261, 2010.

Fiddes, J. and Gruber, S.: TopoSCALE v.1.0: downscaling gridded climate data in complex terrain, Geoscientific Model Development, 7, 387–405, https://doi.org/10.5194/gmd-7-387-2014, 2014.

Freitas, A.: Comprehensible classification models: a position paper, ACM SIGKDD explorations newsletter, 15(1), 1–10, 2014.

Friedman, J. H.: Greedy function approximation: a gradient boosting machine, Annals of statistics, pp. 1189–1232, 2001.

Gascoin, S., Lhermitte, S., Kinnard, C., Bortels, K., and Liston, G. E.: Wind effects on snow cover in Pascua-Lama, Dry Andes of Chile, Advances in Water Resources, 55, 25–39, https://doi.org/0.1016/j.advwatres.2012.11.013, 2013.

Gelaro, R., McCarty, W., Suarez, M., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Da Silva, A., Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Gerber, F., Mott, R., and Lehning, M.: The importance of near-surface winter precipitation processes in complex alpine terrain, Journal of Hydrometeorology, 20, 77–96, https://doi.org/10.1175/JHM-D-18-0055.1, 2019.

GLAMOS: Swiss Glacier Point Mass Balance Observations, release 2021, Glacier Monitoring Switzerland., https://doi.org/10.18750/massbalance_point.r2021.2021, 2021.

Global Modeling and Assimilation Office (GMAO): MERRA-2 inst6_3d_ana_Np: 3d,6-Hourly,Instantaneous,Pressure-Level,Analysis,Analyzed Meteorological Fields V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/A7S6XP56VZWS, accessed: 13.05.2021, 2015a.

Global Modeling and Assimilation Office (GMAO): MERRA-2 tavg1_2d_lnd_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Land Surface Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/RKPHT8KC1Y1T, accessed: 13.05.2021, 2015b.

Global Modeling and Assimilation Office (GMAO): MERRA-2 tavg1_2d_slv_Nx: 2d,1-Hourly,Time-Averaged,Single-Level,Assimilation,Single-Level Diagnostics V5.12.4, Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC), https://doi.org/10.5067/VJAFPLI1CSIV, accessed: 13.05.2021, 2015c.

Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., and Stoffel, M.: 21st century climate change in the European Alps — A review, Science of The Total Environment, 493, 1138–1151, https://doi.org/10.1016/j.scitotenv.2013.07.050, 2014.

Grünewald, T., Stötter, J., Pomeroy, J., Dadic, R., Baños, M., Marturià, J., Spross, M., Hopkinson, C., Burlando, P., and Lehning, M.: Statistical modelling of the snow depth distribution in open alpine terrain, Hydrology and Earth System Sciences, 17, 3005–3021, https://doi.org/10.5194/hess-17-3005-2013, 2013.

Gugerli, R., Gabella, M., Huss, M., and Salzmann, N.: Can weather radars be used to estimate snow accumulation on alpine glaciers? - an evaluation based on glaciological surveys., Journal of Hydrometeorology, 21, 2943–2962, https://doi.org/10.1175/JHM-D-20-0112.1, 2020.

Guidicelli, M., Gugerli, R., Gabella, M., Marty, C., and Salzmann, N.: Continuous spatio-temporal high-resolution estimates of SWE across the Swiss Alps - a statistical two-step approach for high-mountain topography, Frontiers in Earth Science, 9, 399, https://doi.org/10.3389/feart.2021.664648, 2021.

Gupta, A. S. and Tarboton, D. G.: A tool for downscaling weather data from large-grid reanalysis products to finer spatial scales for distributed hydrological applications, Environmental Modelling & Software, 84, 50–69, https://doi.org/10.1016/j.envsoft.2016.06.014, 2016.

Helfricht, K., Kuhn, M., Keuschnig, M., and Heilig, A.: Lidar snow cover studies on glaciers in the Ötztal Alps (Austria): comparison with snow depths calculated from GPR measurements, The Cryosphere, 8(1), 41–57, https://doi.org/10.5194/tc-8-41-2014, 2014.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)., https://doi.org/10.24381/cds.bd0915c6, accessed on 01.06.2021, 2018a.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., and Thépaut, J.-N.: ERA5 hourly data on single levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)., https://doi.org/10.24381/cds.adbb2d47, accessed on 01.06.2021, 2018b.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., , and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., and Steltzer, H.: 2019: High Mountain Areas. In: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)]. In press.

Hugonnet, R., McNabb, R., Berthier, E., Menounos, B., Nuth, C., Girod, L., Farinotti, D., Huss, M., Dussaillant, I., Brun, F., and Kääb, A.: Accelerated global glacier mass loss in the early twenty-first century, Nature, 592, 726–731, https://doi.org/10.1038/s41586-021-03436-z, 2021.

Huss, M., Bauder, A., and Funk, M.: Homogenization of long-term mass-balance time series, Annals of Glaciology, 50(50), 198–206, https://doi.org/10.3189/172756409787769627, 2009.

Huss, M., Bookhagen, B., Huggel, C., Jacobsen, D., Bradley, R. S., Clague, J. J., Vuille, M., Buytaert, W., Cayan, D. R., Greenwood, G., Mark, B. G., Milner, A., Weingartner, R., and Winder, M.: The state and future of the cryosphere in Central Asia, Earth's Future, 5, 418–435, https://doi.org/10.1002/2016EF000514, 2017.

Huss, M., Bauder, A., Linsbauer, A., Gabbi, J., Kappenberger, G., Steinegger, U., and Farinotti, D.: More than a century of direct glacier mass-balance observations on Claridenfirn, Switzerland, Journal of Glaciology, 67(264), 697–713, https://doi.org/10.1017/jog.2021.22, 2021.

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., and Baesens, B.: An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models, Decision Support Systems, 51(1), 141–154, https://doi.org/10.1016/j.dss.2010.12.003, 2011.

Immerzeel, W. W., Lutz, A. F., Andrade, M., Bahl, A., Biemans, H., Bolch, T., Hyde, S., Brumby, S., Davies, B. J., Elmore, A. C., Emmer, A., Feng, M., Fernández, A., Haritashya, U., Kargel, J. S., Koppes, M., Kraaijenbrink, P. D. A., Kulkarni, A. V., Mayewski, P. A., Nepal, S., Pacheco, P., Painter, T. H., Pellicciotti, F., Rajaram, H., Rupper, S., Sinisalo, A., Shrestha, A. B., Viviroli, D., Wada, Y., Xiao, C., Yao, T., and M., B. J. E.: Importance and vulnerability of the world's water towers, Nature, 577, 364–369, https://doi.org/10.1038/s41586-019-1822-y, 2020.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), Journal of Hydrometeorology, 7, 217–234, https://doi.org/10.1175/JHM486.1, 2006.

Liu, Y. and Margulis, S. A.: Deriving Bias and Uncertainty in MERRA-2 Snowfall Precipitation Over High Mountain Asia, Frontiers in Earth Science, 7, 280, https://doi.org/10.3389/feart.2019.00280, 2019.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, RG3003, https://doi.org/10.1029/2009RG000314, 2010.

Marty, C.: Regime shift of snow days in Switzerland, Geophysical Reasearch Letters', 35, L12 501, https://doi.org/10.1029/2008GL033998, 2008.

Mayo, L., Meier, M., and Tangborn, W.: A system to combine stratigraphic and annual mass-balance systems: a contribution to the International Hydrological Decade, J. Glaciol, 11(61), 3–14, https://doi.org/10.3189/S0022143000022449, 1972.

Mölg, T. and Kaser, G.: A new approach to resolving climate-cryosphere relations: Downscaling climate dynamics to glacier-scale mass and energy balance without statistical scale linking, J. Geophys. Res. Atmos., 116, D16 101, https://doi.org/10.1029/2011JD015669, 2011.

Najafi, M. R., Moradkhani, H., and Wherry, S. A.: Statistical Downscaling of Precipitation Using Machine Learning with Optimal Predictor Selection, Journal of Hydrologic Engineering, 16(8), 650–664, https://doi.org/10.1061/(ASCE)HE.1943-5584.0000355, 2011.

O'Neel, S., McNeil, C., Sass, L., Florentine, C., Baker, E., Peitzsch, E., McGrath, D., Fountain, A., 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(253), 850–866, https://doi.org/10.1017/jog.2019.66, 2019.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Pepin, N. C., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., Palazzi, E., Seibert, P., Serafin, S., Schöner, W., Terzago, ., Thornton, J., Vuille, M., and Adler, C.: Climate Changes and Their Elevational Patterns in the Mountains of the World, Reviews of Geophysics, 60(1), 601–606, https://doi.org/10.1029/2020RG000730, 2022.

Rasmussen, L. and Andreassen, L.: Seasonal mass-balance gradients in Norway, Journal of Glaciology, 51(175), 601–606, https://doi.org/10.3189/172756505781828990, 2005.

Rasul, G. and Molden, D.: The Global Social and Economic Consequences of Mountain Cryospheric Change, Frontiers in Environmental Science, 7, 91, https://doi.org/10.3389/fenvs.2019.00091, 2019.

RGI Consortium: Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0, https://doi.org/10.7265/N5-RGI-60, Technical Report, Global Land Ice Measurements from Space, Colorado, USA. Digital Media., 2017.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, Journal of Climate, 24, 3624 – 3648, https://doi.org/10.1175/JCLI-D-11-00015.1, 2011.

Sachindra, D. A., Ahmed, K., Mamunur Rashid, M., Shahid, S., and C., P. B. J.: Statistical downscaling of precipitation using machine learning techniques, Atmospheric Research, 212, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022, 2018.

Salzmann, N. and Mearns, L. O.: Assessing the Performance of Multiple Regional Climate Model Simulations for Seasonal Mountain Snow in the Upper Colorado River Basin, Journal of Hydrometeorology, 13 (2), 539–556, https://doi.org/10.1175/2011JHM1371.1, 2012.

Salzmann, N., Huggel, C., Rohrer, M., and Stoffel, M.: Data and knowledge gaps in glacier, snow and related runoff research – A climate change adaptation perspective, Journal of Hydrology, 518, Part B, 225–234, 2014.

Seiz, G., Foppa, N., Meier, M., and Meister, O.: National Climate Observing System (GCOS Switzerland), Advances in Science and Research, 6, https://doi.org/10.5194/asr-6-95-2011, 2010.

Serifi, A., Günther, T., and Ban, N.: Spatio-Temporal Downscaling of Climate Data Using Convolutional and Error-Predicting Neural Networks, Frontiers in Climate, 3:656479, https://doi.org/10.3389/fclim.2021.656479, 2021.

Sevruk, B., OndrÃ¡s, M., and ChvÃla, B.: The WMO precipitation measurement intercomparisons, Atmospheric Research - ATMOS RES, 92, 376–380, https://doi.org/10.1016/j.atmosres.2009.01.016, 2009.

Sold, L., Huss, M., Machguth, H., Joerg, P., V.G., L., Linsbauer, A., Salzmann, N., Zemp, M., and M., H.: Mass Balance Reanalysis of Findelengletscher, Switzerland; Benefits of Extensive Snow Accumulation Measurements, Frontiers in Earth Science, 4, 18, https://doi.org/10.3389/feart.2016.00018, 2016.

Stone, D., Auffhammer, M., Carey, M., Hansen, G., Huggel, C., Cramer, W., Lobell, D., Molau, U., Solow, A., Tibig, L., and Yohe, G.: The challenge to detect and attribute effects of climate change on human and natural systems, Climate Change, 121, 381–395, https://doi.org/10.1007/s10584-013-0873-6, 2013.

Sun, A. Y. and Tang, G.: Downscaling Satellite and Reanalysis Precipitation Products Using Attention-Based Deep Convolutional Neural Nets, Frontiers in Water, 2:536743, https://doi.org/10.3389/frwa.2020.536743, 2020.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., , and Hsu, K.-L.: A review of global precipitation data sets: Data sources, estimation, and inter-comparison, Geophysical Research Letters, 56, 79–107, https://doi.org/10.1002/2017RG000574, 2018.

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., and de Castro, M.: Global precipitation measurement: Methods, datasets and applications, Atmospheric Research, 104-105, 70–97, https://doi.org/10.1016/j.atmosres.2011.10.021, 2012.

Viviroli, D., Dürr, H. H., Messerli, B., Meybeck, M., and Weingartner, R.: Mountains of the world, water towers for humanity: typology, mapping, and global significance, Water Resources Research, 43 (7), https://doi.org/10.1029/2006WR005653, 2007.

Vorkauf, M., Marty, C., Kahmen, A., and Hiltbrunner, E.: Past and future snowmelt trends in the Swiss Alps: the role of temperature and snowpack, Climatic Change, 165, 44, https://doi.org/10.1007/s10584-021-03027-x, 2021.

Wang, F., Tian, D., Lowe, L., Kalin, L., and Lehrter, J.: Deep learning for daily precipitation and temperature downscaling, Water Resources Research, 57, e2020WR029 308, https://doi.org/10.1029/2020WR029308, 2021.

WGMS: Fluctuations of Glaciers Database. World Glacier Monitoring Service, Zurich, Switzerland, DOI:10.5904/wgms-fog-2020-08. Online access: http://dx.doi.org/10.5904/wgms-fog-2020-08, 2020.

Zandler, H., Haag, I., and Samimi, C.: Evaluation needs and temporal performance differences of gridded precipitation products in peripheral mountain regions, Scientific Reports, 9, 15 118, https://doi.org/10.1038/s41598-019-51666-z, 2019.

Zemp, M., Thibert, E., Huss, M., Stumm, D., Rolstad Denby, C., Nuth, C., Nussbaumer, S. U., Moholdt, G., Mercer, A., Mayer, C., Joerg, P. C., Jansson, P., Hynek, B., Fischer, A., Escher-Vetter, H., Elvehøy, H., , and Andreassen, L. M.: Reanalysing glacier mass balance measurement series, The Cryosphere, 7, 1227–1245, https://doi.org/10.5194/tc-7-1227-2013, 2013.

Zemp, M., Nussbaumer, S., Gärtner-Roer, I., Bannwart, J., Paul, F., and Hoelzle, M. e.: WGMS: Global Glacier Change Bulletin No. 4 (2018-2019), iSC(WDS)/IUGG(IACS)/UNEP/UNESCO/WMO, World Glacier Monitoring Service, Zurich, Switzerland, 278 pp. Based on database version: doi:10.5904/wgms-fog-2021-05, 2021.

Østrem, G. and Brugman, M.: Glacier mass balance measurements. Department of Mines and Technical Surveys, Glaciology Section, 1966.

Østrem, G. and Brugman, M.: Glacier Mass-balance Measurements: A Manual for Field and Office work. NHRI Science Report No. 4. National Hydrological Research Institute, Saskatoon, 1991.

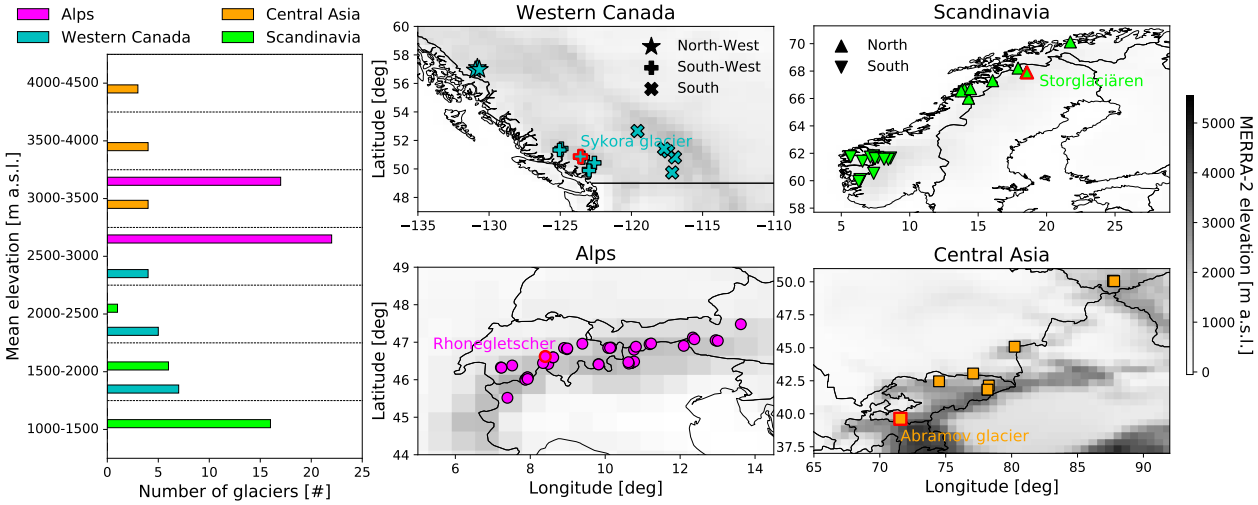

Figure 1. Mean elevation and distribution of the glaciers used in the study (data source: (Zemp et al., 2021)). Glaciers shown in Fig. 8 are highlighted in red.

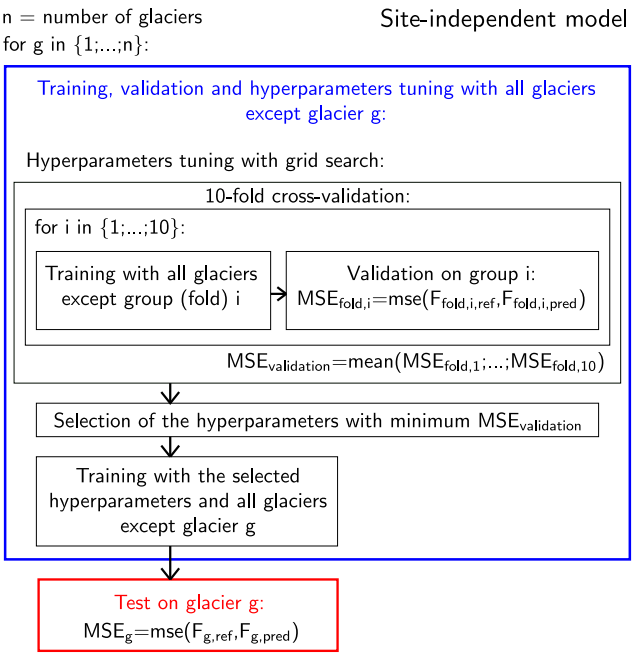

Figure 2. Training, cross-validation with hyperparameters selection and test scheme for the site-independent GBR model. $F_{\mathrm{dB,\,ref}}$ is the reference adjustment factor (Eq. 2), and $F_{\mathrm{dB,\,pred}}$ is the adjustment factor predicted by the GBR model.

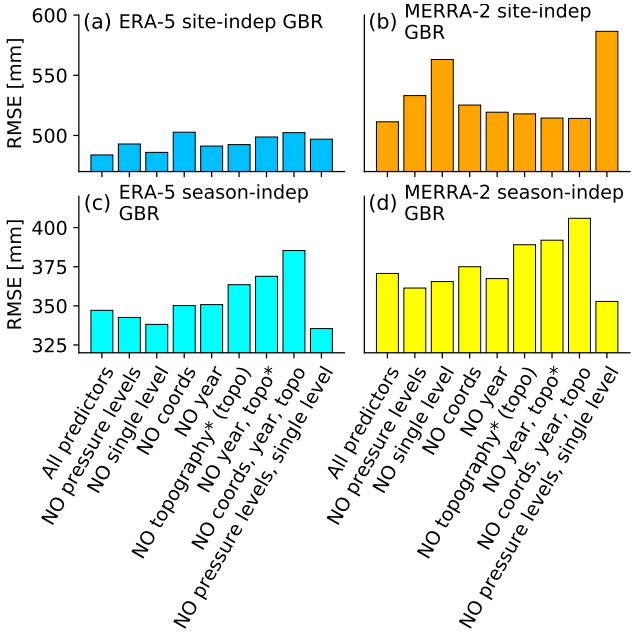

Figure 3. Overall root mean squared error (RMSE) between B_w and GBR models using different groups of predictors, for all analyzed glaciers and years: (a) ERA-5 site-independent GBR, (b) MERRA-2 site-independent GBR, (c) ERA-5 season-independent GBR, and (d) MERRA-2 season-independent GBR. *Topography refers to the topographical parameters describing the reanalysis's subgrid complexity and the average slope and aspect of the glaciers by using the information provided in the Randolph Glacier Inventory version 6 (RGI Consortium, 2017). For the list of variables included in each group of predictors see Tables B1 (original ERA-5 variables), B2 (original MERRA-2 variables) and B3 (downscaled ERA-5 and MERRA-2 variables).

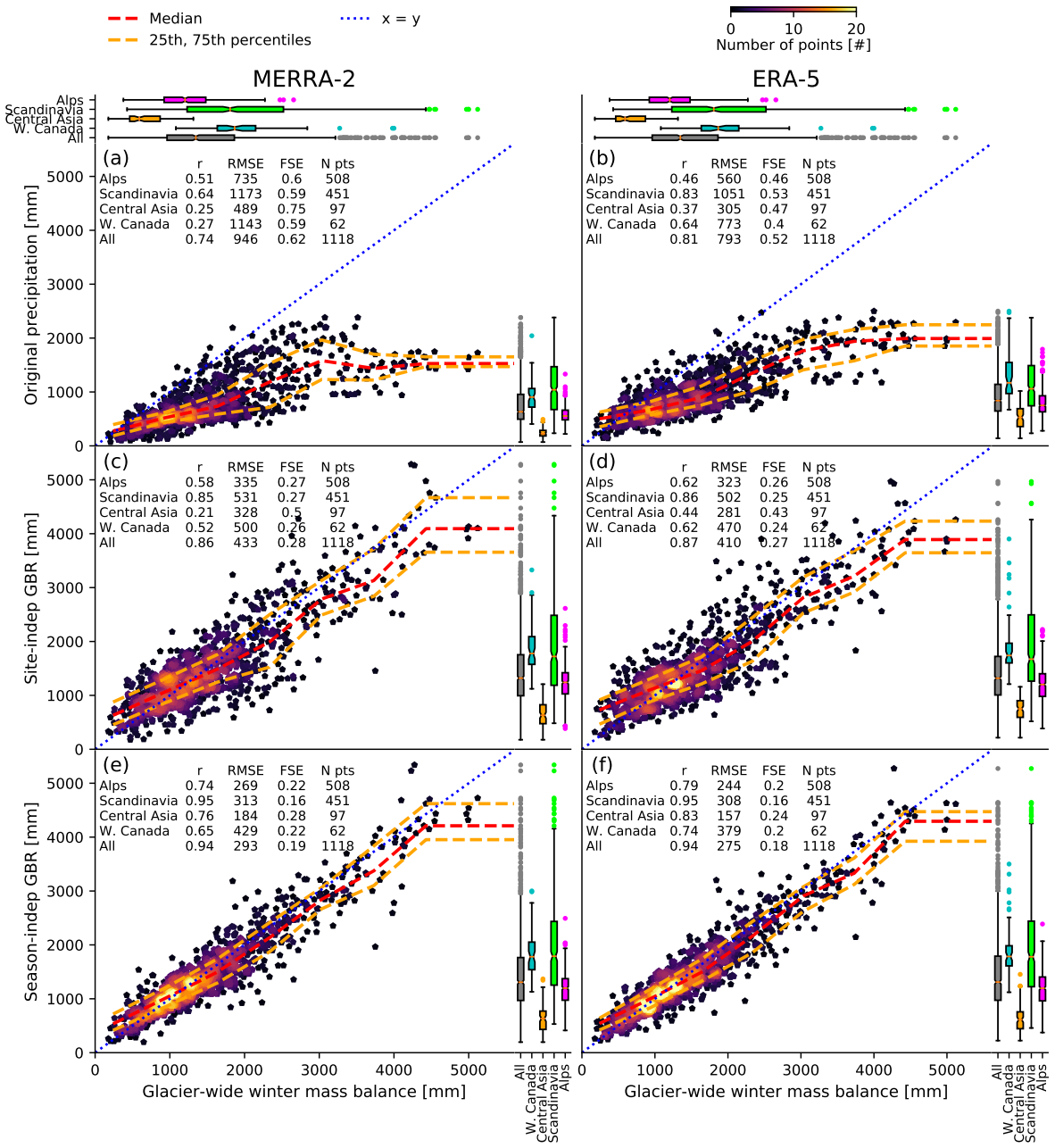

Figure 4. Comparison between all glacier-wide B_w values and the model estimates: (a) original MERRA-2 , (b) original ERA-5, (c) MERRA-2 site-independent GBR, (d) ERA-5 site-independent GBR, (e) MERRA-2 season-independent GBR, (f) ERA-5 season-independent GBR. The Pearson correlation (r), the root-mean-squared error (RMSE), the fraction of standard error (FSE, corresponding to the RMSE divided by the regional mean B_w) and the number of all seasons of all glaciers (N pts) for each region are also reported. The boxplots indicate the distribution of the model's estimates (right) and of the B_w data (top) for each region.

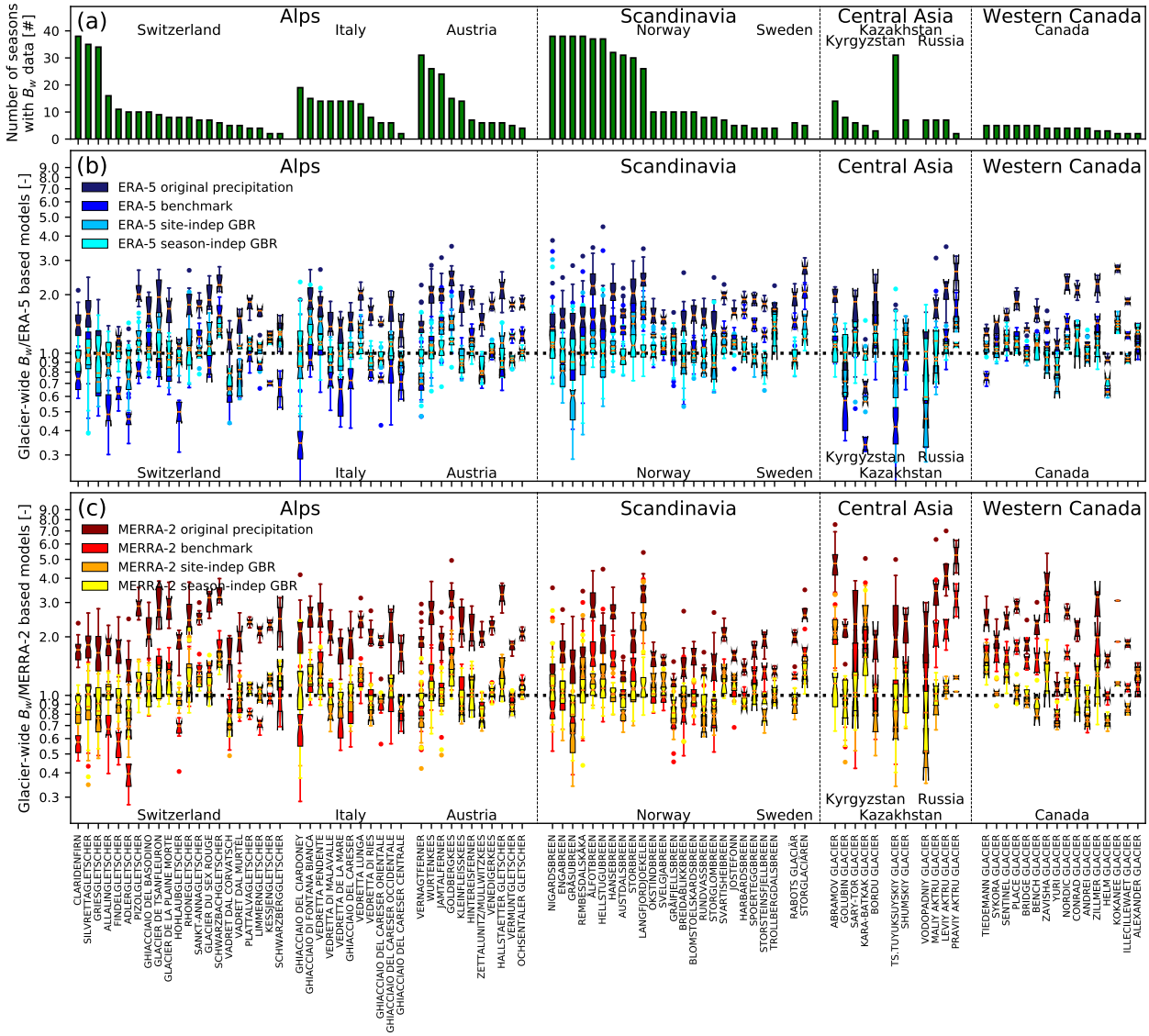

Figure 5. (a) Number of seasons with available B_w data for each glacier. Factors between seasonal glacier-wide B_w and (b) ERA-5-based models and (c) MERRA-2 based models, for each glacier of the study. The variability shown in the boxplots is given by the different seasons of B_w data.

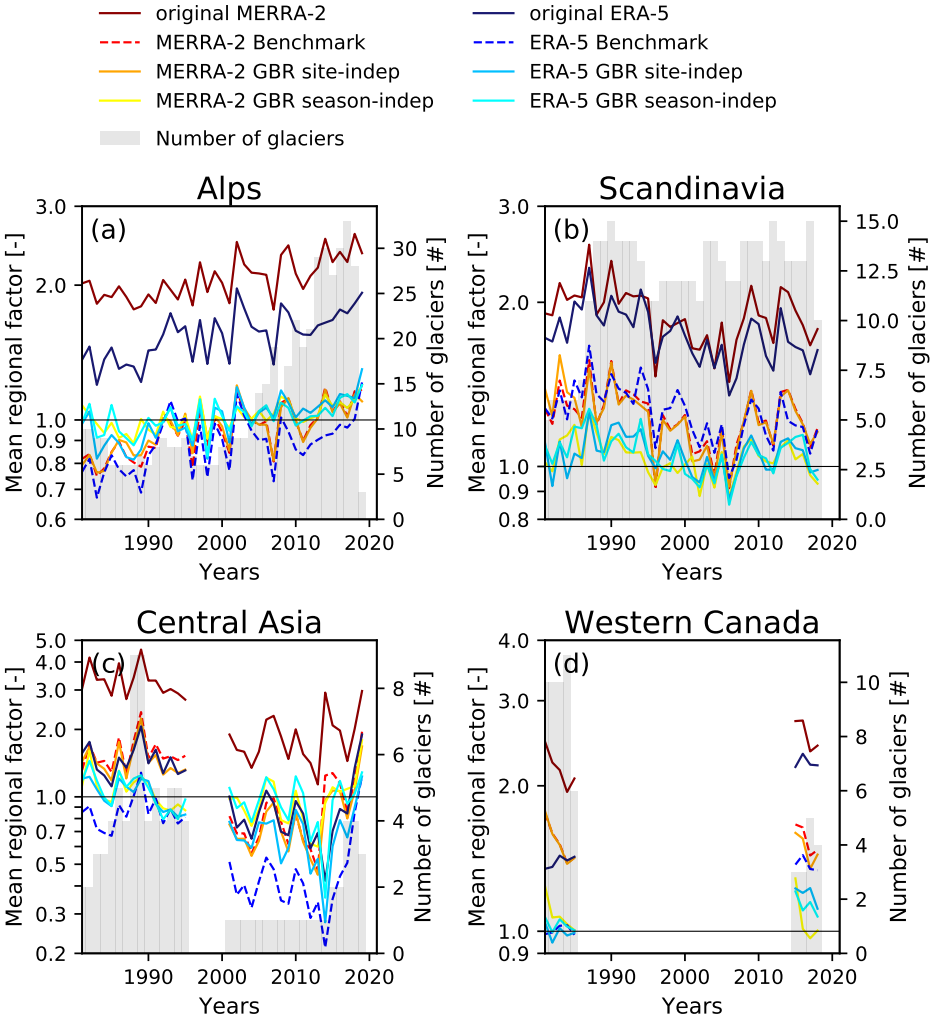

Figure 6. Mean regional factor between the B_w data and the reanalysis-based models' estimates as a function of the accumulation seasons from 1981 to 2019 (last available season): (a) Alps, (b) Scandinavia, (c) Central Asia and (d) Western Canada. For each season, all glaciers with available B_w data were considered (the number of glaciers used to derive the regional factor is indicated by the gray bars).

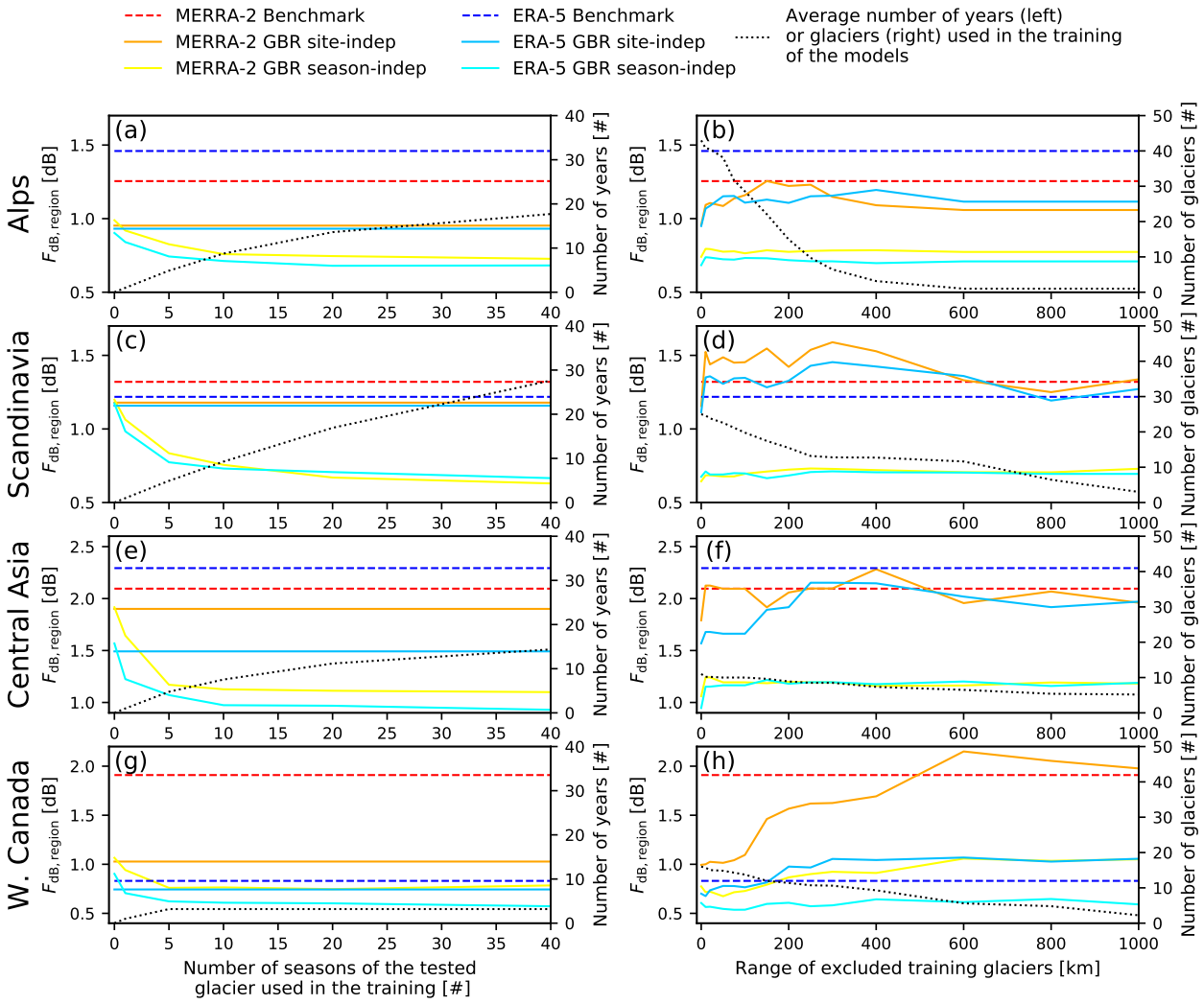

Figure 7. Evaluation of the mean regional factor between the B_w data and the reanalysis-based models' estimates ($F_{\mathrm{dB,region}}$ defined in Eq. 6) depending on different data used in the training of the GBR models. Left column - evaluation of the performance of the season-independent GBR models as a function of the number of seasons of the tested glacier used in the training (from no data to a maximum of 40 years of data of the tested glacier). (a) Model validation depending on the number of training seasons per glacier in the Alps, (c) Scandinavia, (e) Central Asia and (g) Western Canada. Right column - evaluation of the robustness of the GBR models as a function of the number of other glaciers in the same region used in the training. All glaciers located within a range growing from 0 to 1000 km from the tested glacier were excluded from the training. (b) Model evaluation depending on the range of excluded glaciers from the training in the Alps, (d) Scandinavia, (f) Central Asia and (h) Western Canada.

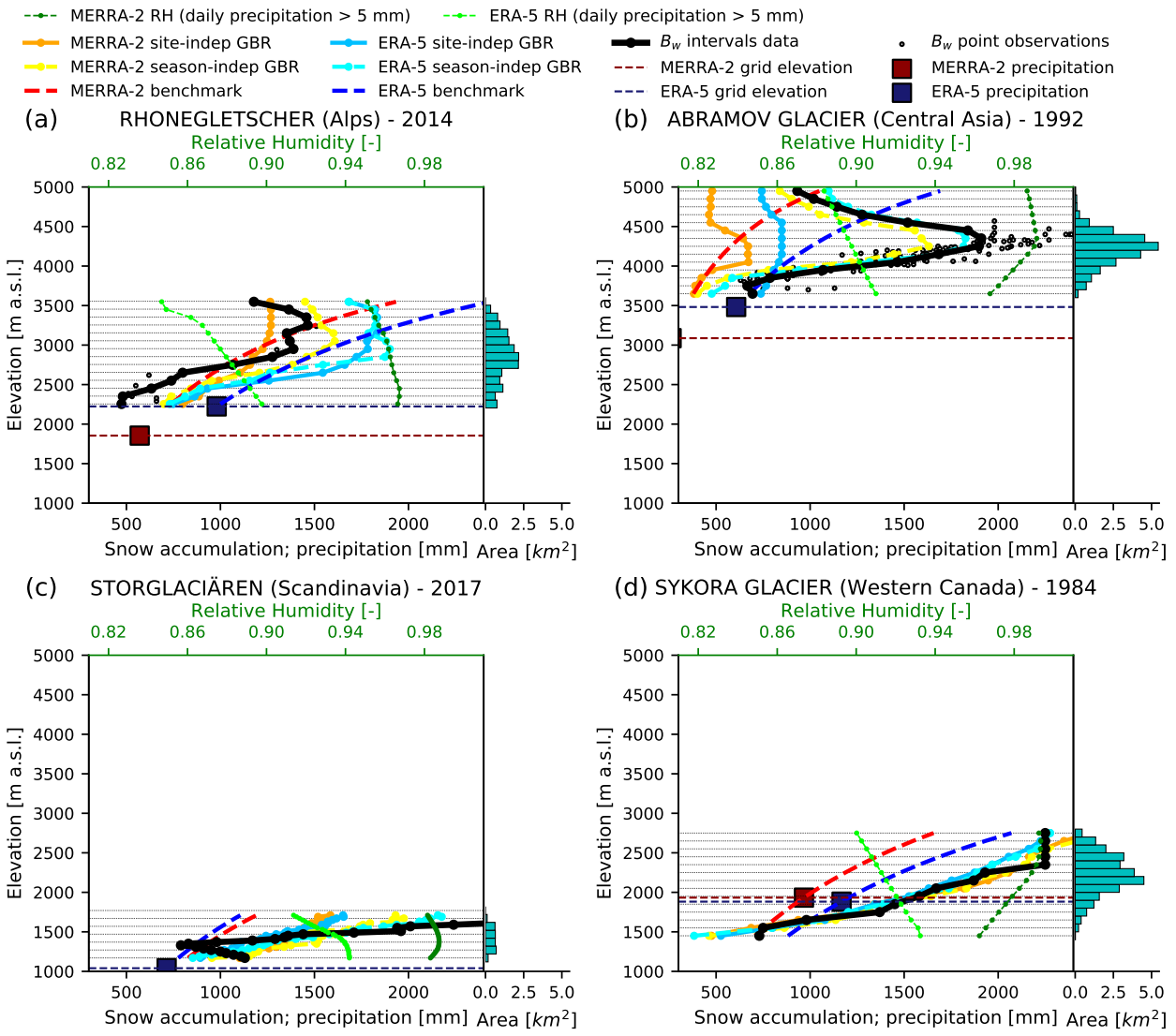

Figure 8. Vertical profiles of B_w data and the modelled SWE at the end of a specific accumulation season: (a) Rhonegletscher (Alps), (b) Abramov glacier (Central Asia), (c) Storglaciären (Scandinavia) and (d) Sykora glacier (Western Canada).

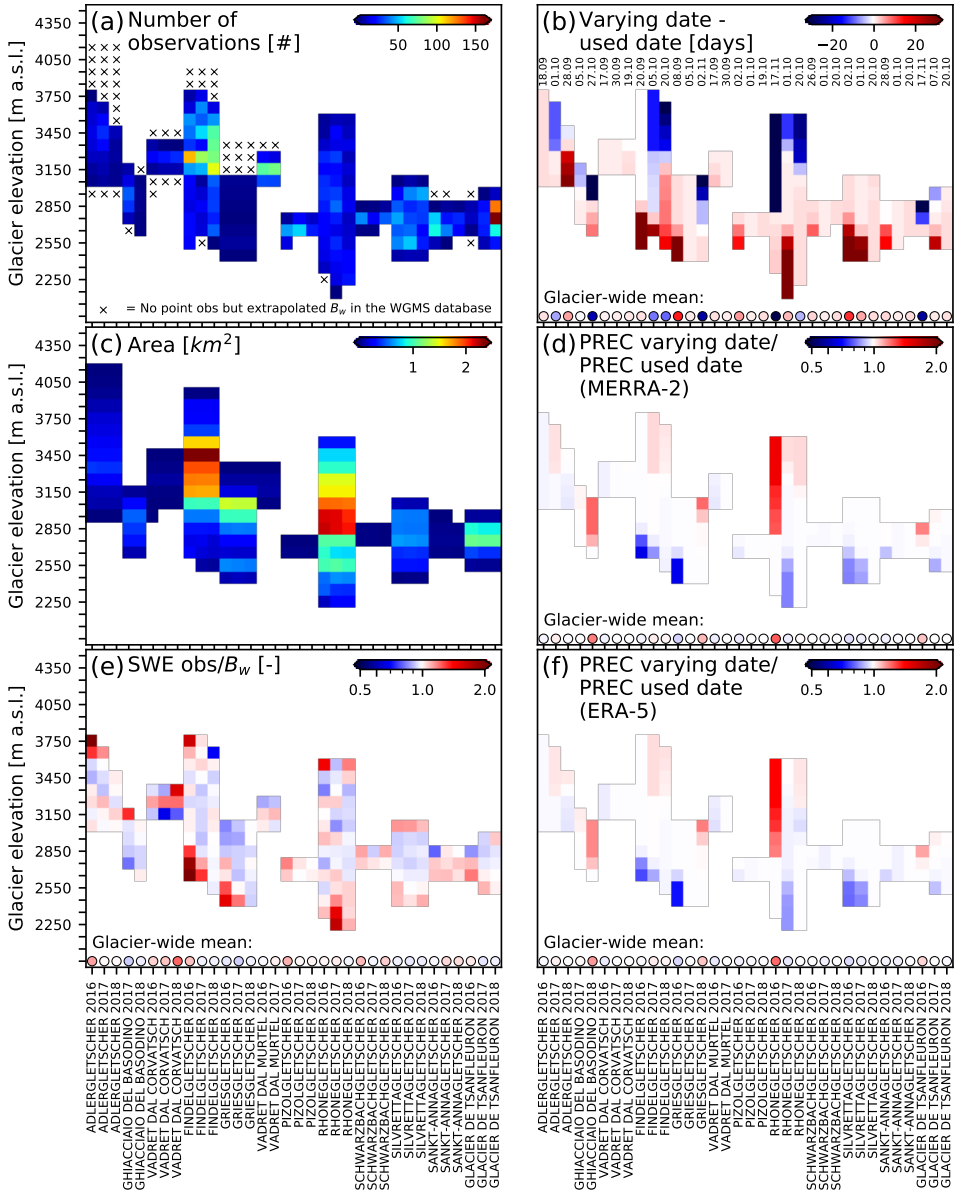

Figure 9. Sensitivity analysis of extrapolated B_w data and used starting dates for 12 glaciers in the Swiss Alps between 2016 and 2018 (GLAMOS, 2021). Left column - difference between B_w data used in this study and point-SWE observations: (a) Number of manual observations performed in the elevation intervals of the glaciers. (c) Area of the glacier according to the elevation interval. (e) Ratio between the observed SWE and the B_w data. Right column - impact of the date considered as beginning of the accumulation season on seasonal precipitation totals: (b) Differences between accurate (varying) dates of the beginning of the accumulation period and the used dates in the study (the day and the month of the used dates are written on the figure (DD.MM)). (d) Ratio between the total precipitation of MERRA-2 according to the accurate dates and the used dates. (f) Ratio between the total precipitation of ERA-5 according to the accurate dates and the used dates.

Table 1. Average hyperparameters of the optimized GBR models of all the studied glaciers. *Subsample* indicates the fraction of samples used for fitting the individual trees.

	Site-independent GBR		Season-independent GBR	
	ERA-5	MERRA-2	ERA-5	MERRA-2
n estimators	86	99	131	123
max. depth	7	7	8	8
min. samples leaf	72	77	10	11
learning rate	0.07	0.07	0.09	0.08
subsample	0.8	0.7	0.8	0.8

Table 2. Pearson correlation (r [-]) between the reanalysis-based models and the glacier-wide B_w over the accumulation seasons (temporal correlation). Only the glaciers with a minimum of 15 seasons with B_w data are shown (i.e. no glacier in Western Canada). The significance of the correlation is based on Student's t-distribution.

	r (p-value)								n seasons
	Original		Benchmark		GB site-indep		GB season-indep		
Glacier	ERA-5	MERRA-2	ERA-5	MERRA-2	ERA-5	MERRA-2	ERA-5	MERRA-2	
Allalingletscher	**0.71**	**0.74**	**0.68**	**0.70**	**0.76**	**0.81**	**0.80**	**0.86**	16
Claridenfirn	**0.71**	**0.83**	**0.74**	**0.85**	**0.69**	**0.83**	**0.74**	**0.82**	38
Griesgletscher	**0.45**	**0.53**	**0.47**	**0.55**	**0.53**	**0.57**	**0.60**	**0.57**	34
Silvrettagletscher	**0.56**	**0.63**	**0.55**	**0.62**	**0.59**	**0.67**	**0.59**	**0.66**	35
Ghiacciaio del Ciardoney	**0.66**	**0.52**	**0.66**	**0.52**	**0.64**	**0.50**	**0.61**	0.36	19
Ghiacciaio di Fontana Bianca	**0.88**	**0.90**	**0.87**	**0.89**	**0.86**	**0.86**	**0.78**	**0.90**	15
Goldbergkees	0.46	0.31	0.46	0.31	**0.63**	**0.62**	**0.58**	0.35	15
Jamtalferner	**0.60**	**0.65**	**0.60**	**0.64**	**0.70**	**0.64**	**0.76**	**0.69**	24
Vernagtferner	**0.64**	**0.66**	**0.64**	**0.65**	**0.65**	**0.64**	**0.52**	**0.59**	31
Wurtenkees	**0.65**	**0.56**	**0.65**	**0.56**	**0.67**	**0.69**	**0.70**	**0.67**	26
Aalfotbreen	**0.80**	**0.78**	**0.81**	**0.79**	**0.90**	**0.91**	**0.93**	**0.93**	37
Austdalsbreen	**0.94**	**0.91**	**0.94**	**0.91**	**0.93**	**0.93**	**0.92**	**0.94**	31
Engabreen	**0.84**	**0.79**	**0.82**	**0.77**	**0.84**	**0.75**	**0.82**	**0.81**	38
Graasubreen	**0.38**	0.28	**0.38**	0.28	**0.49**	**0.56**	**0.58**	**0.68**	38
Hansebreen	**0.88**	**0.85**	**0.88**	**0.85**	**0.92**	**0.92**	**0.94**	**0.94**	32
Hellstugubreen	**0.59**	**0.46**	**0.58**	**0.44**	**0.66**	**0.72**	**0.63**	**0.76**	37
Langfjordjoekelen	**0.74**	**0.72**	**0.74**	**0.72**	**0.85**	**0.80**	**0.81**	**0.70**	26
Nigardsbreen	**0.76**	**0.72**	**0.75**	**0.71**	**0.80**	**0.79**	**0.82**	**0.81**	38
Rembesdalskåka	**0.78**	**0.74**	**0.74**	**0.70**	**0.79**	**0.83**	**0.84**	**0.84**	38
Storbreen	**0.77**	**0.80**	**0.77**	**0.79**	**0.82**	**0.86**	**0.82**	**0.85**	30
Ts. Tuyuksuyskiy glacier	**0.43**	0.25	**0.47**	0.28	**0.42**	0.15	**0.41**	**0.49**	31

Significant correlation (p-value < 0.10), **highly significant correlation (p-value < 0.05)**

Table B1. ERA-5 variables used in the study.

Product type	Variable abbreviation	Variable full name
ERA-5 constants	z	Surface geopotential
	anor	Angle of sub-gridscale orography
	isor	Anisotrpy of sub-gridscale orography
	slor	Slope of sub-gridscale orography
	sdor	Standard deviation of orography
ERA-5 single levels	u100, u10	100, 10 m U wind component
	v100, v10	100, 10 m V wind component
	d2m	2 m dew point temperature
	t2m	2 m temperature
	bld	Boundary layer dissipation
	blh	Boundary layer height
	cp	Convective precipitation
	csf	Convective snowfall
	lsp	Large-scale precipitation
	lspf	Large-scale precipitation fraction
	lsf	Large-scale snowfall
	msl	Mean sea level pressure
	sf	Snowfall
	slhf	Surface latent heat flux
	ssr	Surface net solar radiation
	str	Surface net thermal radiation
	sp	Surface pressure
	sshf	Surface sensible heat flux
	tcrw	Total column rain water
	tcsw	Total column snow water
	tp*	Total precipitation
	p54.162	Vertical integral of temperature
	deg0l	0 degrees C isothermal level
ERA-5 pressure levels	t	Temperature
at 1000, 850, 700, 500, 400, 300 hPa	r	Relative humidity
	w	Vertical velocity

*tp was used as precipitation variable.

Table B2. MERRA-2 variables used in the study.

Product type	Variable abbreviation	Variable full name
MERRA-2 constants	PHYS	Surface geopotential height
	SGH	Isotropic stdv of GWD topography
MERRA-2 land surface diagnostics	PRECSNOLAND	Snowfall
	PRECTOTLAND*	Total precipitation
	TSURF	Surface temperature
MERRA-2 single-level diagnostics	CLDPRS	Cloud top pressure
	CLDTMP	Cloud top temperature
	DISPH	Zero plane displacement height
	H100, H850, H500, H250	Height at 1000, 850, 500, 250 mb
	OMEGA500	Omega at 500 hPa
	PBLTOP	Pbltop pressure
	PS	Surface pressure
	Q850, Q500, Q250	Specific humidity at 850, 500, 250 hPa
	QV10M, QV2M	10, 2 m specific humidity
	SLP	Sea level pressure
	T10M, T2M	10, 2 m air temperature
	T850, T500, T250	Air temperature at 850, 500, 250 hPa
	T2MDEW	Dew point temperature at 2 m
	T2MWET	Wet bulb temperature at 2 m
	TQI	Total precipitable ice water
	TQL	Total precipitable liquid water
	TQV	Total precipitable water vapour
	TROPPB	Tropopause pressure, blended estimate
	TROPPT	Tropopause pressure, thermal estimate
	TROPPV	Tropopause pressure, EPV estimate
	TROPQ	Tropopause specific humidity, blended estimate
	TROPT	Tropopause temperature, blended estimate
	U50M, U10M, U2M	50, 10, 2 m eastward wind
	U850, U500, U250	Eastward wind at 850, 500, 250 hPa
	V50M, V10M, V2M	50, 10, 2 m northward wind
	V850, V500, V250	Northward wind at 850, 500, 250 hPa
MERRA-2 analyzed meteorological fields at 1000 to 700 hPa (25 hPa steps) and 700 to 400 hPa (50 hPa steps)	T	Air emperature
	QV	Specific humidity

*PRECTOTLAND was used as precipitation variable.

Table B3. Variables used by the GBR models. The name "variable_down" refers to the variable downscaled at the elevation of the B_w data (linear interpolation from the pressure levels data), "variable_Ppos" refers to the mean of the variable during the accumulation season considering only days with a minimum precipitation of 5 mm, "delta_variable" refers to the difference between "variable_down" and the variable at the original grid of the reanalysis. P_{grid} is the precipitation at the grid of the reanalysis. The variable names have the same roots as those reported in Tables B1 and B2.

GBR model	Static or seasonal	Original	Downscaled	delta_variable	variable_Ppos (Means with $P_{\mathrm{grid}} > 5mm$)
ERA-5	H_obs, H_grid, lat, lon, slope, cos(aspect), sin(aspect), see constants in Table B1, year	r2m*, Wh10M**, Wh100M**, see single-level in Table B1	w_down, t_down, r_down	H_obs−H_grid t_down−t2m r_down−r2m*	w_down_Ppos, t2m_Ppos, t_down_Ppos, delta_t_Ppos, r2m_Ppos*, r_down_Ppos*, delta_r_Ppos*
MERRA-2	H_obs, lat, lon, slope, cos(aspect), sin(aspect) see constants in Table B2, year	Td10M*, RH10M*, Wh10M**, Wh250**, Wh2M**, Wh500**, Wh50M**, Wh850**, see land surface and single-level in Table B2	T_down, Td_down*, QV_down, RH_down*	H_obs−H_grid T_down−T10M Td_down−T2MDEWM QV_down−QV10M RH_down*−RH10M*	T_down_Ppos, T10M_Ppos, delta_T_Ppos Td_down_Ppos, delta_Td_Ppos QV_down_Ppos, QV10M_Ppos, delta_QV_Ppos RH_down_Ppos*, RH10M_Ppos*, delta_RH_Ppos*

*These variables were derived using the equations reported in Section A. **These variables represent the wind speed derived with the two horizontal components.