

Technical note: The impact of land cover representation in WRF on hydrological modeling in a glaciated catchment

Florentin Hofmeister^{1,2}, Xinyang Fan^{3,4}, Madlene Pfeiffer⁵, Ben Marzeion^{5,6}, Bettina Schaepli³, and Gabriele Chiogna⁴

¹Bavarian Academy of Science and Humanities, Geodesy and Glaciology, Munich, Germany

²Chair of Hydrology and River Basin Management, Technical University of Munich, Munich, Germany

³Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

⁴Department of Geography and Geosciences, GeoZentrum Nordbayern, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany

⁵Institute of Geography, University of Bremen, Bremen, Germany

⁶MARUM – Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany

Correspondence: Xinyang Fan (xinyang.fan@unibe.ch)

Abstract. This technical note addresses several aspects, limitations, and potential solutions in applying a complex modeling chain of a Regional Climate Model (RCM) and a physics-based hydrological model (HM) for climate sensitivity studies in high-elevation alpine regions. We used the Weather Research & Forecasting Model (WRF) to dynamically downscale a global reanalysis product (20CRv3) to a 2 km x 2 km spatial and hourly temporal resolution from 1969 to 2015 as forcings for an HM (WaSiM). This requires transferring the forcing data to the much finer grid resolution (i.e., 25 m) of the HM, while accounting for the complex topography and plausible sub-daily precipitation and temperature lapse rates (TLRs). Thus, we developed a novel workflow for extracting hourly TLRs from different vertical layers of the WRF atmosphere. In addition, we adjusted 2 m WRF relative humidity based on the hourly TLRs and corrected WRF precipitation frequencies using observation data. Our case study demonstrates the impact of TLRs computed from different WRF layers (i.e., 2 m and free atmosphere) on the HM results of a highly glaciated Alpine catchment in the European Alps. In a multi-data evaluation procedure, we found that the TLRs and the HM results are significantly dependent on the coarse surface properties of WRF. By spatially averaging hourly temperature and geopotential height at each WRF pressure level, the calculated TLRs are less dependent on WRF land cover types. Temperature-sensitive processes such as snow and glacier evolution, as well as streamflow response, are more realistically simulated when the HM is forced by TLRs from the WRF free atmosphere rather than by simulated near-surface temperatures. A bivariate correction accounting for temperature and relative humidity further enhanced the simulated magnitude of summer streamflow. The HM results are consistent with observation data over a simulation period beginning in 1969, suggesting the corrected WRF temperature can reliably reproduce the non-stationarity in local temperature observations.

1 Introduction

The European Alps are particularly affected by climate change, as the observed temperature increase is approximately twice the global average, meaning an average warming of 1.8 °C since 1880 (Kotlarski et al., 2023). However, this trend has not been

linear since pre-industrial times, as warming has significantly accelerated in the last 30 years (Nigrelli and Chiarle, 2023), with considerable consequences for the entire Alpine geosystem and its inhabitants (Matiu et al., 2021; Marcolini et al., 2017, 2019; Morin et al., 2021; Magnusson et al., 2020). In addition, the accelerated warming rates of the European Alps are probably leading to heavier rainfall intensities and consequently to greater and sudden floods (Wilhelm et al., 2022). Understanding the drivers of past climate-induced changes in the hydrological response of Alpine catchments is relevant for developing future adaptation strategies for water resources and risk management.

The study of climate-induced long-term changes mainly depends on station observations. However, observations back to the last Little Ice Age (LIA) around 1850 are limited both in space and time in the Alpine region. Homogenized temperature and precipitation records are even available before 1850 (Auer et al., 2005, 2007), but they are not equally distributed and mainly represent lower-laying areas of the Alpine region (van Tiel et al., 2022). The linear extrapolation of these temperature records to higher elevations is problematic because several studies have found different trends in the elevation-dependent warming (e.g., Rangwala and Miller, 2012; Kuhn and Olefs, 2020). High-resolution Regional Climate Models (RCMs) bridge this spatiotemporal gap by dynamically downscaling global reanalysis products to a much finer grid resolution of a few kilometers. To account for the complex orography of the Alps, multiple studies have demonstrated the benefits of high-resolution RCMs for regional climate change studies in the European Alps (Gómez-Navarro et al., 2015; Warscher et al., 2019). A frequently used model for long-term atmospheric simulations is the Weather Research and Forecasting (WRF) model (Powers et al., 2017). The main advantage of dynamical downscaling with atmospheric models is the physical consistency, which means that all fields are downscaled simultaneously, so the interdependencies between downscaled variables are resolved (Berg et al., 2024). The major disadvantage is the high computational effort required to run the WRF model: Even with high-performance computing, centennial simulations over an area of approximately 24,416 km² at 2 km spatial and 1 hour temporal resolution can take several months.

Although many high-resolution WRF climate simulations have a spatial resolution of a few kilometers (Gómez-Navarro et al., 2015; Warscher et al., 2019; Velasquez et al., 2020), they are still far too coarse for resolving localized precipitation patterns or small-scale processes such as snow redistribution in high-elevation areas (Berg et al., 2024). Moreover, the simulated near-surface heat fluxes depend on the spatial resolution of the WRF elevation model and the resolution and accuracy of land cover classes (i.e., land-atmosphere interactions). In addition, elevation-dependent processes, such as temperature and precipitation lapse rates, are not sufficiently resolved by RCMs at the sub-kilometer scale due to their coarser topography, which often under-represents terrain features. Comprehensive correction procedures are therefore necessary to rescale and correct the coarse-scale RCM data to a much finer spatial resolution in mountainous regions, as usually required by distributed hydrological models (HM). Velasquez et al. (2020) developed a three-step procedure for bias-correcting WRF precipitation over a complex terrain in the Swiss Alps. This approach explicitly considers orographic characteristics. The first step is the separation of precipitation concerning different orographic characteristics; the second is the adjustment of daily precipitation with very low intensity; and the third is the application of empirical quantile mapping (EQM). EQM consists of adjusting the quantile values from a simulation to those from the observations through a transfer function (Velasquez et al., 2020). Therefore, the EQM method is based on the availability of meteorological time series with sufficient temporal coverage and quality. How-

ever, meteorological observations are particularly limited in mountainous areas, especially in relation to reliable precipitation records from high altitudes (> 2000 m a.s.l.) (Foehn et al., 2018). In combination with the non-stationarity of meteorological observations (Teutschbein and Seibert, 2012), the evaluation of bias correction methods in data-scarce mountainous regions is particularly challenging.

60 HMs can serve as an additional validation method for the reliability of bias corrections, as adjusted forcing data will affect the quality of HM results (Teutschbein and Seibert, 2012; Shrestha et al., 2017; Hakala et al., 2018; Hofmeister et al., 2023; Meyer et al., 2019). It can be assumed that a bias correction must also yield more realistic and plausible HM results than a model forcing based on the uncorrected RCM (Teutschbein and Seibert, 2012; Meyer et al., 2019). However, the focus should not only be on good and plausible streamflow results for alpine catchments, but also on the reliability of the other hydrological
65 processes (Tuo et al., 2018; Hofmeister et al., 2023; Schaffhauser et al., 2025). Mainly, elevation-dependent processes, such as snowfall and melt generation, depend on a reliable representation of temperature and precipitation lapse rates in mountainous regions (Hakala et al., 2018; Hofmeister et al., 2023; Berg et al., 2024; Meyer et al., 2019). Although most of the previous evaluations of bias correction methods with HMs focused on monthly or daily model results (Teutschbein and Seibert, 2012; Shrestha et al., 2017; Hakala et al., 2018; Meyer et al., 2019), Faghih et al. (2022) found that even the diel cycle of variables
70 simulated by RCMs is biased, which raises issues about the necessity of correcting such biases prior to streamflow simulations at the subdaily timescale. They recommend correcting the diel cycle of RCM variables before using them for hydrological simulations to accurately represent the diel cycle of summer streamflow in small catchments (i.e., < 500 km²) (Faghih et al., 2022).

The objective of this technical note is to demonstrate the variability and efficient correction of WRF near-surface temperature by spatially averaging hourly temperature and geopotential height at each WRF pressure level. In addition, we performed a bivariate correction for WRF temperature and relative humidity. Besides statistical evaluations, we assess the spatiotemporal variability of the uncorrected and corrected WRF temperature, as well as of the bivariate correction approach, using a multi-data approach that includes a fully-distributed and physics-based HM (WaSiM) in a highly glaciated catchment in the eastern European Alps. These evaluations are performed at different temporal scales, spanning up to 42 years. Besides the
80 novel temperature bias correction method, this technical note provides comprehensive and valuable insights into the complex interdependencies among the cryo- and hydrosphere in highly glaciated catchments, and the need for accurate spatiotemporal representation of temperature, precipitation, and relative humidity in rugged topography with large elevation gradients for climate sensitivity studies.

2 Material and methods

2.1 Study site

85 The research site is located in the upper Kauner Valley in the Tyrolean Alps (Austria), has an elevation range of 1810-3535 m a.s.l. (Altmann et al., 2020) and is mainly drained by the Fagge River, a tributary of the upper Inn River system. The drainage area of the Fagge at the stream gauge Gepatschalm (GA), south of the Gepatsch reservoir (Fig. 1), covers about 54 km². The

actual drainage in this study is slightly larger (62 km²), as it extends beyond the stream gauge GA and the catchment is closed at the inflow to the Gepatsch reservoir. The streamflow regime of the Fagge is strongly influenced by snow and glacier melt, which results in distinct seasonality with low flows in winter and high flows in summer. The glacier coverage is relatively high with a 30 % fractional cover in 2015 (Buckel and Otto, 2018), but has declined from 39 % in 1969 according to the first Tyrolean Glacier Inventory (Patzelt, 2013). Accordingly, the entire geosystem of the upper Kauner Valley has undergone profound changes over the respective period, which were investigated in multiple studies (e.g., Altmann et al., 2020, 2024; Piermattei et al., 2023). With several existing long-term hydrometeorological observation stations (Fig. 1), the upper Kauner Valley is an ideal study site for HM applications (e.g. Förster et al., 2016; Rogger et al., 2017; Pesci et al., 2023). Although no continuous glacier mass balance data are available, the high-quality hydrometeorological instrumentation, which is often even better than that of the United States Geological Survey Benchmark Glacier Project (Schaffhauser et al., 2025), makes the upper Kauner Valley an ideal case study for cryohydrological research. In 2020, scree (vegetation cover < 20 %) was the dominant land cover at 32 %, closely followed by glacier cover at 27 %. Rocks with vegetation cover about 12 % of the upper Kauner Valley, followed by grassland and dwarf shrubs. Subalpine coniferous forest only covers 1.3 % of the research area along the valley bottom. The land cover classification was carried out by Ramskogler (2023).

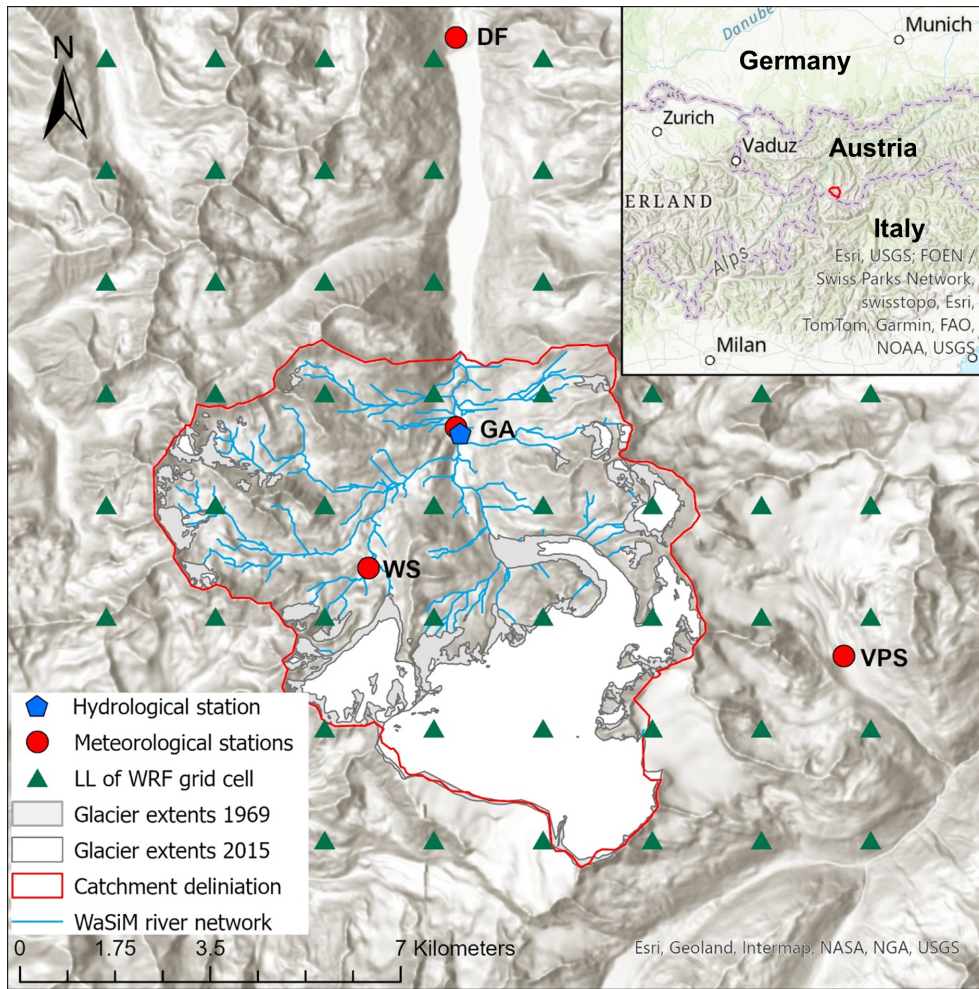


Figure 1. Overview map of the research area upper Kauner Valley in Tyrol (Austria). Green triangles illustrate the lower left coordinates of the 2 km WRF grid cells. Meteorological stations are indicated with red circles. Hydrological stations are represented with a blue symbol. Glacier extents are plotted for 1969 from the first Tyrolian Glacier Inventory (GI 1) and 2015 from the fourth inventory (GI 4). Catchment delineation and river network were computed based on the official digital elevation model of Tyrol from 2006, which is available from the Province of Tyrol and downloadable via the following URL: <https://www.data.gv.at/>.

2.2 Hydrometeorological observation data

High-resolution (i.e., 15 minute intervals) streamflow data at gauge Gepatschalm (GA) were provided by TIWAG-Tiroler
105 Wasserkraft AG, the local hydropower plant operator, for the period 1971 to 2022. Larger data gaps of several months mainly affect the first four years of streamflow observations at GA. Most of the meteorological stations in the upper Kauner Valley (Dammfuss, Gepatschalm, Weisssee), listed in Table 1, are also operated by TIWAG. The TIWAG temperature time series includes only a low fraction (i.e., <1 %) of data gaps. The Bavarian Academy of Science operates multiple stations at the

Table 1. Metadata of the meteorological stations and statistics about the data availability of meteorological observations. The coordinates are given as UTM coordinates for zone 32N (epsg code: 32632). Although several other meteorological variables may be available for the stations, we only indicate the variables used in this technical note. The amount of NAs is given in hourly intervals, except for the VG station, where it is given in daily intervals.

Station	Station code	Easting	Northing	Elevation (m a.s.l.)	Variable	Start	Amount of NA	Share of NAs (%)
Vergoetschen	VG	633168.5350	5211440.0060	1269	P	1895	3598	7.92
Dammfuss	DF	632260.0676	5202225.1060	1770	T	1988	844	0.36
Gepatschalm	GA	632251.1494	5195058.3820	1900	T	2009	18	0.03
Weisssee	WS	630648.4005	5192468.2380	2540	T	2006	15	0.03
Vernagt Pegel	VPS	639386.0800	5190853.3710	2640	T	1975	24550	7.00

nearby Vernagtferner. The main station at Vernagt Pegel has recorded hourly meteorological variables since 1975 and is the highest station (2640 m a.s.l.) with the longest temporal coverage in the region. Due to some failures of the station and sensor, the temperature time series has about 7 % of data gaps for the period 1975-2015. The meteorological station in Vergoetschen (1269 m a.s.l.) is operated by the HD-Tyrol and is located in the lower Kauner Valley. This station was chosen to correct the precipitation frequency of the regional climate model (WRF) due to its high reliability and long temporal coverage (since 1895) of daily precipitation measurements. Longer data gaps at the VG stations are mainly due to World War II. As all stations are still in operation, only the start year of the observation is given in Table 1. Unfortunately, there is no overlap in time with the meteorological station installed in 2017 on the Weissseespitze (Stocker-Waldhuber et al., 2022), which is located at an altitude of 3500 m in the upper Kauner Valley.

2.3 Snow observation data

The snow observation data are used to evaluate the simulation results of the HM at the point scale (with station-based snow water equivalent (SWE) data) and at the catchment scale with binary snow cover data retrieved from an optical remote sensing product (i.e., MODIS).

2.3.1 Calculation of SWE data

Snow depth (SH) measurements within the research site were available only at the TIWAG station Weisssee since 2007. To transform daily SH to daily SWE, we applied the recently developed deltaSnow model (Winkler et al., 2021). SWE is modeled using a process-based multi-layer approach, incorporating empirical regression models based on the relationship between density and diverse at-site parameters (Winkler et al., 2021). Accordingly, the deltaSnow model lies between simple empirical regression models (such as Jonas et al. (2009)) and full thermodynamic models (such as SNOWPACK (Lehning et al., 2002) or Crocus (Vionnet et al., 2012)), offering a physically based but data-sparse SWE estimator. The deltaSnow model does not require meteorological forcing, such as temperature or precipitation data. This is particularly useful where SWE measurements are sparse, but snow depth records are available, as is the case in the upper Kauner Valley. The deltaSnow model (version 1.0.2)

is available in the "nixmass" R package and takes continuous daily SH records without data gaps as input. The deltaSnow model has seven tunable parameters - including new-snow density (ρ_0), maximum density (ρ_{max}), viscous compaction parameters, overburden strain coefficients, and the threshold deviation τ . We used the same model parameters as published by Winkler et al. (2021), which were calibrated and evaluated against station observations from different altitudes and climatic regions of the European Alps.

2.3.2 Snow cover data

Daily and nearly cloud-free MODIS snow cover maps are available for the entire European Alps at 250 m grid resolution since summer 2002 (Matiu et al., 2019a, b). Despite the relatively coarse grid resolution compared to other optical snow products (e.g., Landsat, Sentinel 2), MODIS snow cover maps can provide valuable information on seasonal snow cover evolution at high temporal resolution (i.e., daily). Due to their high temporal resolution and global coverage, MODIS snow cover data have already been used in many studies focusing on hydrological modeling in high-elevation catchments (e.g., Duethmann et al., 2014). After resampling and cropping the MODIS snow cover maps to the respective catchment resolution (i.e., 25 m) and domain, we set glaciated raster cells to no data as they are permanently snow-covered and would bias the comparison with the snow cover simulations of the HM. We use glacier extents from the third Tyrolean glacier inventory from 2006 (Abermann et al., 2012). To ensure consistency between observed and simulated snow cover maps, we set the same grid cells to no data in both products and normalize the fractional snow-covered area (fscs) values from 0 (no snow) to 1 (completely snow-covered). Snow cover data with higher spatial resolution from Sentinel 2 (Hofmeister et al., 2022) could not be used for this case study, as there is not a sufficiently long overlap between the data. The WRF data ends in 2015, the year Sentinel 2 was launched. Therefore, evaluating fractional snow cover with the coarser MODIS product is still better than no evaluation at all.

2.4 Dynamical downscaling with WRF

The model forcing of the HM is generated with the Weather Research and Forecasting (WRF) model (version 4.3), which is based on fully compressible and non-hydrostatic equations (Skamarock et al., 2008). The global Twentieth Century Reanalysis version 3 (20CRv3) data set (Compo et al., 2011; Giese et al., 2016; Slivinski et al., 2019), with a spatial and temporal resolution of $1^\circ \times 1^\circ$ and 3 h is used as driving data (initial and boundary conditions). The simulation is performed in three nested domains, with a grid spacing of 18 km (domain 1), 6 km (domain 2), and 2 km (domain 3). The simulated data in domain 3 are stored at a temporal interval of 1 h. The vertical discretization is based on 49 virtual atmospheric layers, the so-called eta levels, which allows the WRF model to resolve atmospheric flow near complex terrain more accurately. The Yonsei University scheme is selected to calculate the planetary boundary layer (Hong et al., 2006). For further information on the WRF model configuration, refer to Altmann et al. (2024). The considered WRF grid cells of this case study cover an area of about 200 km², as also parts of the neighboring valleys are included (Fig. 1).

We use the Noah land surface model (LSM), which is one of the most widely used land surface components in WRF and plays a crucial role in simulating interactions between the land surface and the atmosphere. The LSM calculates the energy and mass transfer at the ground interface, providing bottom boundary conditions to the atmospheric model WRF (Tomasi

et al., 2017). The Noah LSM is based on 24 land use categories originating from the United States Geological Survey (USGS) (Anderson et al., 1976). From the 24 land cover categories, only seven are represented in the WRF domain of the upper Kauner Valley (Fig. S1 in the supplementary material). Namely, *Woodland Mosaic*, *Grassland*, *Shrubland*, *Evergreen Needleleaf*, *Wooden Wetland*, *Wooded Tundra*, and *Snow or Ice*. The most frequent category of the WRF Kauner Valley domain is *Wooded Tundra* (15 grid cells), followed by *Snow or Ice* (11 grid cells). The latter category means permanent snow- and ice-covered areas where snow and ice cover persist for at least ten months of the year, typically exceeding 60 % of the area. This category signifies a large presence of snow and ice as a permanent landscape feature (Anderson et al., 1976). The category *Snow or Ice* covers most of the higher-elevation research site (i.e., > 2900 m.a.s.l. in Fig. S1), including the Gepatsch glacier. A separate map of the WRF land use categories at 2 km resolution is shown in the supplementary material (Fig. S2).

Figure 2 shows the relation between grid cell surface elevation (m) and simulated temperatures averaged over a test period of three months (i.e., from August to November) stratified according to the land cover per grid cell. The temperatures shown correspond to different eta levels, the 2 m above-ground temperature (T2) and the surface (or skin) temperature (TSK). Eta level 1 (Fig. 2c) corresponds on average to an altitude of about 17 m above the ground surface over the three-month test period. It is computed by subtracting the mean WRF geopotential height of eta level 1 from the mean WRF surface height (2785 m) of the upper Kauner Valley. Eta level 49 (Fig. 2f) has an average altitude of approximately 30 km above the ground surface. The TSK and T2 elevation plots (Fig. 2) show much lower temperatures for snow and ice-covered grid cells and a certain scatter, which indicates that these simulated temperatures are strongly influenced by surface properties, with a snow-related surface cooling effect present up to T2. The elevation gradients (both with and without snow) of TSK and T2 are lower than at the next higher eta levels, which underlines the well-known fact that near-surface temperature gradients are lower than in the free atmosphere (Rist et al., 2020). In this context, a lower gradient means a smaller change of temperature with elevation. The free atmosphere is typically defined in WRF as the part of the atmosphere above the planetary boundary layer, where turbulent mixing from the surface becomes negligible. However, the exact eta level or geopotential height that marks the transition to the free atmosphere is not fixed. It varies dynamically in space and time, depending on surface properties, atmospheric stability, and time of day (Hong et al., 2006).

From eta level 1 to eta level 20 (Fig. 2c and d), the linear gradient between WRF temperatures and surface elevation shows almost no scatter for the first eta levels. Beyond eta level 20, the scatter starts increasing, and there is no relationship anymore at level 49. This can be explained by the fact that eta levels 1 to 20 follow the terrain, while the remaining eta levels up to 49 are pressure levels. The lapse rate at eta level 1 equals $-0.79\text{ }^{\circ}\text{C } 100\text{ m}^{-1}$, i.e., it is higher than in the free atmosphere (which corresponds to $-0.65\text{ }^{\circ}\text{C } 100\text{ m}^{-1}$) for the three-month test period and gradually decreases for higher levels.

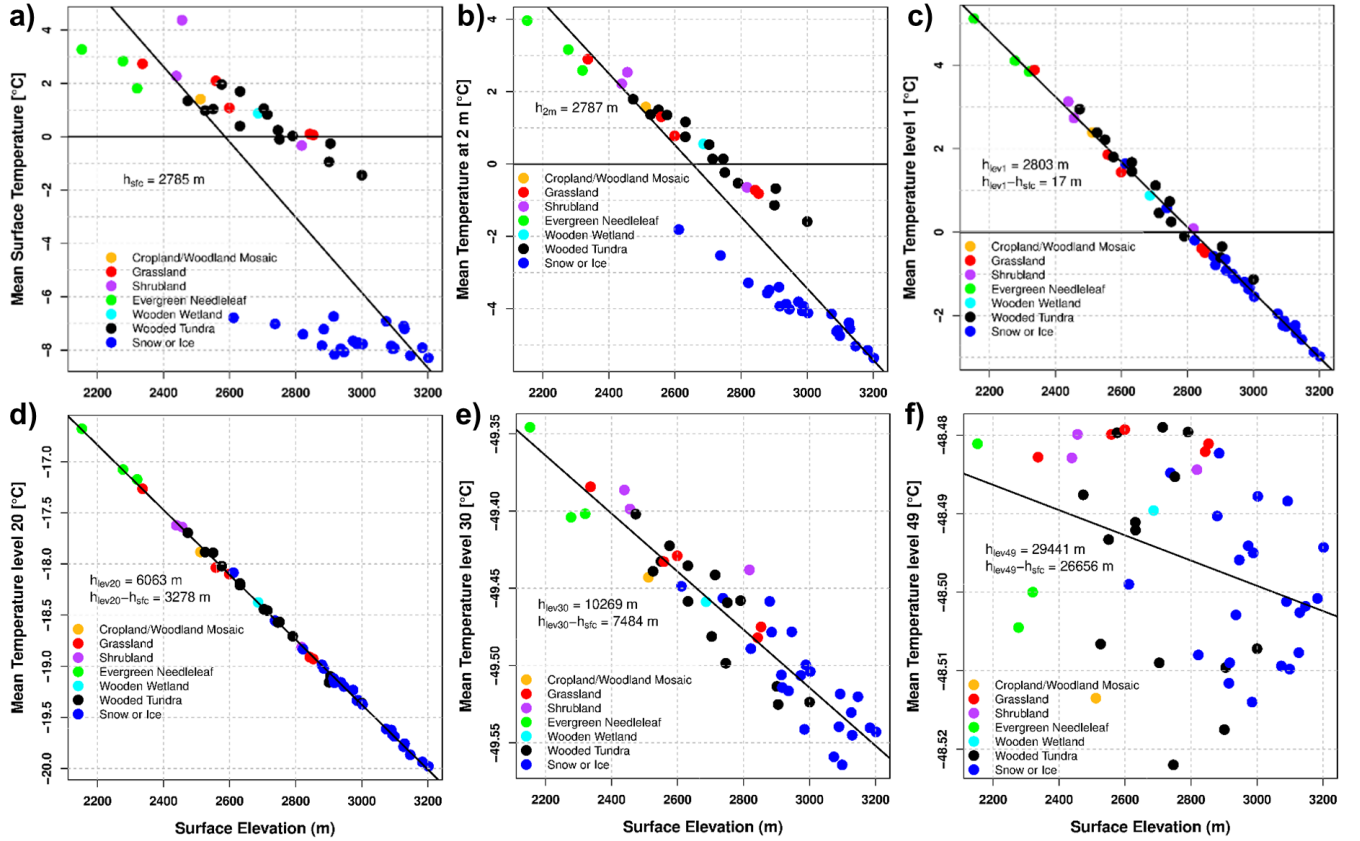


Figure 2. Mean WRF temperature for different land use categories and surface elevations (m) for the Kauner Valley domain covering a test period of three months. The diagonal black line shows the linear gradient between elevation and temperature obtained by least-squares fitting to all data points. The numbers in the subplots indicate the respective geopotential height and the height above ground.

2.4.1 Correction of the WRF temperature

The adjustment of the WRF temperature is done in a three-step approach on hourly time series to consider sub-daily changes in the temperature lapse rates, TLRs. As correcting the WRF temperatures is a complex task due to the multidimensionality of the data, we have summarized the individual processing steps of the developed workflow in Figure 3. First, we compute spatial averages of the hourly temperature of eta levels 1 and 20 over the 200 km² domain of the upper Kauner Valley by applying Equation 1:

$$\bar{T}_{levelX,t} = \frac{1}{N} \sum_{i=1}^M \sum_{j=1}^L T(t,i,j) \quad (1)$$

where $\bar{T}_{levelX,t}$ is the spatial mean of temperature [°C] of the respective eta level (*levelX*) at time *t*. In our case, time *t* has an hourly resolution. $T(t,i,j)$ is the variable value at time *t*, latitude index *i*, and longitude index *j*. *M* is the total number of

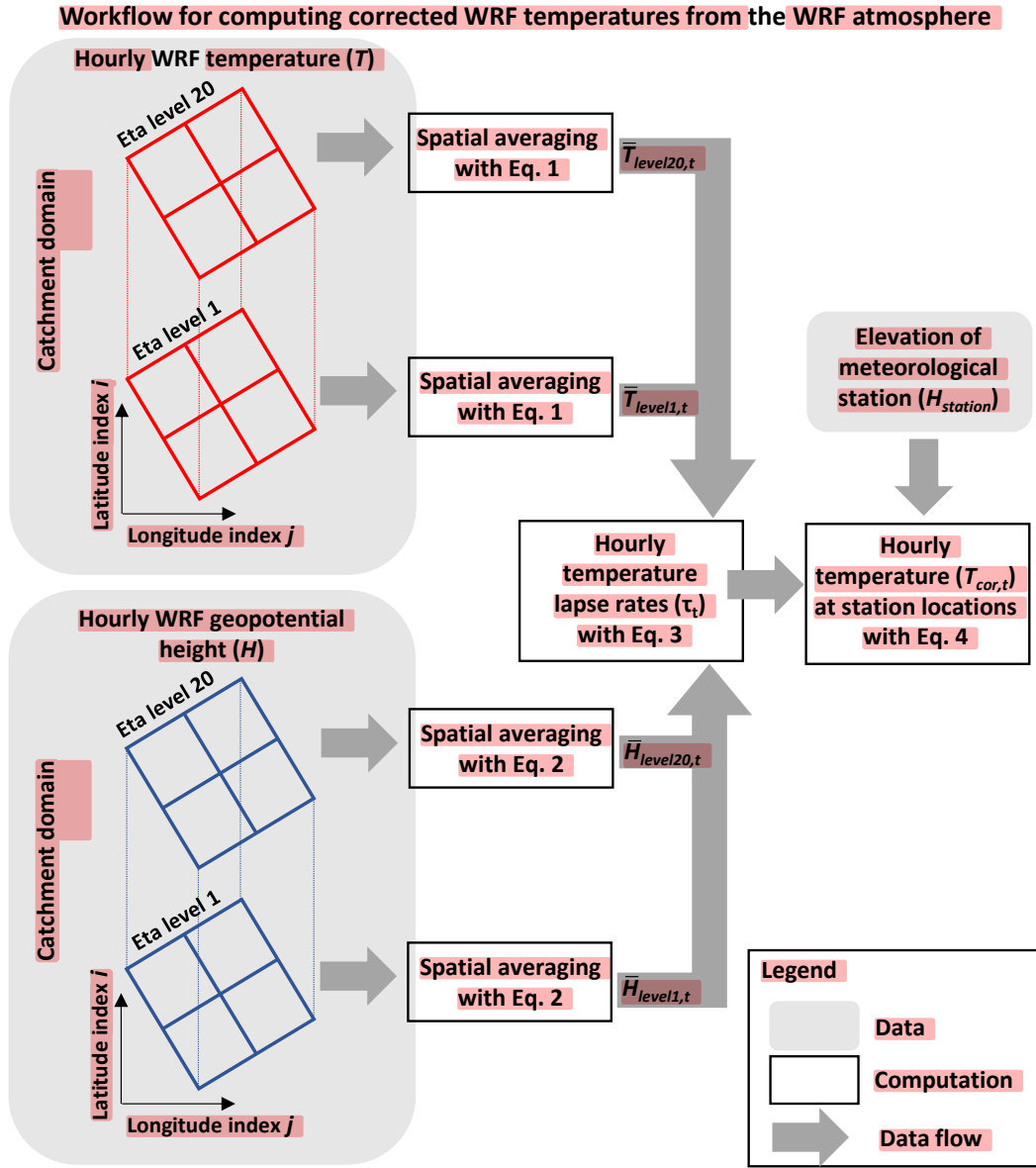


Figure 3. Developed workflow for computing corrected WRF temperatures from the multidimensional WRF atmosphere data.

latitude points, and L is the total number of longitude points. N is the total number of spatial grid points. As the WRF grid of the Kauner Valley has 7 rows (M) and 7 columns (L), a total of 49 grid cells (N) are considered in Equation 1. Spatial averaging reduces the TLRs' dependency on WRF land cover types, especially at lower eta levels.

205 We use the same approach to calculate the spatial averages of the hourly geopotential height following Equation 2:

$$\bar{H}_{levelX,t} = \frac{1}{N} \sum_{i=1}^M \sum_{j=1}^L H(t,i,j) \quad (2)$$

where $\bar{H}_{levelX,t}$ is the spatial mean of geopotential height [m a.s.l.] of the respective eta level (*levelX*) at time t . As the temperature and geopotential height grids have the same spatiotemporal dimensions, the variables of $\bar{H}_{levelX}(t,i,j)$ are identical to $\bar{T}_{levelX}(t,i,j)$.

210 In a second step, we compute the hourly TLRs based on the spatial averages (from Equ.1 and 2) using Equation 3:

$$\tau_t = \frac{\bar{T}_{level_20,t} - \bar{T}_{level_1,t}}{\bar{H}_{level_20,t} - \bar{H}_{level_1,t}} \quad (3)$$

where τ_t [$^{\circ}Cm^{-1}$] is the TLR at time step t , $\bar{T}_{level_20,t}$ [$^{\circ}C$] is the spatial average temperature $\bar{T}$ of eta level 20 at time step t , and $\bar{H}$ is the spatial average altitude of the eta levels in m a.s.l. (i.e., the geopotential height). Eta level 20 corresponds to a geopotential height of about 6000 m a.s.l., which represents a mean height above ground of 3278 m. We have chosen the altitude range from eta level 1 to 20 for two reasons: eta level 1 still includes minor temperature differences due to the land-atmosphere interactions (Fig. 2c), which are more pronounced in summer (Fig. S3a). On the contrary, eta level 20 does not show any influence of the surface properties (Fig. 2d) throughout the season (Fig. S3d). Accordingly, the higher variability and magnitudes from lower levels will be averaged out by the upper levels. Although the transition of mean TLR from eta level 1 to 20 is not linear, we apply the linear function of Equation 3 for computing mean TLR over the entire altitude range.

220 The red line in Figure 4 represents the mean TLR between eta levels 1 and 20. Including eta level 1 in the linear equation ensures that temperature fluctuations are not completely decoupled from the surface and topography of the research area. The TLRs of Figure 4 are computed between two neighboring eta levels and illustrate the sensitivity of TLRs to eta level selection. This calculation was based on the mean temperature of all grid cells in the Kauner Valley for each level. Except for the lapse rate between the first two eta levels, all other lapse rates up to eta level 20 range from -0.52 to -0.67 [$^{\circ}C100m^{-1}$]. Further

225 information for the use of a linear gradient between levels 1 and 20 will be presented in the Results section (3.1).

In the third step, we compute the hourly corrected WRF temperature (T_{cor}) [$^{\circ}C$] at actual altitudes of meteorological stations by applying Equation 4:

$$T_{cor,t} = T_{level_1,t} + \tau_t * (H_{station} - H_{level_1,t}) \quad (4)$$

where $H_{station}$ represents the station elevation in m a.s.l. and $H_{level_1,t}$ the mean geopotential height of the eta level 1 (i.e., 2803 m a.s.l.).

We compute the overall mean TLR using Equation 5:

$$\bar{\tau}_t = \frac{1}{Y} \sum_{t=1}^Y \tau(t) \quad (5)$$

where $\bar{\tau}_t$ is the mean TLR [$^{\circ}Cm^{-1}$] over time and Y is the total number of time steps. Accordingly, the computed mean TLR is -0.57 $^{\circ}C 100 m^{-1}$ from eta level 1 to 20 over the period 1973-2015. The summer TLR (i.e., May to September) is with -0.58 $^{\circ}C$

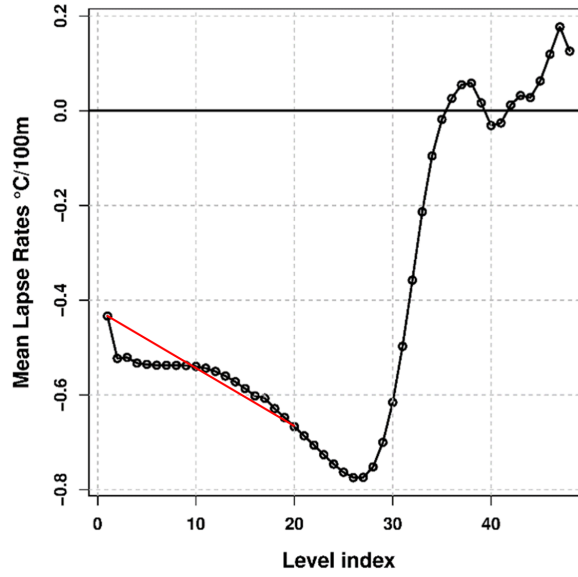


Figure 4. Mean WRF temperature lapse rates in $^{\circ}\text{C} (100 \text{ m})^{-1}$ per eta level. The TLRs are computed for two pairs of consecutive eta levels. The red line represents a linear approximation of the lapse rates based only on the lapse rates from eta levels 1 and 20. The mean elevation of eta level 1 is about 2800 m a.s.l., about 6100 m a.s.l. for eta level 20, and about 10300 m a.s.l. for eta level 30. The elevation range of the Kauner Valley is from 1810 to 3535 m a.s.l..

235 100 m^{-1} slightly larger than the winter TLR ($-0.56 \text{ }^{\circ}\text{C} 100 \text{ m}^{-1}$). We calculate mean TLRs only for direct comparison between the different calculation methods and eta level combinations. However, these mean TLRs are not used in the calculation of hourly TLRs (i.e., Equ. 4 and 6) or in hydrological modeling.

To compare the uncorrected WRF temperature with the corrected WRF temperature (T_{cor}) and the observed station temperature, we account for the height difference between the WRF grid cell (H_{WRF}) and the actual station (H_{station}) by multiplying
 240 the height difference with the hourly TLR $\tau_t [^{\circ}\text{C m}^{-1}]$ at time step t . The altitude adjusted uncorrected WRF temperature $T_{\text{uncor,alt,t}}$ is thus computed by applying Equation 6:

$$T_{\text{uncor,alt,t}} = T_{\text{uncor,t}} + \tau_t * (H_{\text{WRF}} - H_{\text{station}}) \quad (6)$$

where $T_{\text{uncor,t}}$ is the hourly uncorrected 2 m temperature of a WRF cell, H_{WRF} represents the WRF altitude, and H_{station} the actual station altitude in m a.s.l.. Table S1 in the supplementary material lists the metadata for the meteorological stations,
 245 along with the corresponding WRF altitudes and land use categories at each station location. The largest delta between station and WRF altitude is for the lower-elevation stations (i.e., Dammfuss delta of 525 m, Gepatschalm delta of 420 m), meaning the WRF altitude overestimates the actual station altitude. The Weisssee station has an altitude almost identical to that of the Weisssee station, with only a 10 m difference.

2.4.2 Correction of the WRF relative humidity

As relative humidity (RH) is not an independent variable and depends non-linearly on air temperature and moisture, we performed a bivariate correction of RH considering the corrected WRF temperature. For this, we applied the empirical August-Roche-Magnus (ARM) formula, which is often used in meteorology to calculate the saturation vapor pressure of water at a given temperature, helping determine RH (Alduchov and Eskridge, 1996; Huang, 2018). First, we define the hourly saturation vapor pressure $e_s(T_{\text{uncor},t})$ [hPa] over water and ice surfaces using the uncorrected WRF $T_{\text{uncor},t}$ in Equation 7. In a second step, we perform the same computation using the corrected WRF $T_{\text{cor},t}$ for deriving hourly saturation vapor pressure $e_s(T_{\text{cor},t})$ [hPa]. Lastly, we recompute the corrected hourly RH $RH_{\text{cor},t}$ [%] by holding vapor pressure constant and considering the corrected temperature with Equation 7:

$$RH_{\text{cor},t} = RH_{\text{uncor},t} \frac{e_s(T_{\text{uncor},t})}{e_s(T_{\text{cor},t})}, \quad e_s(T_t) = \begin{cases} 6.1094 \exp\left(\frac{17.625 T_t}{243.04 + T_t}\right), & T_t \geq 0^\circ\text{C} \\ 6.1121 \exp\left(\frac{22.587 T_t}{273.86 + T_t}\right), & T_t < 0^\circ\text{C} \end{cases} \quad (7)$$

where $T_{\text{uncor},t}$ is the hourly uncorrected and $T_{\text{cor},t}$ is the hourly corrected 2 m temperature [$^\circ\text{C}$] of a WRF cell. $RH_{\text{uncor},t}$ represents the hourly uncorrected and $RH_{\text{cor},t}$ the corrected 2 m RH [%] of a WRF cell. If the hourly T_t is equal to or larger than 0°C , the ARM formula for the water surface is considered. For temperatures smaller than 0°C , we used the ARM formula for the ice surface. Both ARM coefficients for water and ice are taken from Alduchov and Eskridge (1996), which are commonly used in meteorological data processing and humidity studies (e.g., Huang (2018)). We constrained the corrected RH values to the range 0-100 %, meaning values $< 0\%$ were set to 0% and $> 100\%$ were set to 100% .

2.4.3 Correction of the WRF precipitation

Climate models, including the WRF model, often overestimate the frequency of light rainy days while simultaneously underestimating the total amounts of extreme observed precipitation (Lazoglou et al., 2024). The model's tendency to overestimate the occurrence of light precipitation events is known as the "drizzle bias" (Lazoglou et al., 2024; Velasquez et al., 2020). We follow the approach of Lazoglou et al. (2024) to adjust the number of rainy days based on the assumption that the relationship between observed and simulated rainy days remains the same in time (thresholding). The correction of the dry-days frequency using thresholding has already been applied in several studies before (e.g., Smitha et al., 2018; Velasquez et al., 2020). Thresholding is a widely used, simple, and efficient method for correcting the frequency of wet days. It involves converting all simulated values below a designated threshold to zero (Lazoglou et al., 2024). We adjust the WRF precipitation frequency by computing wet-day frequencies in both monthly observed and simulated datasets. As the "drizzle bias" has a distinct seasonality (Fig. S4) and is consequently more pronounced in the winter period, we computed mean monthly thresholds based on the observed and simulated counts of rainy days. The first step is to compute the mean monthly precipitation days based on daily station data. We select the meteorological station Vergoetschen (1280 m a.s.l.), operated by HD-Tyrol, as it has a long record of daily

Table 2. Monthly mean and standard deviation of observed precipitation days at station Vergoetschen and derived threshold values for correcting the WRF precipitation frequency from 1971 to 2015.

Month	Mean prec days	SD prec days	Threshold (mm h ⁻¹)
1	9	4.83	0.41
2	8	3.81	0.49
3	10	4.06	0.42
4	11	3.93	0.48
5	13	4.37	0.43
6	16	4.52	0.42
7	15	3.6	0.29
8	15	4.58	0.27
9	12	3.96	0.23
10	10	4.43	0.29
11	10	4.17	0.33
12	10	4.32	0.41

precipitation without data gaps from 1971. Observed precipitation days are defined as days with precipitation exceeding 0.01 mm d⁻¹. Observed mean monthly precipitation days from 1971 to 2015 are listed in the second column of Table 2.

280 We estimate monthly-varying precipitation thresholds to identify WRF time steps whose precipitation is to be set to zero for frequency matching. These thresholds are identified based on the absolute minimum error between observed and WRF precipitation days for hourly resolution (Tab. 2). Figure S5 in the supplementary material illustrates the optimization approach. In the second step, the hourly WRF precipitation values below the monthly thresholds (Tab. 2) are set to 0, assuming that the monthly-varying precipitation thresholds are stationary over the period 1971 to 2015, across space, and elevation. The 285 correction of WRF precipitation only leads to a reduction of the annual precipitation over the research site by 4 % compared to the uncorrected, as mainly low precipitation frequencies are set to 0. WRF precipitation intensities (i.e., quantiles) are not evaluated or adjusted due to a lack of reliable precipitation observations from higher elevations (> 2000 m a.s.l.).

2.5 Statistical tests

We use different statistical tests to evaluate the properties of observed and modeled time series. We test the significance of 290 trends with the non-parametric and modified Mann-Kendall test for serially correlated data using the Hamed and Rao (1998) variance correction approach, which is available in the "modifiedmk" R package version 1.6. This approach reduces significant autocorrelation in time series by calculating the effective sample size. The effective sample size is calculated by using the ranks of significant serial correlation coefficients, which are then used to correct the inflated (or deflated) variance of the test statistic (Hamed and Rao, 1998).

295 We apply the non-parametric Kolmogorov–Smirnov (KS) test to evaluate the similarity between empirical cumulative distribution functions (ECDFs). The KS test compares the two ECDFs by quantifying the maximum distance between them. We use the *ks.test* function (version 1.5.1) in R. Besides the p-value, the output of the KS test includes the D-statistic, which is the dimensionless maximum absolute difference between ECDFs. In addition, we compute the maximum vertical difference between the ECDFs.

300 The non-parametric Pettitt test detects a sudden change in the central tendency (i.e., mean) of the time series (Pettitt, 1979). The H0-hypothesis, no change, is tested against the HA-hypothesis, change. The applied Pettitt test is available in the "trend" R package version 1.1.6.

2.6 WaSiM model setup

The grid-based Water Flow and Balance Simulation Model (WaSiM) was initially developed for climate and land use change studies in mountainous regions (Schulla, 2021). In previous studies, WaSiM has proven to be a powerful tool for investigating the spatial and temporal variability of hydrological processes in complex river basins (e.g., Strasser et al., 2019). WaSiM uses physics-based modeling approaches to simulate vertical water flux in the unsaturated zone using the Richards equation and the energy-balance approach for snowmelt. Additionally, WaSiM can consider snow redistribution processes induced by wind or gravity, which is highly relevant for high-elevation alpine catchments (Warscher et al., 2013). A detailed explanation of the snow routine can be found in the WaSiM documentation and the publication of Hofmeister et al. (2022). The model parameters for simulating gravitational slides (Table 3) are identical to those of Hofmeister et al. (2022) as the spatial and temporal resolution of the model setup is the same (i.e., hourly time step and 25 m grid resolution). The range of the main wind direction for wind-induced snow redistribution was derived from monthly mean WRF data for the winter period (from October to May). Accordingly, WaSiM corrects snowfall based on exposition, i.e., snowfall from wind-exposed grid cells is moved to sheltered cells. Consequently, the algorithms for simulating wind-induced snow redistribution and gravitational slides are not event-based, as they try to match snow accumulation at the end of the winter. The temperature threshold (T0R) is set to 2 °C, which is in agreement with other studies conducted in the Alpine region (e.g., Frei et al., 2018). The temperature transition range from snow to rain is set to 2 °C (i.e., half snowfall and half rain), allowing a smooth transition between rain and snowfall from 0 °C (i.e., only snowfall) to 4 °C (i.e., only rain). The correction factors for the incoming and outgoing long-wave radiation are set to 1, meaning no correction is applied.

A variety of algorithms are available in WaSiM for the interpolation of meteorological inputs. Most of the five input variables (i.e., relative humidity, precipitation, wind speed, and short-wave incoming radiation) are interpolated with a bilinear approach from the 2 km WRF to the 25 m grid resolution of WaSiM. Temperature is inter- and extrapolated based on an elevation-dependent regression approach. The interpolation algorithms are applied at an hourly time step, which allows the consideration of variable temperature lapse rates, which is far more realistic than a fixed temperature lapse rate. For consistency, this approach is identical for uncorrected and corrected WRF temperature. Short-wave incoming radiation and temperature are adjusted according to topographic shading following the scheme devised by Oke (2002).

We assume a positive precipitation lapse rate of 0.5 % per 100 m, which is a common assumption for the Ötztal Alps (Schöber et al., 2014; Schmidt et al., 2022), for re-scaling the relatively coarse WRF precipitation from the 2 km grid resolution to the
330 actual topography of the much finer grid resolution of WaSiM (i.e., 25 m). The 25 m grid resolution of the HM is a compromise
between the spatial discretization needed to resolve the small-scale heterogeneity of the research area ($\sim 60 \text{ km}^2$ in size) with
complex topography and the computational time required to run the hydrological model at hourly time steps over longer periods
(~ 40 years). With the activated snow redistribution module of WaSiM, preferential snow erosion and deposition are considered
and simulated.

335 Glacier melt and evolution are particularly important for hydrological studies in glaciated catchments such as the upper
Kauner Valley. For this, WaSiM offers a dynamical glacier module that adds or removes glaciated grid cells dependent on the
specific mass balance (Stahl et al., 2008). The standard threshold value for transforming a grid cell into a glacier cell is 2000
mm SWE. However, WaSiM does not consider the actual ice flow of a glacier. Hence, positive mass gains are not transported
from the accumulation to the ablation area of a glacier, which has certain limitations in areas with high glacier dynamics. A
340 recently published coupling scheme of WaSiM and the Open Global Glacier Model (OGGM) bridges the gap between more
sophisticated glacier evolution models and local catchment hydrology (Pesci et al., 2023). For initialization, WaSiM estimates
the volume of a glacier based on the volume-area scaling (VAS), which offers the main advantage of minimum input data,
as glacier extent data are usually available even in data-scarce regions. According to Bahr et al. (2015), the dimensionless
exponent of the VAS equation is fixed to 1.375, while the scale factor depends on the sample data set. Since the scale factor can
345 only be estimated for larger sample sizes of around 50 or more glaciers to ensure statistical robustness, an already published
scaling factor for the European Alps from Chen and Ohmura (1990) is chosen. Additionally, the dynamic glacier routine in
WaSiM enables modeling of snow metamorphism to firn after one balancing period and to ice after six additional balancing
periods (hydrological years). The initialization of the firn layer is based on the WEchnge parameter that defines the rate of
change in the water equivalent (WE) of firn with each meter of elevation increase (default is 1.8 mm m^{-1}) and the average
350 equilibrium line (ELA), which separates the accumulation and ablation area of a glacier. As no continuous ELA information is
available for the glaciers of the upper Kauner Valley, we use the ELA in m a.s.l. from a neighboring glacier, the Vernagtferner
(WGMS, 2025). Glacier melt is simulated using an extended degree-day approach initially developed by Hock (1999), which
includes information on global radiation at each time step and grid cell to modify the ice and firn melt. Besides a classical
melt factor (MF), one empirical coefficient for minimum radiation and one for maximum radiation are needed. Hence, glacier
355 melt increases when shortwave radiation input is high and vice versa. Snow melt on the glacier is simulated by the same
energy-balance approach as for non-glaciated grid cells to ensure consistency.

In general, we used the standard model parameters from the WaSiM documentation whenever possible (Tab. 3), as the main
focus of this case study is to investigate the sensitivity of WaSiM with respect to changed forcing data. Although calibration
of the main WaSiM parameters can improve model results, we have refrained from parameter optimization to investigate
360 the propagation of the changed forcing to the different model results. This means all model parameters are identical for all
simulation runs. We use WaSiM version 10.07.02 for all simulations in this study.

Table 3. Key processes and main WaSiM parameters.

Process	WaSiM Parameter	Description	Value
Snow accumulation	T0R	Temperature limit for rain (°C)	2
	Ttrans	Temperature-transition range from snow to rain (°C)	2
Gravitational redistribution	l _{lim}	Maximum deposition slope (°)	55
	D _{lim}	Scaling for maximum deposition (mm)	2
	i _{erosion}	Minimum slope for creating slides (°)	50
	f _{erosion}	Fraction of snow pack that forms the slide (-)	0.002
Snow ablation	LWINcorr	Correction factor for incoming long-wave radiation (-)	1.0
	LWOUTcorr	Correction factor for outgoing long-wave radiation (-)	1.0
Wind redistribution	start azimuth	1st quantile of wind direction (°)	179
	end azimuth	3rd quantile of wind direction (°)	217
	cmin	Minimum correction factor (-)	0.5
Glacier initialization	VAscale	Scale factor (-)	28.5
	VAexp	Exponent (-)	1.375
	WEchnge	Change rate of WE per m (mm)	1.8
	ELA	Average equilibrium line elevation (m)	3113 (1969); 3200 (2006)
Glacier melt	MF	Melt factor (mm)	2
	ice _{min}	Minimum radiation coefficient for ice (mm Wh ⁻¹ m ² °C ⁻¹ day ⁻¹)	0.0001
	ice _{max}	Maximum radiation coefficient for ice (mm Wh ⁻¹ m ² °C ⁻¹ day ⁻¹)	0.0007

3 Results

3.1 WRF temperature at station locations

As a first evaluation, we compare the WRF temperature with station observations from the upper Kauner Valley and surrounding
365 (Fig. 1 and Table 1). The uncorrected WRF temperature corresponds to the WRF hourly temperature of the respective 2 km grid
cell, meaning it represents the simulated near-surface (i.e., 2 m) temperature and is adjusted for the height difference between
WRF and station altitude by applying Equation 6. The corrected WRF temperature corresponds to the hourly temperature
computed from the WRF atmosphere applying Equation 4. We compare the general statistics of the time series using ECDFs,
as shown in Figure 5. The periods of the ECDFs refer to all data since the start of measurement recording (Tab. 1) until
370 31 December 2015. No data values (NAs) in the observations are set to no data in the WRF temperature time series. The
lowest station is Dammfuss, at an elevation of 1770 m a.s.l., with continuous temperature records since 1988, followed by the

Gepatschalm station, installed at 1900 m a.s.l. in 2009. The second-highest station is Weisssee, at an elevation of 2540 m a.s.l., with continuous temperature observations since December 2006. The station Vernagt Pegel was installed at 2640 m a.s.l. in 1975. Consequently, the selected stations cover an elevation difference of almost 900 m. In addition, the four stations represent

four different WRF land use categories (Fig. S2 and Tab. S1), which are described in the following section.

The four panels of Figure 5 compare the ECDF of station temperature (black line), of the uncorrected WRF temperature (red line), and the corrected WRF temperature (blue line). We evaluate the similarity between the different ECDFs with the KS test. The results of the KS tests are listed in Table S3 of the supplementary material. At Dammfuss, the uncorrected near-surface WRF temperature agrees with the frequency of observed mean and warmer ($> 5^{\circ}\text{C}$) temperatures but shows a positive bias (higher frequency) for colder temperatures (Fig. 5a). The frequency distribution of the corrected WRF temperature aligns with the observations at the Dammfuss station and shows only a slight cold bias for temperatures $< 0^{\circ}\text{C}$. According to the computed D-statistics (Tab. S3), the maximum distance between the observed and WRF temperature ECDFs is 0.08 for the uncorrected and 0.03 for the corrected temperature. The maximum vertical difference between the ECDFs is at -3.06°C in the case of uncorrected and at 17.46°C for the corrected. The WRF land use category of this particular grid cell is *Woodland Mosaic* (Fig.

At Gepatschalm station, the critical temperature frequencies are more towards warmer temperatures ($> 5^{\circ}\text{C}$) (Fig. 5b). The correction of the WRF temperature reduces the warm bias in the range from 5°C to 15°C , but there is still a slight offset. Conversely, the cold bias in the near-surface WRF temperature is absent in the corrected time series. The maximum distance between the ECDFs is 0.08 in the case of the uncorrected and 0.07 for the corrected WRF temperature (Tab. S3). The WRF land use category of this particular grid cell is *Evergreen Needleleaf* (Fig. S2 and Tab. S1).

We can also observe a similar cold bias ($< 0^{\circ}\text{C}$) between the observation and uncorrected WRF temperature at Weisssee station (Fig. 5c). Using the WRF atmosphere, the cold temperature offset can be corrected, which results in a good agreement with the observed temperature records. The maximum distance between the ECDFs is 0.05 in the case of the uncorrected and 0.03 for the corrected WRF temperature (Tab. S3). The WRF land use category of this particular grid cell is *Wooded Tundra* (Fig. S2 and Tab. S1).

The largest cold bias covering the entire temperature range is present at the highest station, the Vernagt Pegel (Fig. 5d). The corrected WRF temperature shows a more realistic representation of temperature frequencies at 2640 m a.s.l. altitude. Therefore, the maximum distance between the ECDFs is much larger (i.e., D-statistics uncorrected 0.13 and corrected 0.03) compared to the lower-elevation stations (Tab. S3). The respective land cover class of the WRF grid cell is *Snow or Ice* (Fig. S2 and Tab. S1).

Although the comparison with the station data only represents specific elevation ranges and locations of the research area, it indicates that the near-surface WRF temperature has difficulties representing the small-scale heterogeneity in surface-albedo characteristics, especially for the higher altitudes (> 2600 m a.s.l.) that are dominated by the *Snow or Ice* WRF land use category. Hence, the correction approach based on temperature lapse rates from the WRF atmosphere provides more realistic temperature predictions independent of the surface-albedo characteristics.

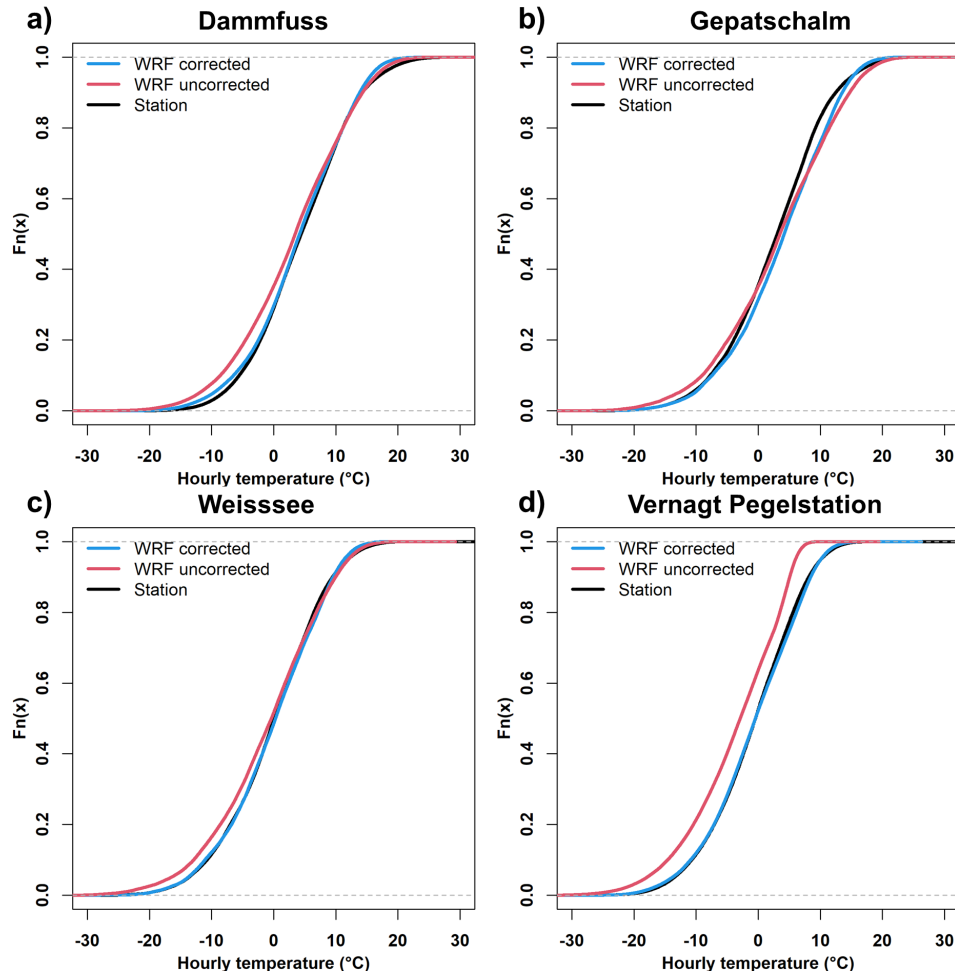


Figure 5. Empirical cumulative distribution functions (ECDF) of observed hourly temperature (black line), uncorrected WRF temperature (red line), and corrected WRF temperature (blue line) at different stations a) Dammfuss (WRF land use category *Woodland Mosaic*), b) Gepatschalm (WRF land use category *Evergreen Needleleaf*), c) Weisssee (WRF land use category *Wooded Tundra*), and d) Vernagt Pegelstation (WRF land use category *Snow or Ice*).

Figure 6 compares seasonal hourly observed and corrected WRF temperatures. The comparison is done for different data subsets based on the availability of hourly temperature records. The seasonal temperature dynamic is well reproduced by the corrected WRF temperature at all observation sites. When comparing the medians of the four boxplots, we observe a slight seasonal tendency of the corrected WRF to overestimate summer temperature. In addition, the standard deviation is under- or overestimated by the corrected WRF temperature for some months. Nevertheless, we cannot observe a systematic bias in the corrected WRF temperatures across different sites and altitudes.

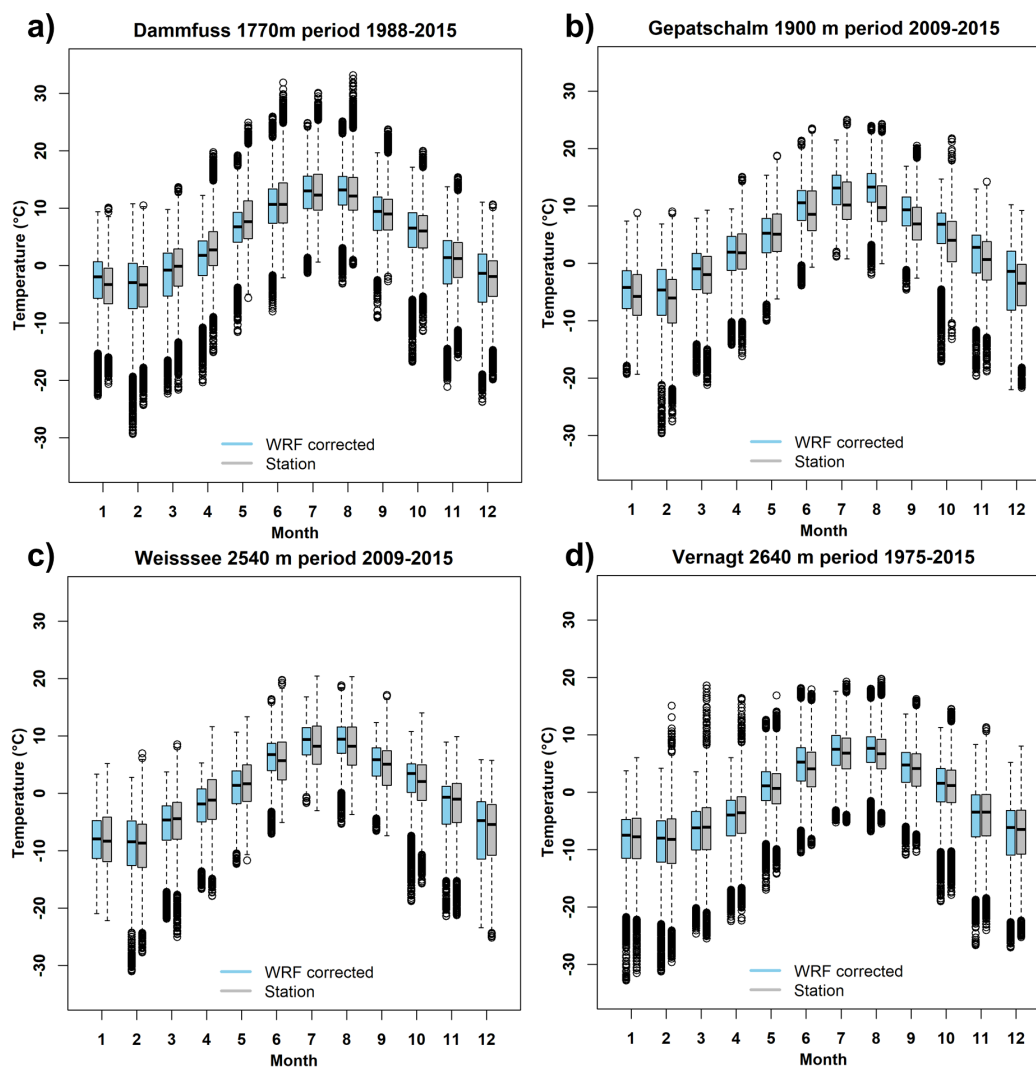


Figure 6. Monthly boxplots of hourly observed temperature (grey) and corrected WRF temperature (blue) at different stations a) Dammfuss, b) Gepatschalm, c) Weisssee, and d) Vernagt Pegelstation. The different temporal coverages are given in the titles of the respective subplots.

In an additional experiment, we determined the influence of ETA level combinations on the calculated hourly temperatures at four observation stations (Fig. S6). For this, we calculated TLRs from neighboring ETA levels up to eta level 25, as shown in Figure 4. Interestingly, we observe a larger temperature spread, indicating a greater dependence of the corrected temperature on the eta level combination at the lower-elevated reference stations (i.e., Dammfuss and Gepatschalm). Although individual eta level combinations can achieve similar or slightly better goodness-of-fit criteria (Tab. S2), the best overall performance is still achieved with the linear gradient between eta levels 1 and 20, with a mean KGE of 0.82 and RMSE of 3.5 °C (Tab. S2).

3.2 Robustness of temperature trends and elevation-dependent warming

In order to evaluate the robustness of the WRF temperature trend, we compare it with a long-term and homogenized observation from the HISTALP data set (Auer et al., 2007). Monthly mean temperature records since 1851 are available at the Obergurgl station, which is approximately 20 km away from the Kauner Valley and located at 1936 m a.s.l.. As this site is located in a neighboring valley (i.e., the Ötz Valley), it also serves as a test of the general transferability of the applied correction method for WRF temperatures. The WRF land use category of the respective grid cell is *Evergreen Needleleaf*. Figure 7a) shows the evolution of monthly and annual temperatures at the station Obergurgl. In general, the WRF temperature follows well the evolution of the observed monthly and annual temperature. The mean temperature at the Obergurgl station is 2.16 °C based on observations from 1973 to 2015. The mean of corrected WRF temperature at the respective elevation (1936 m a.s.l.) is 2.61 °C. Consequently, WRF overestimates the mean temperature by 0.45 °C, as visible in the ECDF plot (Fig. 7b). The mean annual temperature increase is 0.06 °C a⁻¹ at Obergurgl station (Sen's slope of 0.003 °C a⁻¹) and 0.05 °C a⁻¹ for the corrected WRF temperature (Sen's slope of 0.003 °C a⁻¹). Both temperature increases represent a significant positive trend with a confidence interval of 0.95. The Pettitt test detects a probable change point, meaning a sudden change in the central tendency (i.e., mean), in both time series for May 1987. The sudden warming in the European Alps at the end of the 1980s was already reported by several other studies (e.g., Marty, 2008; Marcolini et al., 2019; Hofmeister et al., 2023). These results clearly show that the bias-corrected RCM data are consistent over a longer period (> 30 years) with respect to non-stationarity. However, we also found that the uncorrected WRF temperature has the identical statistical behaviour as the corrected WRF temperature for the Obergurgl WRF grid cell over the same period. Therefore, we did not include the uncorrected WRF time series in Figure 7.

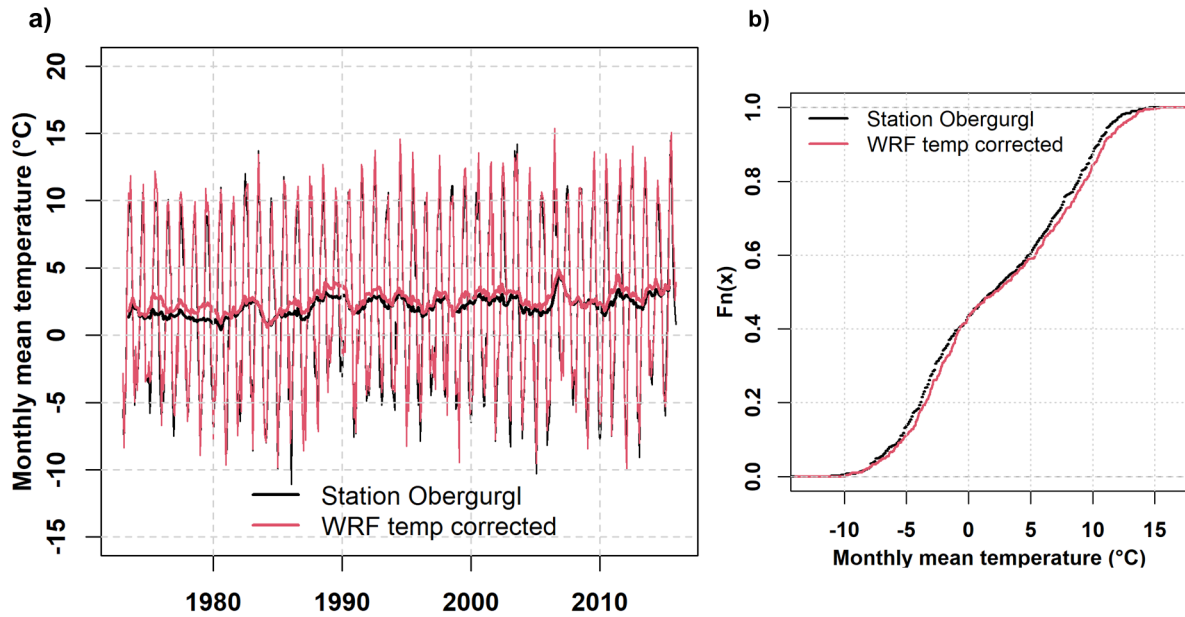


Figure 7. a) Mean monthly temperature at Obergurgl station (thin black line) and corrected WRF temperature (thin red line) at the respective station altitude. The thicker lines illustrate the respective yearly moving average. b) ECDF of the respective time series for the same period (i.e., 1973-2015).

3.3 Bivariate correction of temperature and relative humidity

Due to the cold bias in the uncorrected WRF temperature, the RH adjustment using the ARM formulas resulted in a downward shift in the corrected mean RH across all elevations, as shown in Figure 8. Since the cold bias is more pronounced in the highest glacirized areas of the WRF temperatures, the RH has the most significant impact in the elevation range 2900-3500 m a.s.l..

440 The mean of the uncorrected RH was about 90 % for this distinct elevation range, and had a mean RH of about 73 % after the correction. In addition, the correction reduces the slope of the RH elevation gradient across the upper Kauner Valley from 2.2 % (100 m)⁻¹ for the uncorrected to 0.3 % (100 m)⁻¹ for the corrected RH data. Figure S12 of the supplementary material underlines the local effect of the RH correction by exemplarily showing the ECDFs of hourly uncorrected and corrected RH at three station locations. The largest difference between corrected and uncorrected RH is observed at the Vernagt Pegelstation. It

445 gets obvious that the RH correction affects all parts of the ECDF, with the lower quantiles (i.e., $p < 0.25$) showing the strongest reduction in RH.

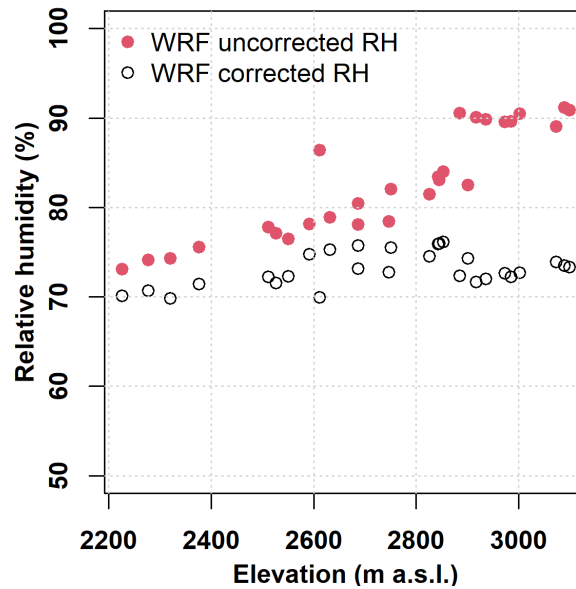


Figure 8. Scatterplot of WRF elevation (m a.s.l.) and mean relative humidity from uncorrected 2 m (red dots) and corrected (black circles) over the period 1970-2015.

3.4 Impact of temperature correction on hydrological model results

We use the HM WaSiM to investigate the impact of WRF temperature (i.e., univariate) and RH (i.e., bivariate) corrections at the catchment scale across different hydrological processes, such as snow, glacier, and streamflow generation. The simulation period is from 1 October 2006 until 30 September 2015, but the analysis period is slightly shorter (eight years) as the first simulation year is considered a spin-up period. The multi-data evaluation of the simulation results is limited by the temporal coverage of the WRF data, which is only available until 31 December 2015 in this case study. Hence, the overlap of observed snow depth and snow cover data is relatively short. Nevertheless, the simulation period is long enough to observe similarities and discrepancies between corrected and uncorrected model runs. In the following, univariate correction refers to the case in which WRF temperature is used as model forcing, and bivariate correction refers to the correction of WRF temperature and relative humidity as forcing for WaSiM.

In Figure 9a, we compare the ECDFs of daily observed SWE at Weisssee station (black line) with simulated SWE at the respective 25 m grid cell of WaSiM. In general, WaSiM tends to overestimate SWE with both corrected and uncorrected model forcing. The simulated SWE appears relatively insensitive to the corrected WRF inputs at this grid cell. The mean daily SWE is 163 mm in the observation, while the simulated SWE with univariate correction is 187 mm, with bivariate correction 185 mm, and 195 mm in the case of no correction applied. Although comparing in situ SWE with simulated SWE involves several assumptions and uncertainties due to the high spatial heterogeneity of snow processes, the similarity in their behavior provides

interesting insights into snow accumulation at 2540 m a.s.l. elevation. We conclude that both correction approaches have little influence on the simulated SWE for this location and elevation.

465 Figure 9b shows the ECDFs of daily observed and simulated fractional snow-covered area (fsca). Daily observed snow cover data are from MODIS (black line). Not surprisingly, the agreement between observed and simulated fsca is largest during the winter period, when the entire research area is covered by snow (Fig. S7). However, the discrepancy between observation and simulation, as well as between the corrected and uncorrected WRF temperature forcing, increases during the low fsca period. In particular, the simulated fsca with uncorrected WRF temperature input overestimates snow cover in the research area, with
470 values < 0.6 fsca. The ECDF of simulated fsca with corrected WRF temperature is more in line with the fsca of MODIS, especially at low to mid fsca values (i.e., fsca < 0.6). The simulated fsca with bivariate correction (cyan line) has a very similar behavior to the one with univariate correction (blue line). The negative snow cover bias in both corrected products during the snow-free period (fsca < 0.3) can result from several factors, including snowpack heterogeneity, the limited MODIS sensor resolution, and the fsca processing. Nevertheless, the fsca data cover the entire research area, unlike the in-situ SWE data, and
475 provide valuable information on the seasonal evolution of the snow cover. We can conclude that both correction approaches yield a more reliable simulation of snow cover, whereas these do not differ significantly from one another, indicating that simulated fsca mainly depends on the temperature correction.

The ECDFs of hourly streamflow at the gauge Gepatschalm are illustrated in Figure 9c. The observed streamflow (black line) shows a typical behavior for a snow and glacier-dominated catchment with an extended low flow period with a median of 0.85
480 $\text{m}^3 \text{s}^{-1}$. The simulated streamflow with uncorrected WRF temperature (red line) significantly underestimates the entire observed ECDF, indicating WaSiM generates too little streamflow in all seasons (Fig. S8a). The WaSiM runs with univariate correction (blue line) and the one considering bivariate correction (cyan line) show a more realistic behavior and follow the ECDF of the observed streamflow. While the mean annual streamflow is well estimated with the univariate correction (i.e., the mean flow of observation and simulation with univariate correction are both $3.1 \text{ m}^3 \text{s}^{-1}$), the bivariate correction simulation run slightly
485 underestimates the mean streamflow ($2.89 \text{ m}^3 \text{s}^{-1}$). For both correction approaches, WaSiM overestimates higher streamflow and underestimates lower streamflow ($p < 0.8$). Hence, there is potential to optimize WaSiM's seasonal streamflow generation (Fig. S8b and c). We compute goodness-of-fit (GOF) criteria to compare daily observed and simulated streamflow for the respective period. Although WaSiM provides hourly streamflow outputs, we compare daily values as they are expected to be less affected by the fact that WRF does not assimilate precipitation station data from that particular region, which limits direct
490 hourly comparisons. For daily values, we assume that model performance depends more on the correct prediction of snow and glacier dynamics than on intense precipitation events. The daily GOF criteria with uncorrected WRF temperature are KGE 0.42, percent bias -46, and RMSE $2.5 \text{ m}^3 \text{s}^{-1}$. The daily performance with univariate correction is much better with KGE 0.71, percent bias -0.7, and RMSE $2.4 \text{ m}^3 \text{s}^{-1}$. Despite the underestimate of the mean streamflow by -6.9 %, the other two performance indices (i.e., KGE 0.78; RMSE $2.14 \text{ m}^3 \text{s}^{-1}$) enhanced when considering bivariate correction in WaSiM. The improvement in
495 model performance is due to lower overestimation of streamflow peaks during the summer period (Fig. S8b and c), resulting from reduced glacier melt contribution. We observe no changes between the two correction products during the low-flow period (Fig. S8b and c). Accordingly, we conclude that the major improvement in streamflow prediction originates from the

temperature correction. Nevertheless, the bivariate correction further enhances WaSiM's performance in runoff generation during the glacier melt period. It is worth mentioning that the only differences between the WaSiM simulations are the applied WRF bias corrections. We did not apply a calibration strategy (i.e., manual or auto calibration) to optimize the model parameters based on the model forcing. This shows the importance of the correct meteorological forcings for model parameterization, including physics-based models, to achieve high model performance.

The simulated cumulated glacier mass balances indicate the highest sensitivity concerning the temperature forcing (Fig. 8d). The cold bias of the near-surface WRF temperature (red line) leads to a constant positive glacier mass balance, which is quite unrealistic for the period 2008 to 2015. In the highest parts of the research area (> 2700 m), the cold bias results in too low snowmelt and, hence, an overestimation of the snow cover (Fig. 8b). As long as the glaciers are snow-covered and not exposed to incoming radiation, WaSiM does not compute melt from firn and ice. When forcing WaSiM with the univariate correction (blue line), as well as with the bivariate correction (cyan line), the simulated glacier mass balances exhibit stronger seasonality and a continuous negative trend, which are more in line with the observed annual mass balances of the nearby Kesselwand glacier (Fig. S9). The Gepatsch and Kesselwand glaciers share parts of the accumulation area. The only exception in the negative trend is the winter period of 2013/14, which was particularly snowy, and the following summer was cooler and wetter. The mass balance year 2013/14 was almost balanced in the Central European Alps with only a slightly negative average balance of -174 mm (WGMS, 2017). At the end of the simulation period, we observe a deficit of almost 6000 mm in the water balance of the catchment when comparing the uncorrected and univariate correction simulation runs. As snow and glacier melt are the dominant streamflow components in the upper Kauner Valley during this period, the deficit in meltwater directly propagates into streamflow generation (Fig. S8a-c).

Table S4 in the supplementary material summarizes the main water balance components simulated with the uncorrected WRF temperatures and the two corrected products. Since the only difference in the model forcing is the temperature and RH input, the total precipitation is identical in all three model runs. The mean annual precipitation sum for the period 2008-2015 is 1487 mm. Due to the cold bias in the uncorrected WRF temperature, the mean snowfall fraction of total precipitation is slightly higher (68%) than that of the corrected simulation runs (62%). Consequently, the fraction of rain in precipitation is smaller for the simulation run with uncorrected WRF temperature. Despite interannual variability in streamflow magnitudes, the mean annual streamflow sum simulated with univariate correction matches observations quite well, with only a slight underestimation of $< 1\%$. The simulation run with bivariate correction underestimates the observed annual streamflow sum by about 7% . The simulated streamflow with the uncorrected WRF temperature underestimates annual streamflow by almost 50% (Fig. S8a). The highest mean annual evapotranspiration (ETR) is generated with bivariate correction (i.e., 84 mm), the ETR with only univariate correction is about 19% smaller (i.e., 68 mm), and even smaller in the case of uncorrected WRF forcing with a mean ETR of 62 mm. Based on these facts, we can assume that the critical aspect of the simulation results with uncorrected WRF temperature is not the snowfall or ETR amount, but rather the timing and magnitude of simulated snowmelt, which requires positive temperatures, especially in higher-elevation parts of the catchment. While covered with snow, WaSiM does not compute firn and ice melt on glaciated cells, leading to a significant discrepancy in the simulated glacier mass balances, as shown in Figure 9d and the calculated mean glacier mass balances in Table S4. From the model results, we conclude

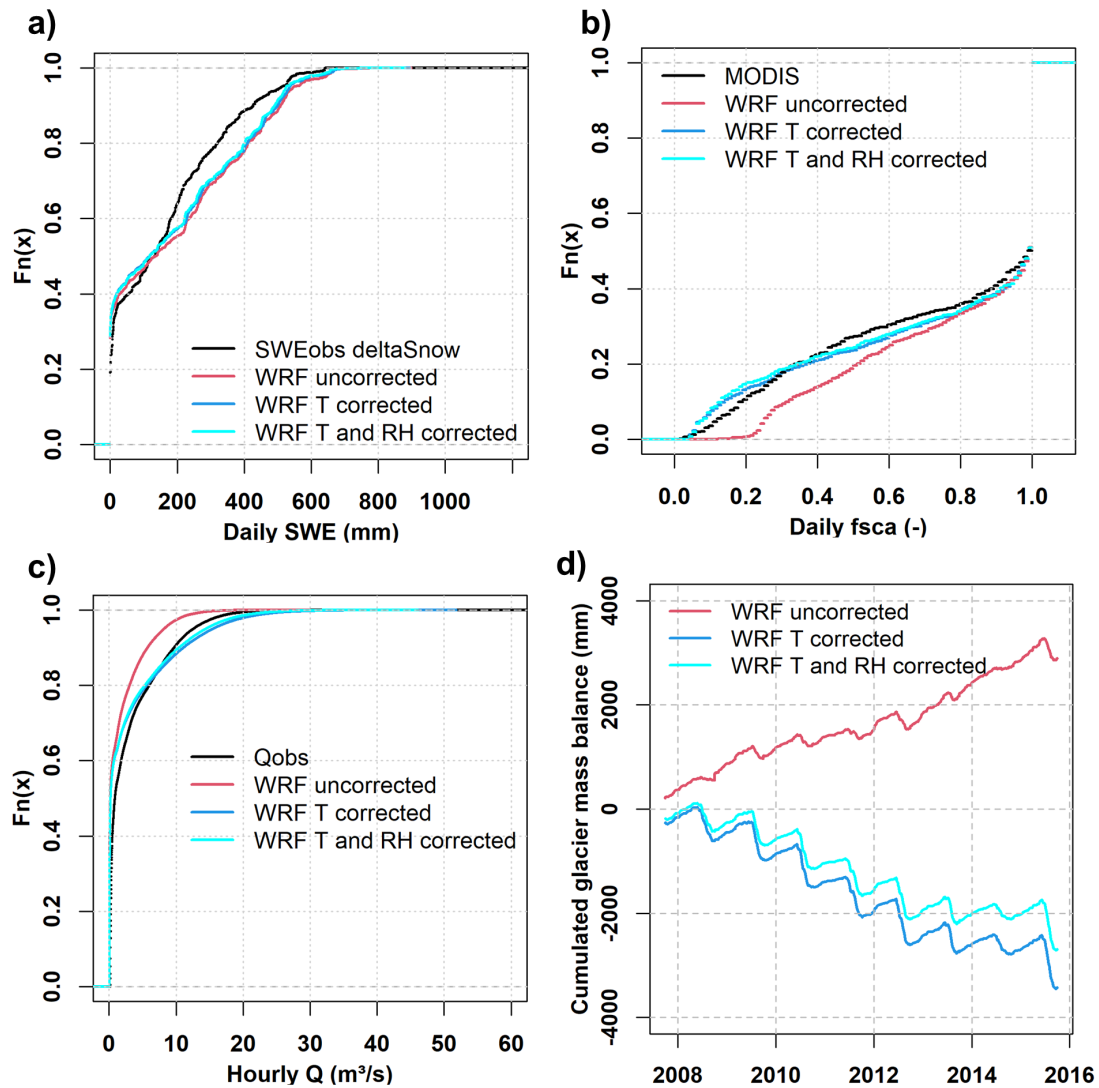


Figure 9. Empirical cumulative distribution functions of observed (black line) and simulated a) SWE at station Weisssee, b) fractional snow covered area (fsca), c) streamflow at gauge Gepatschalm, as well as d) simulated glacier mass balances averaged over the catchment. Simulation results with uncorrected WRF temperature (red line), corrected WRF temperature (blue line), and corrected temperature and RH (cyan line).

that the temperature correction is essential for reliable prediction of seasonal snow and glacier processes and, consequently, for streamflow generation. Nevertheless, the bivariate correction further enhances the magnitude of seasonal streamflow, especially in August and September, when glacier melt contributes at its maximum.

3.5 Longer simulation run (1973-2015)

Given the relatively short simulation period of eight years, the near-surface WRF temperature does not yield plausible HM results, especially for snow and glacier processes (Fig. 9). On the contrary, using TLRs from the WRF atmosphere results in reliable model predictions for several processes, including snow and glacier melt and streamflow generation. The consideration of bivariate correction yields the best model performance for simulated streamflow. In this section, we test the robustness and consistency of the HM results with the two corrected WRF products over a 42-year simulation period. We initialized WaSiM in 1969 using glacier extents from 1968 and considered a spin-up period of three years, so the actual analysis period began in 1973. As almost continuous observed streamflow data are available from 1971 onwards at gauge Gepatschalm, we compare the hourly and daily streamflow over different periods (Fig. 10). The left panels of Figure 10 illustrate the mean of the streamflow in days of the year. The subplots in Figure 10 (a-e) present observed and simulated streamflow for shorter periods of 14 years. Figure 10a shows the earliest period from 1973 to 1987. The blue (univariate correction) and cyan (bivariate correction) lines of the mean simulated streamflow follow well the seasonality and magnitude of the mean observed streamflow (black line). In general, the annual streamflow regime has a typical pattern of a nival regime. The ECDFs of hourly streamflow (Fig. 10b) indicate a good fit in the mean and high flow parts with both correction products. However, WaSiM tends to underestimate low flows ($p < 0.8$). The mean of hourly streamflow is $2.8 \text{ m}^3 \text{ s}^{-1}$ for the observation and $2.5 \text{ m}^3 \text{ s}^{-1}$ for the simulation with univariate correction and $2.3 \text{ m}^3 \text{ s}^{-1}$ with bivariate correction.

The period 1987 to 2001 of snapshot Figure 10c is characterized by slightly higher streamflow magnitudes. The mean observed streamflow is $2.8 \text{ m}^3 \text{ s}^{-1}$ and the simulated mean, considering only univariate correction, is $2.7 \text{ m}^3 \text{ s}^{-1}$ while the simulation with bivariate correction yields a lower mean streamflow of $2.5 \text{ m}^3 \text{ s}^{-1}$. In general, the summer streamflow period is extending into autumn.

In the following period (2001-2015), the high flow period further extends during the summer months (Fig. 10e). Additionally, the streamflow generation shifts towards spring. As a consequence, the mean observed ($3.1 \text{ m}^3 \text{ s}^{-1}$) and both simulated ($3 \text{ m}^3 \text{ s}^{-1}$ with univariate correction; $2.9 \text{ m}^3 \text{ s}^{-1}$ with bivariate correction) streamflow increases. The reason for the increased streamflow is the increased temperature (i.e., mean annual temperature is -3.5°C in 1973-1987 and -2°C in 2001-2015), which leads to a higher glacier melt contribution (Fig. 11b). Although WaSiM well predicts the observed seasonality, it overestimates the mean streamflow in August and September. However, the overestimation is slightly smaller when using the bivariate correction.

The daily mean of streamflow values computed over the entire simulation period provides a good overview of the general tendency of the WaSiM model for underestimating the low flow period while overestimating streamflow at the end of the ablation period in late summer (Fig. 10g and h). Nevertheless, WaSiM reproduces the seasonal streamflow pattern in the upper Kauner Valley well with both correction products. Over the entire simulation period (1973-2015), the daily KGE is 0.76, percent bias -5.5, and RMSE $2.2 \text{ m}^3 \text{ s}^{-1}$ for the univariate correction simulation, and the GOFs for the bivariate correction are KGE 0.8, percent bias -11.3, and RMSE $2 \text{ m}^3 \text{ s}^{-1}$. Considering that the model is not calibrated, meaning the model parameters are not optimized according to the extended simulation period, or for the different corrected forcings, the streamflow results are very satisfying.

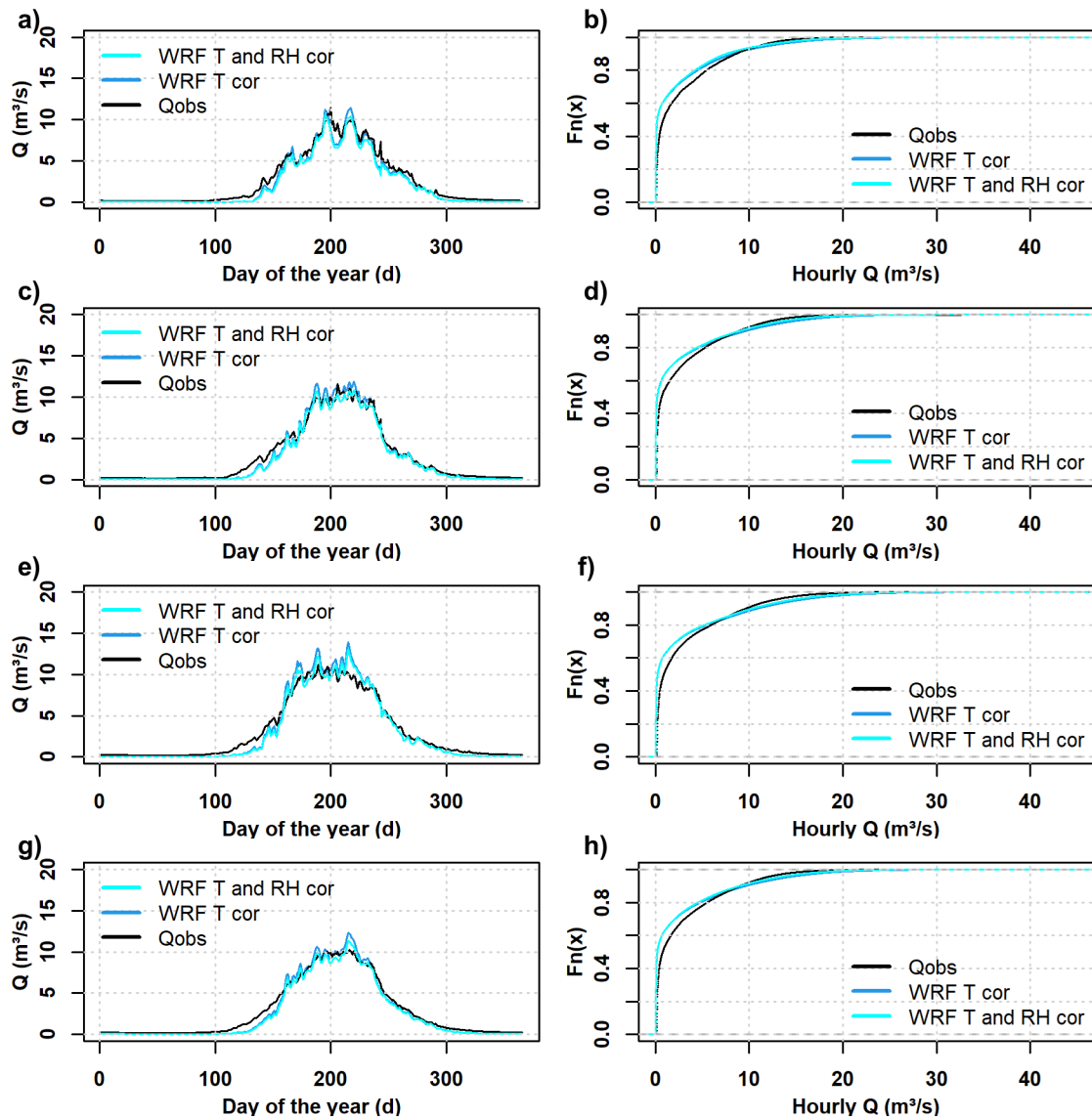


Figure 10. Left panels show mean day of the year of observed (black line) and simulated streamflow with corrected T (blue line) and corrected T and RH (cyan line). The right panels show hourly empirical cumulative distribution functions over four different periods, i.e., 1973-1987 (a, b), 1987-2001 (c, d), 2001-2015 (e, f), and as the overall mean (g, h).

570 The evolution of annual mean temperature and the consequences for snow and glacier development are presented in Figure 11. The first panel of Figure 11a shows the corrected mean annual WRF temperature evolution of the research area. The vertical red line represents a significant change point in the annual temperature time series detected by the Pettitt test in 1987. This particular change point has already been observed in previous studies in different parts of the European Alps (e.g., Marty, 2008; Marcolini et al., 2019; Hofmeister et al., 2023).

575 Both simulated glacier mass balances (i.e., with univariate correction in blue; with bivariate correction in cyan) exhibit a continuous negative trend from the end of the 1980s (Fig. 11b). Before that, the glaciers were in an equilibrium state or had even slight positive mass balances. The positive glacier mass balances align with observed advances of Gepatsch- and Weissseeferner (Altmann et al., 2020), as well as the observed mass balances of the close by Kesselwand glacier (Fig. S9). As the graph shows hourly simulation results, the seasonal variability of snow accumulation and ice ablation is clearly visible in periods with negative glacier mass balance. In general, the simulation run with bivariate correction yields less negative mass balances than the simulation run that considers only temperature correction. Consequently, the Pettitt test detected a significant change point earlier (i.e., 1994) than for the model run with bivariate correction (i.e., 2001). An additional evaluation of the simulated glacier mass balances against observed glacier mass changes from an elevation change product (Hugonnet et al., 2021) revealed an overestimation of cumulated melt rates at the end of the simulation period in 2015 by 36 % in the case of univariate correction and by 22 % in the case of the bivariate correction.

580 The simulated fractional glacier cover (fglc) shows a high similarity between the two corrected methods (Fig. 10c). Mainly during the periods of glacier growth around the 1980s and the strong recession from the late 1990s, differences can be observed. Both correction products lead to a significant change point in the fractional glacier cover (fglc) in 2000. In general, the evolution of the simulated fglc shows a pattern similar to the cumulated glacier mass balances. The black triangles illustrate years with observation data. Both simulated fglc follow the negative trend of the observations but have a positive offset at the end of the simulation period (i.e., observed fglc is 0.3, simulated with univariate correction is 0.35, and with bivariate correction is 0.37). Figure 11d represents the long-term evolution of simulated snow cover duration (scd) forced with the two corrected products, which show an almost identical mean scd. The Pettitt test found a significant change point in 1985 for both simulation runs. Before 1985, the mean scd was fluctuating between 290 and 330 days. After 1985, scd never exceeded 310 days and was even below 260 days in 2002. The significant reduction in scd is a prerequisite for the negative glacier mass balances, as more firm and ice are exposed to the short-wave radiation for a longer period, leading to higher melt rates. As high elevation (> 2700 m a.s.l.) and glaciated areas, which are mainly or permanently covered by snow, the sensitivity of mean scd is rather small. Consequently, we conclude from the evaluation of the long simulation runs that the bivariate correction further enhances the simulated glacier mass balances and the magnitude of summer streamflow, especially in August and September.

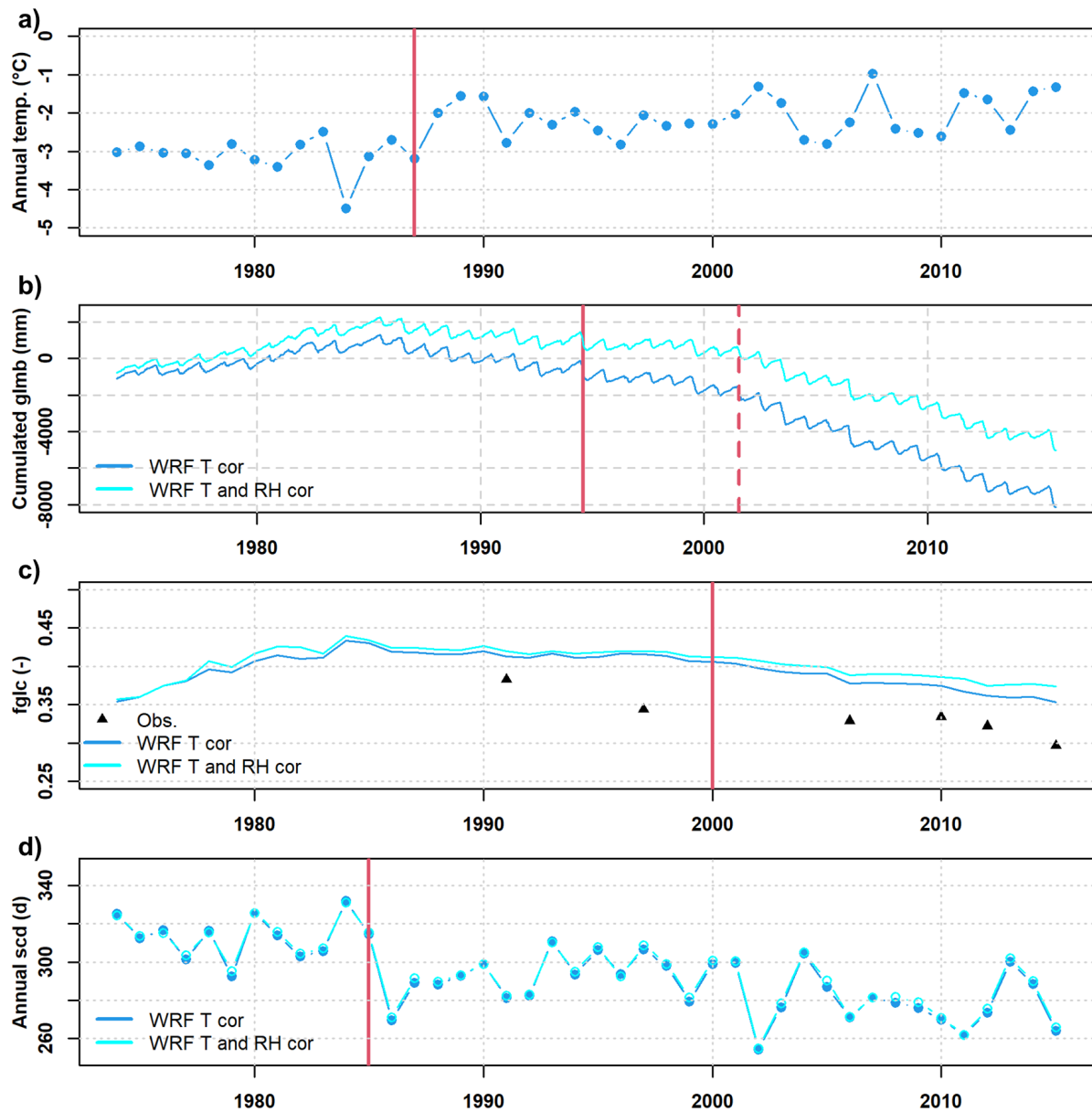


Figure 11. Catchment averages of a) simulated mean annual temperature, b) simulated cumulated glacier mass balances (glmb) with T correction (blue line) and T and RH correction (cyan line), c) observed (black triangles) and simulated fractional glacier cover (fglc) with T correction (blue line) and T and RH correction (cyan line), and d) simulated annual snow cover duration (d) with T correction (blue line) and T and RH correction (cyan line). Significant change points are indicated with a vertical dashed red line in the case of the T and RH corrected simulation run, and as a red solid line for the T corrected simulation run. Different change points are only detected for b) simulated glmb for the other variables, the change points are identical. The years represent hydrological years (i.e., 1st of October to 30th of September).

4.1 Bivariate and multivariate correction

The temperature correction based on TLRs from the WRF atmosphere and on spatial averaging demonstrated to be a robust alternative to the commonly used univariate empirical quantile mapping (EQM), which highly depends on the quality and availability of temperature observations and requires nonstationarity conditions. The univariate WRF temperature correction based on TLRs had the largest impact on the HM results, which is not surprising as the main processes in high-elevation catchments (i.e., snow and glacier dynamics) depend highly on a reliable spatiotemporal temperature distribution (Meyer et al., 2019). However, the HM results across different simulation periods further underscore the need for a bivariate bias correction in physics-based simulations of seasonal cryohydrological processes. Empirical bias correction methods that preserve the absolute moisture and account for temperature-dependent saturation, such as the ARM formula, have been shown to account for nonlinearity in the elevation dependency of the cold bias in the WRF data (Fig. 8). Due to the correction of hourly RH data, even sub-daily cycles of temperature and RH are considered in the bivariate correction. An univariate correction might break the physical links between temperature and RH, affecting energy balance calculation and evaporation estimates in the HM (Meyer et al., 2019). The uncertainty in the bivariate bias correction can be reduced by directly using specific humidity and dew point temperature of WRF. However, these output variables were not stored and are therefore unavailable for further analysis and testing.

We have not applied a bi- or multivariate bias correction, including WRF precipitation. Meyer et al. (2019) showed that simultaneous correction of RCM temperature and precipitation led to more precipitation below air temperatures of 0 °C and therefore more simulated snowfall in two high-elevation catchments in Switzerland than with the data of the univariate approach, where temperature and precipitation are corrected individually. Multivariate bias corrections correct several variables jointly (e.g., T, P, RH, shortwave radiation, wind) while preserving their multidimensional dependence structure, using methods such as copula-based or multivariate quantile mapping. The correlations and co-variability among all climate drivers are essential, especially for simulating compound events with physics-based models. However, multivariate bias corrections require higher computational resources and are methodologically complex. In addition, the demand for sufficiently long, high-quality observations across all variables to estimate joint distributions increases significantly (Meyer et al., 2019).

Another important meteorological variable is shortwave incoming radiation that was dynamically corrected by the WaSiM model concerning topography and day of the year. WRF wind speed originates in our case study from 10 m above ground and was not adjusted to the finer topography of the WaSiM model. Since dynamical downscaling of wind speed and direction requires a complex model approach (e.g., Wagenbrenner et al., 2016; Quéno et al., 2023), we have refrained from doing so, also knowing that wind speed has no influence on snow redistribution in WaSiM. Wind speed mainly influences the simulation of snow-melt and evapotranspiration rates in WaSiM. Since WaSiM considers only one dominant wind direction for wind-induced snow redistribution, time-dependent forcing with actual wind directions is not yet possible.

4.2 Quality of WRF precipitation

Although the main focus of this technical note is the spatiotemporal WRF temperature distribution for high-resolution (i.e., 25 m) distributed hydrological model simulations in alpine catchments, the quality of WRF precipitation is essential for reliable and plausible HM results (Meyer et al., 2019). A key aspect is the removal of the "drizzle bias", meaning the WRF's tendency to overestimate the occurrence of light precipitation events. This bias was already observed and discussed in other studies (e.g., Lazoglou et al., 2024; Velasquez et al., 2020). These low precipitation events have only a minor impact on the annual water balance. In our case study, the removal of low precipitation events results in a reduction of annual precipitation of only 4 %. Since WaSiM considers the presence of precipitation in its cloudiness estimate, precipitation frequency can indirectly affect other processes, such as the computation of long-wave incoming radiation (Schulla, 2021). We remove low precipitation events using monthly thresholds (Tab. 2) under the assumption that these are stationary over the period 1971 to 2015. However, this assumption is only partly true as the observed monthly precipitation days at Vergoetschen station show an increasing trend (Fig. S10) for the first ten years (i.e., 1971 to 1981). After 1981, the total number of monthly precipitation days seems to be stationary at this site until 2015. Nevertheless, the interannual variability of monthly precipitation days is well visible (Fig. S10). In principle, we could remove the low precipitation events for each year separately, as the precipitation observations from Vergoetschen station overlap with the period of the longer simulation run (1973 to 2015). However, we have deliberately opted for global monthly thresholds to be independent of the simulation period. From our perspective, seasonal or even sub-seasonal threshold values are needed to remove low precipitation frequencies due to the clear tendency concerning observed monthly precipitation days between summer and winter (Fig. S11).

Apart from the precipitation frequency, we have not further investigated the quality of the WRF precipitation, which is a key driver and source of uncertainty in the HM (Meyer et al., 2019). The evaluation of precipitation timing, magnitude, and intensity is particularly limited by station observations due to the high spatial heterogeneity and, most importantly, by the large measurement uncertainty in high-elevation regions (> 2000 m a.s.l.). The undercatch of precipitation measurements in cold regions has been demonstrated in several studies (e.g., Kochendorfer et al., 2016; Mair et al., 2016). A potential approach for evaluating WRF precipitation magnitudes would be to evaluate them indirectly by closing the water balance, assuming observed streamflow is available. However, this is only possible when there is no significant contribution from glacier melt and permafrost thawing, which is usually unknown. Due to the many sources of uncertainty, we recommend, in particular, evaluating the seasonal precipitation of RCMs in glaciated high mountain regions with seasonal observation data from glacier mass balance programs, as an evaluation only against streamflow is problematic due to the delayed response and the lack of information on the remaining water balance components, such as evapotranspiration and sublimation. There are only a few glacier observatories in the European Alps with long temporal data availability, including streamflow observations. The Glacier Observatory at Vernagtferner would be a good study site for evaluating seasonal precipitation datasets (WGMS, 2025). Unfortunately, interdisciplinary and long-term observations of cryospheric and hydrological processes are still very limited in high-elevation catchments. Nevertheless, correcting rainfall intensities remains problematic due to limited observational data and non-stationary conditions (Teutschbein and Seibert, 2012; Meyer et al., 2019).

4.3 Representation and evolution of land use categories in WRF

The question arises from the application of more and more high-resolution RCMs in mountainous regions: whether the representation of land cover properties and topography is sufficiently resolved for detailed snow-hydrological studies. Tomasi et al. (2017) observed that WRF reproduces rather poorly near-surface temperature (2 m) over snow-covered terrain in the Adige Valley (eastern Italian Alps), with an evident underestimation, during both daytime and nighttime. The main cause of these errors lies in the miscalculation of the mean grid cell albedo, leading to an inaccurate estimate of the reflected solar radiation calculated by the LSM Noah. Therefore, Tomasi et al. (2017) modified the initialization, the land-use classification, and LSM to improve model results, by intervening in the calculation of the albedo, of the snow cover, and of the surface temperature. By doing so, a significant improvement in the model results compared with the observations is achieved. This means outgoing shortwave radiation is lowered, 2 m temperature maxima increased accordingly, and ground-based thermal inversions are better captured (Tomasi et al., 2017).

Besides the spatial representation of these characteristics, another question is whether the long-term evolution of surface properties is considered or not. Due to the increasing greening effect (i.e., increase in vegetation productivity) and the fast retreat of glacier cover in the proglacial areas (Rumpf et al., 2022), constant land cover classes cannot account for the accelerated change of the surface-albedo properties in high-elevation catchments. Especially for long-term (> 30 years) climate-related studies, a constant land cover parameterization may lead to implausible behavior of some important meteorological variables (e.g., temperature) that are particularly sensitive to the surface albedo. This behavior will impact all following analyses, as demonstrated in this case study. Hence, there is a need for either a robust workaround, as we introduce in this technical note, or the representation of a dynamic land cover. The latter is particularly problematic as land cover information, for example, of the Alpine region, is scarce before 1950 due to the lack of remote sensing data, and the reconstruction of earlier stages is particularly complex (Ramskogler, 2023). Additionally, the re-initialization with updated land cover is a complex task and can lead to unrealistic behavior in RCM and HM simulations. Hence, a robust workaround might be the most suitable approach for most longer RCM applications in snow and glacier-dominated areas.

4.4 Uncertainties in the model cascade

Uncertainties affect all parts of the complex model cascade that was applied in this case study (Jobst et al., 2018; Meyer et al., 2019). The widely used global atmospheric reanalysis product (20CRv3) only assimilates surface pressure data that ensures temporal homogeneity over the entire reanalysis period (i.e., 1836–present), but the limited input data strongly influence variables such as precipitation or near-surface temperature, as they are not directly constrained by observations (Slivinski et al., 2019). Although the ERA5 reanalysis dataset is widely used in Europe for climate change impact studies, it is limited for long-term cryohydrological studies since the LIA in 1850, as ERA5 data are available from 1940.

Another source of uncertainty is the parametrization and handling of the boundary conditions of the WRF model. To reduce uncertainty in dynamical downscaling from the global reanalysis product (20CRv3) to WRF, a nested approach was chosen that considers three different spatial domains and resolutions (18, 6, and 2 km). Although the 2 km WRF can resolve convection and

orographical effects in the central European Alps, the resolution is still too coarse for a plausible representation of temperature and precipitation patterns in high-elevation catchments. As demonstrated in this case study, considerable implausible behavior originates from the direct use of the 2 km WRF data as model forcing for a high-resolution (i.e., 25 m) HM. Based on this observation, we, therefore, conclude that space-related uncertainty (including elevation gradients) between WRF and WaSiM is the dominant uncertainty source in the applied model cascade. One plausible argument is that a statistical bias correction would remove the cold bias in the 2 m WRF temperature. However, this requires long-term and continuous temperature observations at higher altitudes (> 2700 m) that are particularly scarce and represent only in-situ information. In addition, meteorological observations are highly affected by errors and failures at these altitudes (Hofmeister et al., 2023). In fact, higher elevations of mountainous regions have experienced stronger warming trends (e.g., Rangwala and Miller, 2012; Kuhn and Olefs, 2020), as also observable in the WRF temperature data (Tab. S5), which challenges the linear extrapolation of statistical bias corrections to higher elevations.

We have refrained from a comprehensive calibration of the HM parameters, as our intention was to demonstrate the impact of uncorrected and corrected RCM forcing data on the model results. Hence, we used a standard parameterization that relied mainly on values from the literature. Usually, parameter optimization algorithms try to compensate for errors in the model forcing (Meyer et al., 2019) by selecting extreme and implausible values within the parameter ranges (e.g., for the rain-snow threshold), especially when applied in a single-objective calibration. A multi-objective optimization approach will most likely fail to provide reasonable and plausible model results for individual processes, such as snow and glacier dynamics and their interactions with streamflow generation (van Tiel et al., 2022; Schaffhauser et al., 2024; Meyer et al., 2019) under the assumption of biased model forcing.

Another major uncertainty in the HM setup concerns the initialization of the glacier volume, especially in highly glaciated catchments such as the upper Kauner Valley. Due to a lack of glacier data in 1968, we had to rely on published values to parameterize the VA scaling factor (Chen and Ohmura, 1990). Another uncertainty arises from the conceptual glacier routine implemented in WaSiM, which does not account for actual ice flow dynamics in balancing mass between accumulation and ablation areas. This is particularly problematic if the model is applied for longer simulation periods, as demonstrated by Pesci et al. (2023). We observed a delayed recession in the simulated fractional glacier cover (Fig. 11c), which is most likely an effect of the simple glacier representation in WaSiM. In general, it is well known that simplified glacier routines have issues when applied for periods with positive glacier mass balances and glacier advance (Seibert et al., 2018; Schaffhauser et al., 2024).

The date and length of the initialization period slightly impact the simulated annual streamflow. When starting WaSiM in 1969, the annual streamflow was 1.1 % lower at the end of the simulation period compared to the simulation that started in 2006 in the case of the univariate model run. Together with the consistent GOF criteria, this indicates that the WRF WaSiM setup provides reliable results even after 42 years of simulation period. Consequently, these results show that the HM provides essential information on the performance and reliability of downscaled and bias corrected meteorological reanalysis data for high mountain areas. Process-based evaluation of the HM results provides comprehensive information on the quality of bias correction methods and should therefore be an integral part of bias correction frameworks for RCMs (Meyer et al., 2019).

A general disadvantage of the presented model cascade from RCM to HM is the high computational effort required for dynamical downscaling with the WRF model. Although the research catchments of interest are relatively small, about 60 km², a significantly larger area needs to be modeled with the WRF model. Therefore, simplified three-dimensional atmospheric models, such as Berg et al. (2024), are more efficient for the dynamical downscaling of RCM data to the catchment scale. Coupling these simplified atmospheric models with distributed snow-HMs would enable sensitivity and projection runs across multiple ensembles, reducing computational time. This coupling is therefore a promising direction for further development in the field of climate-related studies of the cryo- and hydrosphere interactions.

5 Conclusions

The objective of this technical note is the efficient correction of coarse-resolved 2 km RCM (i.e., WRF) data for high-resolution (i.e., 25 m spatial discretization) hydrological simulations in high-elevation and glaciated catchments. When comparing the hourly 2 m WRF temperature with in-situ stations at the case study site, the upper Kauner Valley, we observed a distinct negative bias at the highest station (> 2600 m a.s.l.). Further analysis revealed that the negative bias is due to the relatively coarse representation of surface properties controlling land-atmosphere interactions, specifically snow and glacier cover, in the WRF model (Fig. 2). Due to a lack of high-altitude meteorological stations (> 2700 m a.s.l.) and nonstationarity in the observation data, univariate empirical quantile mapping approaches, which are commonly used in climate-sensitivity studies, are not suitable for physics-based hydrological modeling of snow- and glacier-dominated catchments. Therefore, we developed a workflow to extract hourly temperature lapse rates from the WRF atmosphere. By spatially averaging hourly temperature and geopotential height at the respective WRF eta levels, the calculated TLRs are less dependent on WRF land cover types. In addition, we corrected the hourly 2 m WRF relative humidity using the corrected WRF temperature, applying a bivariate bias correction to ensure physical consistency between the two variables even for sub-daily cycles. Since most previous evaluations of bias correction methods focused on monthly or daily resolutions, neglecting sub-daily cycles, this technical note introduces a novelty by bias correcting hourly RCM data for a highly glaciated Alpine catchment.

We evaluated the plausibility and reliability of the corrected 2 m WRF temperatures in a two-stage process. First, against a set of four in-situ stations that were covering an altitudinal gradient of almost 900 m. Second, by forcing a physics-based hydrological model (i.e., WaSiM) over two periods with six and 42 years of simulation time with an hourly time step with the uni- and bivariate correction products, and critically evaluating the model performance against several observation products (e.g., streamflow, in-situ SWE, fractional snow-covered area, glacier cover and mass balances).

The comparison of hourly corrected WRF temperature frequencies with in-situ observations proved a high agreement. When forcing the WaSiM model with uncorrected WRF data, WaSiM failed to reliably simulate cryospheric and hydrological processes in the Kauner Valley. The univariate WRF temperature correction based on TLRs had the largest impact on the HM results, as expected, since the main processes in high-elevation catchments (i.e., snow and glacier dynamics) depend highly on a reliable spatiotemporal temperature distribution. Additionally, we observed that the bivariate correction further enhances the magnitude of seasonal streamflow, especially in August and September, when glacier melt contributes at its maximum.

We tested the transferability and consistency of the long-term temperature trend of the corrected WRF temperature by comparing monthly and annual temperatures for the period 1973 to 2015 with homogenized station data from the HISTALP data set. The Obergurgl station is located in a neighboring valley (i.e., the Ötz Valley) approximately 20 km away from the Kauner Valley at 1936 m a.s.l.. Both temperature time series showed a significant positive trend (i.e., station $0.06\text{ }^{\circ}\text{C a}^{-1}$; 770 corrected WRF $0.05\text{ }^{\circ}\text{C a}^{-1}$) over this period, with the corrected WRF product slightly underestimating the annual warming by $0.01\text{ }^{\circ}\text{C}$. However, we also found that the uncorrected WRF temperature shows the same statistical behaviour as the corrected WRF temperature for the Obergurgl WRF grid cell over the same period, further supporting the conclusion that the WRF temperature bias mainly occurs in snow- and ice-dominated areas.

This technical note demonstrates that the newly proposed method for deriving TLRs from the free WRF atmosphere can 775 remove biases arising from WRF's coarse-resolution land-cover representation, is physically consistent, is transferable to other sites, is independent of non-stationary conditions, and is applicable to future climate impact studies. Additionally, the technical note provides comprehensive and valuable insights into the complex interdependencies among the cryo- and hydrosphere in highly glaciated catchments, and the need for accurate spatiotemporal representation of temperature, precipitation, and relative humidity in rugged topography with large elevation gradients for climate sensitivity studies. Nevertheless, further research is 780 needed, for example, on the impact of spatial averaging TLRs at larger spatial scales ($> 200\text{ km}^2$) and for catchments with different topography, land cover, and geographical location.

Code and data availability. A compiled version of the hydrological model *WaSiM* is available at <http://www.wasim.ch/en/> for Windows and Linux OS (no source code available). The Δ Snow model (v1.0.2) used for transforming the snow height data to snow water equivalent is accessible in the R package "*nixmass*" (Winkler et al., 2021). The modified Mann-Kendall test after Hamed and Rao (1998) is available in 785 the "*modifiedmk*" R package (version 1.6 <https://cran.r-project.org/web/packages/modifiedmk/modifiedmk.pdf>). The Kolmogorov–Smirnov (KS) test is available in the base R code (version 1.5.1 <https://www.rdocumentation.org/packages/dgof/versions/1.5.1/topics/ks.test>). The Pettitt test is available in the "*trend*" R package (version 1.1.6 <https://www.rdocumentation.org/packages/trend/versions/1.1.6/topics/pettitt.test>). The glacier outlines are available from the Austrian Glacier Inventories for 1969 (GI 1), 1998 (GI 2), 2006 (GI 3), and 2015 (GI 4) on PANGAEA. The digital elevation model of the entire Tyrol can be downloaded in 5 m or 10 m spatial resolution (https://www.data.gv.at/katalog/dataset/land-tirol_tirgelnde#additional-info). The homogenized temperature data at the Obergurgl station are obtained from 790 the HISTALP data set (<https://www.zamg.ac.at/histalp/>). Precipitation records from the Vergoetschen station were downloaded from the Austrian data portal eHYD (<https://ehyd.gv.at/>). The temperature records of the Vernagt Pegel station are available on PANGAEA (<https://doi.pangaea.de/10.1594/PANGAEA.829530>). The other hydrometeorological data of the upper Kauner Valley are only available from TIWAG on request. The WRF data of the Kauner Valley domain for the period 1969 to 2015 will be uploaded to an online repository as soon 795 as the manuscript is accepted for publication.

Author contributions. All authors participated in the conceptualization, methodology, and design of the technical note. MP performed the WRF simulations and conducted the initial analysis of the WRF data. FH conducted the formal analyses and investigations, including

hydrological modeling. FH wrote the original manuscript and visualized all results. All co-authors contributed to the editing and review of the manuscript. BS and GC supervised the work and acquired the project funding.

800 *Acknowledgements.* The authors acknowledge the valuable discussions with Marco Möller about the methodology and results of this study. The authors thank the project partners at Eurac Research for providing the information on the catchment land cover. The authors thank the TIWAG-Tiroler Wasserkraft AG, HD-Tyrol, and GeoSphere Austria for the hydrometeorological observation data. FH, XF, and GC acknowledge the support of the German Research Foundation (DFG) research unit (FOR2793/2) investigating the "Sensitivity of High Alpine Geosystems to Climate Change since 1850" (SEHAG) under grant CH981/3-2. MP and BM are also part of the SEHAG DFG
805 research unit and acknowledge the grant numbers MA 6966/4-2 and LA4426/1-2. XF and BS acknowledge the Swiss National Science Foundation (SNSF) grant (200020E-204030) for "SEHAG Subproject 2 -Impact of climate change on groundwater storage in high Alpine catchments: from observation to model predictions".

Competing interests. The authors declare that there is no competing interest in this work.

References

- 810 Abermann, J., Seiser, B., Meran, I., Stocker-Waldhuber, M., Goller, M., and Fischer, A.: The third glacier inventory of North Tyrol, Austria, for 2006, <https://doi.org/10.1594/PANGAEA.806960>, artwork Size: 6250 data points Pages: 6250 data points, 2012.
- Alduchov, O. A. and Eskridge, R. E.: Improved Magnus Form Approximation of Saturation Vapor Pressure, 35, 601–609, [https://doi.org/10.1175/1520-0450\(1996\)035<0601:IMFAOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1996)035<0601:IMFAOS>2.0.CO;2), 1996.
- Altmann, M., Piermattei, L., Haas, F., Heckmann, T., Fleischer, F., Rom, J., Betz-Nutz, S., Knoflach, B., Müller, S., Ramskogler, K., Pfeiffer, M., Hofmeister, F., Ressler, C., and Becht, M.: Long-Term Changes of Morphodynamics on Little Ice Age Lateral Moraines and the Resulting Sediment Transfer into Mountain Streams in the Upper Kauner Valley, Austria, 12, 3375, <https://doi.org/10.3390/w12123375>, 2020.
- 815 Altmann, M., Pfeiffer, M., Haas, F., Rom, J., Fleischer, F., Heckmann, T., Piermattei, L., Wimmer, M., Braun, L., Stark, M., Betz-Nutz, S., and Becht, M.: Long-term monitoring (1953–2019) of geomorphologically active sections of Little Ice Age lateral moraines in the context of changing meteorological conditions, 12, 399–431, <https://doi.org/10.5194/esurf-12-399-2024>, 2024.
- 820 Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E.: A land use and land cover classification system for use with remote sensor data, <https://doi.org/10.3133/pp964>, 1976.
- Auer, I., Böhm, R., Jurković, A., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Brunetti, M., Nanni, T., Maugeri, M., Briffa, K., Jones, P., Efthymiadis, D., Mestre, O., Moisselin, J.-M., Begert, M., Brazdil, R., Bochnicek, O., Cegnar, T., Gajić-Čapka, M., Zaninović, K., Majstorović, S., Szalai, S., Szentimrey, T., and Mercalli, L.: A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002, 25, 139–166, <https://doi.org/10.1002/joc.1135>, eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1135>, 2005.
- 825 Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplová, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, 27, 17–46, <https://doi.org/10.1002/joc.1377>, 2007.
- 830 Bahr, D. B., Pfeffer, W. T., and Kaser, G.: A review of volume-area scaling of glaciers, 53, 95–140, <https://doi.org/10.1002/2014RG000470>, number: 1, 2015.
- 835 Berg, J., Reynolds, D., Quéno, L., Jonas, T., Lehning, M., and Mott, R.: A seasonal snowpack model forced with dynamically downscaled forcing data resolves hydrologically relevant accumulation patterns, 12, 1393 260, <https://doi.org/10.3389/feart.2024.1393260>, 2024.
- Buckel, J. and Otto, J.-C.: The Austrian Glacier Inventory GI 4 (2015) in ArcGis (shapefile) format, <https://doi.org/10.1594/PANGAEA.887415>, artwork Size: 4 MBytes Medium: application/zip Pages: 4 MBytes, 2018.
- Chen, J. and Ohmura, A.: Estimation of Alpine glacier water resources and their change since the 1870s, 193, 127–135, https://www.researchgate.net/profile/Atsumu_Ohmura/publication/241663882_Estimation_of_Alpine_Glacier_Water_Resources_and_their_Change_Since_the_1870s/links/53fcea990cf22f21c2f52215.pdf, 1990.
- 840 Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, , Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Quarterly Journal of the Royal Meteorological Society, 137, 1–28, <https://doi.org/10.1002/qj.776>, 2011.
- 845

- Duethmann, D., Peters, J., Blume, T., Vorogushyn, S., and Güntner, A.: The value of satellite-derived snow cover images for calibrating a hydrological model in snow-dominated catchments in Central Asia, 50, 2002–2021, <https://doi.org/10.1002/2013WR014382>, number: 3, 2014.
- 850 Faghih, M., Brissette, F., and Sabeti, P.: Impact of correcting sub-daily climate model biases for hydrological studies, *Hydrology and Earth System Sciences*, 26, 1545–1563, <https://doi.org/10.5194/hess-26-1545-2022>, 2022.
- Foehn, A., García Hernández, J., Schaeffli, B., and De Cesare, G.: Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments, *Journal of Hydrology*, 563, 1092–1110, <https://doi.org/10.1016/j.jhydrol.2018.05.027>, 2018.
- 855 Frei, P., Kotlarski, S., Liniger, M. A., and Schär, C.: Future snowfall in the Alps: projections based on the EURO-CORDEX regional climate models, 12, 1–24, <https://doi.org/10.5194/tc-12-1-2018>, 2018.
- Förster, K., Oesterle, F., Hanzer, F., Schöber, J., Huttenlau, M., and Strasser, U.: A snow and ice melt seasonal prediction modelling system for Alpine reservoirs, 374, 143–150, <https://doi.org/10.5194/piahs-374-143-2016>, 2016.
- Giese, B. S., Seidel, H. F., Compo, G. P., and Sardeshmukh, P. D.: An ensemble of ocean reanalyses for 1815–2013 with sparse observational input, *Journal of Geophysical Research: Oceans*, 121, 6891–6910, <https://doi.org/10.1002/2016JC012079>, 2016.
- 860 Gómez-Navarro, J. J., Raible, C. C., and Dierer, S.: Sensitivity of the WRF model to PBL parametrisations and nesting techniques: evaluation of wind storms over complex terrain, 8, 3349–3363, <https://doi.org/10.5194/gmd-8-3349-2015>, 2015.
- Hakala, K., Addor, N., and Seibert, J.: Hydrological Modeling to Evaluate Climate Model Simulations and Their Bias Correction, *Journal of Hydrometeorology*, 19, 1321 – 1337, <https://doi.org/10.1175/JHM-D-17-0189.1>, 2018.
- Hamed, K. H. and Rao, A. R.: A modified Mann-Kendall trend test for autocorrelated data, 204, 182–196, [https://doi.org/10.1016/S0022-1694\(97\)00125-X](https://doi.org/10.1016/S0022-1694(97)00125-X), number: 1-4, 1998.
- 865 Hock, R.: A distributed temperature-index ice- and snowmelt model including potential direct solar radiation, 45, 101–111, <https://doi.org/10.3189/S0022143000003087>, number: 149, 1999.
- Hofmeister, F., Arias-Rodriguez, L. F., Premier, V., Marin, C., Notarnicola, C., Disse, M., and Chiogna, G.: Intercomparison of Sentinel-2 and modelled snow cover maps in a high-elevation Alpine catchment, 15, 100 123, <https://doi.org/10.1016/j.hydroa.2022.100123>, 2022.
- 870 Hofmeister, F., Graziano, F., Marcolini, G., Willems, W., Disse, M., and Chiogna, G.: Quality assessment of hydrometeorological observational data and their influence on hydrological model results in Alpine catchments, p. 02626667.2023.2172335, <https://doi.org/10.1080/02626667.2023.2172335>, 2023.
- Hong, S.-Y., Noh, Y., and Dudhia, J.: A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes, 134, 2318–2341, <https://doi.org/10.1175/MWR3199.1>, 2006.
- 875 Huang, J.: A Simple Accurate Formula for Calculating Saturation Vapor Pressure of Water and Ice, 57, 1265–1272, <https://doi.org/10.1175/JAMC-D-17-0334.1>, 2018.
- 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, 592, 726–731, <https://doi.org/10.1038/s41586-021-03436-z>, 2021.
- 880 Jobst, A. M., Kingston, D. G., Cullen, N. J., and Schmid, J.: Intercomparison of different uncertainty sources in hydrological climate change projections for an alpine catchment (upper Clutha River, New Zealand), *Hydrology and Earth System Sciences*, 22, 3125–3142, <https://doi.org/10.5194/hess-22-3125-2018>, 2018.
- Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, 378, 161–167, <https://doi.org/10.1016/j.jhydrol.2009.09.021>, number: 1, 2009.

- Kochendorfer, J., Rasmussen, R., Wolff, M., Baker, B., Hall, M. E., Meyers, T., Landolt, S., Jachcik, A., Isaksen, K., Brækkan, R., and Leeper, R.: The Quantification and Correction of Wind-Induced Precipitation Measurement Errors, preprint, *Hydrometeorology/Instruments and observation techniques*, <https://doi.org/10.5194/hess-2016-415>, 2016.
- Kotlarski, S., Gobiet, A., Morin, S., Olefs, M., Rajczak, J., and Samacoïts, R.: 21st Century alpine climate change, 60, 65–86, <https://doi.org/10.1007/s00382-022-06303-3>, 2023.
- Kuhn, M. and Olefs, M.: *Elevation-Dependent Climate Change in the European Alps*, Oxford University Press, <https://doi.org/10.1093/acrefore/9780190228620.013.762>, 2020.
- Lazoglou, G., Economou, T., Anagnostopoulou, C., Zittis, G., Tzyrkalli, A., Georgiades, P., and Lelieveld, J.: Multivariate adjustment of drizzle bias using machine learning in European climate projections, 17, 4689–4703, <https://doi.org/10.5194/gmd-17-4689-2024>, 2024.
- Lehning, M., Bartelt, P., Brown, B., Fierz, C., and Satyawali, P.: A physical SNOWPACK model for the Swiss avalanche warning, 35, 147–167, [https://doi.org/10.1016/S0165-232X\(02\)00073-3](https://doi.org/10.1016/S0165-232X(02)00073-3), 2002.
- Magnusson, J., Nævdal, G., Matt, F., Burkhart, J. F., and Winstral, A.: Improving hydropower inflow forecasts by assimilating snow data, 51, 226–237, <https://doi.org/10.2166/nh.2020.025>, 2020.
- Mair, E., Leitinger, G., Chiesa, S. D., Niedrist, G., Tappeiner, U., and Bertoldi, G.: A simple method to combine snow height and meteorological observations to estimate winter precipitation at sub-daily resolution, *Hydrological Sciences Journal*, 61, 2050–2060, <https://doi.org/10.1080/02626667.2015.1081203>, number: 11, 2016.
- Marcolini, G., Bellin, A., Disse, M., and Chiogna, G.: Variability in snow depth time series in the Adige catchment, 13, 240–254, <https://doi.org/10.1016/j.ejrh.2017.08.007>, 2017.
- Marcolini, G., Koch, R., Chimani, B., Schöner, W., Bellin, A., Disse, M., and Chiogna, G.: Evaluation of homogenization methods for seasonal snow depth data in the Austrian Alps, 1930–2010, 39, 4514–4530, <https://doi.org/10.1002/joc.6095>, number: 11, 2019.
- Marty, C.: Regime shift of snow days in Switzerland: REGIME SHIFT OF SNOW DAYS, 35, n/a–n/a, <https://doi.org/10.1029/2008GL033998>, number: 12, 2008.
- Matiu, M., Jacob, A., and Notarnicola, C.: Daily MODIS Snow Cover Maps for the European Alps from 2002 onwards at 250 m Horizontal Resolution Along with a Nearly Cloud-Free Version, 5, 1, <https://doi.org/10.3390/data5010001>, 2019a.
- Matiu, M., Jacob, A., and Notarnicola, C.: Daily MODIS snow cover maps for the European Alps from 2002 onwards at 250m horizontal resolution along with a nearly cloud-free version, <https://doi.org/10.5281/ZENODO.3601891>, 2019b.
- Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., De Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., Notarnicola, C., Petitta, M., Scherrer, S. C., Strasser, U., Winkler, M., Zebisch, M., Cicogna, A., Cremonini, R., Debernardi, A., Faletto, M., Gaddo, M., Giovannini, L., Mercalli, L., Soubeyroux, J.-M., Sušnik, A., Trenti, A., Urbani, S., and Weigluni, V.: Observed snow depth trends in the European Alps: 1971 to 2019, 15, 1343–1382, <https://doi.org/10.5194/tc-15-1343-2021>, 2021.
- Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, *Hydrology and Earth System Sciences*, 23, 1339–1354, <https://doi.org/10.5194/hess-23-1339-2019>, 2019.
- Morin, S., Samacoïts, R., François, H., Carmagnola, C. M., Abegg, B., Demiroglu, O. C., Pons, M., Soubeyroux, J.-M., Lafaysse, M., Franklin, S., Griffiths, G., Kite, D., Hoppler, A. A., George, E., Buontempo, C., Almond, S., Dubois, G., and Cauchy, A.: Pan-European meteorological and snow indicators of climate change impact on ski tourism, 22, 100 215, <https://doi.org/10.1016/j.cliser.2021.100215>, 2021.

- Nigrelli, G. and Chiarle, M.: 1991–2020 climate normal in the European Alps: focus on high-elevation environments, *Journal of Mountain Science*, 20, 2149–2163, <https://doi.org/10.1007/s11629-023-7951-7>, 2023.
- Oke, T. R.: *Boundary layer climates*, Routledge, 2nd ed edn., OCLC: 228443090, 2002.
- 925 Patzelt, G.: Austrian glacier inventory 1969 (G I), <https://doi.org/10.1594/PANGAEA.807098>, artwork Size: 38382 data points Pages: 38382 data points, 2013.
- Pesci, M. H., Schulte Overberg, P., Bosshard, T., and Förster, K.: From global glacier modeling to catchment hydrology: bridging the gap with the WaSiM-OGGM coupling scheme, 5, 1296–1344, <https://doi.org/10.3389/frwa.2023.1296344>, 2023.
- Pettitt, A. N.: A Non-Parametric Approach to the Change-Point Problem, *Applied Statistics*, 28, 126, <https://doi.org/10.2307/2346729>, number: 2, 1979.
- 930 Piermattei, L., Heckmann, T., Betz-Nutz, S., Altmann, M., Rom, J., Fleischer, F., Stark, M., Haas, F., Ressler, C., Wimmer, M. H., Pfeifer, N., and Becht, M.: Evolution of an Alpine proglacial river during 7 decades of deglaciation, 11, 383–403, <https://doi.org/10.5194/esurf-11-383-2023>, 2023.
- Powers, J. G., Klemp, J. B., Skamarock, W. C., Davis, C. A., Dudhia, J., Gill, D. O., Coen, J. L., Gochis, D. J., Ahmadov, R., Peckham, S. E., Grell, G. A., Michalakes, J., Trahan, S., Benjamin, S. G., Alexander, C. R., Dimego, G. J., Wang, W., Schwartz, C. S., Romine, G. S., Liu, Z., Snyder, C., Chen, F., Barlage, M. J., Yu, W., and Duda, M. G.: The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions, 98, 1717–1737, <https://doi.org/10.1175/BAMS-D-15-00308.1>, 2017.
- 935 Quéno, L., Mott, R., Morin, P., Cluzet, B., Mazzotti, G., and Jonas, T.: Snow redistribution in an intermediate-complexity snow hydrology modelling framework, <https://doi.org/10.5194/egusphere-2023-2071>, 2023.
- 940 Ramskogler, K.: Vegetation development in proglacial areas of the European Alps : patterns, trends, and driving factors, <https://ulb-dok.uibk.ac.at/ulbtirolhs/content/titleinfo/9098480>, 2023.
- Rangwala, I. and Miller, J. R.: Climate change in mountains: a review of elevation-dependent warming and its possible causes, 114, 527–547, <https://doi.org/10.1007/s10584-012-0419-3>, 2012.
- Rist, A., Roth, L., and and, H. V.: Elevational ground/air thermal gradients in the Swiss inner Alpine Valais, Arctic, Antarctic, and Alpine Research, 52, 341–360, <https://doi.org/10.1080/15230430.2020.1742022>, 2020.
- 945 Rogger, M., Chirico, G. B., Hausmann, H., Krainer, K., Brückl, E., Stadler, P., and Blöschl, G.: Impact of mountain permafrost on flow path and runoff response in a high alpine catchment, 53, 1288–1308, <https://doi.org/10.1002/2016WR019341>, 2017.
- Rumpf, S. B., Gravey, M., Brönnimann, O., Luoto, M., Cianfrani, C., Mariethoz, G., and Guisan, A.: From white to green: Snow cover loss and increased vegetation productivity in the European Alps, 376, 1119–1122, <https://doi.org/10.1126/science.abn6697>, 2022.
- 950 Schaffhauser, T., Tuo, Y., Hofmeister, F., Chiogna, G., Huang, J., Merk, F., and Disse, M.: SWAT-GL: A new glacier routine for the hydrological model SWAT, 60, 755–766, <https://doi.org/10.1111/1752-1688.13199>, 2024.
- Schaffhauser, T., Hofmeister, F., Chiogna, G., Merk, F., Tuo, Y., Machnitzke, J., Alcamo, L., Huang, J., and Disse, M.: Merits and limits of SWAT-GL: application in contrasting glaciated catchments, *Hydrology and Earth System Sciences*, 29, 3227–3256, <https://doi.org/10.5194/hess-29-3227-2025>, 2025.
- 955 Schmidt, L. K., Francke, T., Rottler, E., Blume, T., Schöber, J., and Bronstert, A.: Suspended sediment and discharge dynamics in a glaciated alpine environment: identifying crucial areas and time periods on several spatial and temporal scales in the Ötztal, Austria, *Earth Surface Dynamics*, 10, 653–669, <https://doi.org/10.5194/esurf-10-653-2022>, number: 3, 2022.
- Schulla, J.: Model Description WaSiM (Water balance Simulation Model), http://www.wasim.ch/downloads/doku/wasim/wasim_2021_en.pdf, 2021.

- 960 Schöber, J., Schneider, K., Helfricht, K., Schattan, P., Achleitner, S., Schöberl, F., and Kirnbauer, R.: Snow cover characteristics in a glacierized catchment in the Tyrolean Alps - Improved spatially distributed modelling by usage of Lidar data, *Journal of Hydrology*, 519, 3492–3510, <https://doi.org/10.1016/j.jhydrol.2013.12.054>, 2014.
- Seibert, J., Vis, M. J. P., Kohn, I., Weiler, M., and Stahl, K.: Technical note: Representing glacier geometry changes in a semi-distributed hydrological model, *Hydrology and Earth System Sciences*, 22, 2211–2224, <https://doi.org/10.5194/hess-22-2211-2018>, 2018.
- 965 Shrestha, M., Acharya, S. C., and Shrestha, P. K.: Bias correction of climate models for hydrological modelling—are simple methods still useful?, *Meteorological Applications*, 24, 531–539, 2017.
- Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Wang, W., Huang, X.-Y., and Duda, M.: A Description of the Advanced Research WRF Version 3, Tech. rep., UCAR/NCAR, <https://doi.org/10.5065/D68S4MVH>, artwork Size: 1002 KB Medium: application/pdf, 2008.
- Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Blancq, F. L., Lee, T., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszynski, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, *Quarterly Journal of the Royal Meteorological Society*, 145, 2876–2908, <https://doi.org/10.1002/qj.3598>, 2019.
- 970 Smitha, P., Narasimhan, B., Sudheer, K., and Annamalai, H.: An improved bias correction method of daily rainfall data using a sliding window technique for climate change impact assessment, 556, 100–118, <https://doi.org/10.1016/j.jhydrol.2017.11.010>, 2018.
- Stahl, K., Moore, R. D., Shea, J. M., Hutchinson, D., and Cannon, A. J.: Coupled modelling of glacier and streamflow response to future climate scenarios: MODELLING OF GLACIER AND STREAMFLOW, 44, <https://doi.org/10.1029/2007WR005956>, number: 2, 2008.
- 980 Stocker-Waldhuber, M., Seiser, B., and Fischer, A.: Meteorological data Weißseespitze/Austria, 2017 et seq., <https://doi.org/10.1594/PANGAEA.939830>, artwork Size: 7 datasets Medium: application/zip Pages: 7 datasets, 2022.
- Strasser, U., Förster, K., Formayer, H., Hofmeister, F., Marke, T., Meißl, G., Nadeem, I., Stotten, R., and Schermer, M.: Storylines of combined future land use and climate scenarios and their hydrological impacts in an Alpine catchment (Brixental/Austria), 657, 746–763, <https://doi.org/10.1016/j.scitotenv.2018.12.077>, 2019.
- 985 Teutschbein, C. and Seibert, J.: Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods, *Journal of Hydrology*, 456–457, 12–29, <https://doi.org/https://doi.org/10.1016/j.jhydrol.2012.05.052>, 2012.
- Tomasi, E., Giovannini, L., Zardi, D., and De Franceschi, M.: Optimization of Noah and Noah_MP WRF Land Surface Schemes in Snow-Melting Conditions over Complex Terrain, 145, <https://doi.org/10.1175/MWR-D-16-0408.1>, 2017.
- 990 Tuo, Y., Marcolini, G., Disse, M., and Chiogna, G.: A multi-objective approach to improve SWAT model calibration in alpine catchments, *Journal of Hydrology*, 559, 347–360, <https://doi.org/10.1016/j.jhydrol.2018.02.055>, 2018.
- van Tiel, M., Stahl, K., Freudiger, D., and Seibert, J.: Glacio-hydrological model calibration and evaluation, 7, <https://doi.org/10.1002/wat2.1483>, 2022.
- Velasquez, P., Messmer, M., and Raible, C. C.: A new bias-correction method for precipitation over complex terrain suitable for different climate states: a case study using WRF (version 3.8.1), 13, 5007–5027, <https://doi.org/10.5194/gmd-13-5007-2020>, 2020.
- Vionnet, V., Brun, E., Morin, S., Boone, A., Faroux, S., Le Moigne, P., Martin, E., and Willemet, J.-M.: The detailed snowpack scheme Crocus and its implementation in SURFEX v7.2, 5, 773–791, <https://doi.org/10.5194/gmd-5-773-2012>, 2012.

- Wagenbrenner, N. S., Forthofer, J. M., Lamb, B. K., Shannon, K. S., and Butler, B. W.: Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja, 16, 5229–5241, <https://doi.org/10.5194/acp-16-5229-2016>, 2016.
- 1000 Warscher, M., Strasser, U., Kraller, G., Marke, T., Franz, H., and Kunstmann, H.: Performance of complex snow cover descriptions in a distributed hydrological model system: A case study for the high Alpine terrain of the Berchtesgaden Alps, 49, 2619–2637, <https://doi.org/10.1002/wrcr.20219>, number: 5, 2013.
-
- Warscher, M., Wagner, S., Marke, T., Laux, P., Smiatek, G., Strasser, U., and Kunstmann, H.: A 5 km Resolution Regional Climate Simulation for Central Europe: Performance in High Mountain Areas and Seasonal, Regional and Elevation-Dependent Variations, 10, 682, 1005 <https://doi.org/10.3390/atmos10110682>, 2019.
- WGMS: Fluctuations of Glaciers Database, <https://doi.org/10.5904/WGMS-FOG-2017-10>, 2017.
- WGMS: Vernagtferner, Alps – world glacier monitoring service, https://wgms.ch/products_ref_glaciers/vernagtferner-alps/, 2025.
-
- Wilhelm, B., Rapuc, W., Amann, B., Anselmetti, F. S., Arnaud, F., Blanchet, J., Brauer, A., Czymzik, M., Giguet-Covex, C., Gilli, A., Glur, L., Grosjean, M., Irmeler, R., Nicolle, M., Sabatier, P., Swierczynski, T., and Wirth, S. B.: Impact of warmer climate periods on flood hazard 1010 in the European Alps, 15, 118–123, <https://doi.org/10.1038/s41561-021-00878-y>, number: 2, 2022.
- Winkler, M., Schellander, H., and Gruber, S.: Snow water equivalents exclusively from snow depths and their temporal changes: the Δ_{snow} model, 25, 1165–1187, <https://doi.org/10.5194/hess-25-1165-2021>, 2021.