

## RESEARCH ARTICLE

# A quantile-based approach to improve homogenization of snow depth time series

Gernot Resch<sup>1</sup>  | Roland Koch<sup>2</sup> | Christoph Marty<sup>3</sup>  | Barbara Chimani<sup>2</sup>  |  
Michael Begert<sup>4</sup>  | Moritz Buchmann<sup>3,5,6</sup>  | Johannes Aschauer<sup>3</sup>  |  
Wolfgang Schöner<sup>1</sup> 

<sup>1</sup>University of Graz, Department of Geography and Regional Science, Graz, Austria

<sup>2</sup>Zentralanstalt für Meteorologie und Geodynamik ZAMG, Vienna, Austria

<sup>3</sup>WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

<sup>4</sup>Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

<sup>5</sup>Institute of Geography, University of Bern, Bern, Switzerland

<sup>6</sup>Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

## Correspondence

Gernot Resch, Institute of Geography and Regional Science, University of Graz, Heinrichstraße 36, 8010 Graz, Austria.  
Email: [gernot.resch@uni-graz.at](mailto:gernot.resch@uni-graz.at)

## Funding information

FWF Fonds zur Förderung der wissenschaftlichen Forschung, Project Hom4Snow, Grant/Award Number: I 3692

## Abstract

Austrian observations of snow depth date back to 1895 and are thus among the longest available quantitative snow information from hydrometeorological networks worldwide. It is well known that such long-term observations are prone to inhomogeneities, which may not only affect climatologies and trends, but derived products used in research or practice. While the reliability of available methods for detecting breaks in snow time series has been shown before and could also be confirmed by our work, we focused on improving the adjustment method. Conventional methods often refer to the median of difference or quotient series (INTERP), whereas our proposed method also uses a quantile-wise adjustment (InterpQM), which is useful to minimize a bias on the tails of the frequency distribution. We demonstrated the success of the new method by using Swiss parallel snow depth observations. Errors of the analysed indicators could be reduced in 68% of the cases when compared with INTERP. The results were best for large snow depths, being up to 19% better. Overall, InterpQM was better in 75% of validation cases for the daily large, 72% of all observations and 56% of mean seasonal snow depth cases. We describe the performed homogenization procedure in detail, including quality control, gap filling, homogeneity testing, break detection, calculation of and improvements to the adjustment method. Our results show that snow depth time series generally have a lower number of breaks compared with station data of other climate variables. This underlines their high quality, even if measuring snow presents challenges. Using Austrian snow depth series as an example, the effects of the new adjustment method on trends were analysed using the Mann–Kendall and Sen's Slope. Homogenization may have a significant effect on derived trends: Two of the six adjusted series were changed from nonsignificant to significant and one vice versa.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

**KEYWORDS**

Alps, climate change, homogenization, mountain, observational data analysis, quantile matching, snow, statistical methods

**1 | INTRODUCTION**

Deposited snow is an important component of the global climate system. In Northern Hemisphere winters, it can cover more than 50% of the land surface (Armstrong and Brun, 2008). For mountain regions, snow on the ground is an essential and widespread component of the cryosphere and plays an essential role in feeding glaciers and energy budget and thus as an insulating and reflective layer of the Earth's surface (Hock et al., 2019). Deposited snow is influenced by melting, sublimation/deposition and settling in addition to precipitation processes and exhibits a relatively high spatial and particularly high temporal autocorrelation. Presence or absence and the thickness of a snow cover have a major impact on plant and animal species (e.g., Lamprecht et al., 2018; Steinbauer et al., 2018; Johnston et al., 2019) and for the society of Alpine countries, for example, via its influence on the availability of water in early spring via snowmelt. It also has a large socio-economic dimension in regions heavily dependent on winter tourism (Ballotta et al., 2020). Several studies have highlighted the changes in snow cover in the Alpine region due to climate change (e.g., Hock et al., 2019; Schöner et al., 2019; Matiu et al., 2021). Although there are indications that, due to the decrease in snow, the summer season is becoming more attractive in the affected regions (Statistik Austria 2019), the role of winter tourism is important for the economy of these regions, especially since winter tourists usually spend more money and, in some regions, ~80% of the workforce is employed in the tourism sector (Pröbstl-Haider et al., 2021).

A recent analysis of more than 2000 stations in the European Alps (Matiu et al., 2021) found that 85% of the stations showed a negative trend for mean monthly snow depth for winter (NDJFMAM) in the period 1971–2019. 27% of the negative trends were significant, but none of the positive trends. Pulliainen et al. (2020) found a significant decrease in snow-mass in the Northern Hemisphere, although there are areas such as Siberia and coastal regions with an increase in snow-mass in the period 1980–2018.

The two most frequently observed variables, snow depth (HS, the total height of the snow cover from the base to the snow surface) and depth of snowfall (HN), are measured at about 7000 stations worldwide (Armstrong et al., 2009). In Austria, snow is measured at about 160 stations by the Zentralanstalt für Meteorologie und

Geodynamik (ZAMG) and at about 690 stations by the Hydrographischer Dienst (HD). The snow records of the HD are usually much longer and date back to 1895. The reason for this difference is, that a large part of the Austrian observation data collected by ZAMG was destroyed during the Second World War. While the density of the measurement network for weather stations in inhabited areas is quite good, the station network for snow measurements is patchy, especially in mountainous regions (Hock et al., 2019). Although measurement techniques have changed over time and automatic stations measuring snow depth by ultrasound or laser are becoming more common (Matiu et al., 2021), the method of snow observation has not changed for most stations in Austria and is still done manually at 07:00 CET (Haberkorn et al., 2019). This is a major advantage for the analysis of long-term snow trends, as changes in the measurement technique of climate variables often introduce systematic errors.

As Auer et al., 2007 noted, longer climate measurements not only show pure climate variability but also non-climatic influences. These so-called inhomogeneities cause a systematic change in the time series and are usually the result of a relocation, changing observers, changes in the station's environment or instruments and measurement procedures. Effects like these must be considered when analysing time series of the data, although it is difficult to determine the respective influence of these effects for snow depth, mainly because of the lack of trustworthy metadata (Buchmann et al., 2021). It is also well known that wind-induced errors can greatly affect snow observations (Sevruk, 1985; Goodison et al., 1998; Lehning et al., 2008). Due to the many challenges in making accurate and representative snow observations, consistent long-term records of snow are rare (Vaughan et al., 2013). Homogenization, now a standard procedure for processing climate data (Auer et al., 2007), is used to remove these nonclimatic influences with statistical methods. It consists of detecting and adjusting shifts, often called breaks, in the time series by relative homogeneity methods (relative to neighbouring stations). While this is well researched for variables such as temperature and precipitation (Venema et al., 2012), fewer methods have been developed and tested for homogenizing snow series (Marcolini et al., 2017, 2019; Schöner et al., 2019). A new study focusing on improvements in break detection is currently in review (Buchmann et al., 2022, under review). Regarding

the adjustments, the SNHT (Marcolini et al., 2017, 2019) and a modified version of INTERP (Vincent et al., 2002; Schöner et al., 2019), which is implemented in HOMOP (Nemec et al., 2013), have been applied to daily snow depth series. INTERP uses a single adjustment (correction) factor per break. Although this improved the overall data quality, the method can artificially bias the tails of the frequency distribution of snow depths. Therefore, Marcolini et al. (2019) recommended to only homogenize monthly or seasonal snow depth values to obtain robust results.

To overcome these limitations and make the homogenization of daily values possible, we extended INTERP with quantile matching. For testing and evaluating the improved method, different inhomogeneity scenarios were created using selected Swiss parallel series as a basis. The impact of the new method on trends was then compared with INTERP-homogenized and original Austrian time series.

## 2 | DATA AND METHODS

### 2.1 | Swiss parallel dataset

Swiss snow observations are mainly carried out by two institutions: the Federal Office of Meteorology and Climatology (MeteoSwiss) and the WSL Institute for Snow and Avalanche Research (SLF). 28 pairs with independent manual snow depth observations for at least 20 years, located at a maximum horizontal distance of 3 km and a vertical distance of fewer than 100 m were compiled to a parallel dataset, including quality checks and gap-filling. The stations are between 490 m a.s.l. (Payerne) and 1770 m a.s.l. (Bivio) and have high correlations: The smallest correlation based on daily values is 0.84 (Payerne), the average correlation is 0.95 and the highest is 0.99 (Santa Maria, 1415 m a.s.l.). More details about the dataset can be found in Buchmann et al., 2021.

The quality and availability of the corresponding metadata vary over time. The metadata of the SLF stations at least include relocations and observer changes. Information on MeteoSwiss stations is generally more detailed, but specific snow information is still sparse or not accurate enough. For example, sometimes there is a sequence of very similar observer names, and it is unclear whether this is a change of observer or a correction of the name. Also, nonspecific names have been used, for example, “Swiss Border Force”, which makes an unnoticed change of observers very likely. Independence of observations can be deceptive and difficult to (dis)prove. For example, stations are listed in two separate databases under different names, different coordinates suggest two independent stations or missing periods were filled in

with the parallel station sometime in the past. All these points were considered when creating the Swiss parallel dataset.

To validate the success of a new homogenization method, appropriate validation data sets are needed, which in a perfect world would represent the truth (e.g., true snow depth). In the real world, however, these are not available, so artificial datasets have been created for this purpose in homogenization studies. Such approaches have been shown for temperature and precipitation (e.g., Venema et al., 2012) to evaluate the respective homogenization methods. We follow a different approach in our study by not only using artificial snow series but also datasets from parallel measurements (two independent measurement series at almost the same location). The parallel measurements provide a best-possible setup for homogenization studies (Venema et al., 2020), which means in our case testing and validating the effectiveness of the adjustment methods. However, such data are generally rare, especially those covering longer periods. Luckily, a new dataset containing daily snow depth observations from Switzerland (Aschauer and Marty, 2020; Buchmann et al., 2021) was available and was therefore chosen for testing and improving our adjustment methods for snow homogenization. Selected parallel series were then altered to construct artificial series of snow depth with less (SCless) or more snow (SCmore) by subtracting or adding between 5–30 cm snow as described in Section 2.3.4.

### 2.2 | Austrian dataset

Snow depth observations are carried out at about 1000 stations in Austria (Haberkorn et al., 2019), most of which are operated by ZAMG and HD. About 87% of these observations, especially at HD stations, are still measured manually at 07:00 CET. This is advantageous for the homogeneity of a time series, as observations with the same methodology cover a period from the late nineteenth century until today. First automatic observations were introduced around 2010. The ZAMG stations use laser sensors, while the HD relies on ultrasonic devices. Depending on the network, observers follow slightly different guidelines and procedures (Haberkorn et al., 2019) but the modalities remain very similar, allowing the two data sources to be merged. From these networks, the SNOWPAT dataset (Schöner et al., 2019; Olefs and Koch, 2020) was created, which is the basis for this work. It consists of daily quality-checked (HS and HN) and homogenized (HS) manual snow observations, currently from 40 ZAMG and 46 HD stations. For this study it was extended by 21 stations, mainly from southeastern Austria in order to improve the spatial coverage. Since so

many snow stations in Austria are still having manual measurements, only those were considered in our study in order to exclude inhomogeneities resulting from automatization. The stations are distributed all over Austria (Figure 1), with a denser network in alpine areas. The lowest station is located at 121 m a.s.l. (Podersdorf am See), the highest at 2140 m a.s.l. (Villacher Alpe). Other high-altitude stations are currently not included. As the digitization of historical observations is time-consuming, many of the HD's historical observations have not yet been fully digitized, but there are continuous efforts to do so. As a result, most of the HD time series available and used for analyses (e.g., Matiu et al., 2021) start with 1971, but thanks to some HD province offices there has been access to digitized and quality-checked long-term observations, with the earliest records included starting with 1895. The mean length of the HS series in the dataset is 79 years for ZAMG and 105 years for HD. Most of the stations contain both daily HS and HN observations, with HS usually starting earlier. Interestingly, this contrasts with Switzerland (Scherrer et al., 2013).

The extent of available metadata depends on the station network operator. ZAMG has an extensive collection of information on its stations, which increases with time. It contains information on station relocations and changes in the observation system, such as observer changes. In contrast, less information is available from HD, where only station relocations are documented.

Before use, the data set was already subjected to comprehensive quality control by the operating institutions. The aim was to reduce the uncertainty in the data and to produce consistent, physically plausible time series of snow depth and snowfall depth. Time series used had been quality controlled by the station supervisors and were subjected to further checks, for example, for

internal and external consistency. The identification of errors was done by several plausibility checks. Whereas the ZAMG series had already been quality checked for the entire series length, most of the stations provided by the HD had only been checked back to the early 1970s, with earlier digitized observations being raw data. With additional checks, remaining errors which usually resulted from data-digitization could be identified and excluded. An uncertainty remained in the data set: a suspicious number of observations ending with 5 or 0, for example, 150 or 165, which occurred mainly at higher snow depths and on stakes with a 5 cm scale. This is likely a rounding error, caused by the human observers. The same error was also found in the Swiss dataset (Aschauer and Marty, 2020).

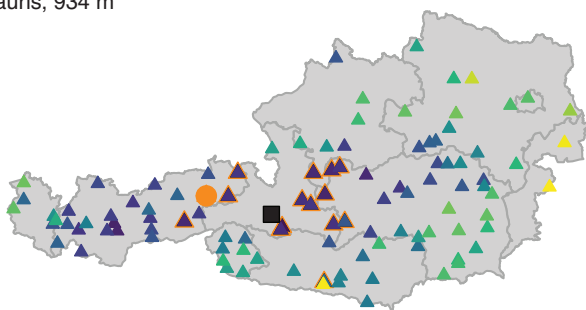
Gaps with a length of up to 6 years were filled using the gradient-plus-inverse-distance-squared method GIDS if the necessary conditions were fulfilled (Nalder and Wein, 1998). Unfortunately, gaps between 1940 and 1946 often could not be filled due to the low number of reference stations. GIDS is a combination of multiple linear regression (MLR) and inverse distance weighting (IDW). In regression (1),  $X$  is the snow depth (to be predicted),  $Lon$  (longitude),  $Lat$  (latitude) and  $h$  (station height) are the predictors. The resulting coefficients  $b_1$ ,  $b_2$  and  $b_3$  were used to adjust the snow depth of each reference station to the geographical location of the candidate station.

$$X = b_0 + b_1 Lon + b_2 Lat + b_3 h \quad (1)$$

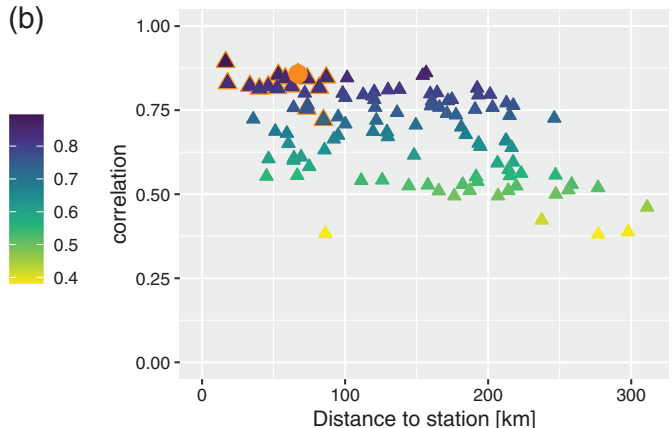
In the next step, the adjusted snow depth series was weighted with IDW (2). The HS value ( $HS_p$ ) to be predicted at a given location ( $Lon_p$ ,  $Lat_p$ ,  $h_p$ ),  $n$  is the number of reference stations used and  $d_i$  is the respective distance to the candidate station location:

(a)

Rauris, 934 m



(b)



**FIGURE 1** Correlations between the Austrian candidate station Rauris (black square) and all other stations in the dataset on a map (a) and against horizontal distance (b). The orange circle indicates the selected reference station, the orange borders show that the necessary conditions for selection as a reference station are fulfilled. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7742)]

$$HS_p = \left[ \sum_{i=1}^n \left( \frac{1}{d_i} \right)^2 \right]^{-1} * \sum_{i=1}^n \left\{ \left[ (HS_i + b_1(Lon_p - Lon_i) + b_2(Lat_p - Lat_i) + b_3(h_p - h_i)) \left( \frac{1}{d_i} \right)^2 \right] \right\} \quad (2)$$

The successful application of GIDS strongly depends on the number and representativeness of the reference stations used. Best results were obtained with 10–20 stations at <100 km distance. After filling, all results were again checked for plausibility. With these steps, the method provided satisfactory results for both the HS and HN series. For evaluation, a three-year synthetic gap was created in selected stations, filled and the results tested with LOOCV (leave one out cross-validation), excluding one station at a time. The results showed increasing uncertainty with increasing altitude, which is a known problem for observations at high elevations (e.g., Gulpepe, 2015; Prein and Gobiet, 2017). The mean absolute error (MAE) and the mean bias error (MBE) were calculated for assessing the results, which were also visually inspected. In this way, stations that were unsuitable for filling gaps could be easily identified and excluded from the calculations. A recently performed comparison of gap-filling methods for snow depth time series (Aschauer and Marty, 2021) showed that there are better methods than GIDS available, less impacted by station density. However, with carefully performed plausibility checks the method produced useful results.

## 2.3 | Homogenization of snow depth series

### 2.3.1 | Reference series

To detect breaks in the time series of a candidate station, the observations are compared with either one or several other stations, so called reference stations, which should be subject to the same climate as the candidate time series. Reference stations for snow depth homogenization face several challenges: As shown in Figure 1 using the example of the Austrian station Rauris (934 m), the correlation based on daily values with neighbouring stations does not change linearly with distance. Due to complex orographic situations and related influences such as exposure to shortwave radiation, wind speed and direction, amount of snowfall and temperature, closest stations do not necessarily show the highest correlation. Thus, it can be quite low for stations at a smaller distance, but above an acceptable threshold for stations

>200 km away. The amount of snow changes with the vertical distance as well, which is a good reason to exclude stations exceeding a certain elevation distance as it has a strong influence on the adjustment factor.

As stated in Venema et al. (2020), two independent observation series at the same location, one of which can be used as a reference for the other, would be the best possible solution for the task. However, since this is hardly the case for most stations, other solutions must be found. These can either be a highly correlated neighbouring station (e.g., Nemec et al., 2013) or a computed composite of several neighbouring stations, which can also be weighted differently (e.g., Szentimrey, 1999; Domonkos, 2015). Both approaches have specific advantages, disadvantages and requirements. It is well known in climate research that the probability for inhomogeneities in long-term climate observational series is high (Auer et al., 2007; Venema et al., 2020). Consequently, the homogeneity of reference series must also always be questioned and cannot be assumed. It is important to note that a series incorrectly assumed to be homogeneous implies the risk of introducing inhomogeneities and a bias to the series to be homogenized. To keep this risk low when using fully automated homogenization packages, for example, ACMANT (Domonkos, 2015), it is therefore common to use either weighted or unweighted mean composites. The decision which series to include or exclude for this can have a noticeable impact on the homogenization result, as no metadata is used in these approaches.

In the case of the Swiss parallel dataset, each pair's MeteoSwiss station served as candidate and the SLF station as a reference, which was decided out of practical reasons and to have consistency for the comparisons. For the Austrian dataset, an unweighted composite and a single reference approach were compared. To ensure that the candidate and reference station had similar climate conditions and to avoid false correlations, the selection criteria from Marcolini et al., 2019 were applied: Minimum correlation of 0.7, within 100 km horizontal and 300 m vertical distance. Time series meeting these criteria were selected as possible reference series and ranked according to the correlation. For the single reference variant, the station with the highest correlation was used. If at least four series met the conditions, an

unweighted average was calculated as a composite. The comparison did not show any significant advantage for one or the other variant in the analysed cases and since the former had already been used for snow data (Marcolini et al., 2019; Schöner et al., 2019), this approach was pursued. Most of the time, the highest correlated reference station was chosen, but if its data quality or representativeness was doubtful, the second highest was used instead. Reasons for this were either identified breaks in the reference series within 10 years of a break in the candidate station or due to major differences in the immediate station locations or local climate (narrow valley in the candidate station versus a wide plain in the reference station). This was the case for Rauris and Bad Gastein (both in Salzburg, Austria) due to a break in the respective reference series and Oed (Lower Austria) due to the station environment.

### 2.3.2 | Break detection

After Mestre et al. (2011), each time series consists of a climatic and a station effect as well as random white noise. With a constant station effect, a time series is seen as homogeneous. With break detection, constant shifts of station effects are identified and adjusted afterwards.

Metadata is considered as an important source for deciding whether a detected break in the data can be accepted or not (Auer et al., 2007) and is a valuable source of information for homogenization to understand sources of inhomogeneities. However, as noted by Venema et al. (2020), all metadata are subject to a degree of uncertainty and may also contain errors that can be misleading. Therefore, the available metadata was only used as an indicator for the uncertainty of observations: It was assumed that time series from stations that have experienced major shifts in the past may poorly represent the “true” temporal variability. Shifts due to changed observation practices can be excluded for the analysed Austrian and Swiss manual snow depth time series. In our case, all breaks were caused either by station relocations or observer changes, which mostly also included relocations.

Snow depth shows a strong interannual variability, especially compared with long-term trends (Marke et al., 2015). This is caused by its sensitivity to different atmospheric conditions, location and altitude. To overcome these high-frequency variations and to provide robust trends at larger temporal and spatial scales, previous studies of (at least to our knowledge) break-detection in snow time series were based on seasonal data (NDJFMA, DJF). Although there are several different methods for detecting systematic shifts (breaks) in

climate data series, to our knowledge only two are known today to be used for detecting breaks in snow height (HS) series: SNHT (Alexandersson, 1986; Alexandersson and Moberg, 1997; Moberg and Alexandersson, 1997) and PRODIGE (Causinus and Mestre, 2004), which is implemented in ZAMG's homogenization package HOMOP. A comparison of the two methods showed good overall performance for both, with equally credible results. Differences occurred mainly for suspicious fractures at very low snow depths, which were detected by HOMOP but not by SNHT (Marcolini et al., 2019). Due to the assumption of homoscedasticity in the homogeneous time series, both tests are not able to detect inhomogeneities in trends, where inhomogeneities in the trend often show up as a string of single jumps.

For the creation of the evaluation-dataset, homogeneous subperiods were identified in the HS series of the Swiss parallel dataset using the SNHT. The quotient series (candidate/reference) were split into shorter sequences until either only homogeneous sub-periods remained, or the periods were too short (< 10 years, Moberg and Alexandersson, 1997) to provide stable results and were excluded from further analysis. As with SNOWPAT (Schöner et al., 2019; Olefs et al., 2020), the test was carried out with quotients of mean seasonal snow depth (HSmean, NDJFMA). Since homogeneity tests have difficulty interpreting changes at the beginning and end of a time series (Toreti et al., 2011), corresponding results <10 years after the beginning/before the end of a time series were flagged as suspicious. Alexandersson (1986) developed the SNHT to detect a single break in a time series. Therefore, it has difficulties with time series that contain two or more breaks (Moberg and Alexandersson, 1997). In this case, the significance of the largest break may be biased. As also emphasized by Alexandersson (1986), if two equal but opposite shifts occur, both may not be detected as would be the case for multiple shifts in the same direction. The latter would be considered as natural variability. In addition, the SNHT is not the optimal method to correctly detect trends (Moberg and Alexandersson, 1997). If inhomogeneities in trends occur, they would be detected as a series of shifts.

### 2.3.3 | Adjustment methods

#### *INTERP*

Once the breakpoints were identified, the inhomogeneities in the daily snow observations were adjusted with INTERP (Vincent et al., 2002). It is a median-based approach that can do reasonable adjustments to climate time series and improve their homogeneity and has been used to calculate daily adjustments for several climate

variables, for example, temperature (Vincent et al., 2002) or relative humidity (Chimani et al., 2018). Since ratios between the candidate series and the reference series are generally used for the homogenization (adjustment), it is thus predefined that the number of days with snow cover in the candidate series would not be increased by the homogenization. The adjustment factor is calculated according to (3), where  $C/R$  are the daily time series of the candidate/reference station,  $a/b$  the time after/before the detected inhomogeneity. All data of the inhomogeneous subsection of the candidate series are multiplied by the same adjustment factor, which must not be correct in the case of very small or very large snow depths.

$$\text{Adjustment factor} = \frac{\left(\frac{\text{median}(C_a)}{\text{median}(R_a)}\right)}{\left(\frac{\text{median}(C_b)}{\text{median}(R_b)}\right)} \quad (3)$$

### InterpQM

INTERP is known to provide useful results and improve a time series homogeneity (Vincent et al., 2002; Chimani et al., 2018). Schöner et al. (2019) used it to homogenize snow depth time series but emphasized that only monthly or seasonal values should be analysed, as it may introduce a bias for small and very high values in the daily observations.

This constraint was the motivation for implementing improvements to INTERP, which was extended by the concept of quantile matching. For this purpose, the data was first split into several subsets/interquantile ranges (IQR), for example, from the 0th to 94th and 95th to 100th percentile, as shown in Figure 2. A “raw” adjustment factor was then calculated for each IQR (called “InterpQMraw” in Figure 2) using Formula 3. To avoid artificial jumps in snow depths, the “raw” factors of two neighbouring IQRs were then linearly interpolated on a percentile scale between the centres of the two IQRs. No interpolation was performed to the lower (upper) half of the smallest (largest) IQR. The resulting factors (called “InterpQM” in Figure 2) were then applied to each corresponding percentile of the time series. If the number of unique snow depth values in a time series was <100, the interpolation was based on the number of available unique values. To determine the optimum interquantile ranges (optimal in terms of the number of values per IQR, its position in the percentiles and the improvement achieved), several versions of InterpQM were compared with the performance of INTERP on the evaluation data. The best performing version (see Section 3.2) was then selected for homogenization of the SNOWPAT dataset.

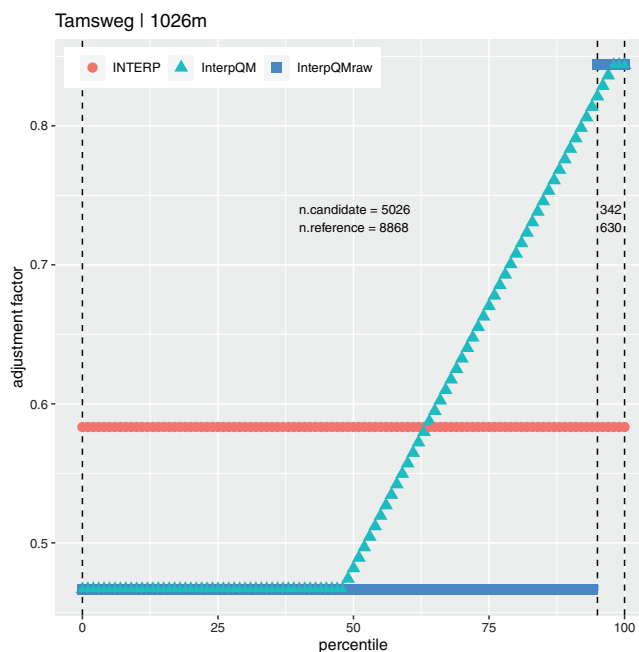


FIGURE 2 Calculated adjustment factors for the second break of the Austrian station Tamsweg based on interquantile ranges (IQR). Red circles: INTERP, blue squares: Adjustment factors of the IQRs (InterpQMraw), turquoise triangles: Interpolated adjustment factors (InterpQM), vertical dashed lines: Separation of the IQRs.  $n_{\text{candidate}}$  and  $n_{\text{reference}}$  are the numbers of observations in each IQR in the candidate and reference series. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

### 2.3.4 | Evaluation of adjustment methods

Based on the results of the SNHT applied to the Swiss dataset, eight station-pairs were selected to evaluate the performance of the different adjustment methods. The station pairs are listed in Table 1, together with their altitude, the correlation of the daily snow depth of each station pair and the chosen evaluation period, with the break at the end of the period, respectively. The presence and timing of the detected breaks were decisive for the selection of suitable station pairs, so that no break would influence the results. If a break was found in the series, it had to be at least 20 years before the end of the observations so there would be no periods to be tested shorter than 10 years. The next steps were to select a 10-year period of the MeteoSwiss station, add artificial inhomogeneities, adjust them and evaluate the effectiveness of the adjustment methods (INTERP, InterpQM). To account for a broader range of potential inhomogeneities, a set of three different scenarios was created for each pair: SCrep, SCless and SCmore. In the former, all values in the test period were replaced by a highly correlated neighbouring MeteoSwiss station. To simulate the relocation of a station from a location with less (SCless) or more snow

Name	Altitude (m)	corrHS	Test period	Break
Adelboden	1325	0.92	1984–1993	31 October 1993
Andermatt	1440	0.99	1999–2008	31 October 2008
Bivio	1770	0.97	1984–1993	31 October 1993
Davos	1560	0.97	1977–1986	31 October 1986
Einsiedeln/Sihlsee	910	0.97	1970–1979	31 October 1979
Küblis	815	0.96	1997–2006	31 October 2006
Santa Maria	1415	0.99	1980–1989	31 October 1989
Zuoz/S-chanf	1660	0.97	1955–1964	31 October 1964

**TABLE 1** List of the selected Swiss station pairs for testing the efficiency of the adjustment methods, together with its altitude (defined as the mean of each pair's altitudes), the correlation between candidate and reference, the selected 10-year test period and date of artificial break.

(SCmore) than at its actual location, the snow depths were adjusted by subtracting or adding between 5–30 cm snow. The adjustments were linearly interpolated between the 0th and 100th percentiles and then added to or subtracted from the corresponding values, for example, five was added to the lowest, 13 to the middle and 30 to the highest snow depths of a station. No new snow days were added to SCmore, but some were removed from SCless through this process. The time series were restricted to the beginning of the 10-year test period and the end of the observations.

To assess the effectiveness of the applied adjustments, the mean arctangular absolute percentage error MAAPE (4) (Kim and Kim, 2016) and RMSE (5) were calculated for three snow indicators: All daily snow depths (HSdaily), seasonal (NDJFMA) means of snow depth (HSmean) and the largest 5% of daily snow depths (HS95). HS95 was preferred here over the seasonal maximum snow depths (HSmax) as a more robust quantity, because it analyses not only the single largest but a whole group of large values. Since HS95 is not a common snow indicator, it was not used later but the more common HSmax was used for further comparisons. The RMSE based on seasonal values was also used to quantify the Efficiency (6) (Domonkos et al., 2012) of the performed homogenization. MAAPE was used due to its more stable results in datasets with many zeros and small numbers, which is inherent in snow observations. Efficiency has been used also in other studies on time series homogenization (e.g., Gubler et al., 2017; Chimani et al., 2018). HSorig is the original snow depth series, HShom is the homogenized evaluation series, RMSEval is the RMSE from the evaluation versus original time series and RMSEhom is the RMSE from the homogenized versus original time series.

$$\text{MAAPE} = \frac{1}{n} \sum_{i=1}^n \arctan_i \left( \left| \frac{HSorig_i - HShom_i}{HSorig_i} \right| \right) \quad (4)$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^n (HSorig_i - HShom_i)^2} \quad (5)$$

$$E = \frac{\text{RMSE}_{\text{val}} - \text{RMSE}_{\text{hom}}}{\text{RMSE}_{\text{val}}} \times 100 \quad (6)$$

### 2.3.5 | Impact on time series

The effects of the homogenization on long-term trends in the snow series were assessed with the nonparametric, rank-based two-sided Mann–Kendall test (Mann, 1945; Kendall, 1975; Lettenmaier et al., 1994) applied on various snow-climate indicators. For this purpose HSmean and HSmax, the mean and maximum seasonal snow depth, were selected as together they cover a significant portion of the winter snow depth distribution. The significance level used for rejecting the H0 (no trend in the time series) was a  $p$  value  $< 0.05$ . To assess the strength of the trend, the Theil–Sen (Sen, 1968) was calculated for the two indicators. This involves sorting the pairs of time and observation from smallest to largest and calculating the slopes between each pair. The median of all slopes is then referred to as the “Sens Slope” and used as a measure to assess the overall strength of a trend.

In order to analyse to what extent homogenized time series differ from the original ones, the Wilcoxon rank-sum test and Kruskal–Wallis test, followed by Dunn's test at a 95% confidence interval were performed to HSdaily, HSmean and the number of days per season with snow depth  $> 1$  (dHS1) of the original and homogenized time series. The mean of the adjustments and their standard deviation were compared with the estimated uncertainty of manual snow depth observations for assessing if the adjustments were larger. The uncertainty of manual snow observations depends mainly on each observer and site, the scale on the poles (1 or 5 cm) and the weather



conditions during the observation. Based on SLFs and ZAMGs long experience, it is estimated to <5 cm on a pole with a 5 cm scale and 1–2 cm with a 1 cm scale.

### 3 | RESULTS

#### 3.1 | Break detection

Inhomogeneities in the Swiss dataset of parallel observations were detected applying the SNHT on the quotients of HSmean (NDJFMA) between the candidate and the respective parallel station series: About 66% of the station pairs showed breaks. An overview is presented in Table 2. Between one and two breaks were detected with PRODIGE and confirmed with the SNHT and metadata in six of the 86 Austrian stations (Table 4, Marcolini et al., 2019): Bad Gastein (1092 m), Galtür (1577 m), Oed (400 m), Rauris (934 m), St. Leonhard im Pitztal (1335 m) and Tamsweg (1026 m). A possible explanation for the large difference in the detected number of breaks between SNOWPAT and the parallel dataset are significantly higher correlations between the time series of the station pairs for the latter. This indicates a lower noise of the quotient series, which otherwise would mask smaller

breaks. For additional prove, break detection was also performed for the Swiss HN pairs, indicating that for HS and HN different breaks were identified for almost every station pair. This finding could be explained by different influences on HS and HN observations (e.g., ablation processes for HS), different measurement devices and observation and procedures. If compared with variables like temperature (Venema et al., 2012), a lower number of breaks was detected in the analysed snow depth time series.

#### 3.2 | Evaluation of InterpQM

To compare the performance of INTERP and InterpQM in homogenizing snow depth series and to identify the best performing InterpQM version, the three inhomogeneity scenarios (SCmore, SCless and SCrep), which were artificially created from the Swiss parallel dataset, were adjusted with both methods for the breaks detected, as described in Section 3.1. For this purpose, daily snow depths (HSdaily), seasonal means of snow depth (HSmean) and the largest 5% of daily snow depths (HS95) were analysed. In a first step, the percentage of cases with a better performance of InterpQM than INTERP was calculated. This

TABLE 2 Overview of the stations of the Swiss parallel dataset where breaks were detected using the SNHT on seasonal data.

Station	Correlation	Altitude (m)	Break
Klosters	0.89	1190	1999
Adelboden	0.92	1325	1983
Andermatt	0.99	1440	1998
Bever	0.95	1750	1968
Bivio	0.97	1770	1983
Bosco-Gurin	0.97	1525	1980
Cavaglia	0.93	1690	1999
Crans-Montana	0.97	1590	1976
Gadmen	0.96	1190	1995
Göschenen	0.95	1110	1989, 2009
Einsiedeln/gross	0.97	910	2009
Einsiedeln/Sihlsee	0.97	910	1968
Lauterbrunnen	0.95	800	2000
Mürren	0.97	1650	1988
Samedan	0.93	1750	1995
Sedrun	0.95	1420	1984, 2001
Splügen	0.95	1457	1985, 2004
Zermatt	0.97	1600	1990
Zernez	0.92	1475	1996

Note: The detected break is shown together with each pair's correlation and altitude, which is defined as the mean of each pair's altitudes.

InterpQM ranges	SCmore (%)	SCless (%)	SCrep (%)	Overall (%)
0, 25, 50, 75, 100	40	50	67	52
0, 25, 50, 75, 95, 100	36	50	65	50
0, 50, 100	36	56	72	55
0, 50, 75, 100	36	54	74	54
0, 50, 75, 95, 100	36	54	71	54
0, 50, 95, 100	36	54	72	54
0, 75, 100	67	53	67	62
0, 75, 95, 100	69	53	61	61
0, 95, 100	79	56	68	68

Note: The first column shows the used percentiles for defining the interquartile ranges (IQR).

TABLE 3 Percentage of the analysed Swiss evaluation cases per InterpQM version with better results than INTERP.

TABLE 4 Results of the break detection and adjustment factor determination of the Austrian stations together with their altitude, the correlation based on daily values between candidate and reference and the likely reason for the break.

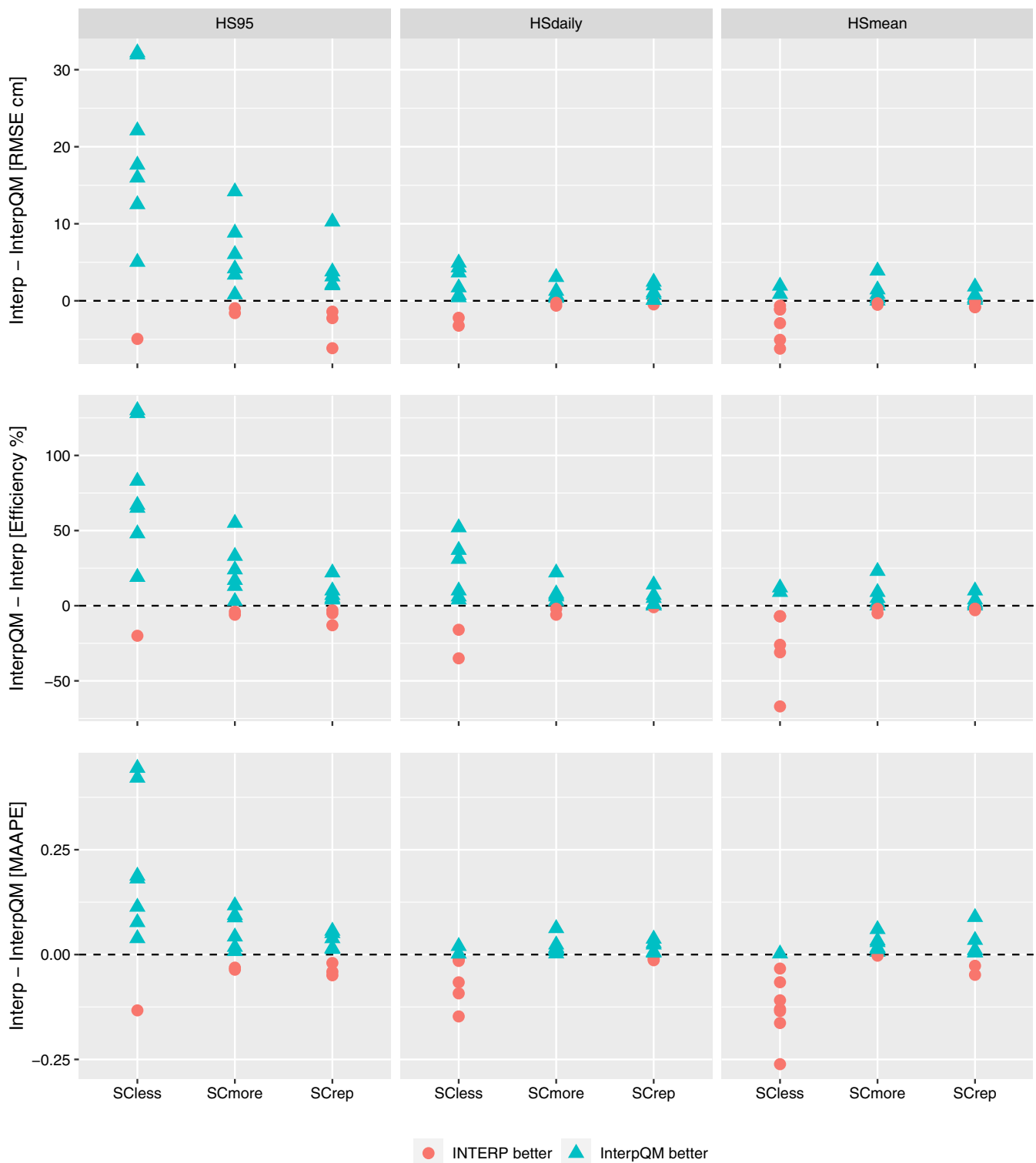
Station	Altitude (m)	Year	INTERP	InterpQM 1–94, 95–100	Correlation	Meta-information
Bad Gastein	1092	1972	1.24	1.18, 1.24	0.89	Change of observer
Galtür	1577	1988	1.01	0.99, 0.94	0.91	Relocation of station
Rauris	934	1973	1.16	1.19, 1.37	0.86	Relocation of station
		1993	1.51	1.56, 1.36		Relocation of station
Tamsweg	1026	1983	0.88	0.61, 1.09	0.85	Relocation of station
		1998	0.58	0.47, 0.84		Relocation of station
Oed	400	1996	1.23	1.75, 1.46	0.81	Relocation of station
St. Leonhard im Pitztal	1335	1985	0.94	1, 1.03	0.92	Relocation of station

evaluation (Table 3) showed that using more IQRs did not necessarily improved the results. The overall best results for all different InterpQM-versions were gained in the more complicated SCrep-dataset, regardless of amount and distribution of the IQRs. The results of SCless slightly favour all different InterpQM versions, but many of them had problems with SCmore, where only three were clearly better than INTERP. From these, the version with two IQRs (0th to 94th, 95th to 100th) showed the overall best results (79% in SCmore, 56% in SCless and 68% in SCrep) and was therefore chosen for the homogenization of the Austrian dataset. In the following, it is referred to as InterpQM and its results are analysed in more detail. For presenting the results in a more intuitive way, a compilation is presented in Figure 3 as a station-by-station difference of the two methods, separated by the analysed scenario and snow indicator. The differences between INTERP–InterpQM were evaluated by the RMSE, the MAAPE and the Efficiency (with a higher value indicating better performance).

The differences between the two methods were largest for SCless and smallest for SCrep. For comparison, the absolute values of the evaluation measures of all

InterpQM versions against INTERP are presented in Figure S1. The results were quite different for the tested scenarios: For both methods, homogenization efficiency was lowest for SCrep and highest for SCmore. The largest differences were found for very large snow depths (HS95), where the RMSE, the Efficiency and the MAAPE for InterpQM were better than for INTERP at seven (out of eight stations) for SCless, six for SCmore and five for SCrep. The differences between the two methods in Efficiency were up to 130% for SCless, 55% for SCmore and 22% for SCrep. For the MAAPE, they ranged between  $-0.13$  and  $0.44$ , with the smallest differences found for SCrep.

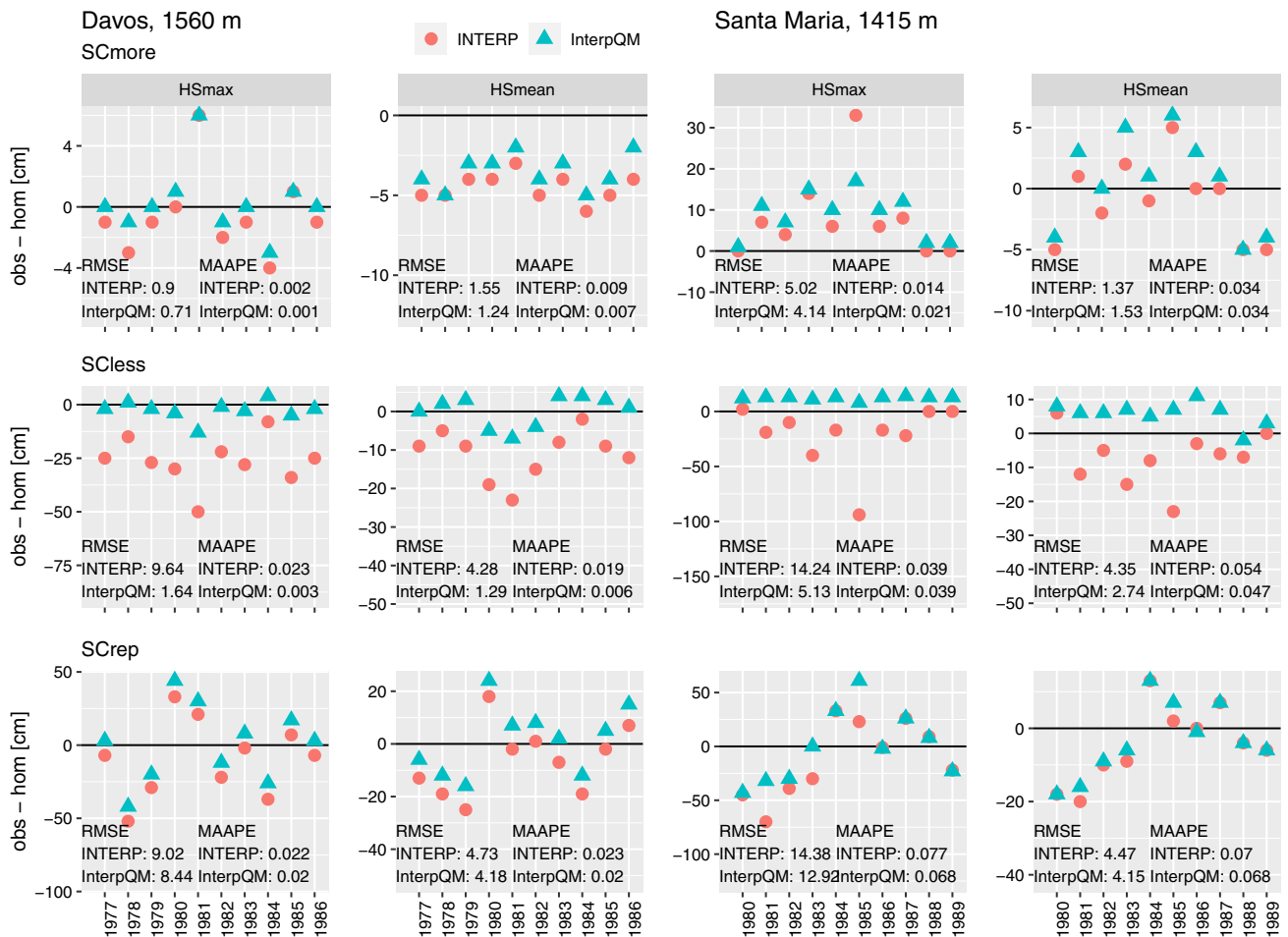
In HSdaily, the RMSE was lower for InterpQM in six out of eight stations in all three evaluation datasets, Efficiency was higher in six stations for SCless and SCmore and seven for SCrep. The differences (INTERP–InterpQM) were smaller than for HS95, ranging from  $-3.2$  to  $+4.9$  cm and  $-52$  to  $+35\%$ . The MAAPE of InterpQM was lower in two stations for SCless, for all stations in SCmore and five stations in SCrep. The results for HSmean showed a better performance of INTERP for six



**FIGURE 3** Comparison of the error measures (RMSE, efficiency and MAAPE) of the homogenization performed for the three Swiss inhomogeneity scenarios SCless, SCmore and SCrep. Triangles and dots show the difference between the two methods for each station, with triangles showing better performance of InterpQM and circles showing better performance of INTERP. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

stations of the SCless scenario for RMSE and Efficiency and seven for MAAPE, but in contrast lower RMSE and Efficiency in six (5) and lower MAAPE in seven

(6) stations with InterpQM for SCmore (SCrep). Since the MAAPE can be roughly considered as a percentage error, it can be summarized that InterpQM reduced the relative



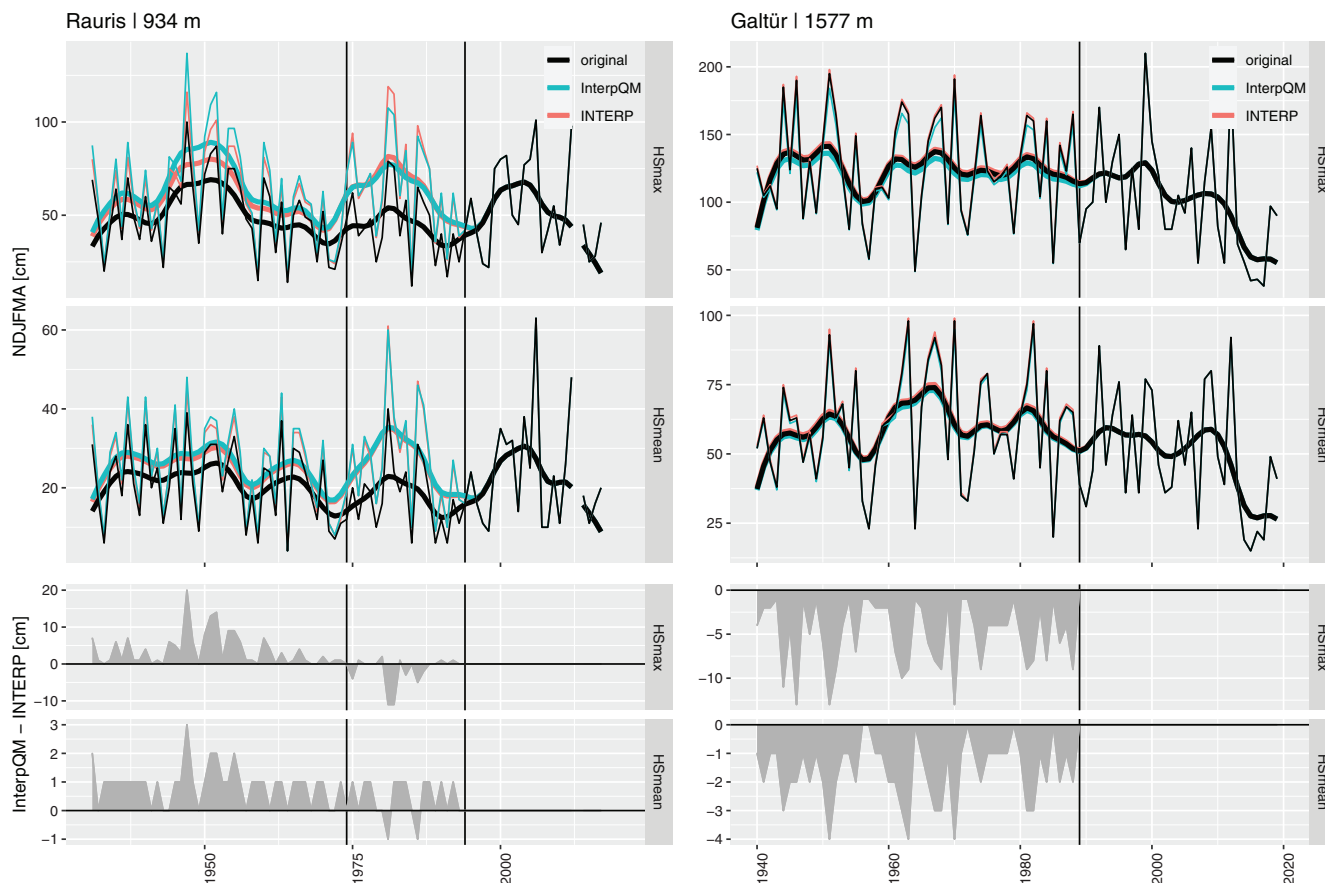
**FIGURE 4** Difference between observed and homogenized data for HSmax (maximum seasonal snow depth) and HSmean (mean seasonal snow depth) of the evaluation data set for the Swiss stations Davos and Santa Maria. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/joc.7712)]

errors of HS95 in SCless between 7% and 19%, between 1% and 4% in SCmore and between 2% and 3% in SCrep better than INTERP. On the contrary, errors of HSmean were 3–26% larger using InterpQM in SCless, but smaller for SCmore (between 1% and 3%) and for SCrep (3–9%). For HSdaily errors were also between 1% and 8% larger with InterpQM for SCless but 1–4% smaller for SCmore and 1–3% for SCrep.

As mentioned, either a single adjustment factor is calculated for all data (INTERP) or one for each IQR that is interpolated to each percentile (InterpQM) of a time series. Individual seasons that differ greatly from the others, for example, with extremely large snow depths, therefore, affect the INTERP adjustment factor and thus all adjusted values, while the results for InterpQM are less influenced. Two examples of this can be seen in Figure 4: The season 1980/1981 in the SCless dataset of Davos and 1984/1985 in Santa Maria. Like this case, a strong year-to-year variability of snow depth also poses a challenge for INTERP.

### 3.3 | Homogenization of Austrian snow depth series

Given the evaluation results, the adjustments with InterpQM should give a more realistic picture of the real snow depths, especially for large and small values. As presented in Table 4, the homogenization with InterpQM increased the snow depth for three stations (Bad Gastein, Oed, Rauris), decreased it for two (Galtür, St. Leonhard im Pitztal) while lowering the smaller and increasing the larger values in Tamsweg. INTERP raised the snow depth in Bad Gastein, Rauris and Oed but as well in Galtür, while decreasing it in St. Leonhard and Tamsweg. As an example, the seasonal mean (HSmean) and maximum snow depth (HSmax) as well as the differences of INTERP and InterpQM of Rauris and Galtür are shown in Figure 5. In Rauris, adjustments prior to the first break in 1973 were up to 21 cm lower with INTERP than those of InterpQM, while it was the other way round, with up to 11 cm larger values between 1973 and the second



**FIGURE 5** Comparison of the original time series and the homogenized versions with INTERP and InterpQM for the Austrian stations Rauris and Galtür for seasonal maximum and mean snow depth. The thick lines are Gaussian filtered with an 11y window. The vertical black lines indicate the breaks in Rauris (1973 and 1993) and Galtür (1988). The lower part of the graph shows the difference between the two homogenization methods. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

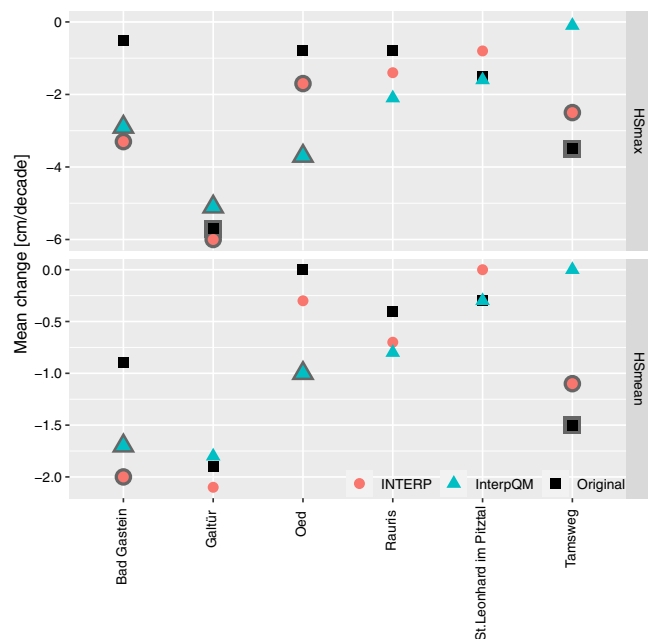
break in 1993. Galtür represents an interesting case, with opposite sign of the adjustments: The snow depth was slightly increased with INTERP but decreased with InterpQM. The difference between the two methods is up to 14 cm for HSmax and 4 cm for HSmean.

### 3.4 | Impact of homogenization

The Wilcoxon and Kruskal–Wallis Test revealed the different impact of the two adjustment methods to the homogenized series: While the adjustments did not cause a significant change for dHS1 for any of the stations, they had a significant effect for HSdaily of all stations with InterpQM, but only for three with INTERP (Galtür, Rauris and Tamsweg). HSmean was only significant for Bad Gastein and St. Leonhard with InterpQM and Bad Gastein, St. Leonhard, Oed and Tamsweg with INTERP. Figure 6 shows long-term trends of snow depth as the mean change per decade and its significance, calculated for the entire time series for HSmax and HSmean. In the

original time series, all stations showed a negative trend for HSmax, only Galtür ( $-5.7$  cm/decade) and Tamsweg ( $-3.5$ ) were significant. When homogenized with INTERP/InterpQM, the negative trend of Bad Gastein was increased by 2.8/2.4 cm/decade, of Oed by 0.9/2.9 and of Rauris by 0.6/1.3. INTERP weakened the trend of St. Leonhard by 0.7 cm/decade while InterpQM slightly intensified it by 0.1. Both methods reduced it in Tamsweg (1/3.4), where the adjustments by InterpQM were so strong that the trend ( $-0.1$  cm/decade) was almost removed. The trends of Bad Gastein and Oed were also significantly changed by both methods, while only InterpQM removed both the negative trend and its significance of Tamsweg.

The HSmean trends were generally smaller and only for Tamsweg significant in the original data. Both INTERP/InterpQM strengthened the trends of Bad Gastein by 1.1/0.8 cm/decade, of Oed by 0.3/1 and of Rauris by 0.3/0.4. In Galtür, INTERP intensified it by 0.2, while it was weakened with InterpQM by 0.1. In St. Leonhard, INTERP reduced it by 0.3, while no changes were done



**FIGURE 6** Summary of trend and significance tests performed for HSmean and HSmax for the adjusted Austrian stations. Colours and symbols denote the original (black), INTERP (red circle) or the InterpQM version (light blue triangle); outlined symbols indicate a significant ( $p$  value < 0.05) trend shown as a mean change per decade. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

by InterpQM. In Tamsweg, INTERP weakened the trend by 0.4, the stronger adjustments by InterpQM (1.5 cm/decade) completely removed it. Due to strong adjustments, both methods changed the trend in Bad Gastein to significant. The changes in Galtür were only minor, but with the large reduction of snow depth, InterpQM turned Oed to significant. Despite the decreased trend in Tamsweg with INTERP it remained significant but was completely removed with InterpQM.

A comparison of the calculated adjustment values, both for INTERP and InterpQM, with the uncertainties of the snow depth observations shows that the mean adjustments of the time series ranged between 2 and 14 cm and the corresponding standard deviation between 0.9 and 6.9 cm. The measurement uncertainty of snow depth can be assumed <2 cm for most cases and <5 cm for special cases, for example, snow stakes with a 5 cm scale. This is lower than the value of the adjustments, or put in another way, the adjustments exceed the inaccuracy of the measurements.

## 4 | CONCLUSIONS

Adjustments to a time series by homogenization can have a strong impact on climatological analysis. Careful

evaluation and application of homogenization is therefore very important. A general disadvantage of homogenized data is that their physical consistency cannot be guaranteed, so they should not be used for all kinds of analyses. The choice between homogenized and non-homogenized data depends on the intended analysis and the cause of the inhomogeneity, for example, a systematic error due to sensor bias or a station relocation. As for trend and most other analyses, homogenized data are clearly preferable, for the analysis of individual extreme values, for example, the highest observed snow depths at a station, nonhomogenized data may be a good choice.

Based on the results of a previous homogenization of snow depth time series (Schöner et al., 2019), and a recently available Swiss parallel dataset, the potential of improving snow homogenization by introducing a quantile-based adjustment scheme was investigated. For this purpose, the widely used method based on a median adjustment (INTERP) was extended with quantile matching (InterpQM). A comparison of the methods showed that the quantile approach generally provides better and more realistic results, especially for very large snow depths. Despite remaining uncertainties, InterpQM provides more robust data, for example for the derivation of trends or analysis of decadal-scale variability. Interestingly, the results of the various evaluation datasets made with the Swiss parallel data series, showed that the implementation of quantile matching improved the results in most but not all cases. This is surprising in that one might expect the quantile adjustment (InterpQM) to be always better or at least equally good as a median-based adjustment method (INTERP).

Since the true measured value is always unknown in homogenization, the results of the evaluations carried out are an important guide for assessing the improved adjustment method. In order to evaluate the impacts of possible station changes on snow depth series, we tried to simulate different inhomogeneity scenarios and then compared the performance of the two methods for homogenization. The comparison of the different inhomogeneity scenarios for different indices shows that Quantile Matching generally leads to the expected improvement of the homogenization of snow depth. Only in exceptional cases, such as the seasonal mean snow depth and a reduced snow depth before the break, INTERP and InterpQM perform similarly, in most other cases InterpQM was better. Overall, the analysed InterpQM-variant performed better in 68% of all evaluation cases, in 56% of cases with reduced pre-break snow depth, 79% of the increased snow-depth- and 68% of the replacement variant cases. When comparing indicators across all scenarios, InterpQM performed better in 75% of all cases for HS95, in 58% for HSdaily and 56% for

HSmean. The analysis of the adjustments for very large snow depths (HS95) showed that the advantages of the quantile matching method were strongest here. This confirms the main idea of our study of implementing quantile matching for homogenizing time series of snow depth: Improving the quality of large snow depths in the time series, as they can have a strong influence on long-term trends. From our results we conclude that this was successful, even if the improvements were not as clear-cut as postulated. The evaluation of InterpQM shows that it achieved by far the best results compared with INTERP on the most difficult dataset, where an entire period was replaced by values from another station. This inhomogeneity was also considered to be the most realistic, since the exact impact of each station relocation is largely unknown and was only simulated in the other scenarios. We tried not to overfit InterpQM to specific inhomogeneities and to simulate different possible impacts as realistically as possible. However, this cannot be avoided completely and a possible reason for the partly very similar results of INTERP and InterpQM could be that the artificial inhomogeneities could also be adjusted very well by the median approach of INTERP.

After the evaluation using the Swiss parallel time series, InterpQM was applied to the Austrian SNOWPAT dataset, where Schöner et al. (2019) had detected and adjusted breaks at six stations. As shown for the example of Galtür, INTERP and InterpQM can lead to considerable differences in the results, even in the sign of the adjustment. While this had no effect on the sign or significance of the trend for Galtür, for other station time series the significance varied with the different methods while the sign of the adjustment was the same (e.g., HSmean in Oed). The evaluation carried out with the Swiss parallel data set and the comparison of the error measures for different InterpQM variants also showed that the use of a higher number of adjustment factors (higher number of IQRs) does not necessarily increase the quality of the results.

### AUTHOR CONTRIBUTIONS








The study concept was defined by Gernot Resch, Wolfgang Schöner and Christoph Marty. Gernot Resch, Roland Koch, Michael Begert, Moritz Buchmann were involved in the data curation. Used scripts were written by Gernot Resch, formal analysis was conducted by Gernot Resch with input for methodology from Wolfgang Schöner, Christoph Marty, Michael Begert, Barbara Chimani and Johannes Aschauer. The visualization was done by Gernot Resch with input from Christoph Marty and Wolfgang Schöner. The original draft was written by Gernot Resch with assistance from Roland Koch. The paper review was performed by Gernot Resch, Roland

Koch, Christoph Marty, Wolfgang Schöner, Barbara Chimani, Moritz Buchmann and Johannes Aschauer. Funding acquisition was performed by Christoph Marty and Wolfgang Schöner.

### ACKNOWLEDGEMENTS

The authors want to thank the Austrian Water Budget Department for the access to their datasets. This research was developed with the financial support of the FWF (FWF Fonds zur Förderung der wissenschaftlichen Forschung, Project Hom4Snow, Grant Number: I 3692) and the University of Graz. For data juggling, quality checks, evaluation, homogenization and different other tasks R 4.1.2 (R Core Team, 2021) on a Linux and Mac-System with the tidyverse package (Wickham et al., 2019) were used. We would also like to thank all snow observers. They made this study possible through their continuous efforts. We would like to thank the two anonymous reviewers for their valuable time on the manuscript. Their ideas and constructive comments helped to improve and polish the paper.

### ORCID

Gernot Resch  <https://orcid.org/0000-0002-4734-1158>  
 Christoph Marty  <https://orcid.org/0000-0002-0398-6253>  
 Barbara Chimani  <https://orcid.org/0000-0003-0298-4181>  
 Michael Begert  <https://orcid.org/0000-0002-9569-6842>  
 Moritz Buchmann  <https://orcid.org/0000-0002-1710-5496>  
 Johannes Aschauer  <https://orcid.org/0000-0002-2605-8003>  
 Wolfgang Schöner  <https://orcid.org/0000-0001-6546-0639>

### REFERENCES

- Alexandersson, H. (1986) A homogeneity test applied to precipitation data. *Journal of Climatology*, 6(6), 661–675. <https://doi.org/10.1002/joc.337060607>.
- Alexandersson, H. and Moberg, A. (1997) Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology*, 17(1), 25–34. [https://doi.org/10.1002/\(SICI\)1097-0088\(199701\)17:1%3C25::AID-JOC103%3E3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(199701)17:1%3C25::AID-JOC103%3E3.0.CO;2-J).
- Armstrong, R., Steffen, K., Monteduro, M., & Sessa, R. (2009). Assessment of the status of the development of the standards for the terrestrial essential climate variables: Snow. In: *Global terrestrial observing system*, Vol. 10. <http://www.fao.org/gtos/doc/ecvs/t12/t12.pdf>. Accessed 23 September 2021.
- Armstrong, R.L. and Brun, E. (2008) *Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling*. Cambridge, UK: Cambridge University Press.
- Aschauer, J., & Marty, C. (2020). Providing data provision for a sensitivity analysis of snow time series. <https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/>

- Internationale-Zusammenarbeit/GCOS/doc/Final\_report\_Poviding\_Data\_Provision\_for\_a\_Sensitivity\_Analysis\_of\_Snow\_Time\_Series.pdf. Accessed 23 September 2021.
- Aschauer, J. and Marty, C. (2021) Evaluating methods for reconstructing large gaps in historic snow depth time series. *Geoscientific Instrumentation, Methods and Data Systems*, 10(2), 297–312. <https://doi.org/10.5194/gi-10-297-2021>.
- 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., et al. (2007) HISTALP: historical instrumental climatological surface time series of the greater alpine region. *International Journal of Climatology*, 27(1), 17–46. <https://doi.org/10.1002/joc.1377>.
- Ballotta, L., Fusai, G., Kyriakou, I., Papapostolou, N.C. and Pouliasis, P. K. (2020) Risk management of climate impact for tourism operators: an empirical analysis on ski resorts. *Tourism Management*, 77, 104011. <https://doi.org/10.1016/j.tourman.2019.104011>.
- Buchmann, M., Begert, M., Brönnimann, S. and Marty, C. (2021) Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series. *International Journal of Climatology*, 41(S1), E2553–E2563. <https://doi.org/10.1002/joc.6863>.
- Buchmann, M., Coll, J., Aschauer, J., Begert, M., Brönnimann, S., Chimani, B., Resch, G., Schöner, W. and Marty, C. (2022) Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-) automatic homogenisation methods. *The Cryosphere Discuss.* <https://doi.org/10.5194/tc-2022-48>.
- Caussinus, H. and Mestre, O. (2004) Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(3), 405–425. <https://doi.org/10.1111/j.1467-9876.2004.05155.x>.
- Chimani, B., Venema, V., Lexer, A., Andre, K., Auer, I. and Nemeč, J. (2018) Inter-comparison of methods to homogenize daily relative humidity. *International Journal of Climatology*, 38(7), 3106–3122. <https://doi.org/10.1002/joc.5488>.
- Domonkos, P. (2015) Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology*, 122(1–2), 303–314. <https://doi.org/10.1007/s00704-014-1298-5>.
- Domonkos, P., Venema, V., Mestre, O., & France, M. (2012). Efficiencies of homogenization methods: our present knowledge and its limitation. Proceedings of the 7th Seminar and Quality Control in Climatological Databases, 19–32.
- Goodison, B., Louie, P.Y.T. and Yang, D. (1998) WMO solid precipitation measurement intercomparison. *Final Report*, 67, 318.
- Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. and Rosas, G. (2017) The influence of station density on climate data homogenization. *International Journal of Climatology*, 37(13), 4670–4683. <https://doi.org/10.1002/joc.5114>.
- Gultepe, I. (2015) In: Dmowska, R. (Ed.) *Chapter Three - Mountain weather: Observation and modeling*, Vol. 56. Elsevier, pp. 229–312. <https://doi.org/10.1016/bs.agph.2015.01.001>.
- Haberkorn, A., Helmert, J., Leppänen, L., López-Moreno, J.I. and Pirazzini, R. (2019) *European Snow Booklet - an Inventory of Snow Measurements in Europe*. <https://doi.org/10.16904/envidat.59>.
- Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. (2019). High mountain areas. In: IPCC special report on the ocean and cryosphere in a changing climate. <https://www.ipcc.ch/srocc/>
- Johnston, A.N., Bruggeman, J.E., Beers, A.T., Beever, E.A., Christophersen, R.G. and Ransom, J.I. (2019) Ecological consequences of anomalies in atmospheric moisture and snowpack. *Ecology*, 100(4), 1–12. <https://doi.org/10.1002/ecy.2638>.
- Kendall, M. (1975) In: Griffin, C. (Ed.) *Rank correlation methods*, 4th edition. San Francisco, CA: Charles Griffin.
- Kim, S. and Kim, H. (2016) A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679. <https://doi.org/10.1016/j.ijforecast.2015.12.003>.
- Lamprecht, A., Semenchuk, P.R., Steinbauer, K., Winkler, M. and Pauli, H. (2018) Climate change leads to accelerated transformation of high-elevation vegetation in the Central Alps. *New Phytologist*, 220(2), 447–459. <https://doi.org/10.1111/nph.15290>.
- Lehning, M., Löwe, H., Ryser, M. and Raderschall, N. (2008) Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resources Research*, 44(7), 1–19. <https://doi.org/10.1029/2007WR006545>.
- Lettenmaier, D.P., Wood, E.F. and Wallis, J.R. (1994) Hydroclimatological trends in the continental United States, 1948–88. *Journal of Climate*, 7, 586–607. [https://doi.org/10.1175/1520-0442\(1994\)007%3C0586:HCTITC%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007%3C0586:HCTITC%3E2.0.CO;2).
- Mann, H.B. (1945) Nonparametric tests against trend. *Econometrica*, 13(3), 245–259.
- Marcolini, G., Bellin, A. and Chiogna, G. (2017) Performance of the standard Normal homogeneity test for the homogenization of mean seasonal snow depth time series. *International Journal of Climatology*, 37, 1267–1277. <https://doi.org/10.1002/joc.4977>.
- Marcolini, G., Koch, R., Chimani, B., Schöner, W., Bellin, A., Disse, M. and Chiogna, G. (2019) Evaluation of homogenization methods for seasonal snow depth data in the Austrian Alps, 1930–2010. *International Journal of Climatology*, 39(11), 4514–4530. <https://doi.org/10.1002/joc.6095>.
- Marke, T., Strasser, U., Hanzer, F., Stötter, J., Wilcke, R.A.I. and Gobiet, A. (2015) Scenarios of future snow conditions in Styria (Austrian Alps). *Journal of Hydrometeorology*, 16(1), 261–277. <https://doi.org/10.1175/JHM-D-14-0035.1>.
- 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., et al. (2021) Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere*, 15(3), 1343–1382. <https://doi.org/10.5194/tc-15-1343-2021>.
- Mestre, O., Gruber, C., Prieur, C., Caussinus, H. and Jourdain, S. (2011). SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50(11), 2343–2358. <https://doi.org/10.1175/2011jamc2641.1>.
- Moberg, A. and Alexandersson, H. (1997) Homogenization of Swedish temperature data. Part II: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology*, 17(1), 35–54. [https://doi.org/10.1002/\(sici\)1097-0088\(199701\)17:1<35::aid-joc104>3.3.co;2-6](https://doi.org/10.1002/(sici)1097-0088(199701)17:1<35::aid-joc104>3.3.co;2-6).
- Nalder, I.A. and Wein, R.W. (1998) Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest.



- Agricultural and Forest Meteorology*, 92(4), 211–225. [https://doi.org/10.1016/S0168-1923\(98\)00102-6](https://doi.org/10.1016/S0168-1923(98)00102-6).
- Nemec, J., Gruber, C., Chimani, B. and Auer, I. (2013) Trends in extreme temperature indices in Austria based on a new homogenized dataset. *International Journal of Climatology*, 33(6), 1538–1550. <https://doi.org/10.1002/joc.3532>.
- Olefs, M., & Koch, R. (2020). SNOWPAT: long-term homogenized snow observations in Austria. [https://data.ccca.ac.at/dataset/snowpad\\_2019-v01/resource/36f3a5ac-fde2-492f-8248-adf79b26ee8e](https://data.ccca.ac.at/dataset/snowpad_2019-v01/resource/36f3a5ac-fde2-492f-8248-adf79b26ee8e). Accessed 24 September 2021.
- Olefs, M., Koch, R., Schöner, W. and Marke, T. (2020) Changes in snow depth, snow cover duration, and potential snowmaking conditions in Austria, 1961–2020: a model based approach. *Atmosphere*, 11(12), 1330. <https://doi.org/10.3390/atmos11121330>.
- Prein, A.F. and Gobiet, A. (2017) Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, 37(1), 305–327. <https://doi.org/10.1002/joc.4706>.
- Pröbstl-Haider, U., Lund-Durlacher, D., Olefs, M. and Pretenthaler, F. (2021) Tourismus und Klimawandel. In: Pröbstl-Haider, U., Lund-Durlacher, D., Olefs, M. and Pretenthaler, F. (Eds.) *Tourismus und Klimawandel*. Berlin Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-61522-5>.
- Pulliaainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T. and Norberg, J. (2020) Patterns and trends of northern hemisphere snow mass from 1980 to 2018. *Nature*, 581(7808), 294–298. <https://doi.org/10.1038/s41586-020-2258-0>.
- R Core Team. (2021) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Scherrer, S.C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R. and Appenzeller, C. (2013) Snow variability in the Swiss Alps 1864–2009. *International Journal of Climatology*, 33(15), 3162–3173. <https://doi.org/10.1002/joc.3653>.
- Schöner, W., Koch, R., Matulla, C., Marty, C. and Tilg, A.-M. (2019) Spatiotemporal patterns of snow depth within the Swiss-Austrian Alps for the past half century (1961 to 2012) and linkages to climate change. *International Journal of Climatology*, 39(3), 1589–1603. <https://doi.org/10.1002/joc.5902>.
- Sen, P.K. (1968) Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. <https://doi.org/10.2307/2285891>.
- Sevruk, B. (1985) Correction of precipitation measurements: Swiss experience. *Zürcher Geogr. Schriften*, 23, 187–196.
- Statistik\_Austria. (2019). Tourismus in Österreich 2018: Ergebnisse der Beherbergungsstatistik. [https://www.statistik.at/web\\_de/services/publikationen/13/index.html?includePage=detailedView&ion-Name=Tourismus&pubId=569](https://www.statistik.at/web_de/services/publikationen/13/index.html?includePage=detailedView&ion-Name=Tourismus&pubId=569)
- Steinbauer, M.J., Grytnes, J.-A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., Rixen, C., Winkler, M., Barni, M.B.-D.E., Bjorkman, A.D., Breiner, F.T., Burg, S., Czortek, P., Dawes, M. A., Delimat, A., Dullinger, S., Erschbamer, B., Felde, V.A., Fernández-Arberas, O., et al. (2018) Mountain summits is linked to warming. *Nature*, 556, 231–234.
- Szentimrey, T. (1999). Multiple Analysis of Series for Homogenization (MASH). Proceedings of the second seminar for homogenization of surface climatological data, Budapest, Hungary. WMO, 41, 27–46.
- Toreti, A., Kuglitsch, F.G., Xoplaki, E., Della-Marta, P.M., Aguilar, E., Prohom, M. and Luterbacher, J. (2011) A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *International Journal of Climatology*, 31(4), 630–632. <https://doi.org/10.1002/joc.2088>.
- Vaughan, D., Comiso, J., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Konrad, S. and Zhang, T. (2013) Observations: Cryosphere. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press. [https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5\\_Chapter04\\_FINAL.pdf](https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter04_FINAL.pdf).
- Venema, V., Trewin, B., Wang X.L., Tamás, S., Lakatos, M., Aguilar, E., Auer, I., Guijarro, J. A., Menne, M., Oria, C., Louamba, W. S. R. L., Argiriou, A., Hechler, P., Vertačnik, G., & Yosef, Y. (2020). Guidelines on Homogenization 2020 edition (Issue WMO-No. 1245). [https://library.wmo.int/?lvl=notice\\_display&id=21756](https://library.wmo.int/?lvl=notice_display&id=21756)
- Venema, V.K.C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnick, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., et al. (2012) Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1), 89–115. <https://doi.org/10.5194/cp-8-89-2012>.
- Vincent, L.A., Zhang, X., Bonsal, B.R. and Hogg, W.D. (2002) Homogenization of daily temperatures over Canada. *Journal of Climate*, 15(11), 1322–1334. [https://doi.org/10.1175/1520-0442\(2002\)015<1322:HODTOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2).
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin, T., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., et al. (2019) Welcome to the Tidyverse. *The Journal of Open Source Software*, 4(43), 1–6. <https://doi.org/10.21105/joss.01686>.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Resch, G., Koch, R., Marty, C., Chimani, B., Begert, M., Buchmann, M., Aschauer, J., & Schöner, W. (2022). A quantile-based approach to improve homogenization of snow depth time series. *International Journal of Climatology*, 1–17. <https://doi.org/10.1002/joc.7742>