~~Homogenizing Swiss~~ The benefits of homogenising snow depth series - ~~Impact~~ Impacts on decadal trends and extremes for Switzerland

Moritz Buchmann[1,2,*], Gernot Resch[3,*], Michael Begert[4], Stefan Brönnimann[2,5], Barbara Chimani[6], Wolfgang Schöner[3], and Christoph Marty[1]

[1]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[4]Federal Office of Meteorology and Climatology (MeteoSwiss), Zurich Airport, Switzerland
[5]Institute of Geography, University of Bern, Bern, Switzerland
[3]Department of Geography and Regional Science, University of Graz, Graz, Austria
[6]Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
[*]These authors contributed equally to this work.

Correspondence: Gernot Resch (gernot.resch@uni-graz.at)

Abstract. Our current knowledge ~~on~~ of spatial and temporal snow depth trends is based almost exclusively on ~~these non-homogenized data. Long-term observations of deposited snow are well suited as indicator of climate change.~~ time-series of non-homogenised observational data. However, like ~~all~~ other long-term series from observations, they are prone to inhomogeneities that can influence and even change trends if not taken into account. ~~We~~ In order to assess the relevance of homogenisation for time-series analysis of daily snow depths, we investigated the effects of ~~removing~~ adjusting inhomogeneities in the ~~large~~ extensive network of Swiss snow depth observations ~~on trends and~~ for trends and changes in extreme values of commonly used snow indices, such as snow days, seasonal averages or maximum snow depth in the period 1961-2021. ~~For this task, three homogenization methods were applied~~Three homogenisation methods were compared for this task: Climatol and HOMER, which ~~use a median based adjustment method, and interpQM, which applies quantile based adjustments~~apply median based adjustments, and the quantile based interpQM. All three were run using the same ~~break points and input data . This allowed us to investigate and quantify the effects of these methods on the homogenization results.~~ input data with identical breakpoints. We found that ~~all three methods~~ they agree well on trends ~~in~~ of seasonal average snow depth, while differences are ~~visible for seasonal maximum snow depth~~ detectable for seasonal maximums and the corresponding extreme values. ~~Here, the~~ Differences between homogenised and non-homogenised series result mainly from the approach for generating reference series. The comparison of homogenised and original values for the 50-year return level of seasonal maximum snow depth showed that the quantile-based method ~~performed slightly better than the two median-based methods, as it~~ had the smallest number of stations outside the 95 % confidence interval~~for 50-year return periods of maximum snow depth. These differences are mainly caused by the way the reference series are selected. The combination of a high minimum~~ . Using a multiple criteria approach as e.g. thresholds for series correlation (> 0.7) ~~and restrictions in~~ as well as for vertical (< 300 m) and horizontal (< 100 km) distances proved to be better suited than ~~only using correlations or distances respectively as criteria. The adjustments removed~~ using correlation or distances alone. Overall, the homogenisation of snow depth series changed all positive trends for ~~snow days in the original data and strengthened~~ derived series of snow days to either no trend or negative trends and amplifying the negative mean trend, especially for stations > 1500

m. ~~In addition, the number of significant negative stations was~~ The number of stations with a significant negative trend increased between 7 - 21 % depending on the method, with the strongest changes ~~at higher~~ occurring at high snow depths. The reduction in the 95 % confidence intervals of the absolute maximum snow depth of each station indicates a decrease in variation and an increase in confidence in the results.

~~Climatol~~Homogenisation, ~~HOMER~~Alps, ~~interpQM, trends,~~ extreme values, ~~Switzerland~~manual measurements, ~~Alps~~snow, snow depth, ~~manual measurements~~Switzerland, trends

1 Introduction

During winter in the northern hemisphere, more than 50 % of the earth's surface can be covered with snow (Armstrong and Brun, 2008) . The thickness and duration of a snow cover play an important role for many animal and plant species (Johnston et al., 2019), but also have an important socio-economic dimension: For example, the timing and amount of snow melt are important for hydropower companies, the number of days with a certain minimum snow depth (e.g. 30 cm) is a metric widely used for the profitability of ski resorts (Abegg et al., 2020). Accurate and reliable measurements of solid precipitation, e.g. total height of fallen snow (snow depth) or the amount of freshly fallen snow (~~new snow height~~depth of snowfall), are difficult to obtain but important (Nitu et al., 2018) ~~. They~~ as they are used for various purposes, e.g. as ground evidence for large-scale grid-based forecasts of snow depth (Olefs et al., 2013) or the operational assessment of snow models used ~~in the operational chain of~~ for avalanche hazard forecasting (Morin et al., 2020). Long-term measurements are key to climate monitoring and related analyses. They are not only used for climatological analyses ~~(e.g. Matiu et al., 2021; Pulliainen et al., 2020)~~ (Matiu et al., 2021; Pulliainen et al., 2020), but also for creating ~~and bias-correction gridded datasets (e.g. Cornes et al., 2018; Hersbach et a~~ bias-corrected models (Maraun et al., 2017) and gridded datasets (Cornes et al., 2018; Hiebl and Frei, 2018; Hersbach et al., 2020; Li et al.
.

~~The longer a~~ All climate time series comprise a climate signal, a station signal and white noise (Caussinus and Mestre, 2004) . The station signal includes the characteristics of the environment, observers and instruments of each station. If the station signal changes over time, it can alter the climate signal, e.g. by amplifying or weakening trends. It should therefore be adjusted before further analysis are done. According to the approach of relative homogeneity testing and adjusting this is possible as long as neighbouring stations follow an identical climatic signal (variability and trend). The longer the time series, the ~~more likely for itto contain inhomogeneities due to station relocation,~~ higher the probability of large changes causing breaks/breakpoints in it. There are many reasons for this, such as changes in instruments, observers, the station environment, or a combination of the above factors ~~(e.g. Auer et al., 2007). These inhomogeneities can significantly influence~~ (Auer et al., 2007; Venema et al., 2020). Alexandersson and Moberg (1997) even found that multi-decadal time series without breaks are rare. Breaks can significantly alter derived trends (Begert et al., 2005; Gubler et al., 2017; Resch et al., 2022) and extreme values (Kuglitsch et al., 2009). ~~To~~ Therefore, to address this problem, ~~time series are usually homogenized in two steps~~climate time series should be homogenised, which does not always happen or is possible, usually by a two-step procedure: First, immanent ~~break points or inhomogeneities are identifiedand then adjusted in the subsequent step. Homogeneity tests that are used for the identification of break points mostly assess~~ breakpoints are identified. Relative homogeneity tests used for breakpoint identification mostly assess significant changes of ratios or differences between the station to be ~~homogenized~~ homogenised (candidate series) and ~~highly correlated~~ neighbouring stations (reference series)~~that are selected using~~. Reference series are selected on the basis of several criteria, mostly ~~the~~ correlation and horizontal ~~distances. In the~~ as well as vertical distance. In a second step, the ~~reference series are used to calculate adjustments for the period before a break, which change the observations as if the current station environment had also existed in the past and the measurement techniquehad also been used~~

~~in the past.This~~ candidate series is homogenised to the present state, thus compensating for previous non-climatic deviations, e.g. changes in observation procedures, sensors or measurement technique, in the time series.

Today, this is a standard ~~process~~ procedure for climate data like temperature and precipitation (Venema et al., 2020), but has only recently been adopted for snow depth time series~~(Schöner et al., 2019; Resch et al., 2022).~~ : First steps towards detecting and adjusting breaks were made by Marcolini et al. (2017). Schöner et al. (2019) used the HOMOP tool (Nemec et al., 2013) to homogenise seasonal depth time series and for calculating trends and identifying snow regions in Austria and Switzerland. Marcolini et al. (2019) compared the performance of two homogenisation methods (HOMOP and SNHT (Alexandersson, 1986; Alexanders) and their effects on trends in seasonal mean snow depth. Their results showed the need for improving adjustment methods in order to (i) enable the application to data with higher temporal resolution (e.g. daily data) and (ii) to improve the adjustment of extreme values. Taking up these needs, an adjustment method using quantile matching was introduced by Resch et al. (2022).

Widespread used metrics to describe the snow ~~pack include snow days and maximum and average snow depths .~~ cover include average and maximum snow depths and days with a snow depth above a certain threshold, referred to here as snow days. This commonly used index is defined as the number of days within a certain time period (e.g. season) with a certain snow depth, usually between 1 - 50 cm. (Abegg et al., 2020; Schmucki et al., 2017). Snow days are relevant for ecology ~~(e.g. Stone et al., 2002; Jonas et al., 2008) and climatology (e.g. Scherrer et al., 2004; Marty, 2008)~~(Stone et al., 2002; Jonas et al., 2008) , climatology (Scherrer et al., 2004; Marty, 2008) or the ski tourism industry (Abegg et al., 2020), whereas the average and maximum snow depths are particularly applicable to climatology and engineering applications. ~~Trends~~ Trend and extreme value analyses of snow ~~variables (e.g. Scherrer et al., 2013; Matiu et al., 2021) are commonly used~~ indices (Scherrer et al., 2013; Matiu et al., 20 are common methods in climate monitoring ~~(Bocchiola et al., 2008; Marty and Blanchet, 2012; Buchmann et al., 2021a), model verification (e.g. Brown et al., 2003; Essery et al., 2013), whereas extreme value analysis is important for the definition of~~ (Bocchiola et al., 2008; Marty and Blanchet, 2012; Buchmann et al., 2021a) and model verification (Brown et al., 2003; Essery et al., 2013 , while extreme value analyses are important for defining snow loads and limits for building-codes ~~(e.g. Croce et al., 2021; Schellander et al.~~ (Croce et al., 2021; Schellander et al., 2021; Al-Rubaye et al., 2022).

~~In spite of recent efforts of studying the homogenization of snow depth (Marcolini et al., 2019; Schöner et al., 2019; Buchmann et al., 20 , it is still an open question whether the use of one of the methods is more advantageous than the others, and how the methods impact trends and extreme values. So far, Swiss snow depth time series have not been homogenized and an impact assessment has not been carried out neither for Switzerland or any other region.~~

We ~~are using the set of break points,~~ use the breakpoints recently identified by Buchmann et al. (2022) for manual Swiss snow series with a joint ~~use of three state-of-the-art~~ application of three widely used breakpoint-detection and ~~homogenization~~ homogenisation methods: ACMANT (Domonkos, 2011), Climatol (Guijarro, 2018) and HOMER ~~(Caussinus and Mestre, 2004; Domonkos . This break point set is then used as input for three homogenization~~ (Mestre et al., 2013) for three homogenisation methods to calculate and apply adjustments: Climatol, HOMER, and interpQM (Resch et al., 2022). ACMANT, which is fully automatic, does not allow manual breakpoint input and was therefore not used for our analyses. The first two methods apply median based adjustments, the latter uses a ~~quantile based~~ quantile-based approach. All three methods are then applied to the network of Swiss snow depth time series. This ~~enables~~ allows us to assess the impact of ~~homogenization~~ homogenisation (dependent on the method used) on ~~trends of~~ the trends in seasonal mean and maximum snow depths, ~~and~~ days with snow on the ground ~~as well as~~ and extreme values of maximum snow depths. ~~To focus our study, our~~ Our research questions are the following:

1. How do the ~~homogenized~~ homogenised series compare across the three methods ~~applied~~used?

2. What ~~is the impact of homogenization on decadal trends of~~ influence does homogenisation have on the decadal trends in average and maximum snow depth?

3. How do the three ~~homogenization methods influence widely used~~ homogenisation methods affect widely used snow indices?

4. To what extent are the maximum snow depths with a 50-year return period (as an example for snow metrics used by practitioners) affected by the ~~various homogenization~~ different homogenisation methods?

The article is structured as follows: Section 2 describes the data and Section 3 introduces the various methods used. Results are shown in Section 4 and discussed in Section 5. Conclusions are drawn in Section 6.

Table 1. Summary of autocorrelations for lag (year) 1 for all stations (n = 40)

	Minimum	**Maximum**	**IQR**	**Mean**	**Median**
HSmean	-0.34	0.36	0.22	0.1	0.1
HSmax	-0.25	0.42	0.24	0.03	0
dHS5	-0.1	0.48	0.14	0.18	0.21
dHS30	-0.29	0.36	0.15	0.06	0.06
dHS50	-0.3	0.26	0.16	0.03	0.06

2 Data

Daily manual snow depth measurements (HS) from 184 Swiss stations ~~are used in this study~~serve as the basis for quantifying the benefit of data homogenisation for snow depth series. Seasonal (November to April) and monthly average (HSavg), maximum snow depths (HSmax), and the number of snow days >= 5 cm (dHS5) are calculated from the daily snow ~~depth values~~ depths measured at 07:00 o'clock each day. For obvious reasions, daily snow depth time series inherit a strong autocorrelation. We used seasonal indicators of snow depth, which imply no to low autocorrelation with exception of cases when the snowcover did not completely melt over the summer. However, this is neither the case for any of the selected stations nor for any of the seasons analysed. This is shown in Table 1 for lag 1 - 10 years autocorrelation, with strongest correlation for lag 1. The results showed that autocorrelation is very low (mean 0.03 - 0.18, interquartile range 0.14 - 0.24). Consequently, a Trend-Free-prewhitening of snow depth series (Yue et al., 2002) or the application of a modified MK-test was not necessary. Figure 1 shows the station distribution ~~. The stations , maintained by either~~ of the 184 Swiss stations used. They are maintained either by the Federal Office of Meteorology and Climatology (MeteoSwiss) or the WSL Institute for Snow and Avalanche Research (SLF)~~provide data within~~ , covering the period from 1931 to 2021 and span from 200 to 2500 m a.s.l. (shown in the right panel of Figure 1). Only stations with complete data coverage between November and April for each year and at least 30 years of data are considered. ~~For a~~ A detailed description of the ~~data set, please see~~ dataset can be found in Buchmann et al. (2022).

2.1 ~~Break points~~The set of pre-identified breakpoints

We use the set of 45 ~~break points~~ breakpoints (found in 40 ~~series), detected with a joint use of~~ snow depth series) identified by Buchmann et al. (2022) for our analyses. Two series (stations Bernina Hospiz and Gütsch) were removed from the original 42-station-subset due to insufficient data quality between 1961 and 2021. Breakpoints were detected using ACMANT (Domonkos, 2011), Climatol (Guijarro, 2018), and HOMER ~~(Caussinus and Mestre, 2004; Domonkos, 2011; Picard et al., 2011; Guijarro, 2018), where break points are~~ (Mestre et al., 2013), with breakpoints accepted as valid if ~~they are at least identified by 2-out-of-3 methods~~ detected within two years by at least 2 of 3 methods. For details, e.g. on the differences between methods in the detected breaks and motivation for the criteria of break acceptance, see Buchmann et al. (2022). ~~Break points are detected~~

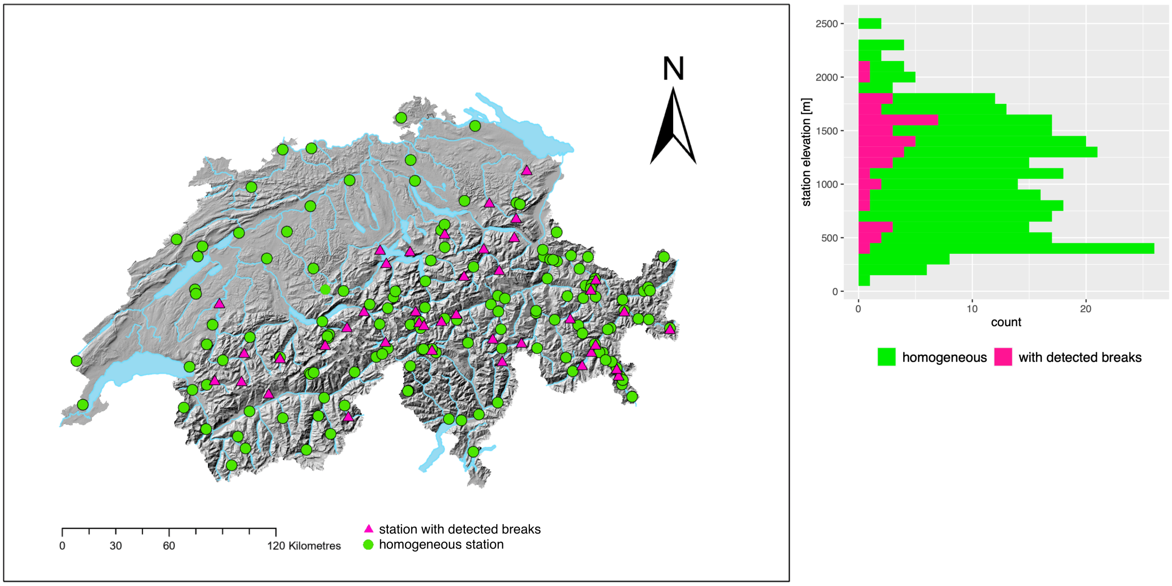

Figure 1. Left panel: Map of Switzerland ~~showing~~ with all 184 Swiss stations used in this study. The ~~42~~ 40 identified inhomogeneous stations ~~(identified in this study)~~ with valid ~~break points~~ breakpoints are highlighted in pink triangles. ~~Green~~ The green circles are series that are considered homogeneous. Right panel: Elevation distribution of the homogeneous stations and those with detected breaks.

~~using~~ Breakpoints are identified based on seasonal series. ~~Available metadata was used to further verify break points where applicable~~Where appropriate, available metadata has been used for verification. However, as our metadata is neither perfect nor complete, it is only used as an additional ~~clue~~ source of information and not as stand-alone evidence. ~~The analysis described in Buchmann et al. (2022) is re-run with the three additional Austrian data series along the eastern Swiss border to increase the number of available reference stations for Eastern Swiss series, thus increasing the valid break points to 45 (from the original 43 mentioned in Buchmann et al. (2022))~~To improve the station density near the Swiss eastern border, three Austrian stations were added to the data base. Figure 1 shows the location of all 184 ~~series, the inhomogeneous ones are coloured pink~~ Swiss series in the left panel, stations with detected breaks are marked with pink triangles. The right panel shows the elevation distribution of the stations. ~~Two series (stations Bernina Hospiz and Gütsch) with insufficient data quality between 1961 and 2021 were removed from the subset of 42 series reported in Buchmann et al. (2022).~~

3 Methods

Each breakpoint of a candidate series is adjusted by a multiplicative approach to the most recent status of the snow station. This is in agreement for all three adjustment methods applied. Adjustment factors are based on statistical measures of the candidate and reference series, respectively, and applied to the monthly (Climatol, HOMER) or daily (interpQM) values. These statistical measures (e.g. median, quantiles) applied for adjustment are different for the three methods and are described below. Important to know, the reference series used for adjustment by the three methods are not identical and selected on different criteria. For interpQM and HOMER, they are known to the user.

All methods compared use the same dataset to select suitable reference stations for the calculation of the adjustment factors based on the pre-determined breakpoints, which in our case are provided by Buchmann et al. (2022) and used by each method via importing a file containing the breakpoints. Although it is possible to manually select suitable reference stations for each series and use only these for each method, we have chosen to let the methods themselves select their reference stations based on their internal criteria.

3.1 ~~Homogenization~~ Adjustment methods

Climatol (Guijarro, 2018) is a fully-automatic ~~homogenization function~~ homogenisation method based on SNHT (Alexandersson, 1986) ~~and uses a step-wise detection of homogeneous sub-periods.~~ for break detection and a linear regression approach after Easterling and Peterson (1995) for the adjustments. It uses composite reference series that are constructed as a weighted ~~mean~~average, using the horizontal and vertical distance between suitable reference ~~series~~ and the candidate series as weight. ~~It allows the use of a manually adjusted break file, which in our case is provided by Buchmann et al. (2022). Candidate~~ We used the default settings, i.e. 100 km, where the horizontal distance weight is set to 0.5 and the vertical distance scaling to 0.1. As for all adjustment methods, candidate series are adjusted ~~stepwise backwards~~ back in time starting from the most recent homogeneous sub-period~~, by adjusting every~~ . Doing so, each detected break (sub-period between breaks) ~~with an annual adjustment factor (see (Guijarro, 2018)~~ is adjusted applying an adjustment factor derived from annual values (see Guijarro (2018) for details). Which is calculated for Climatol as follows:

~~HOMER (Caussinus and Mestre, 2004; Domonkos, 2011; Picard et al., 2011; Guijarro, 2018) is a toolbox which provides several methods for break detection and adjustments, like~~ The adjustment factor of a time series z is calculated as follows:

$$Climatol = \frac{\sigma Q \bar{z}_b + \bar{Q}}{\sigma Q \bar{z}_a + \bar{Q}} \quad (1)$$

Where $\bar{z}_b$ and $\bar{z}_a$ are the mean snow depth between the beginning of the measurements of z and the breakpoint (before) and from the breakpoint to the end (after), respectively. σ_Q and $\bar{Q}$ are the standard deviation and mean of the non-standardized ratio time series $Q = Reference/Candidate$ (Alexandersson and Moberg, 1997).

HOMER (Mestre et al., 2013) is an interactive semi-automatic toolbox that provides various methods for detection and adjustment of breaks, such as pairwise comparison (Caussinus and Mestre, 2004), a fully automatic detection and correction joint-segmentation (Picard et al., 2011) and ACMANT-detection (Domonkos, 2011). ~~Like Climatol, HOMER can be run with an independently derived break file and also uses a manual added metadata-file~~For our purposes, the pairwise comparison was chosen, as it accepts the use of independently derived breakpoint metadata-files, like Climatol and interpQM. ~~In HOMER,~~ The series are adjusted with a single annual factor for the entire period before a ~~break point~~breakpoint. The adjustment factor is ~~determined by means of~~ derived from variance analysis ANOVA ~~(Caussinus and Mestre, 2004; Mestre et al., 2013)~~ (Caussinus and Mestre, 2004) based on the ~~reference sub-networks, which in our case have a minimum correlation of~~ selected reference stations. These are defined either on the basis of the horizontal distance or the first-difference correlations. Due to the large vertical distances between stations, even for short horizontal distances, the latter was chosen with a minimum Spearman ρ of $>$ 0.8~~with the candidate~~.

$$HOMER = O_{ij} - \hat{v}^{C*}_{jh(i,j)} + \hat{v}^{C*}_{j,k_j+1} \tag{2}$$

Where O_{ij} is the matrix of original time series j with time index i, $\hat{v}^{C*}_{jh(i,j)}$ the estimation of the correction for a set of breaks per candidate station C_j in a homogeneous subperiod $h_{i,j}$ and $\hat{v}^{C*}_{j,k_j+1}$ the estimation of the adjusted station signal with the number of breakpoints of a a station k_j.

InterpQM (Resch et al., 2022) is an extension of INTERP (Vincent et al., 2002), ~~using~~ that uses quantile matching to improve the adjustments, ~~considering~~ taking into account the frequency distribution of ~~adjustment factors. It yields daily homogenized data~~ the daily values to be adjusted. It provides homogenised data on a daily basis, which then allows the analysis of the subsequently derived snow indicators. For this purpose, the daily measurements of the candidate and reference series are split into ~~several interquantile ranges (IQR~~two interquantile subsets (IQS), which are then compared. An adjustment factor

$$interpQM = \frac{\left(\frac{\tilde{C}_a}{\tilde{R}_a}\right)}{\left(\frac{\tilde{C}_b)}{\tilde{R}_b)}\right)} \tag{3}$$

is calculated for each ~~IQR~~ IQS and then linearly interpolated between neighbouring ~~IQR's~~ IQSs to avoid artificial jumps ~~.~~ ~~Reference series~~ in the data. $\tilde{C}$ and $\tilde{R}$ are the median of the daily time series of the candidate/reference station before (b) or after (a) the detected breakpoint to be adjusted. The reference series can either be selected manually or ~~calculated using a weighted mean of a network of~~ be a composite series calculated from a weighted average of selected stations (< 100 km horizontal and < 300 m vertical distance, > 0.7 correlation, no detected breaks), which was chosen here. The selection can be manually refined and ~~optimized by making use of~~ optimised, using local knowledge and experience. The distribution of ~~the weighting between the stations of a network~~ weights between these stations can either be exponential or linear. ~~For reducing the strong impact of single~~ To reduce the strong influence of individual highly correlated stations on the result, a linear distribution of the weighting was chosen. ~~Identified breaks from a break filewere used~~Breakpoints are derived from a pre-defined breakpoint file.

3.2 Detection of trends and changes in snow depth series

The use of ~~homogenization~~ homogenisation techniques that adjust daily values allows the analysis of the impact on derived ~~parameters (indicators) which~~ indicators that require daily data for their calculation, e.g. snow days. ~~These common used indicators are defined as the number of days within a time period with a snow depth greater or equal to a specified threshold, typically between 1 - 50 cm. (e.g. Abegg et al., 2020; Schmucki et al., 2017). Since Climatol and HOMER calculate either seasonal or monthly parameter values, only interpQM could be used for their analysis.Since it was decided that interpQM does not change the total number of snow days~~ Only the original data and interpQM are compared here, as HOMER only provides monthly or seasonal data and Climatol kept crashing when using the full daily dataset. Since we did not want to pre-select stations as this would influence the results, we decided not to use it for this purpose.

InterpQM does not add new days with snow ($HS > 0$)~~, the thresholds~~ . To avoid a possible negative bias and because almost no changes were expected for the homogenised series of days with HS > 1 cm, the threshold values of 5, 30 and 50 cm~~(dHS5, dHS30~~, ~~dHS50) were used and summed up for each hydrological year~~ are clearly more meaningful. Snow days are accumulated based on a temporal reference between November and April ~~.~~

~~Trends are detected using the~~ each year (hydrological year). Trends are determined using the standard non-parametric Mann-Kendall trend test ~~(Mann, 1945; Kendall, 1975). As we are working with seasonal (November-April) values, the large auto-correlation of daily snow depth does not apply here. Trends are treated significant if their corresponding p-values are less than 0.05 and not significant if greater or equal to 0.05 (5 % assumption)~~(Mann, 1945; Kendall, 1975) and are considered significant if they are above the 95 % level.

Theil-Sen slopes (Theil, 1950; Sen, 1968) are used to estimate the strength of the trends. ~~Decadal~~ The decadal trends are expressed as change in [cm/decade] ~~. To compare the homogenized~~ or [days/decade]. For the comparison of the homogenised subsets of 40 stations, we focus on the period from 1961 to 2021~~since~~ , as most stations have data ~~within~~ for this period. ~~Trends~~ The trends for all available decades are provided in the supplement (Figure A2).

To ~~study the impact of homogenization~~ investigate the effects of homogenisation on extreme snow depths, return levels for the seasonal maximum snow depth (HSmax) (Marty and Blanchet, 2012) are calculated for a fixed return period of 50 years (Buchmann et al., 2021a; Marty and Blanchet, 2012) based on original and ~~homogenized~~ homogenised data (R50HSmax). ~~Calculations are done~~ This approach was chosen because the international standards for maximum snow load on buildings are based on R50HSmax (see e.g. Schellander et al. (2021)). The calculations were performed with the R package extRemes (Gilleland and Katz, 2016) ~~run in standard~~ in default settings (GEV, estimation method MLE, and 95 % confidence intervals). ~~To assess the significance of the homogenization performed, a~~ In order to determine to what extent homogenised and original time series differ in their distribution and to assess the differences between the results of the applied adjustment methods, a two-sample Kolmogorov-Smirnov ~~test was conducted~~ (in the following referred to KS-test) and the non-parametric two-sample Wilcoxon test (in the following referred to as W-test) were performed with seasonal data for all derived indices.

3.3 Intercomparison experiment of ~~homogenization~~ adjustment methods

We use the subset of 40 stations with identified breaks as input and ~~homogenize it. Both~~ adjust them with the three methods. While Climatol and HOMER use monthly values as input and ~~are therefore run with monthly HSavg as well as HSmax values. InterpQM~~ thus only provide monthly HSavg and HSmax values, interpQM works with daily snow depth values~~, from which~~. From these the analysed seasonal HSavg and HSmax are then derived after the successful ~~homogenization~~adjustment. Decadal trends are calculated for seasonal ~~HSavg, aggregated either from~~ dHS (snow days) of several thresholds, HSavg aggregated from either monthly HSavg (HOMER and Climatol) or daily HS (interpQM). ~~Furthermore, trends for snow days are compared for original and interpQM.~~ The largest HSmax value per station, calculated over the entire period, (absolute maximum snow depth, maxHSmax) is compared for ~~homogenized~~ homogenised and original values. ~~Return~~ The return levels for seasonal HSmax with a 50-year return period (R50HSmax) are ~~retrieved from either daily homogenized HS aggregated into~~ determined either from daily homogenised HS aggregated to seasonal HSmax (interpQM) or ~~monthly homogenized~~ from monthly homogenised HSmax (HOMER and Climatol). ~~Trends~~ All calculated trends and R50HSmax ~~for the subset~~ of the different methods are then compared~~across the methods.~~ .

Climatol automatically fills ~~in missing values, thus artificially increasing the length of the series. However, for the trend analysis, only the time period available in the original series are considered for the homogenized data in order to make it comparable. ACMANT, being fully automatic, does not allow for manual break point input and was therefore not used for our analyses. We are using Climatol v3.1.2 (Guijarro, 2019), and InterpQM (Resch et al., 2022) on R 4.2 (R Core Team, 2022) and HOMER v2.6 on R 2.15 (R Core Team, 2012). To evaluate the differences between the used homogenization methods, a Kolmogorov-Smirnov test was performed with seasonal data.~~ eventually existing missing dates and interpolates their corresponding values, resulting in an artificially increased length of these series. It also automatically adjusts outliers in the homogeneous period in the default settings.

4 Results

In the following section, we compare the results of different adjustment methods on the one hand and the homogenised data with the non-homogenised data on the other. In this way we can show both the effect of homogenisation and the ~~impact of~~ dependence of the ~~performed homogenization on indices of snow depth series is shown. The figures and tables demonstrate that it mostly lowered the original snow depths and increased the number of time series with a significant negative trend in the time period studied. The differences between homogenization methods were more pronounced for R50HSmax than for trends of HSavg. The Kolmogorov-Smirnov test performed with seasonal values showed that the homogenized data was significantly different from the original in five stations for HSmax and eight stations for HSavg. While there were no significant differences between the methods for HSmax, there were for four stations for HSavg.~~ results on the method used. In chapter 4.1 we show this as an example for the number of snow days and in particular for the effects on the trend (for interpQM only). Similarly, this is also shown for the maximum snow depth in chapter 4.2 Finally, in chapter 4.3 a particular example is given for the magnitude and frequency of extreme snow depth.

4.1 Trends of snow days

Table 2. Statistics for snow days (dHS) for the period 1961 to 2021 on a seasonal basis with thresholds of 5 (dHS5), 30 (dHS30) and 50 (dHS50) cm for both original (Orig) and ~~interpQM-homogenized~~ interpQM-homogenised data (iQM).

	dHS5				**dHS30**				**dHS50**			
	<1500 m		**>1500 m**		**<1500 m**		**>1500 m**		**<1500 m**		**>1500 m**	
	Orig	iQM	Orig	iQM	Orig	iQM	Orig	iQM	Orig	iQM	Orig	iQM
Median trend [days/decade]	-4.8	-6	-2.4	-2.8	-5.7	-5.5	-3.3	-3.7	-3.2	-2.5	-4	-5.5
Mean trend [days/decade]	-5.6	~~-5.8~~ -5.9	-3.3	-3.4	-5.7	-4.9	-4.3	-4.7	-4.7	-3.5	-5.4	-5.8
Positive [%]	0	0	0	0	4	0	14	0	4	0	7	0
No trend [%]	0	0	7	7	19	19	0	0	42	35	0	0
Negative [%]	100	100	93	93	77	81	86	100	54	65	93	100
Significant Positive* [%]	0	0	0	0	0	0	0	0	0	0	0	0
Significant Negative* [%]	81	85	69	77	70	67	33	43	86	53	31	50
Positive to no trend [%]	-	0	-	0	-	0	-	0	-	0	-	0
Positive to Negative [%]	-	0	-	0	-	7	-	14	-	7	-	7
Negative to no trend [%]	-	0	-	0	-	0	-	4	-	0	-	0
Negative to Positive [%]	-	0	-	0	-	0	-	0	-	0	-	0
Significant to not significant [%]	-	4	-	0	-	12	-	0	-	15	-	0
Not significant to significant [%]	-	8	-	7	-	15	-	7	-	12	-	21

Percentages for significant negative and significant positive, indicated with an asterisk, are calculated based on the total number of negative/positive values respectively.

~~The snow days of each season were examined in two subsets of stationsbelow and above~~

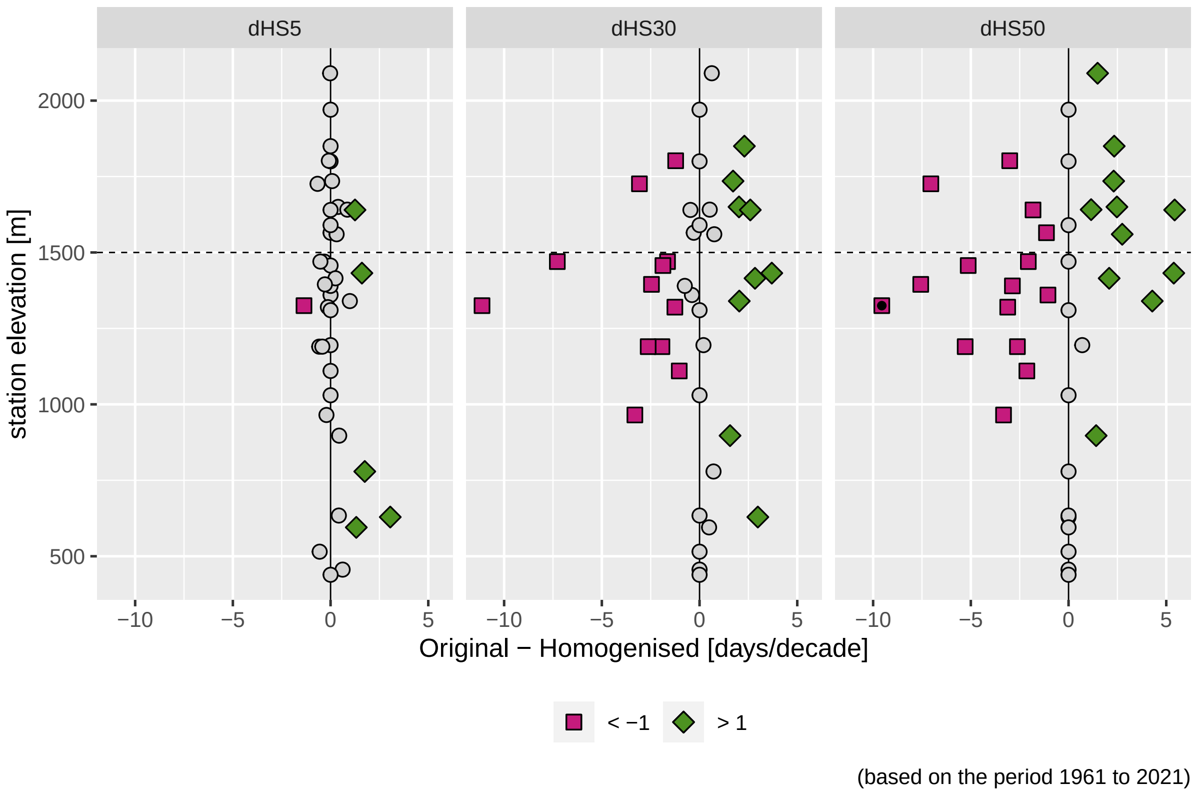

Figure 2. Difference in snow day trends between original and interpQM adjusted series for thresholds 5, 30 and 50 cm (dHS5, dHS30 and dHS50). Purple squares indicate stations with a result of < -1, green diamonds of > 1 day/decade. Black dots indicate a significant difference. Positive values indicate a stronger negative trend by homogenisation.

The number of snow days per season was examined for two subgroups of stations, below (n = 26) and above (n = 14) 1500 m ~~,~~ a.s.l., referred to in this chapter ~~called~~ as "~~lower altitude~~low elevation" and "~~higher altitude~~high elevation" stations. This ~~altitude was chosen as the dividing line because the area above it is currently normally covered with snow in winter.A~~ threshold was also used by e.g. Auer et al. (2007). Additionally, a strong decrease in snow depth ~~in the altitude range from~~ between 1500 to 2500 m ~~has been identified~~ was determined for the coming decades ~~(e.g. Marke et al., 2015; Marty et al., 2017), which makes this altitude interesting for the analysis of the~~ (Marke et al., 2015; Marty et al., 2017). This makes this elevation range interesting for analyses of changes that have already ~~occurred. Thresholds~~ taken place. We analysed thresholds of 5, 30 and 50 cm (dHS5, dHS30, dHS50) ~~where analyzed~~ for the original and ~~homogenized data. As shown in Table 2, the number of snow days per season is declining for the vast majority of stations for all analyzed thresholds and altitude levels.~~ homogenised data.

~~In the original data, no station has a positive trend for~~ The adjustments made had the strongest effect on dHS30 and dHS50 at stations above 1000 m, as can be seen in Figure 2. The percentage of significant negative time series has increased for all indices above 1500 m and dHS5 ~~, but between 4 - 14 % of stations~~ below 1500 m while it was reduced by 3 % for stations below 1500 m for dHS30 and by 33 % (from 86 to 53 %) for dHS50. The ~~homogenization with interpQM removed all positive snow day trends. Due to the warmer temperatures at lower elevations, the smaller thresholds of dHS5 and dHS30 show a stronger decline there. In the original data, the mean trend for stations at lower elevations is -5.6 days~~difference between the trend strength of the original and homogenised time series was more than one day/decade at six out of 40 stations for dHS5~~and -5.7~~ , at 21 for dHS30 ~~. For stations >1500 m it is -3.3 days/decade for dHS5 and -4.3 for dHS30. The percentage of stations with a significant negative trend~~ and at 26 for dHS50. Negative trends were strengthened at 5 stations for dHS5~~and dHS30 is also higher for lower elevations: 81 % of the lower and 69 % of the higher altitude stations~~, 9 for dHS30 and 11 stations, while they were weakened at 1 station for dHS5~~and 70 %, respectively 33 %~~ , 12 for dHS30 ~~. This is the other way round for the 50 cm threshold: The mean trend for lower elevations is -4.7 days/decade and -5.4 for higher situated stations. Only 54 % of lower stations have a negative trend for~~ and 15 for dHS50~~, but 93 % of stations at higher elevations. This may be irritating, but it has several reasons: Firstly, the variability of greater snow depths is strong. Secondly, days with a snow depth of >50 cm are possible at lower altitudes, but generally rare. Therefore, 89 % of stations <1000 m and 18 % of stations between 1000 - 1500 m show no trend for~~ . To detect significant differences between the original and homogenised time series, the non-parametric Wilcoxon test was applied. As can be seen in Figure 2, this was only the case at the Adelboden station for dHS50~~. With the exception of Ilios, all other stations below 1500 m show a strong mean decrease between -2.9 to -14.7 days/decade.~~ .

~~As mentioned above, no station in the original data shows a~~ The number of snow days per season is declining for the vast majority of stations for all analysed thresholds and elevation levels, as shown in Table 2. In the original dataset, none of the stations investigated has a positive trend for dHS5, three show a slightly positive trend for ~~dHS5. For~~ dHS30 ~~, one lower station~~ (Unterwasser-Iltios at 1340 m ~~(Iltios) shows a small positive trend of~~ with +0.9~~days/decade, and two at higher elevations (~~, Mürren at 1650 m with +0.3 and St.Moritz at 1850 m with +1.7 days/decade) ~~. For~~ and two for dHS50 ~~, the stations Iltios (~~(Unterwasser-Iltios with +1.2 ~~)~~ and St.Moritz ~~(~~with +1.1 ~~) are positive. Due to the warmer temperature at lower altitudes, the smaller threshold values dHS5 and dHS30 show stronger negative trendsthere.The mean trend for dHS5 is -5.6~~ days/~~decade for low and -3.3 for higher elevations. For~~ decade).

Overall, the homogenisation removed all positive trends and, depending on the threshold for snow depth and elevation subset, either did not change or reduced the number of stations without trends: E.g. 86 % of the high elevation stations had a negative trend for dHS30 ~~it is -5.7~~ before, and ~~-4.3, respectively. Both the colder temperature and the greater amount of precipitation lead to more frequent greater snow depths at higher elevations. The~~ 100 % after the homogenisation. The percentage of low elevation stations with no trend for dHS50 ~~is stronger at higher (-5.4 days/decade) than at lower elevations (-4.7). Also,~~ changed from 42 % to 35 % after the homogenisation, while the percentage of stations with ~~decreasing dHS50 is almost twice as high for larger altitudes (93 % at >1500 m and~~ a negative trends was raised from 54 % ~~for <1500 m)~~to 65 %.

~~Homogenization with interpQM removed all positive trends~~ In general, the adjustments changed the median and mean trends of both subsets for dHS5 and the higher elevation subsets for dHS30 and ~~increased the percentage of stations with a decline in snow days for both~~ dHS50 to more negative, the lower elevation subsets of dHS30 and dHS50 to less negative. The mean trends ~~for lower altitudes~~ of the lower elevations changed from -5.6 to ~~-5.8~~ -5.9 days/decade for dHS5, ~~weakened~~ from -5.7 to -4.9 for dHS30 and from -3.7 to -3.5 for dHS50. For higher ~~altitudes it~~ elevations they changed from -3.3 to -3.4 for dHS5, ~~strengthened~~ from -4.3 to -4.7 for dHS30 and from -5.4 to -5.8 for dHS50.

~~All analysed stations above 1500 m show a decrease in days per season with snow depths of more than 30 or 50 cm after the adjustments made. It reinforced the negative trend for~~ The percentage of low-elevation stations with no trend is different for the larger thresholds than for dHS5~~in both altitudesubgroups. The mean trend of~~ , where it increases from 0 to 7 % with increasing altitude, but decreases for both dHS30 ~~at lower altitudes got weakened from -5.7 to -4.9 days/decade, while it strengthened from -4.3 to -4.7 at higher altitudes. For~~ (from 19 to 0 %) and dHS50 (from 42 to 0 %). Homogenisation changed these figures only for dHS50, ~~the negative trend weakened from -4.7 to -3.5 days/decade at lower elevations, but strengthened at higher elevations (from -5.4 to -5.8). The percentage of stations, whose negative trend was significant was increased for all stations for~~ where instead, where instead of 42 % only 35 % of the lower elevation stations don't show a trend. The number of stations with a negative trend decreased for both dHS5 ~~, but only at higher altitudes~~ (from 100 to 93 %) and the lower stations for dHS30 (from 77 to 81 %). However, the numbers increased at the higher elevations for dHS30 ~~. For dHS50, it decreased sharply~~ (from 86 % to 100 %) and at all elevations for dHS50 (from 54 to 65 % for the lower ~~altitudes but increased from 31 % to 50~~ and from 93 to 100 % for the higher ~~altitudes. The stronger impact of the homogenization with interpQM to larger thresholds is also shown in Fig. 2, with changes between -11 to +4 days/decade~~ elevations). A similar pattern is seen in the significant negative trends: An increase at all higher elevation stations (between 8 and 19 %), but a decrease at lower elevations for dHS30 ~~and -10 to +5 days/decade for dHS50. In contrast, the changes from homogenization for dHS5 were minor, between -1 to +~~(3 ~~days/decade. All in all,~~ %) and dHS50 (33 %). Overall, interpQM weakened the dHS5-trends for 35 % of all stations, strengthened ~~it~~ them for 30 % and did not change ~~it~~ them for 35 %. For dHS30, 38 % of all stations had weaker trends after the adjustments, 40 % had ~~a stronger one~~ stronger trends and for 22 % it did not change. For dHS50, the trend ~~was~~ weakened for 30 %, strengthened for 38 % and ~~did not change~~ remained unchanged for 32 % of all stations. The adjustments changed the trend of one station to non-significant for dHS5 and of 12 to significant. Six stations for dHS30 were changed to non significant and 10 to significant. For dHS50, the trends of 10 stations were changed to non significant and of 12 to significant.

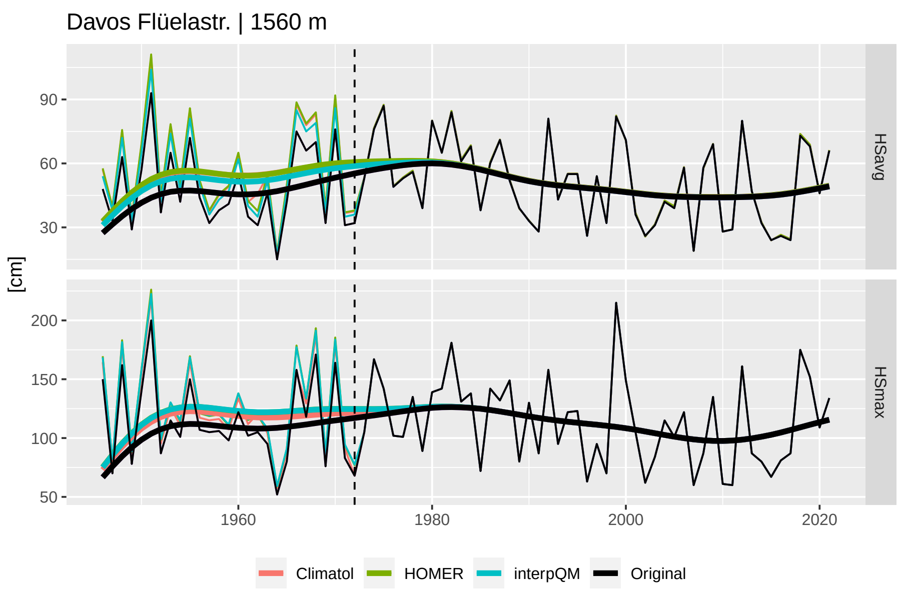

Figure 3. ~~Difference~~ Comparison of ~~trends for snow days between Original~~ original and ~~interpQM~~ homogenised seasonal mean (HSavg) and maximum (HSmax) snowdepth for ~~5, 30~~ the SLF-station in Davos. Panel (a) shows original and ~~50 cm thresholds~~ adjusted seasonal time series (~~dHS5, dHS30~~ thin lines) and ~~dHS50~~Gaussian filtered with an 30 y window (thick lines). The vertical dashed line indicates the identified break in 1972.

The KS-test did not reveal significant differences between the original and the interpQM-adjusted time series in the distribution of the dHS5-, dHS30- or dHS50-time series for any of the stations analysed. A comparison with the W-test also showed no significant differences for dHS5 and dHS30, but at one station (Adelboden) for dHS50.

4.2 Trends of mean and maximum snow depth

4.3 ~~Trends of mean and maximum snow depth~~

The effect of homogenisation on the mean (HSavg) and maximum snow depths (HSmax) is illustrated using the example of Davos in Figure 3. The adjustments made increased the seasonal mean snow depth before the break in 1972 between 2 - 11 cm with interpQM, 3 - 17 cm with Climatol and 3 - 18 cm with HOMER. The impact on the seasonal maximum snow depth range from 2 - 19 cm with Climatol, 7 - 23 cm with interpQM and 7 - 26 cm with HOMER.

To assess the impact of ~~homogenization~~ homogenisation on trends of HSavg and HSmax~~of snow depth series~~, decadal trends are calculated for each ~~homogenization~~ homogenisation method and the original data respectively. Figure 4 shows the trends for HSavg in the left and for HSmax in the right panel. Trends are expressed as cm/decade for the period from 1961 to 2021 for each method and station of the ~~inhomogeneous~~ nonhomogeneous subset, black dots indicate significant trends. For

HSavg, we found an overall similar pattern across the methods. Figure A3 shows the trends as differences between original and ~~homogenized~~ homogenised values for Climatol, HOMER, and interpQM for both HSavg and HSmax~~. Two~~ , separately. Two of the original series (St. Moritz and ~~Iltios~~Unterwasser-Iltios) show positive trends, whereas HOMER displays positive trends for Simplon and Glarus. No trends are positive with interpQM or Climatol. Except Glarus (HOMER), none of the positive trends are significant. Homogenisation made the HSavg trends of 17 (HOMER) and 18 (Climatol, interpQM) of the 40 stations either negative or more negative, and of 21 (interpQM), 22 (Climatol) or 23 (HOMER), less negative, respectively. Table 3 describes the mean and median trends across all stations, as well as the change from positive to negative and significant to not significant and vice-versa for both HSavg and HSmax. The mean trends of HSavg for Climatol and HOMER appear to be weaker than for the original and interpQM homogenised.

Figure 4 further reveals that the ~~homogenized~~ homogenised trends for HSavg mimic the pattern of the original trends, which shows almost zero trends for stations below 500 m, strong negative and significant trends for the group between 1000 and ~~1400~~ 1500 m, followed by mostly not significant trends for stations between 1500 and 1600 m a.s.l. This suggests that the various intrinsic ways of building reference series and sub-networks of the underlying ~~homogenization~~ homogenisation methods do not have a significant impact on decadal trends of HSavg.

~~When focusing on trends of HSmax, the vast majority of series show negative trends. However, in contrast to HSavg, some are positive across the methods, see~~ The vast majority of trends for HSmax, 37 of the original series and 39 for all homogenisation methods, show a negative trend, as shown in the right side of Table 3 for details. The number of significant trends is about 20 % lower than for HSavg~~. The order of the methods is maintained with interpQM still~~ , with interpQM showing the largest and HOMER the lowest number of significant trends. The most striking difference between the ~~pattern in panel (a) and (b) of~~ patterns of HSavg and HSmax in Fig. 4 is the area ~~with no~~ without significant trends. ~~For HSavg, this is located~~ This is between 1500 and 1600 m a.s.l. ~~, whereas for HSmax it concerns stations~~ for HSavg for all homogenisation methods, and below 1000 m a.s.l. ~~The exception is interpQMwhich shows significant trends for stations below 1000 m a. s.l. There seems to~~ for HSmax with the exception of time series adjusted by interpQM. There seems be no particular altitudinal pattern~~other than that~~ , except that the trends below 1000 m a.s.l. are weak ~~across the~~ for all methods and increase in strength ~~from~~ between 1200 ~~to~~ and 1400 m a.s.l. This suggests that the trends for HSmax, in contrast to HSavg, ~~trends for HSmax seem~~ appear to be more sensitive to the underlying ~~homogenization methods.~~ homogenisation methods.

The performed KS-test for revealing noticable differences between the original and adjusted HSavg time series showed significant differences for four stations for HOMER (Meien, Klosters, Sils-Maria, Stans), two for Climatol (Meien, Sils-Maria) and interpQM (Klosters, Stans). The W-test showed similar results with six stations for HOMER (Meien, Klosters, St.Moritz, Glarus, Sils-Maria and Stans), five for Climatol (Meien, Klosters, St.Moritz, Glarus, Sils-Maria and Stans) and one for interpQM (Klosters). For a comparison of the results of the adjustment methods, the homogenised time series were compared against each other with the KS- and W-test. With the KS-test, significant differences were found for all methods for two stations (Glarus and Stans). The W-test-results were significant between HOMER and Climatol also for two stations (Luzern and Stans). For HSmax, the KS-test showed significant differences between the original and adjusted time series for three stations for HOMER and interpQM (Klosters, St.Moritz and Elm) and two stations for Climatol (Klosters and St.Moritz). The

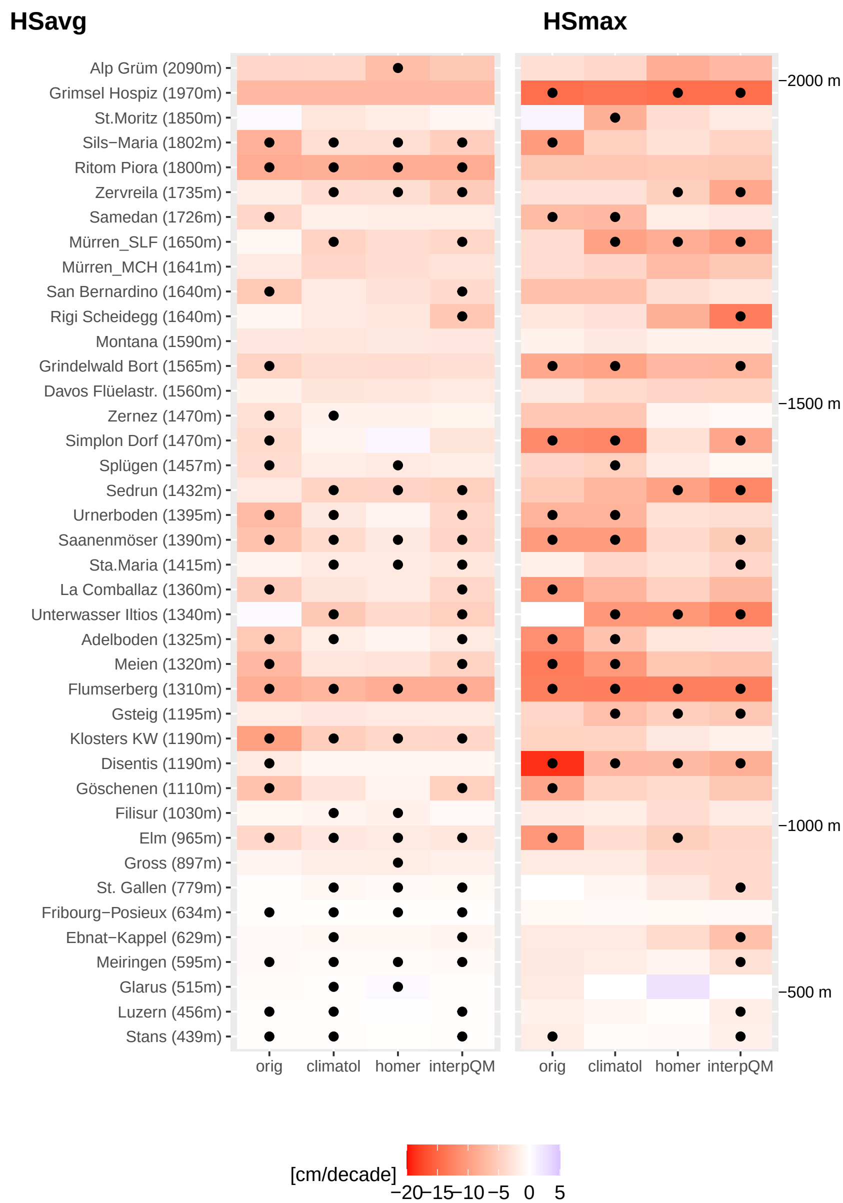

Figure 4. Comparison of trends calculated with original and ~~homogenized~~ homogenised data (Climatol, HOMER, and interpQM) for the period 1961-2021 for HSavg (left side) and HSmax (right side). Stations are ordered according to elevation. Black dots indicate statistical significance with p-values below 0.05.

W-test was significant for four stations with HOMER and interpQM (Klosters, St.Moritz, Elm, Sils-Maria) and three with Climatol (Klosters, St.Moritz, Sils-Maria). The adjustment methods were significantly different only with the W-test for three stations (La Comballaz, Saanenmöser, Samedan) between HOMER and Climatol.

Table 3. Statistics for trends of HSavg and HSmax for the period 1961 to 2021.

	HSavg				**HSmax**			
	Original	Climatol	HOMER	interpQM	Original	Climatol	HOMER	interpQM
Median trend [cm/decade]	-2.4	-2.5	-2.3	-2.7	-4.3	-4.9	-3.8	-4.5
Mean trend [cm/decade]	-3.7	-2.8	-2.6	-3.2	-5.8	-5.6	-4.6	-5.4
Positive [%]	5	0	5	0	8	3	3	3
Negative [%]	95	100	95	100	98	98	98	98
Significant [%]	55	55	43	60	38	35	23	45
Significant negative* [%]	58	55	42	60	41	36	23	46
Significant positive* [%]	0	0	50	0	0	0	0	0
Positive to Negative [%]	-	5	8	5	-	5	5	5
Negative to Positive [%]	-	0	5	0	-	0	5	0
Significant to not significant [%]	-	23	33	15	-	15	25	20
Not significant to significant [%]	-	23	20	20	-	13	13	28

Percentages for significant negative and significant positive, indicated with an asterisk, are calculated based on the total number of negative/positive values respectively.

4.3 Impact on ~~maximum~~ extreme snow ~~depth~~depths

To investigate a possible influence of the ~~homogenization procedure on~~ homogenisation on the magnitude and frequency of extreme snow depths, the absolute maximum snow depths (maxHSmax) recorded at each station over the entire period, the year with the absolute maximum snow depth ~~, as well as~~ and the difference between original and ~~homogenized~~ homogenised maxHSmax are plotted for each station and ~~homogenization~~ homogenisation method. Figure 5 shows the results. Here we found that for the majority of series, the differences are 0. The differences are generally left-skewed, except for the largest differences observed with Climatol (~~Fig. 4~~ Panel (d) of Figure 5). This again suggests that, in contrast to the trends ~~of HSavg, differences across the~~ for HSavg, the differences between methods are more ~~visible~~ apparent for HSmax. Furthermore, ~~Fig. 4~~ Panel (c) of Figure 5 clearly highlights the four known snow-rich winters of 1951, 1968, 1975, and 1999.

~~Return~~ The return levels for 50-year return periods of maximum snow depth (R50HSmax) are calculated from ~~homogenized~~ homogenised data and compared ~~to~~ with the values obtained from the original data including the 95 % confidence intervals. Figure 6 shows the original values in grey with ~~their~~ the associated 95 % confidence intervals and the ~~homogenized~~ homogenised values in colour. ~~Here we found a similar pattern for a~~ A pattern was found to occur in all methods for the majority of stations~~across the methods. However, between six (interpQM) and 13 (Climatol), with HOMER in-between with 11, stations have R50HSmax outside~~ . For Climatol, seven stations are above the 95 % confidence intervals of the original values ~~. For Climatol, seven are above~~ for R50HSmax and six below, ~~HOMER shows~~ for HOMER there are four above and seven below, ~~whereas for interpQM the numbers~~ while for interpQM there are three above and three below, see Table 4 for details. This again suggests that the differences between the ~~homogenization~~ homogenisation methods are more pronounced for R50HSmax than for trends of HSavg, with interpQM performing slightly better than Climatol or HOMER. An additional analysis (not shown here) of the change ~~of~~ in the 95 % confidence intervals ~~reveals~~ shows that the 95 % confidence intervals of the ~~homogenized~~ homogenised values are smaller than the original ~~. Mean values over~~ ones. The mean values of R50HSmax across all 40 stations range from 89 cm for the original to 75 cm (HOMER), with Climatol (87 cm) and interpQM (83 cm) ~~in-between~~in between.

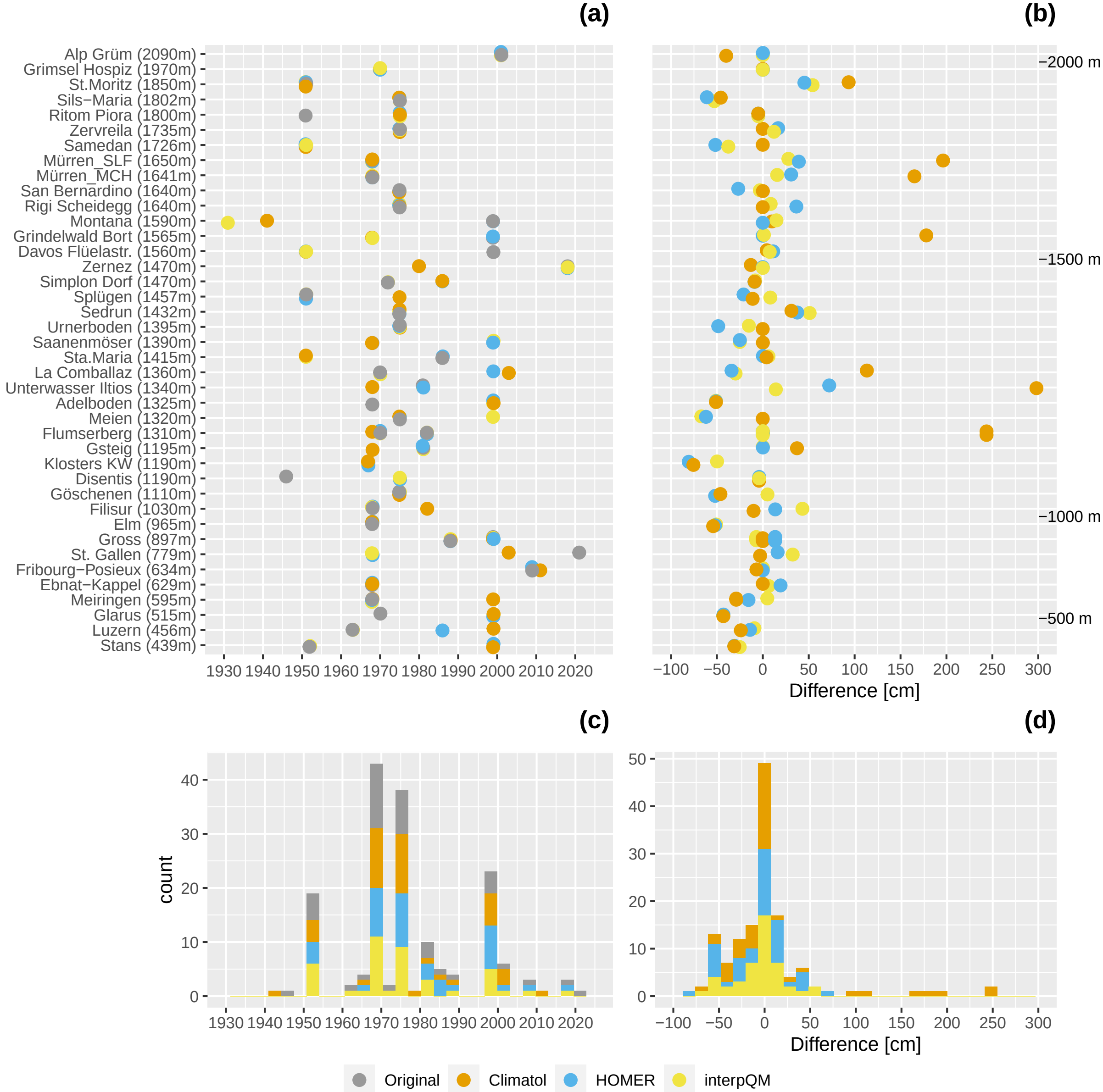

Figure 5. Maximum values of HSmax recorded for each station and method over the entire period (1961-2021). Panel (a) shows the year for which the absolute maximum snow depth is recorded. Panel (b) displays the differences between original and ~~homogenized~~ homogenised values. Panel (c) and (d) are the corresponding histograms.

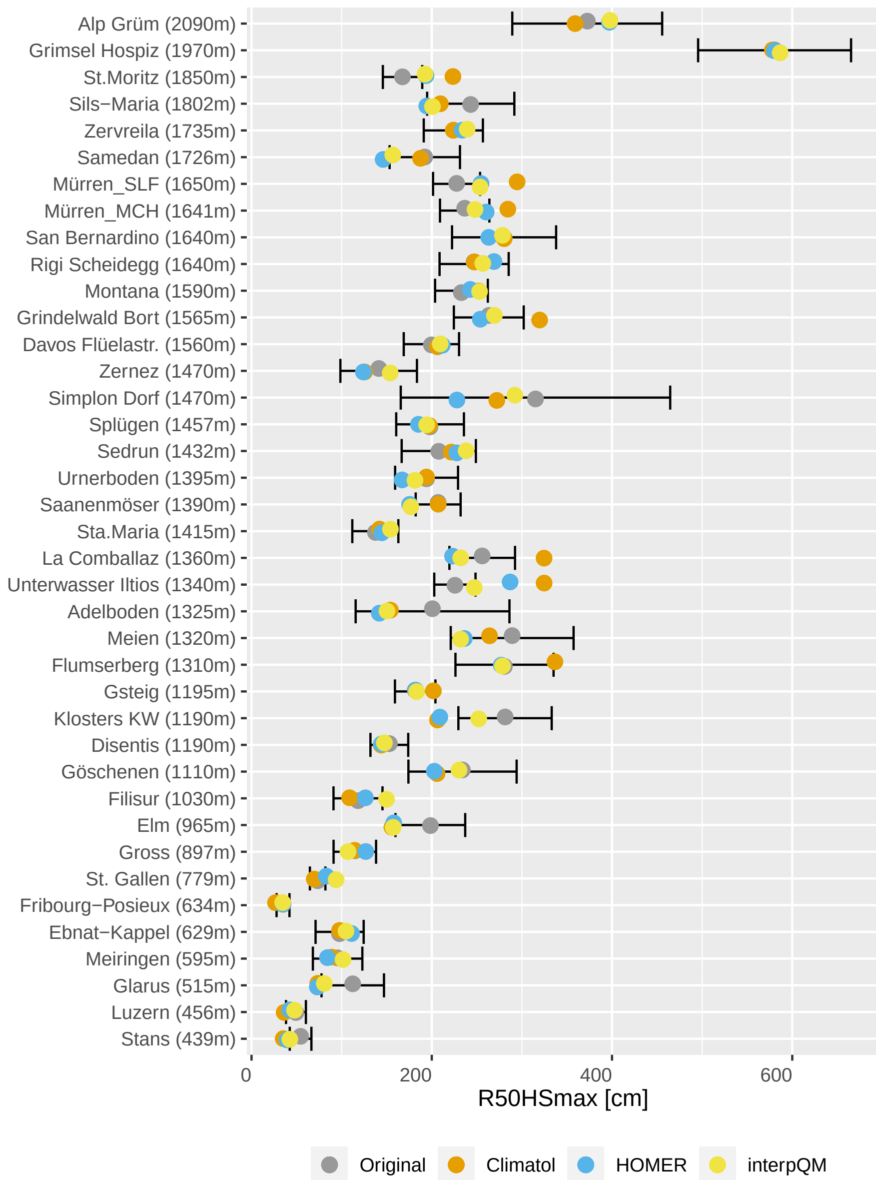

Figure 6. HSmax with 50-year return periods and 95 % confidence intervals for both original (grey) and ~~homogenized~~ homogenised data using Climatol (orange), HOMER (blue), and interpQM (yellow). The whiskers represent the 95 % confidence interval for the original values. Stations are ordered ~~left to right~~ according to elevation.

Table 4. Statistics for R50HSmax: Number and percentage of stations that are outside the origininal's 95 % confidence intervals for each ~~homogenization~~ homogenisation method.

	R50HSmax		
	Climatol	HOMER	interpQM
Outside 95 % conf interv	13 (32.5 %)	11 (27.5 %)	6 (15 %)
Above	7 (17.5 %)	4 (10 %)	3 (7.5 %)
Below	6 (15 %)	7 (17.5 %)	3 (7.5 %)

5 Discussion

The three methods agreed in decreasing the snow depth in the time prior to the breaks for 19 (48 %) of the 40 stations, while increasing it for 17 (43 %). For 4 (9 %) stations, the methods had different signs for the adjustments. The differences between the homogenisation methods were more pronounced for R50HSmax and HSmax than for HSavg.

~~The analysis of dHS5, which~~ In contrast to the larger thresholds of the snow day analysis, dHS5 shows almost no differences between the original and the ~~homogenized series, confirms~~ homogenised series, confirming the stability of this metric as ~~reported in~~ described by Buchmann et al. (2021b). The ~~altitude pattern~~ elevation-dependent pattern with the strongest adjustment effects for dHS30 and dHS50 between 1000 - 1700 m can be explained by the fact that, firstly, at stations below 1000 m a.s.l. ~~hardly any dHS30 and dHS50 occur andtherefore small changes can have a strong influence~~there are few days with a snow depth of 30 cm or more due to the generally warmer temperature and lower snowfall amount and, secondly, that above 1700 m winter temperatures are low enough and therefore less sensitive to warming in winter so that the trends are smaller. A similar pattern can be seen in the absolute values (Appendix A1). Those high-elevation stations that show large differences ~~between~~ in the trends before and after the ~~homogenization~~ homogenisation in Fig. 2 (~~SIA, x.Sam, 6SB, ZNZ, 4SM, and 5SP~~Sils-Maria, Samedan, San Bernardino, Zernez, Simplon Dorf, and Splügen) are all located at ~~location highly influenced by southern flows. Especially the Engadin, a high-elevated,~~ sites strongly influenced by southerly flows. In particular, the Engadine in the southeast, a high elevation inner alpine valley with ~~its~~ a dry and cold climate, is ~~normally~~ often not associated with large snow depths or many days with dHS30 or dHS50.~~Furthermore, these large positive differences occur either from (a) a change of trend from positive to negative after homogenization or (b) from no trend to negative. Moreover, the size of the adjustment factor is not directly responsible for these large differences, as the impact on the trends is more sensitive on the length and location of the adjusted sub-period than the magnitude. Even more so, as the actual adjustment factors for x.SAM, SIA, ZNZ, anf ABO are all smaller than 1, and are not among the largest adjustment factors in the subset, even when considering absolute values.~~

All but two of the trends for HSavg (~~both in original and homogenized~~ in both the original and homogenised data) are negative, ~~corresponding to~~ which is consistent with the findings from previous snow studies ~~(e.g. Laternser and Schneebeli, 2003; Marty, 2008; S~~ (Laternser and Schneebeli, 2003; Marty, 2008; Scherrer et al., 2013; Fontrodona Bach et al., 2018; Matiu et al., 2021). Marcolini et al. (2019) report an increase in series showing significant trends for HSavg after ~~homogenization~~ homogenisation (40 to 44 %). ~~We can see the same effect~~ The same effect is observed here for interpQM, but not for Climatol (no change) and HOMER~~, which shows a decrease of significant trends after homogenization~~ . Both show a decrease in significant negative trends after homogenisation (Table 3). The same increase ~~can be observed for trends in~~ in the number of significant negative trends is observed for snow days and HSmax. ~~There are no striking differences across the homogenization method for HSavg, whereas for HSmax, the results from the various homogenization methods differ for stations below 1000 m a.s.l. with interpQM having stronger and more significant trends. A performed Kolmogorov-Smirnov-Test showed no significant differences for the results~~

~~of three homogenization methods for HSavg except 4 of the 40 stations(GLA, GRA, LUZ, SNS)~~The adjustments decreased the snow depth prior to a break at 55 % and increased it at 45 % of the stations.

For ~~the majority of stations,~~ most stations, the R50HSmax ~~from homogenized~~ of the homogenised data are still within the 95 % confidence intervals of the original values. However, depending on the ~~homogenization~~ homogenisation method, between ~~3 to 7~~ three to seven of the investigated 40 stations (see Table 4) ~~stations show~~ have R50HSmax ~~exceeding~~ that exceed the original values beyond the 95 % confidence intervals, with potential implications for engineering applications and building codes. Values ~~well~~ that are significantly above the 95 % intervals are predominantly from Climatol. The reference networks in Climatol are ~~built with the help of~~ created using the Euclidean distances between candidate and reference series, with an optional ~~scale-factor~~ scaling-factor for the vertical component. We set this threshold to $wz = 100$ ~~, as in that configuration, the station pair~~ to avoid the selection of stations that are close together horizontally but far apart vertically, e.g. the stations Davos (1570 m a.s.l.) and Weissfluhjoch (2535 m a.s.l.)~~cannot be selected~~, which are only 4 km apart horizontally. However, it ~~maybe the case~~ may also be that this threshold is ~~just~~ simply not low enough to ~~inhibit~~ prevent further station combinations with a similarly large gradient. Unfortunately, ~~Climatol does not allow the user to see which series have been used as references for a given~~ the user cannot see in Climatol which series were used as reference for a particular station. The ~~decrease~~ reduction of the 95 % confidence intervals ~~across the methods after the homogenization~~ for all methods after homogenisation indicates a decrease in variation and an increase in confidence in the results for very large snowdepths, as shown with the absolute maximum snowdepth.

The observed differences ~~across the methods~~ between the three methods compared can be explained by the ~~various intrinsic means of handling the construction of the~~ respective methods used to construct the reference series sub-networks ~~of reference series~~ and the adjustments. HOMER adjusts the entire period before an identified ~~break point with~~ breakpoint using a single factor, ~~whereas~~ while Climatol uses multiple factors dependent on the reference series constructed using ~~homogenized~~ homogenised sub-periods. InterpQM, on the other hand, uses multiple adjustment factors based on quantile matching for the entire inhomogeneous period, similar to HOMER. The range of the applied adjustments for interpQM is shown in Appendix A4.

The selection of suitable reference series is the crucial part of the ~~homogenization procedure,~~ homogenisation procedure, both for the detection of breaks ~~as well as~~ and for the adjustment step. HOMER can ~~either be run~~ be run either in correlation or distance mode, ~~meaning~~ i.e. the sub-networks are ~~built~~ compiled based on thresholds for either correlation or horizontal distances. In Climatol, the sub-networks are ~~built using~~ formed based on the Euclidean distance between series with a ~~scale~~ scaling parameter for the vertical component. ~~InterpQM allows the user to~~ In InterpQM, the user can choose correlation and ~~both horizontal and~~ horizontal as well as vertical distance thresholds. For ~~an elevation dependent~~ a height-dependent variable such as snow depth, the ~~possibility~~ ability to select the sub-networks by ~~means of~~ setting thresholds for vertical and horizontal distances separately proves invaluable. It is possible, albeit cumbersome, to ~~manually~~ define the sub-networks manually and use them as input for HOMER. The ~~possibility~~ ability in HOMER to visually ~~scrutinize~~ inspect the set of reference series used for each candidate station can ~~give a useful hint~~ provide a useful indication of how accurately the reference series reflect local climatic or topographic characteristics; ~~e.g. is~~ for example does a majority of the reference series ~~drawn~~ come from a completely different micro-climate? This is ~~especially~~ particularly important for a study area with ~~a complex Alpine~~ complex alpine topography, where neighbouring valleys ~~can~~ may have completely different climates: North/South, inner-Alpine, or ~~pre-Alps~~pre-Alpine. Furthermore, these lists of reference series can also be used to identify stations with suspicious reference series that are probably not ~~suited for homogenization~~suitable for homogenisation.

The analysis of the sub-networks for HOMER and interpQM shows that due to the distance ~~constraint~~ restriction in interpQM, reference series are drawn from a more similar region, whereas ~~for HOMER ,~~ in HOMER distant stations with high correlations are frequently included. To avoid selecting close-by, but unsuitable reference series due to ~~local-scale~~ local climatic variations, the correlation criterion in interpQM works well.

Both Marcolini et al. (2019) and Buchmann, M. et al. (2022) found that relocations where responsible for by far the most detected breakpoints in snow depth time series. The metadata of many stations are sparse and therefore often do not provide enough information to give a sufficient answer to the question, why a relocation caused a break. A change in elevation within +/- 150 m is not necessarily a cause of a break, but moving a station either below or above the typical height of a site's inversion is highly likely. Significant changes in the station environment are also very likely to cause a break, e.g. moving a station to an area with fewer buildings or fewer and smaller trees and vice versa.

6 Conclusions

This study is the first in-depth comparison of different ~~homogenization~~ homogenisation methods applied to a large network of snow depth series between 500 and 2500 m. The focus is on their influence on ~~extreme snow depths, decadal trends in~~ the decadal trends of the number of snow days~~of different thresholds~~, i.e. days with a snow depth above a certain threshold (5, 30 and 50 cm), ~~and~~ the seasonal mean and maximum snow depths (HSavg, HSmax) and extreme snow depths. The results ~~confirm the significant importance of homogenizing long time series (e.g. Auer et al., 2007; Venema et al., 2020)~~underpin the relevance of homogenising long term snow depth series for trend and extreme values analysis. Due to the impact of homogenisation on derived trends, this is especially true for conclusions drawn from ~~single series. For~~ individual series. In our analyses, for the long-term trends of HSavg and dHS5, the overall picture does not change ~~with homogenized or original data~~ through homogenisation of original data by median/mean based adjustment methods. However, ~~it becomes clearer when a quantile based homogenization~~ the picture becomes different when a quantile-based homogenisation approach (interpQM) is applied, which in the case of Swiss snow depth series, shows the strongest effect with ~~exclusively~~ only negative trends for HSavg and ~~, as expected,~~ a slight increase in the number of significant trends. The differences between the methods ~~become clearer when considering the~~ increase when looking at seasonal maximum values: The trends for HSmax, where trends of low elevation stations were significant only with interpQM, absolute maximum snow depths and extreme values. The ~~performed homogenization~~ homogenisation performed with interpQM increases the confidence in the derived extreme values based on the 95 % confidence interval, which is particularly ~~useful~~ relevant for engineering applications. As far as snow days are concerned, the quantile based adjustments had the strongest impact on the larger snow depth thresholds.

~~Separating~~ Our results support a homogenisation approach that separates the breakpoint detection from the adjustment procedure, ~~i.e.using~~ e.g. to use the robust combined detection approach described in Buchmann et al. (2022) in combination with the adjustment procedure from Resch et al. (2022)~~has proven to be promising. The ability to easily~~ . However, the ability to manually adjust the automatic selection ~~process of the (reference~~ of the reference (sub-network) stations used for ~~comparison with the candidate station is crucial to achieve the best-possible~~ homogenisation is crucial for optimising the results. A combination of ~~several selection~~ criteria such as correlation, horizontal and vertical distances as well as manual interventions seems to be more advantageous ~~for snow in combination with~~ (given the complex topography ~~prevalent in mountainous regions such as the Alpsthan using only~~ in mountain regions like the Alps) for snow depth than the use of a single selection criterion.

~~Since the break detection~~ So far, the homogenised snow depth time series show no evidence of a bias in the methods towards increasing or decreasing snow depths due to the adjustments made, neither in Austria nor in Switzerland. In this study, depending on the homogenisation method, the mean snow depth before a break was increased at about 52 - 57 % of the stations and decreased at between 42 - 45 %. 95 % of the 40 inhomogeneous stations show a negative trend for seasonal mean snow depth in the original data, which is significant for 58 %. These figures are lower for the 144 homogeneous stations in the dataset, where 78 % show a negative trend that is significant for 50 %.

As pointed out, break detection for snow depth is preferably done ~~separately, there~~ using the described two-out-of-three method. From our experience, there there is no incentive or ~~obvious benefit to rely only on automatic homogenization methods ,~~ advantage to use automatic homogenisation methods such as HOMER and Climatol. On the contrary, automatic methods open the door to unintended automatic outlier corrections or adjustments based on the selection of reference series that are sufficiently correlated but cannot be assigned in a climatologically meaningful way. To achieve reasonable results, ~~even~~ these methods require ~~some~~ a certain degree of user intervention, e.g. ~~use a pre-defined~~ the use of a predefined selection of reference stations, ~~using~~ thresholds for correlation, horizontal and vertical distances. Therefore~~it would be best to use the results of the described 2-out-of-3 method for break point~~ , it seems promising to separate the detection and adjustment of breaks using the the described two-out-of-three method for detection and interpQM for the ~~adjustments, albeit being laborious~~adjustment, as it provides reliable results especially for larger snow depths and yields daily data.

Funding information FWF (Fonds zur Förderung der wissenschaftlichen Forschung, Project Hom4Snow, Grant/Award Number: I 3692) and Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (Grant/Award Number:SNF:200021L 175920)

Code and data availability. Input data for the various homogenisation methods are available on EnviDat https://doi.org/10.16904/envidat.336

Author contributions. The study was devised by MBU and CM with input from GR, WS and MBT. Snow day analysis was performed by GR, HSavg and HSmax by MBU. Figures were produced by GR and MBU. GR and MBU discussed the results with input from CM, SB and WS. MBU and GR wrote the initial draft. The article was finalised by GR with contributions from all co-authors.

Competing interests. We do not have any competing interests.

Acknowledgements. The authors want to thank both MeteoSchweiz, SLF, ZAMG and the Austrian Water Budget Department for access to their datasets. For data juggling, ~~homogenization~~ homogenisation and evaluation, R 4.2 (R Core Team, 2022) with the tidyverse package (Wickham et al., 2019) as well as R 2.15 was used. We would also like to thank all snow observers. Without their continuous effort, this study would not have been possible. 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 Schweizerischer Nationalfonds zur Förderung der WissenschaftlichenForschung (Grant/Award Number:SNF: 200021L 175920)

ORCID

Moritz Buchmann https://orcid.org/0000-0002-1710-5496

Gernot Resch https://orcid.org/0000-0002-4734-1158

Michael Begert https://orcid.org/0000-0002-9569-6842

Stefan Brönnimann https://orcid.org/0000-0001-9502-7991

Barbara Chimani https://orcid.org/0000-0003-0298- 4181

Wolfgang Schöner https://orcid.org/0000-0001-6546-0639

Christoph Marty https://orcid.org/0000-0002-0398-6253

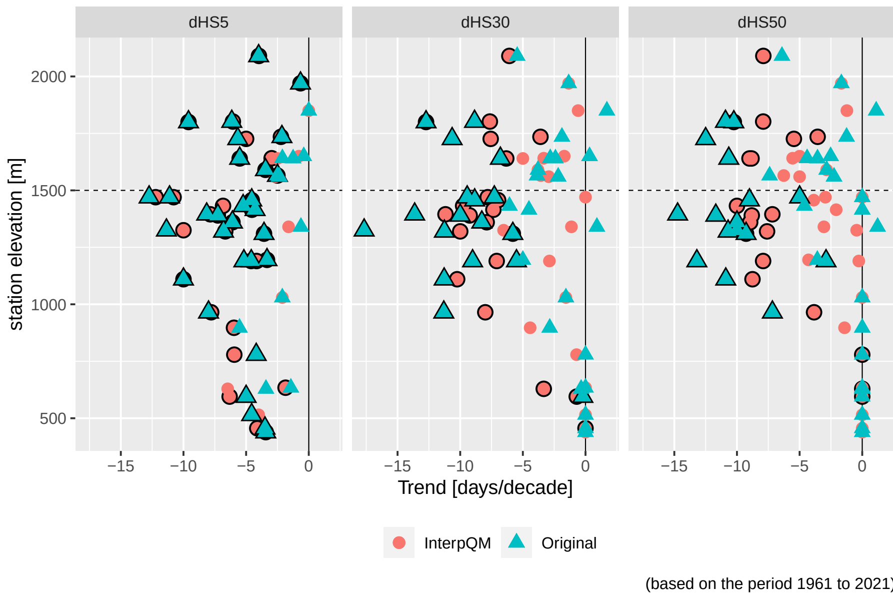

Figure A1. Absolute trends for days with snow depth of at least 5, 30 and 50 cm per season. Significant trends are marked by a black border.

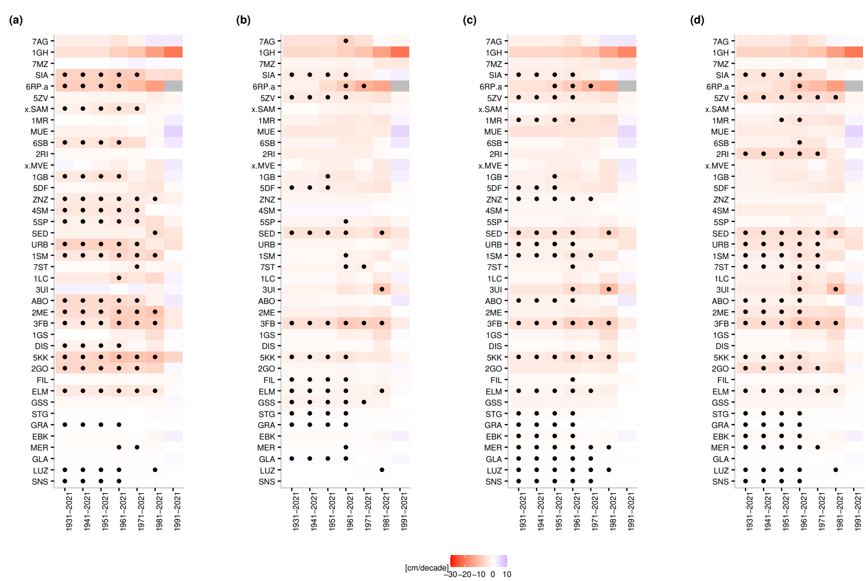

Figure A2. Trends for HSavg: Shown are all methods and all decades. Original (a), HOMER (b), Climatol (c), and interpQM (d).

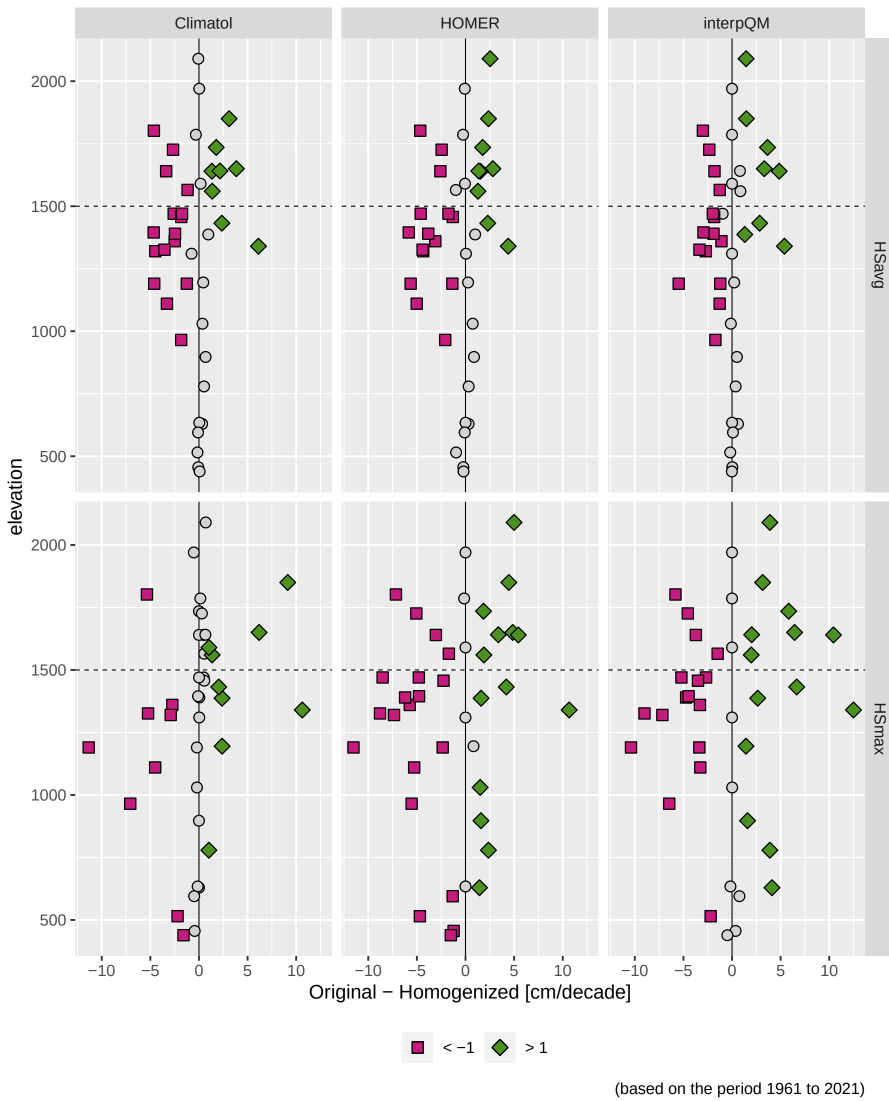

Figure A3. Comparison of differences of trends calculated with Climatol, HOMER, and interpQM for the period 1961-2021 for HSavg and HSmax. Differences are ~~expressed~~ calculated as original ~~minus homogenized~~- homogenised. ~~Stations are ordered according to elevation (low-high, bottom-top) with an corresponding altitudinal scale in panel (b). Black dots~~ Purple squares indicate ~~statistical significance~~ stations with ~~p-values below 0.05~~a result of < -1, green diamonds of > 1 days/decade.

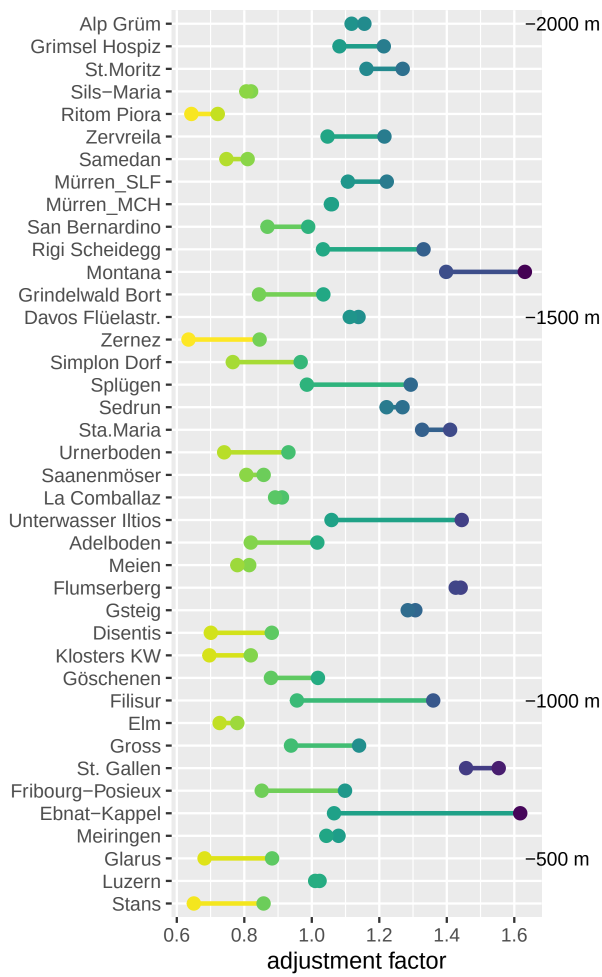

Figure A4. Range of the applied adjustment factors of interpQM for the different quantiles

References

Abegg, B., Morin, S., Demiroglu, O. C., François, H., Rothleitner, M., and Strasser, U.: Overloaded! Critical revision and a new conceptual approach for snow indicators in ski tourism, International Journal of Biometeorology, https://doi.org/10.1007/s00484-020-01867-3, 2020.

Al-Rubaye, S., Maguire, M., and Bean, B.: Design Ground Snow Loads: Historical Perspective and State of the Art, Journal of Structural Engineering, 148, 03122 001, https://doi.org/10.1061/(ASCE)ST.1943-541X.0003339, 2022.

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 6, 661–675, https://doi.org/10.1002/joc.3370060607, 1986.

Alexandersson, H. and Moberg, A.: HOMOGENIZATION OF SWEDISH TEMPERATURE DATA. PART I: HOMOGENEITY TEST FOR LINEAR TRENDS, International Journal of Climatology, 17, 25–34, https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J, 1997.

Armstrong, R. and Brun, E.: Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling, Cambridge University Press, https://doi.org/10.3189/002214309788608741, 2008.

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 Nieplova, E.: HISTALP—historical instrumental climatological surface time series of the Greater Alpine Region, International Journal of Climatology, 27, 17–46, https://doi.org/https://doi.org/10.1002/joc.1377, 2007.

Begert, M., Schlegel, T., and Kirchhofer, W.: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000, International Journal of Climatology, 25, 65–80, https://doi.org/10.1002/joc.1118, 2005.

Bocchiola, D., Bianchi Janetti, E., Gorni, E., Marty, C., and Sovilla, B.: Regional evaluation of three day snow depth for avalanche hazard mapping in Switzerland, Natural Hazards and Earth System Sciences, 8, 685–705, https://doi.org/10.5194/nhess-8-685-2008, 2008.

Brown, R. D., Brasnett, B., and Robinson, D.: Gridded North American monthly snow depth and snow water equivalent for GCM evaluation, Atmosphere-Ocean, 41, 1–14, https://doi.org/10.3137/ao.410101, 2003.

Buchmann, M., Begert, M., Brönnimann, S., and Marty, C.: Local-scale variability of seasonal mean and extreme values of in situ snow depth and snowfall measurements, The Cryosphere, 15, 4625–4636, https://doi.org/10.5194/tc-15-4625-2021, 2021a.

Buchmann, M., Begert, M., Brönnimann, S., and Marty, C.: Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series, International Journal of Climatology, 41, E2553–E2563, https://doi.org/10.1002/joc.6863, 2021b.

Buchmann, M., Coll, J., Aschauer, J., Begert, M., Brönnimann, S., Chimani, B., Resch, G., Schöner, W., and Marty, C.: Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-)automatic homogenization methods, The Cryosphere, 16, 2147–2161, https://doi.org/10.5194/tc-16-2147-2022, 2022.

Buchmann, M., Aschauer, J., Begert, M., and Marty, C.: Input data for break point detection of Swiss snow depth series, EnviDat [data set], https://doi.org/10.16904/envidat.297, 2022.

Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, Journal of the Royal Statistical Society: Series C (Applied Statistics), 53, 405–425, https://doi.org/10.1111/j.1467-9876.2004.05155.x, _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9876.2004.05155.x, 2004.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., and Jones, P. D.: An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, https://doi.org/10.1029/2017JD028200, 2018.

Croce, P., Formichi, P., and Landi, F.: Extreme Ground Snow Loads in Europe from 1951 to 2100, Climate, 9, https://doi.org/10.3390/cli9090133, 2021.

Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), International Journal of Geosciences, 02, 293–309, https://doi.org/10.4236/ijg.2011.23032, 2011.

Easterling, D. R. and Peterson, T. C.: A new method for detecting undocumented discontinuities in climatological time series, International Journal of Climatology, 15, 369–377, https://doi.org/10.1002/joc.3370150403, 1995.

Essery, R., Morin, S., Lejeune, Y., and Ménard, C. B.: A comparison of 1701 snow models using observations from an alpine site, Advances in Water Resources, 55, 131–148, https://doi.org/10.1016/j.advwatres.2012.07.013, 2013.

Fontrodona Bach, A., van der Schrier, G., Melsen, L. A., Klein Tank, A. M. G., and Teuling, A. J.: Widespread and Accelerated Decrease of Observed Mean and Extreme Snow Depth Over Europe, Geophysical Research Letters, 45, 12,312–12,319, https://doi.org/https://doi.org/10.1029/2018GL079799, 2018.

Gilleland, E. and Katz, R. W.: extRemes 2.0: An Extreme Value Analysis Package in R, Journal of Statistical Software, 72, 1–39, https://doi.org/10.18637/jss.v072.i08, 2016.

Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C., and Rosas, G.: The influence of station density on climate data homogenization, International Journal of Climatology, 37, 4670–4683, https://doi.org/10.1002/joc.5114, 2017.

Guijarro, J. A.: Homogenization of climatic series with Climatol, Tech. rep., AEMET, https://doi.org/10.13140/RG.2.2.27020.41604, 2018.

Guijarro, J. A.: climatol: Climate Tools (Series Homogenization and Derived Products), https://CRAN.R-project.org/package=climatol, r package version 3.1.2, 2019.

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

Hiebl, J. and Frei, C.: Daily precipitation grids for Austria since 1961—development and evaluation of a spatial dataset for hydroclimatic monitoring and modelling, Theoretical and Applied Climatology, 132, 327–345, https://doi.org/10.1007/s00704-017-2093-x, 2018.

Johnston, A. N., Bruggeman, J. E., Beers, A. T., Beever, E. A., Christophersen, R. G., and Ransom, J. I.: Ecological consequences of anomalies in atmospheric moisture and snowpack, Ecology, 100, 1–12, https://doi.org/10.1002/ecy.2638, 2019.

Jonas, T., Rixen, C., Sturm, M., and Stoeckli, V.: How alpine plant growth is linked to snow cover and climate variability, Journal of Geophysical Research: Biogeosciences, 113, https://doi.org/10.1029/2007JG000680, 2008.

Kendall, M.: Rank Correlation Methods, Charles Griffin, 1975.

Kuglitsch, F. G., Toreti, A., Xoplaki, E., Della-Marta, P. M., Luterbacher, J., and Wanner, H.: Homogenization of daily maximum temperature series in the Mediterranean, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011606, 2009.

Laternser, M. and Schneebeli, M.: Long-term snow climate trends of the Swiss Alps (1931–99), International Journal of Climatology, 23, 733–750, https://doi.org/10.1002/joc.912, 2003.

Li, Q., Yang, T., and Li, L.: Evaluation of snow depth and snow cover represented by multiple datasets over the Tianshan Mountains: Remote sensing, reanalysis, and simulation, International Journal of Climatology, 42, 4223–4239, https://doi.org/10.1002/joc.7459, 2022.

Mann, H.: Nonparametric tests against trend, Econometrica, pp. 245–259, 1945.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nature Climate Change, 7, 764–773, https://doi.org/10.1038/nclimate3418, 2017.

Marcolini, G., Bellin, A., and Chiogna, G.: Performance of the Standard Normal Homogeneity Test for the homogenization of mean seasonal snow depth time series: PERFORMANCE OF SNHT FOR SNOW DEPTH TIME SERIES, Int. J. Climatol, 37, 1267–1277, https://doi.org/10.1002/joc.4977, 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, International Journal of Climatology, 39, 4514–4530, https://doi.org/10.1002/joc.6095, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6095, 2019.

Marke, T., Strasser, U., Hanzer, F., Stötter, J., Wilcke, R. A. I., and Gobiet, A.: Scenarios of Future Snow Conditions in Styria (Austrian Alps), Journal of Hydrometeorology, 16, 261 – 277, https://doi.org/10.1175/JHM-D-14-0035.1, 2015.

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

Marty, C. and Blanchet, J.: Long-term changes in annual maximum snow depth and snowfall in Switzerland based on extreme value statistics, Climatic Change, 111, 705–721, https://doi.org/10.1007/s10584-011-0159-9, 2012.

Marty, C., Schlögl, S., Bavay, M., and Lehning, M.: How much can we save? Impact of different emission scenarios on future snow cover in the Alps, The Cryosphere, 11, 517–529, https://doi.org/10.5194/tc-11-517-2017, 2017.

Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C. M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., and De Gregorio, L.: Observed snow depth trends in the European Alps: 1971 to 2019, The Cryosphere, 15, 1343–1382, https://doi.org/10.5194/tc-15-1343-2021, 2021.

Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Boehm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER : a homogenization software - methods and applications, IDOJARAS, 117, 47–67, https://www.researchgate.net/publication/281471961, 2013.

Morin, S., Horton, S., Techel, F., Bavay, M., Coléou, C., Fierz, C., Gobiet, A., Hagenmuller, P., Lafaysse, M., Ližar, M., Mitterer, C., Monti, F., Müller, K., Olefs, M., Snook, J. S., van Herwijnen, A., and Vionnet, V.: Application of physical snowpack models in support of operational avalanche hazard forecasting: A status report on current implementations and prospects for the future, Cold Regions Science and Technology, 170, 102 910, https://doi.org/10.1016/j.coldregions.2019.102910, 2020.

Nemec, J., Gruber, C., Chimani, B., and Auer, I.: Trends in extreme temperature indices in Austria based on a new homogenised dataset, International Journal of Climatology, 33, 1538–1550, https://doi.org/10.1002/joc.3532, 2013.

Nitu, R., Roulet, Y., Wolff, M., Earle, M., Reverdin, A., Smith, C., Kochendorfer, J., Morin, S., Rasmussen, R., Wong, K., Alastrué, J., Arnold, L., Baker, B., Buisan, S., Collado, J. L., Colli, M., Collins, B., Gaydos, A., Hannula, H.-R., Hoover, J., Joe, P., Kontu, A., Laine, T., Lanza, L., Lanzinger, E., Lee, G. W., Lejeune, Y., Leppänen, L., Mekis, E., Panel, J., Poikonen, A., Ryu, S., Sabatini, F., Theriault, J., Yang, D., Genthon, C., van den Heuvel, F., Hirasawa, N., Konishi, H., Nishimura, K., and Senese, A.: WMO Solid Precipitation Intercomparison Experiment (SPICE) (2012-2015), techreport 131, WMO, https://library.wmo.int/index.php?lvl=notice{_}display{&}id=20742, 2018.

Olefs, M., Schöner, W., Suklitsch, M., Wittmann, C., Niedermoser, B., Neururer, A., and Wurzer, A.: SNOWGRID – A New Operational Snow Cover Model in Austria, International Snow Science Workshop Grenoble – Chamonix Mont-Blanc - October 07-11, 2013, pp. 38–45, https://www.researchgate.net/publication/281377426, 2013.

Picard, F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam, B., and Robin, S.: Joint segmentation, calling, and normalization of multiple CGH profiles, Biostatistics, 12, 413–428, https://doi.org/10.1093/biostatistics/kxq076, 2011.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., and Norberg, J.: Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018, Nature, 581, 294–298, https://doi.org/10.1038/s41586-020-2258-0, 2020.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2012.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2022.

Resch, G., Koch, R., Marty, C., Chimani, B., Begert, M., Buchmann, M., Aschauer, J., and Schöner, W.: A quantile-based approach to improve homogenization of snow depth time series, International Journal of Climatology, n/a, https://doi.org/10.1002/joc.7742, 2022.

Schellander, H., Winkler, M., and Hell, T.: Towards a reproducible snow load map - An example for Austria, Advances in Science and Research, 18, 135–144, https://doi.org/10.5194/asr-18-135-2021, 2021.

Scherrer, S. C., Appenzeller, C., and Laternser, M.: Trends in Swiss Alpine snow days: The role of local- and large-scale climate variability, Geophysical Research Letters, 31, https://doi.org/10.1029/2004GL020255, 2004.

Scherrer, S. C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R., and Appenzeller, C.: Snow variability in the Swiss Alps 1864-2009, International Journal of Climatology, 33, 3162–3173, https://doi.org/10.1002/joc.3653, 2013.

Schmucki, E., Marty, C., Fierz, C., Weingartner, R., and Lehning, M.: Impact of climate change in Switzerland on socioeconomic snow indices, Theoretical and Applied Climatology, 127, 875–889, https://doi.org/10.1007/s00704-015-1676-7, 2017.

Schöner, W., Koch, R., Matulla, C., Marty, C., and Tilg, A.-M.: Spatiotemporal patterns of snow depth within the Swiss-Austrian Alps for the past half century (1961 to 2012) and linkages to climate change, Int J Climatol, 39, 1589–1603, https://doi.org/10.1002/joc.5902, 2019.

Sen, P.: Estimates of the regression coefficient based on Kendall's tau, Journal of the American Statistical Association, 63, 1379–1389, 1968.

Stone, R. S., Dutton, E. G., Harris, J. M., and Longenecker, D.: Earlier spring snowmelt in northern Alaska as an indicator of climate change, Journal of Geophysical Research: Atmospheres, 107, ACL 10–1–ACL 10–13, https://doi.org/10.1029/2000JD000286, 2002.

Theil, H.: A rank-invariant method of linear and polynomial regression analysis, Proc. Konink. Nederl. Akad. Wetensch. Ser. A Math. Sci., 53, 1950.

Venema, V., Trewin, B., and Wang, X.: Guidelines on Homogenization 2020 edition, Tech. rep., World Meteorological Organization Issue WMO-No. 1245, https://library.wmo.int/?lvl=notice_display&id=21756, 2020.

Vincent, L. A., Zhang, X., Bonsal, B. R., and Hogg, W. D.: Homogenization of Daily Temperatures over Canada, Journal of Climate, 15, 1322 – 1334, https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2, 2002.

Wickham, H., Averick, M., Bryan, J., Chang, W., Mcgowan, L. D. A., François, R., Grolemund, 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., Takahashi, K., Vaughan, D., Wilke, C., and Woo, K.: Welcome to the Tidyverse, The Journal of Open Source Software, 4, 1–6, https://doi.org/10.21105/joss.01686, 2019.

Yue, S., Pilon, P., Phinney, B., and Cavadias, G.: The influence of autocorrelation on the ability to detect trend in hydrological series, Hydrological Processes, 16, 1807–1829, https://doi.org/10.1002/hyp.1095, 2002.