Local-scale ~~uncertainty~~ variability of seasonal mean and extreme values of in-situ snow depth and snow~~-~~fall measurements

Moritz Buchmann[1,3], Michael Begert[2], Stefan Brönnimann[3,4], Christoph Marty[1]

1WSL-Institute for Snow- and Avalanche Research SLF, Davos, Switzerland

5   2Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

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

4Institute of Geography, University of Bern, Bern, Switzerland

Correspondence to: Moritz Buchmann (moritz.buchmann@slf.ch)

Abstract.

10   Daily ~~M~~measurements of snow depth and snowfall ~~on the daily scale~~ can vary strongly over short distances. However, it is not clear if there is a seasonal dependence in these variations and how they ~~–~~impact common snow climate indicators based on mean values, as well as estimated return levels of extreme events based on maximum values.

To analyse the impacts of local-scale variations we compiled a unique set of parallel snow measurements from the Swiss Alps consisting of 30 station pairs with up to 77 years of parallel data. Station pairs are mostly located in the same villages (or

15   within 3 km horizontal and 150 m vertical distances).

Investigated snow climate indicators include average snow depth, maximum snow depth, sum of new snow, days with snow on the ground, days with snowfall as well as snow onset and disappearance dates, which are calculated for various seasons (December to February (~~DFJ~~DJF), November to April (NDJFMA), and March to April (MA)). We computed relative and absolute error metrics for all these indicators at each station pair to demonstrate the potential ~~uncertainty~~variability. We found

20   the largest relative inter-pair differences for all indicators in spring (MA) and the smallest in DJF. Furthermore, there is hardly any difference between DJF and NDJFMA which show median ~~uncertainties~~ variabilities of less than 5% for all indicators. Local-scale ~~uncertainty~~ variability ranges between less than 24% (DJF) and less than 43% (MA) for all indicators and 75% of all station pairs. The ~~H~~highest ~~(lowest)~~ percentage (90%) of station pairs with ~~uncertainty~~ variability less than 15% is observed for days with snow on the ground with 90~~% (average snow depth, 30%).~~%. The lowest percentage (30%) of station pairs with

25   variability less than 15% is observed for average snow depth. Median differences of snow disappearance dates are rather small (three days) and similar to the ones found for snow onset dates (two days). An analysis of potential sunshine duration could not explain the higher ~~uncertainties~~ variabilities in spring.

To analyse the impact of local-scale variations on the estimation of extreme events, 50-year return levels were quantified for maximum snow depth and maximum 3-day new snow sum, which are often used for avalanche prevention measures. The

30   found return levels are within each other's 95% confidence intervals for all (but two) station pairs, revealing no striking differences.

The findings serve as an important basis for our understanding of ~~uncertainties~~ variabilities of commonly used snow indicators and extremal indices. Knowledge about such ~~uncertainties~~ variabilities in combination with break-detection methods is the groundwork in view of any homogenization efforts regarding snow time series.

35

1 Introduction

Snow, in all its forms, is of great social and environmental importance. Implications can be found in many fields as diverse as ecology, climatology, hydrology, tourism, and natural hazards. All measurements of ~~the~~ snow cover are dependent on the local characteristics of the site. ~~Especially,~~: i.e. exposure to wind or solar radiation. Furthermore, ~~, but as well as~~ nearby buildings

40 or trees may have an impact on the measured snow quantities. In an ideal world this would not matter as the basic guidelines (e.g. (World Meteorological Organization et al., 2018) ~~tell~~ recommend to measure snow in a flat, not wind-exposed measurement field, which is at least ~~in~~ at the same distance from a building or tree as the height of the obstacle. In reality, manual measurement locations ~~do~~ sometimes do not fulfil these basic ~~guidelines~~ requirements for many reasons, ~~like~~ such as the availability of suitable terrain, ~~of possible~~ and observers or easy access to the site. Due to this fact, the variability of the

45 measured snow quantities on the 1-km scale (i.e. next to an open field) may be smaller than the variability on the 10-meter scale (e.g. south or north of a house). In theory, this bias introduced by sometimes not ideal measuring locations may be season dependent as we hypothesize that the wind impact is mostly relevant during the accumulation season (availability of loose snow) and the solar impact during the ablation season (more available melt energy). The knowledge about such a possible seasonally dependent bias is important, as it helps to understand existing inhomogeneities. Furthermore, this information is

50 invaluable in view of homogenization efforts of snow data series.

Moreover, climatological applications and studies usually focus either on the meteorological winter DJF ~~(~~e.g. (Scherrer and Appenzeller, 2006)~~)~~ or on the ~~period~~ six months NDJFMA ~~(~~e.g.(Marcolini et al., 2017) ~~)~~. However, winter in the Alps is not simply restricted to these three or six months. In contrast, for many ecological and hydrological applications, the melting spring snow cover is the main interest ~~(~~e.g. (Brown and Robinson, 2011; Livensperger et al., 2016; Zampieri et al., 2015)~~)~~ and for

55 some applications even the onset of the snow cover (Roland et al., 2021). We therefore analyse the variations of important and commonly used snow climate indicators for seasonal effects. Additionally, snow onset and disappearance dates are introduced, as they are important for snow phenology, which is especially crucial for ecological purposes e.g. (Vorkauf et al., 2021). ~~Extending~~ Using and extending the data set of parallel time series introduced by Buchmann et al. (2021) in available number of stations, months and years, enables ~~to investigate~~ the investigation of the impact of the ~~above mentioned~~ above-mentioned

60 bias introduced by sometimes not ideal measuring locations, hereafter referred to as "local bias" or variability~~,~~. Investiagting snow onset and disappearance dates, as well as extreme value analyses, we strive to answer the following questions:

1) ~~Inivestigate~~ Is the ~~B~~behaviour of the "local bias" ~~as~~ a function of time series length (~~sections~~ Sect. 4.1 ~~and 5.1~~), i.e.~~:~~ ~~to check whether~~ are relative percentage deviations ~~are~~ dependent on ~~the~~ length of ~~the~~ parallel series?~~.~~

2) Is there any seasonal dependence of the "local bias"? To what extent do the various seasons contribute to variations observed in indicators derived from parallel snow series (sections 4.2)? To further explore the effects at the beginning and end of the snow season, snow onset and disappearance dates are used to put the seasonal variations into context and to test the hypothesis that the beginning and especially the end of a snow season are most sensitive to the "local bias" for snow climate indicators (Sections 4.3 and 4.4).

3) What is the impact of the "local bias" on return levels based on a commonly calculated return period of extreme events? Such values and their uncertainties are often used to answer engineering questions for prevention measures (Sect. 4.5).

The paper is organised as follows: Section 2 introduces the data set, Sect. 3 covers the statistical methods used for the analyses, results are presented and discussed in Sect. 4, and conclusions are drawn in Sect. 5.

2 Data

Our data consist of daily manually measured snow depth (HS) and height of new snow (HN) maintained by two separate institutions in Switzerland; Institute for Snow and Avalanche Research (SLF) and the Federal Office for Meteorology and Climatology (MeteoSwiss). Measurements are taken from a fixed graduated stake and HN-measurements from a board on the top of yesterday's snow cover every morning at 6:00 UTC at least between November and April (for details refer to Haberkorn (2019) and Buchmann et al. (2021)). To obtain parallel series, station pairs were constructed by combining stations into pairs which are located within a distance of 3 km horizontally and 150 m vertically of each other. The mean horizontal (vertical) distance in the data set is 1 km (50 m). Most, but not all pairs consist of one SLF and one MeteoSwiss station. The so defined set consists of 30 (24 for HN) station pairs between 490 and 1770 m a.s.l. with complete data between October and May (September and June in extreme cases). It includes one station pair with 77 years of parallel data (1944-2020) and 10 station pairs with more than 50 years of parallel data, and incorporates a total of 1338 station-years covering the time period from 1943 to 2020. However, not all station pairs cover the same length. Table A1 shows the various available time periods for each station pair. Six station pairs are excluded from all calculations involving HN due to irregular measurement procedures in the past, manifested by clusters of cases where HN equals HS minus HS from the previous day, thus neglecting the settling of the snow pack.

For our analyses, we focused on derived snow climate indicators as seasonal values. The main variables are defined in Table 1: average snow depth (HSavg), maximum snow depth (HSmax), sum of new snow (HNsum), maximum 3-day new snow sum (HN3max), days with snow on the ground, defined by HS of at least 1 cm (dHS1) and days with snowfall of at least 1 cm (dHN1). Figure A1 gives a quantitative impression of the data set by depicting elevation and mean values for HSavg and dHS1. Availability and quality of the corresponding metadata records (coordinates, observer) is an issue. Although we managed to compile complete metadata records, there is no guarantee that they are always precise and correct. Theoretically, the exact locations of the snow measurements are known, however, until about two decades ago, only

approximate coordinates ~~are~~ were recorded. The main reason is the general ~~unconsciousness~~ lack of awareness for the importance as to where the snow measurements are actually conducted. Further, in the case of some MeteoSwiss sites, the coordinates refer to the main meteorological measurements and the snow measurements may have been conducted on a slightly different spot. Also, sometimes decades have passed between station visits thus resulting in ~~lacking~~ missing information.

100 Potential sunshine duration~~s~~ ~~are~~ is obtained with the help of Swisstopo's digital elevation model DHM25 which has an accuracy in Alpine terrain of 5-8 metres.

~~The basis for all our analyses, the indicator data set, is available from Buchmann et al. (2021).~~

Table 1: Definitions and abbreviations of the investigated snow climate indicators.

Snow climate indicator	Abbreviation
~~Seasonal a~~Average snow depth	HSavg
~~Seasonal M~~maximum snow depth	HSmax
~~Seasonal n~~New snow sum	HNsum
Maximum 3-day new snow sum	HN3max
Days with snow on the ground, HS>0 cm	dHS1
Days with ~~snowfall~~new snow, HN>=1 cm	dHN1
Snow onset date, first day of longest period with HS>0 cm	Dstart
Snow disappearance date, day after longest period with HS>0 cm	Dend

3 Methods

To be able to compare and quantify the differences of the various snow climate indicators, we use relative percentage differences (RPD), calculated according to eq. 1 for each indicator (i) and station pair (X-Y) with the number of years denoted by n and k indicating the actual year. These RPDs are expressed as seasonal mean values for DJF, NDJFMA, and MA or monthly mean values. A potential influence of observational length on RPD is investigated by plotting the number of available parallel years against mean RPD for each indicator and station pair in the data set.

$$RPD := \frac{1}{n} \sum_{k=1}^{n} \left| \frac{X_k^i - Y_k^i}{mean(X_k^i, Y_k^i)} \right| \tag{1}$$

Absolute differences (absD) are calculated accordingly (see equation 2).

$$absD := \frac{1}{n} \sum_{k=1}^{n} \left| X_k^i - Y_k^i \right| \tag{2}$$

To be able to capture all onset and disappearance dates, ~~W~~we defined the current hydrological year as the period from 1 Sep of the previous year to 31 Aug of the current year, as it is often used in glaciology.

There are various definitions for ~~S~~snow onset (Dstart) and disappearance dates (~~Dstop~~Dend) depending on the application in hand e.g. Foster (1989), Kirdyanov et al. (2003), Peng et al. (2013), Stone et al. (2002), and Klein et al. (2016). However, as none of them suits our purpose and for sake of simplicity, we defined them as ~~are defined as~~ the first day (Dstart) and day after the last day (~~Dstop~~Dend) of the longest period with continuous snow cover. For the purposes in this study, we additionally allowed gaps of up to three consecutive days with no snow cover during the season. The chosen gap length allows ~~to include~~ the inclusion of full seasons in case they were fragmented in the middle of the winter by a maximum of three days without snow. Such an approach corresponds much more to the experience of the biotic world than just using the duration of the longest continuous snow cover. ~~This definition guarantees a continuous snow cover throughout the snow season which is important for applications involving ecology, albedo, or climatological analyses.~~

In order to assess the impact of the local scale differences on the long-term temporal changes of snow onset and disappearance dates Theil‑Sen linear slopes (Sen PK, 1968; Theil H, 1950) are calculated for each station pair. For each station we calculated absolute changes (AC) defined as the difference between the fitted value at the end and the fitted value at the beginning of the time series. ~~defined according to eq. 2 as:~~

~~AC := fitted value at end − fitted value at beginning~~ ~~(3)~~

To further investigate the potential impact of the beginning and end of the snow season on local bias, median differences of daily snow depth measurements for each station pair are calculated, separately for the first (accumulation) and last (ablation) 60-day periods of the hydrological year.

~~Median difference := median HS(X) − median HS(Y)~~ ~~(4)~~

To investigate a potential influence of the "local-bias" on the differences in snow disappearance dates, we compared them to the difference in potential MA sunshine duration hours (Sdur) for a selection of station pairs with good quality metadata. Sdur are obtained as daily values with the help of a digital elevation model and geographical information system (GIS) software. This calculation depends on the accuracy of the coordinates of the measurement sites.

140

To analyse the impact of the "local bias" on potential extreme events based on annual maximum 3-day new snow sum HN3max (e.g. Bocchiola et al., 2008) and annual maximum snow depth HSmax (e.g. Marty and Blanchet, 2012), return levels for fixed (50-year) return periods are calculated for each station and indicator using the R-package extRemes and standard settings (GEV, estimation method MLE, and 95% confidence intervals). For three station pairs (CAV, KUB, ZER), Gumbel instead of

145   GEV is used due to bad model fit. The analysis is based on data for NDJFMA.

4 Results and discussions

4.1 Time series length and relative percentage differences

To investigate a possible relationship of relative percentage differences (RPD) with time series length, we plotted the mean RPD (season NDJFMA) against the length of the underlying parallel time series for each indicator (Figure(Fig. 1). Here we

150   found no clear pattern. RPD are between 10 and 50% for all station pairs and indicators. The values are in line with findings from Buchmann et. al. (2021) for a 25-year common period. However, the variations within each indicator show some visual differences; with average snow depth (HSavg) having the largest (20%) and days with snow on the ground (dHS1) the smallest (8%) variations. Outliers are caused by the same station pairs for all indicators. For a majority of station pairs, all HN variables are below 20%.There is no difference between the shorter and the longer time series. This suggests that the lengths of the time

155   series have no effect on the RPD values. This further implies that time series of different lengths can be compared and compiled into one dataset for the purpose of this study. Moreover, this highlights the possible large differences among the various station pairs involved.

The absence of any clear relationship between observational length and RPD (Figure(Fig. 1) justifies the combined use of station pairs with varying lengths. However, it could be possible that the overall variations are too large to disentangle the

160   signal of possible breaks from the noise.

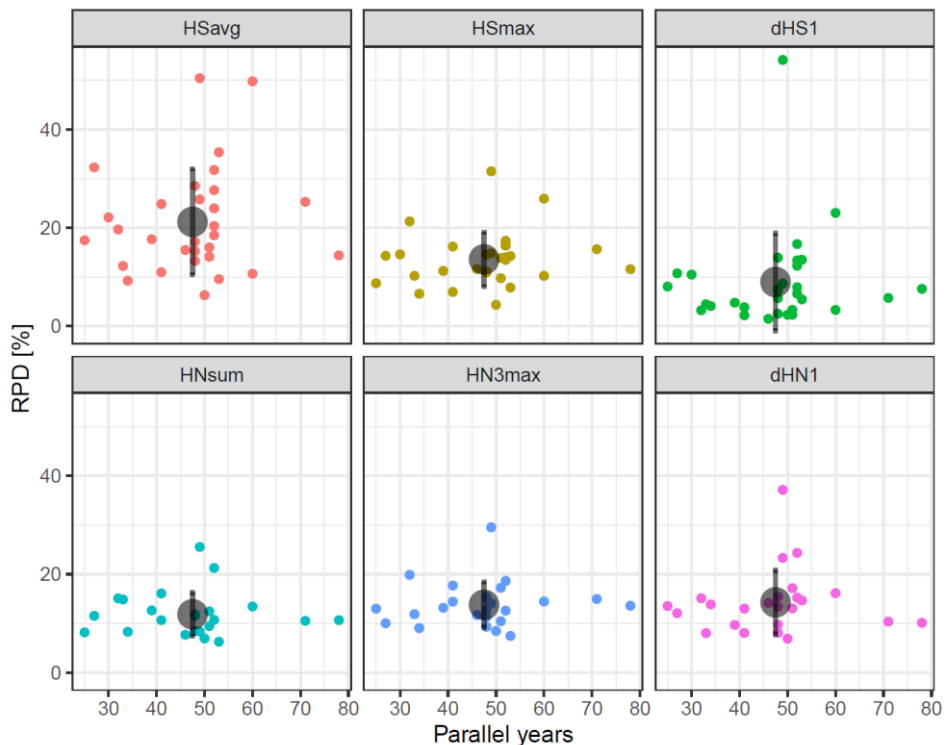



Figure 1: Relative percentage differences (RPD) as a function of number of parallel years for six snow climate indicators. Black dots and error bars indicate mean and standard deviations. The variables are defined in Table 1. Snow depth (HS) and snowfall (HN) indicators are based on 30 and 24 station pairs, respectively.

165

4.2 Seasonal influence

To explore the effects of various seasons on the snow climate indicators, we calculated RPD and absolute differences (absD)
for three commonly used seasons: winter (DJF), spring (MA), and the ~~entire season~~six months November to April (NDJFMA),
170 as well as for each individual month. Figure 2 summarizes the results. Here we found that RPD for DJF and NDJFMA are
similar (difference never exceeds 5%). However, the main differences are visible in the MA period, which shows the largest
RPD values for all indicators (~~Figure~~(Fig. 2 ~~plot~~ panel A).

Furthermore, dHS1 has the lowest median RPD for all seasons, whereas HSavg is the one with the largest median RPD for all
three seasons. Overall median RPD values for NDJFMA are between 7 and 18% and correspond to findings in Buchmann et
175 al. (2021). Figure 2 (~~plot~~ panel B) further reveals that largest relative variations occur early (November) and late (MA) in the
season and the patterns look similar for all indicators.

However, the picture looks somewhat different when we look at absD (~~Figure~~(Fig. 2 ~~plots~~ panels C and D). Here we see that
the largest median differences are not found in the same seasons for all indicators. Figure 2 (~~plot~~ panel D) reveals that absD

for HSavg and maximum snow depth (HSmax) seem to be largest in March and smallest in November, whereas for sum of new snow (HNsum), the largest absD values are found in January and February and the smallest values in November and April. The daily HS snow depth values are generally largest in March and lowest in November thus explaining the pattern of absD for HSavg and HSmax. The same applies to HNsum; the largest monthly HN snowfall sums are observed in January and February and where aswhereas November and April are months with usually the smallest HN snowfall sums. dHS1 shows the largest absD at the beginning and end of the snow season (first and last two months) likely due to larger differences in thermal conditions (ground and local air temperature) in these transition months. dHN1, in contrast, shows hardly anyfewer monthly differences than the rest, because the number of snowfall events is only marginally dependent on local measurement conditions. These monthly patterns transform to indicator dependent seasonal absD patterns (Figure(Fig. 2 plot panel C). The different seasons reveal much smaller variations compared to RPD. For NDJFMA, absD varies between 5 and 10 cm or 5 and 8 days for all indicators with the exception of HNsum. Due to the cumulating nature of the snowfall sum, the indicator HNsum always shows the largest absD values, which increases with the number of included months (NDJFMA). The same is true for the other HN snowfall count indicator dHN1, but on a much lower level. Additional information of absD allows to put the RPD into context. The lower absD values for dHS1 during DJF are no surprise, as the ground tends to be snow-covered for almost all station pairs during this period, hence the difference only amounts inapplies to cases where the snow cover is quite low (couple of cm). NDJFMA shows the highest absD value for dHS1 because this period contains the beginning and end of the season.

These findings imply that the higher RPD values in the MA period for all indicators are mainly caused by smaller absolute values in this period. The only exceptions are HSavg and HSmax, which show also high absD values in the MA period. This indicates that the "local-bias" is indeed larger for snow depth during MA.

To further test the seasonal influence from a more general point of view, we just looked at the snow depth (HS) evolution, comprising of accumulation and ablation for each station pair irrespective of the actual season. The two station pairs below 700 m a.s.l. were excluded from this analysis due to too short median winter seasonsmedian winter seasons shorter than 60 days. We calculated the climatology of the daily HS snow depth differences for each station pair. We then constructed the difference series as the combination of the first (and last) 60 days for each station pair. The median of all these difference series is the thick black line in FigureFig. 3. These two periods stand for accumulation (first 60 days) and ablation (last 60 days). The 60-day period is an empirical value. Here we found that the differences and variabilities observed in the first period seem to be smaller than in the last period (Figure(Fig. 3). This suggests that the ablation period shows more variation than the accumulation period, which corroborates the pattern found in the absD analysis.

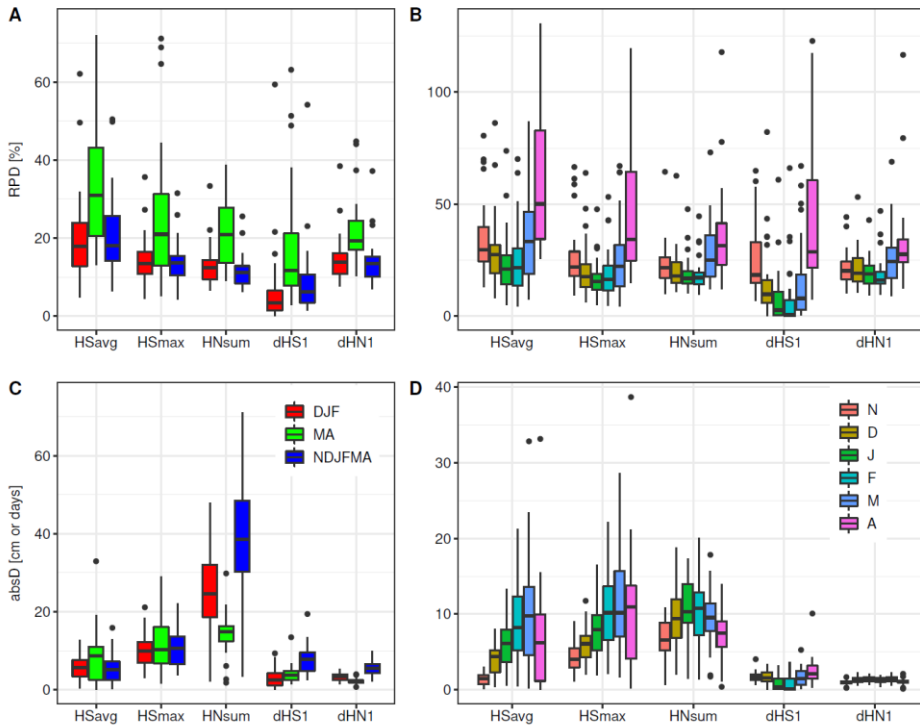



Figure 2: Shown are ~~R~~relative percentage deviations (RPD) and absolute differences (absD) for five snow climate indicators. Depicted are (~~left~~panels A and B) three seasons December to February DJF (red), March-April MA (green), and November to April NDJFMA (blue) and (~~right~~panels C and D) all months individually. Snow depth (HS) and snowfall (HN) indicators are based on 30 and 24 station pairs, respectively. The variables are defined in Table 1.~~Snow depth (HS) indicators consists of 30, snowfall (HN) indicators of 24 station pairs.~~

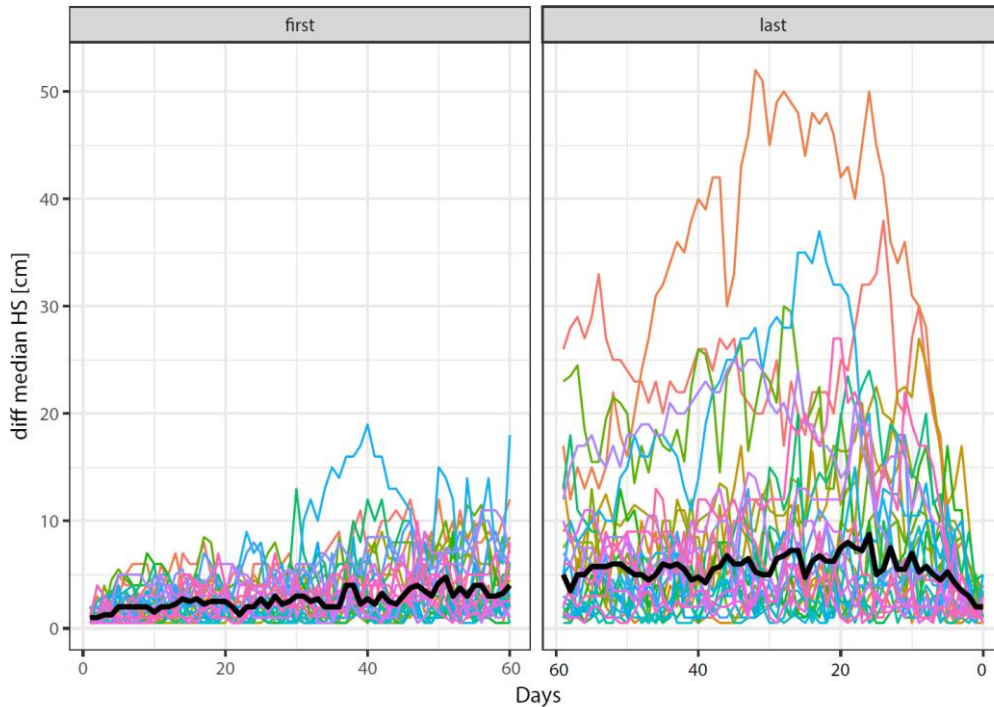



Figure 3: Median daily absolute differences in snow depth for each station pair for the first and last 60 days of each snow season. The bold line highlights the median of all station pairs.

4.3 Snow onset and disappearance dates

To further investigate the larger impact of the "local-bias" towards the end of the snow season compared to the beginning, differences between station pairs in mean snow onset (Dstart) and snow disappearance dates (~~Dstop~~Dend) are analysed. Figure 4 (left) highlights the computed absolute differences for Dstart and ~~Dstop~~Dend. We found that for 20 ~~(10)~~ out of 30 station pairs, differences of ~~Dstop~~Dend are larger ~~(smaller)~~ than differences of Dstart. In contrast to the general impression that Dstart should be the same for parallel stations, our data shows the differences of several days are not uncommon, which ~~is probably~~could be caused by the different thermal conditions of the soil of the measurement field. Figure 4 (right) depicts the mean inter-pair differences for Dstart and ~~Dstop~~Dend. The median values over all station pairs are 2 days for snow onset and 3 days for snow disappearance dates. This corroborates the previous findings of ~~Figure~~Fig. 3 that spring (ablation) shows larger variations than autumn (accumulation) with PIO, ZER, GSS, and RIE being the exceptions. We found that for 75% of the station pairs, Dstart varies between 0 and 4 days, whereas ~~Dstop~~Dend shows slightly larger differences (0 to 6 days). A possible explanation for the large differences associated with station 5KK is discussed in section 4.6.

Additionally, we use absolute temporal changes of snow onset and disappearance dates, expressed as days per decade, as yet another indicator to test the ~~uncertainty~~ variability within the station pairs. For this purpose, the temporal ~~trend~~ change in days for Dstart and ~~Dstop~~Dend are calculated for each station pair ~~(Figure~~(Fig. 5A2). ~~As not all station pairs cover the same time~~

periods, Tthe pairs are aligned from short low (left) to long high (right) parallel periodselevation. Here we see that for a majority of station pairs, the direction of changes for Dstart and DstopDend is the same, Dstart tend to be associated with positive (later) and DstopDend with negative (earlier) values.

Although inter-pair differences occur more pronounced and more frequently during the decline phase (spring) compared to the

235   accumulation period (Figure(Fig. 3), the actual end of the snow season (snow disappearance date) does not show these huge larger variations (Figure(Fig. 4). DstopDend appears to be a rather stable indicator, varying for 50% of the station pairs on average between 2 and 6 days. In contrast to what is observed in the much more complex topography on the catchment scale, this is a good result as the relative changes derived from trend analysis of station (point) data from flat measurement fields can also be transferred to the catchment scale, even though the absolute values may be different (Grünewald and Lehning (2015)).

240   Our values of temporal changes in Dstart and Dend (Fig. 5) correspond to values obtained by Klein et al. (2016). Although the definition of Dstart and Dend are different and also the time periods are not exactly the same, the absolute changes in Dstart and Dend are mostly similar for the few stations analysed by both studies. This suggests that the absolute changes of Dstart and Dend are in general quite robust and "local bias" not a main issue for the majority of stations.

Depending on the application, snow onset and disappearance dates are differently defined in spite being used frequently in

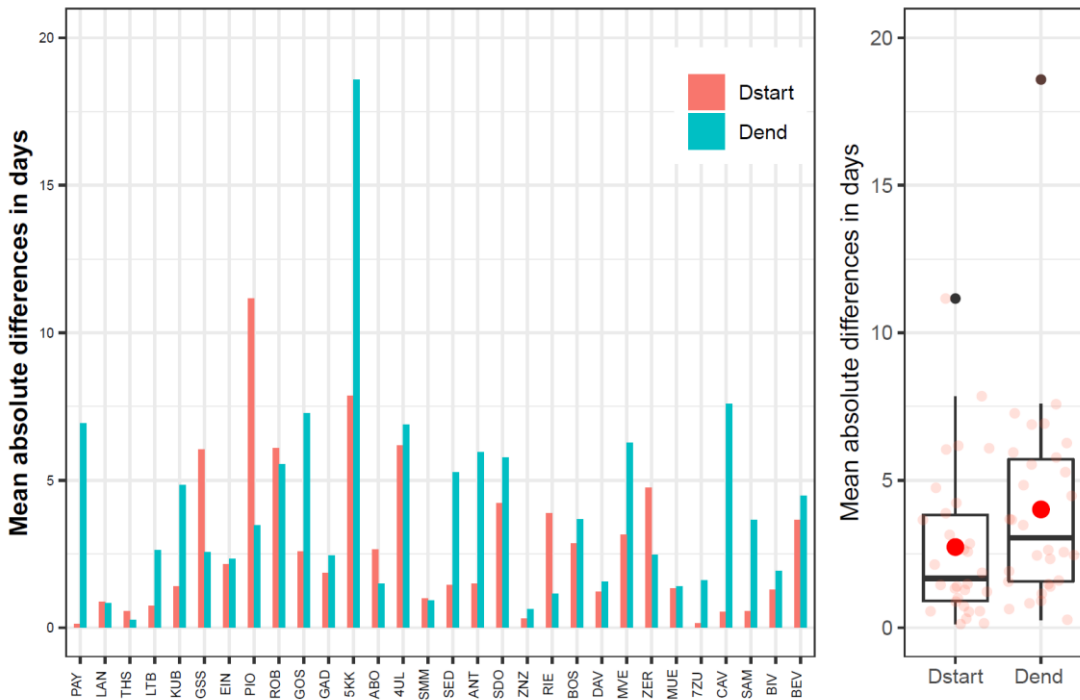
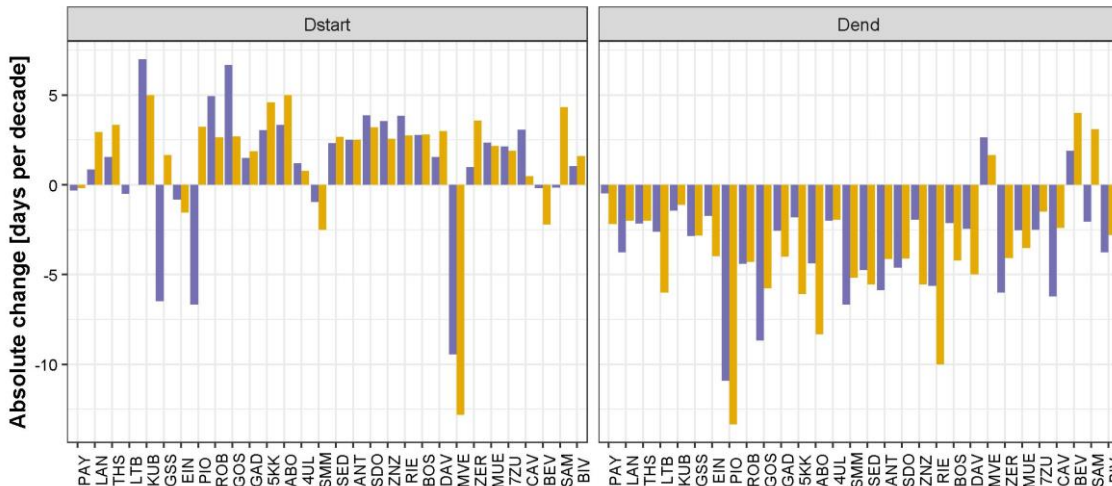
245   various studies focusing on snow climate and ecology (e.g. (Foster, 1989; Kirdyanov et al., 2003; Klein et al., 2016; Peng et al., 2013; Stone et al., 2002). Among the few stations analysed by Klein et al. (2016) are four that feature in our data set as well. Although the time period is not exactly the same, the absolute changes in Dstart and Dstop are very similar (see Table A2). As Klein et al. (2016) focused on single stations we can now see that the values are similar for the corresponding parallel stations as well. This suggests that the absolute changes of Dstart and Dstop are in general quite robust.

250

Figure 4: Left: mean absolute differences in onset (Dstart) and disappearance (~~Dstop~~Dend) dates for all station pairs. The station pairs are ordered left to right according to their elevation. Right: Variations of mean absolute differences of Dstart and ~~Dstop~~Dend for all station pairs and available years. The light red dots mark the actual values and the bold red dot indicates the mean.

255 **Figure 5: Absolute changes of snow onset (Dstart) and disappearance (Dend) dates calculated for all available parallel periods and expressed as days per decade. Station pairs are ordered according to elevation from left to right. Yellow and purple bars indicate individual stations within a pair.**

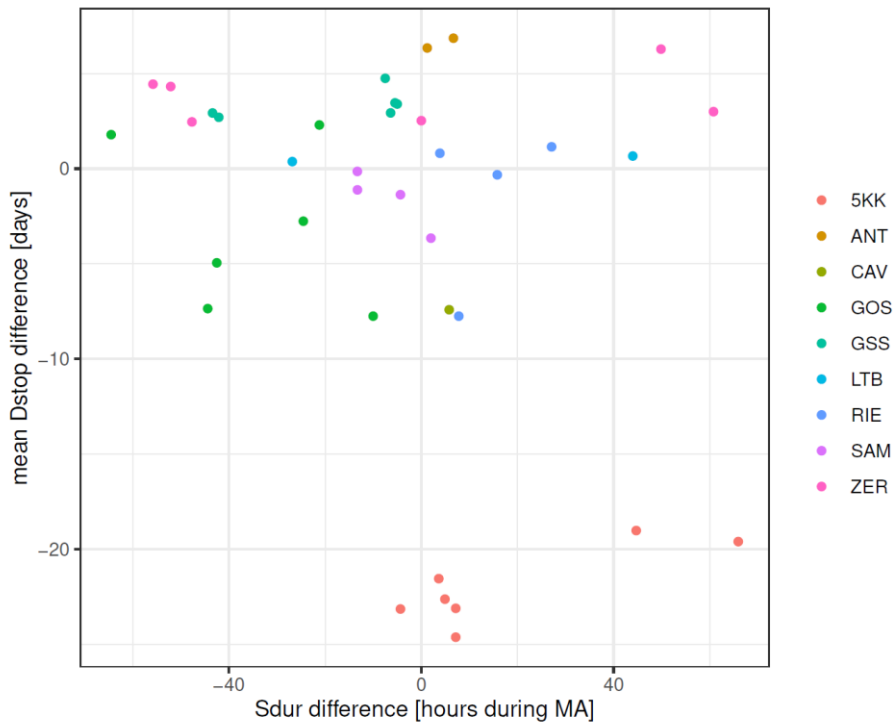
4.4 Influence of potential sunshine duration on snow disappearance dates

260 ~~To look~~When looking for a ~~possibility~~ possible explanation ~~to explain~~for the inter-pair differences in ~~Dstop~~Dend with available local-scale variables we compared them to differences in potential sunshine durations. Figure ~~A3~~6 displays the relationship of differences in ~~Dstop~~Dend and differences in potential sunshine duration during March and April for a selection of station pairs. The largest difference in a station pair ~~of~~ (64 hours during MA) amounts to approximately one hour per day. Mean and standard deviations of all unpaired combinations of this subset are -5 and 124 hours. Here we found no relationship between

265 potential sunshine duration and differences in ~~Dstop~~Dend. As outlined in section 2, our metadata is not perfect and although we limited this analysis to a selection of measurement sites with reliable coordinates, the influence of local scale obstacles cannot be detected with such an approach. As the accuracy is true for the 50 m scale, but not for 10 m, the conundrum "in front or behind a house" still remains. An example is given in Sect. 4.6.

~~We use the station pair Klosters to illustrate said issue: At a first glance, no striking difference is visible between the two~~

270 ~~stations. Elevation and surroundings are similar. One station is situated at a train station, the other next to a power station. However, when looking closer, it is revealed that the one at the train station is located on a first floor roof deck cutting. The second one sits just next (3 m) to the huge turbine house. Both stations are affected by their surroundings and corresponding specific exposure to wind, solar radiation and temperature is definitely different. Such differences are difficult or impossible to detect if the coordinates or metadata are not accurate. But even if these issues are known, measuring the actual impacts is~~

275 ~~virtually impractical.~~

280

4.5 Extreme value analyses

To test whether estimated extreme events based on annual maximum values of snow depth and snowfall differ significantly within station pairs, return levels for two indicators HSmax and HN3max are calculated for a 50-year return period using the NDJFMA season (Fig. 7). Inter-pair differences of the 50-year return levels are small (7-8% for both indicators). We found

285   that for all but two station pairs (both for HN3max), the return levels of the individual stations are usually within the 95% confidence intervals, indicated by error bars in FigureFig. 57. This suggests that in spite of obvious differences, in terms of return levels for 50-year return periods, the station pair values are similar or at least within each other's 95% confidence intervals.

Analysis of estimated 50-year return periods based on annual extreme values demonstrates that tThe relative differences of

290   return levels for HSmax and HN3max are less than 30% for all station pairs and less than 15% for 75% of the station pairs. For all but twothree station pairs these values are within each other's 95% confidence intervals, which is a good result for engineers oftensometimes may be useful for applications where there is normally only a single time series available only

depending on one single station series. On the other hand, the analysis clearly proves that just taking the return value of one station, without considering the estimation uncertainty, drastically limits the validity of the results.

295

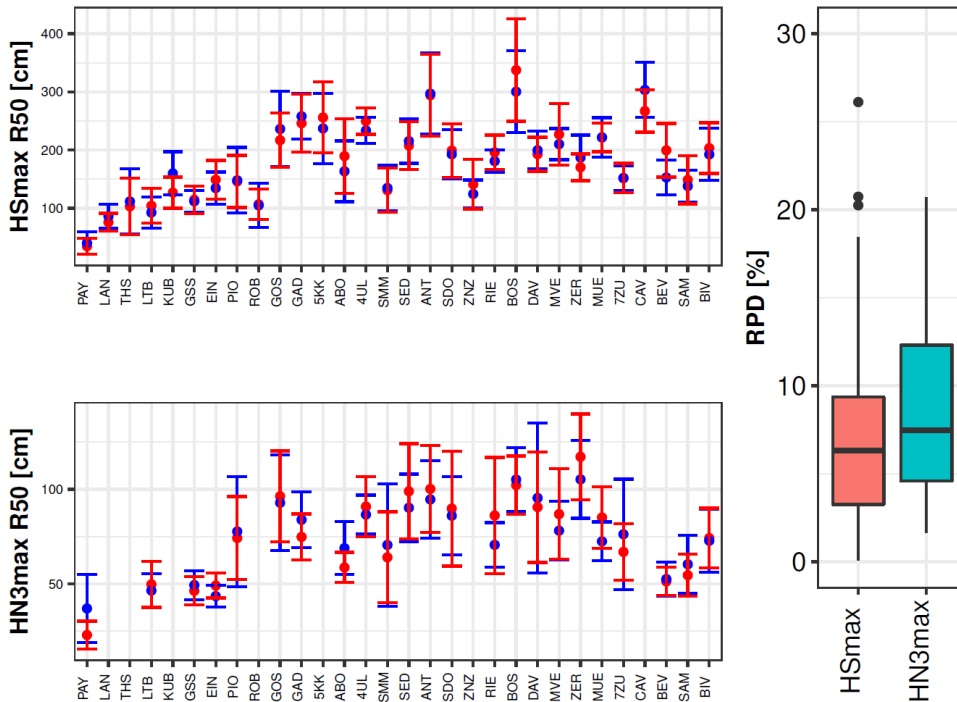



Figure 75: Return levels for seasonal maximum snow depth (HSmax) (top left) and maximum 3-day new snow sum (HN3max) (bottom left) for 50-year return periods. MeteoSwisseh (slfSLF) stations are coloured red (blue). The error bars indicate the 95% confidence intervals. Station pairs are ordered left to right from low to high elevation. Relative percentage differences (RPD) of
300 **return levels for HSmax and HN3max are depicted on the right.**

4.6 Metadata, possible explanations, and limitations

To illustrate and explain the issues associated with metadata, we focus on the station pair Klosters (5KR, 5KK). We use the station pair Klosters to illustrate said issue: Concerning metadata, Aat a first glance, no striking difference is visible between the two stations. Elevation and surroundings are similar. One station is situated at a train station, the other next to a power
305 station. However, when looking closer, it is revealed that the one at the train station is located on a first-floor roof deck cutting. The second one sits just 35 m awayto the east of from thea largehigh turbine house (with corresponding shading) at the power station.next (3 m) to the huge turbine house. Both stations are affected by their surroundings and corresponding specifie exposure to wind, solar radiation and temperature is definitely different. Such differences are difficult or impossible to detect if the coordinates or metadata are not accurate. But even if these issues are known, measuring the actual impacts is virtually
310 impractical.

Even though potential sunshine durations can vary on a local scale, especially in mountainous areas, the influence on specific stations is virtually impossible to determine. The influence of a tree or house in close proximity to a station (not only regarding sunshine duration) ~~on a station~~ remains an important but unaccountable factor, especially because the exact location of the snow measurement stake in the past is often not known. It is therefore not possible to attribute any local factors in the first place. For example, a measurement field in front of a south facing wall is influenced completely differently than a station located behind said wall on the north facing side. These limitations inhibit the simple use of terrain indicators to explain the variations. Furthermore, the influence of the observer on the snow onset or disappearance date cannot go unnoticed. Officially, the ground ought to be declared snow-free if at least 50 % of the measurement field is snow-free. However, as simple as the instruction may sound, the interpretation can vary and may easily account for a difference of a couple of days. Unfortunately, it is impossible to prove.

Ideally, station metadata would include the exact coordinates of the snow stake and panoramic pictures of its surroundings. Also, ~~frequent~~regular visits to sites and observers would be beneficial.

5 Conclusions

We presented the first assessment of local-scale ~~uncertainties~~ variability of common extremal indices and snow climate indicators based on a unique dataset of long-term parallel snow measurements. ~~Analysis of estimated 50-year return periods based on annual extreme values demonstrates that the relative differences of return levels for HSmax and HN3max are less than 15% for 75% of the station pairs. For all but two station pairs these values are within each other's 95% confidence intervals, which is a good result for engineers often only depending on one single station series. On the other hand, the analysis clearly proves that just taking the return value of one station, without considering the estimation uncertainty, drastically limits the validity of the result.~~

Our term variability or "local bias" is only valid for the parallel analysis (two point measurements). But even so, the results give an indication of possible variations for various indices.

Regarding common snow climate indicators, the results revealed relatively small median differences for the majority of the station pairs and indicators. However, spring months (MA) have the highest relative differences (16-30%) and show more inter-pair variation (8-43%) for all indicators than DJF (2-24%) or NDJFMA (3-25%). Relative differences for these two seasons varied between 3% (~~dHS1~~days with snow on the ground) and 20% (~~HSavg~~average snow depth) for all indicators, whereas ~~HSavg~~ average snow depth also demonstrated the largest spread of all indicators. ~~HSavg~~ Average snow depth also revealed in absolute terms the highest "local-bias" (ca. 10 cm) in spring. Additionally, inter-pair median differences of the snow disappearance date (per definition a proxy for spring) displayed again more variation and higher absolute values than the

snow onset date. However, the differences between the "local-bias" of the snow disappearance date (2 days) and the snow onset date (3 days) are small; Nevertheless, there is considerable variation for some station pairs.

345 As the station pairs are constructed with stations that are on average located within 1 km, the variations are most likely down to local-scale influences. This suggests that seasonal differences are likely caused by "local-bias", which in isolated cases can have ~~huge~~ large effects and that "local-bias" is probably amplified in MA (and Nov~~ember~~), likely because of the larger impact of radiation (and thus temperature). However, insufficient metadata prohibits further analyses on the influence of local-scale factors. But being able to quantify season dependent biases is important in itself as it increases our knowledge about existing inhomogeneities; especially in view of homogenization efforts of snow data series. The larger differences found in spring may

350 be an indication to preferably search for inhomogeneities (breaks) in these months or to determine correction factors separately for the ablation season.

The analysis of estimated 50-year return periods based on annual extreme values demonstrates~~,~~ that the return levels for HSmax and HN3max ~~are~~, for the large majority of ~~the~~ station pairs, are within each other's 95% confidence intervals~~,~~; corroborating that the knowledge ~~that~~of considering the uncertainty in return level estimation is indispensable.

355 ~~Even though potential sunshine durations can vary on a local scale, especially in mountainous areas, the influence on specific stations is virtually impossible to determine. The influence of a tree or house in close proximity (not only regarding sunshine duration) on a station remains an important but unaccountable factor, especially because the exact location of the snow measurement stake in the past is often not known. It is therefore not possible to attribute any local factors in the first place. For example, a measurement field in front of a south facing wall is influenced completely differently than a station located~~

360 ~~behind said wall on the north facing side. These limitations inhibit the simple use of terrain indicators to explain the variations. Furthermore, the influence of the observer on the snow onset or disappearance date cannot go unnoticed. Officially the ground ought to be declared snow-free, if at least 50 % of the measurement field is snow-free. However, as simple as the instruction may sound, the interpretation can vary and may easily account for a difference of a couple of days.~~

365 Generally, "local-bias" is often ~~negligible~~small and negligible for many applications, at least for the large majority of the stations.
The problem is, that stations exhibiting "local-bias" are not easily detectable in the record without the availability of parallel measurements.
~~The problem is that problematic stations are not easily detectable without parallel measurements.~~ Therefore, a larger number

370 of neighbouring stations are needed to find such problematic stations and to be able to develop stable homogenizations procedures. Assuming the same problem also exists in other parts of the world, implies ~~This means~~ that the current number of available long-term snow measurement sites ~~in Switzerland~~ should at least be maintained.

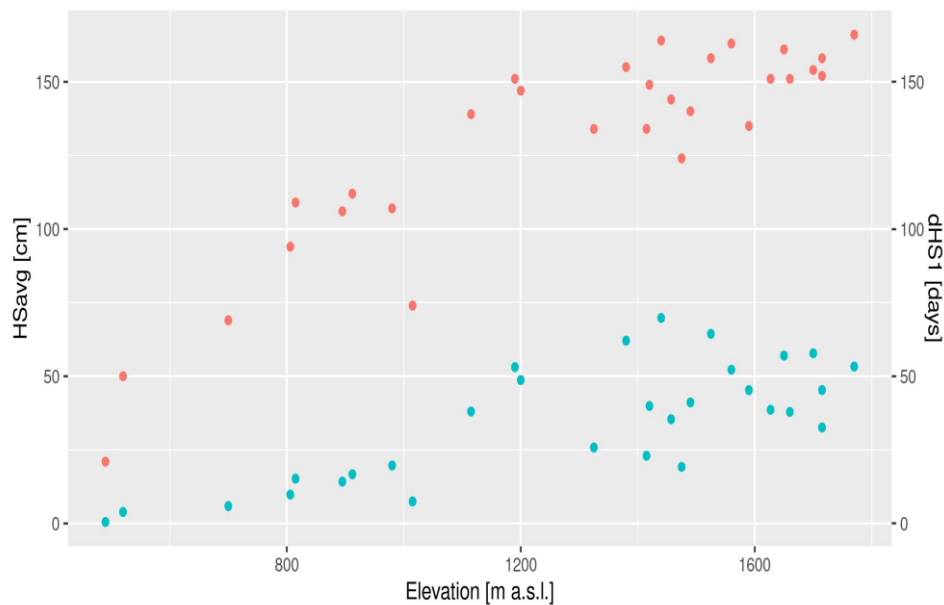
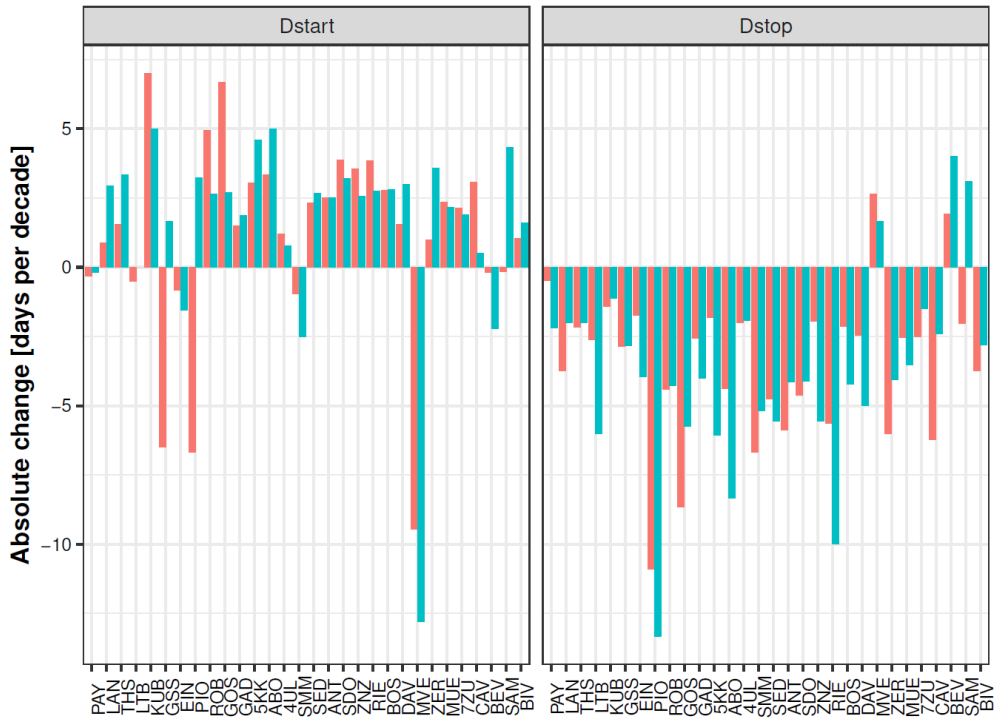
Appendix

375

Figure A 1: Shown ~~are~~ **in green** _are_ **mean values of average snow depth (HSavg) and** _shown_ **in red** _are_ **mean values of days with snow on the ground (dHS1) for each station pair and NDJFMA. Station pairs are ordered according to their elevation.**

380 Figure A 2: Absolute changes of snow onset (Dstart) and disappearance (Dstop) dates calculated for all available parallel periods. Station pairs are ordered according to elevation. Red and green bars indicate individual stations within a pair.

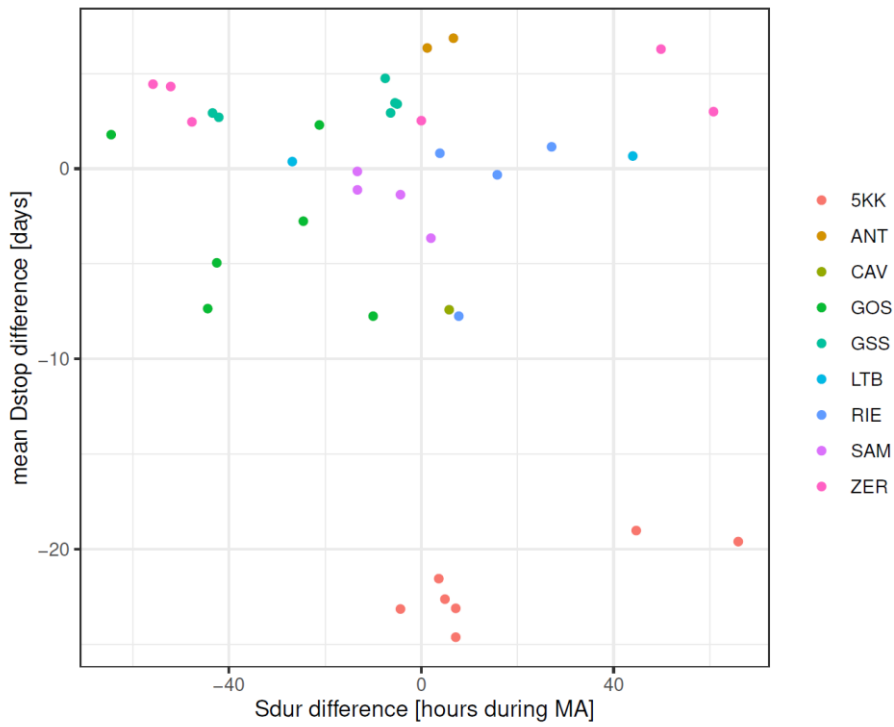



Figure A 3: Relationship between differences in snow disappearance dates (Dstop) and differences in potential sunshine durations (Sdur) for months March and April (MA), calculated for a selection of station pairs with manually checked coordinates during different time periods.

385

mch slfStation pair		start	stop	Elevation [m a.s.l.]	Length [years]	HN	mDstart	mDstopDend	mHSavg [cm]	mdHS1 [days}
PAY	PAV	1970	2018	490	48		15 Jan	26 Jan	0.5	21
LAN	5LQ	1969	2020	520	51	no	03 Jan	06 Feb	3.9	50
THS	5TH	1969	2020	700	51	no	25 Dec	19 Feb	5.9	69
LTB	1LB	1969	2020	806	51		20 Dec	05 Mar	9.8	94
KUB	5KU	1991	2020	815	29	no	07 Dec	15 Mar	15.2	109
GSS	EIN	1973	2020	895	47		18 Dec	09 Mar	14.2	106
EIN	SSE	1943	2020	912	77		13 Dec	12 Mar	16.7	112
PIO	6AM	1979	2003	980	24		18 Dec	20 Mar	19.7	107
ROB	7PV	1961	2020	1015	59	no	27 Dec	24 Feb	7.5	74
GOS	2GO	1969	2020	1115	51		03 Dec	06 Apr	38	139
GAD	1GA	1969	2019	1190	50		23 Nov	16 Apr	53.1	151
5KK	5KR	1968	2020	1200	52	no	22 Nov	15 Apr	48.7	147
ABO	1AD	1966	2013	1325	47		03 Dec	29 Mar	25.8	134
4UL	4MS	1950	2020	1380	70		17 Nov	21 Apr	62.1	155
SMM	7ST	1979	2012	1415	33		25 Nov	31 Mar	23	134
SED	5SE	1970	2020	1420	50		20 Nov	12 Apr	39.9	149
ANT	2AN	1967	2016	1440	49		14 Nov	01 May	69.8	164
SDO	5SP	1973	2020	1457	47		24 Nov	10 Apr	35.4	144
ZNZ	7ZN	1972	2020	1475	48		01 Dec	28 Mar	19.2	124
RIE	4WI	1975	2007	1490	32		26 Nov	10 Apr	41.1	140
BOS	6BG	1962	2014	1525	52		19 Nov	26 Apr	64.4	158
DAV	5DF	1966	2006	1560	40		14 Nov	27 Apr	52.2	163
MVE	4MO	1952	1978	1590	26		06 Dec	10 Apr	45.3	135
ZER	4ZE	1966	2004	1627	38		17 Nov	13 Apr	38.6	151
MUE	1MR	1972	2019	1650	47		20 Nov	29 Apr	57	161
7ZU	7SC	1951	2010	1660	59		16 Nov	14 Apr	37.9	151
CAV	7CA	1969	2020	1700	51	no	24 Nov	25 Apr	57.8	154
BEV	7SD	1951	1982	1715	31		12 Nov	19 Apr	45.3	158
SAM	7SD	1980	2020	1715	40		19 Nov	17 Apr	32.6	152
BIV	5BI	1969	2014	1770	45		11 Nov	30 Apr	53.3	166

Table A 1: Station pairs, available parallel periods, mean elevation, and number of parallel years. HN indicates whether a station pair was used for HN analyses. mDstart and mDstopDend are mean onset and disappearance dates, mHSavg indicates the mean average snow depth, and mdHS1 the mean number of days with snow on the ground. Station indicative starting with number refer to SLF stations and station indicative without a number refer to MeteoSwiss stations.

395

Station	Dstart_Klein	Dstart	Dstop_Klein	Dstop	Period
ANT	2.4	2.5	-5	-5.9	1967-2016
2AN		2.5		-4.2	1967-2016
BOS	2.4	2.8	-7.3	-2.1	1962-2014
6BG		2.8		-4.2	1962-2014
DAV	3.1	1.6	-4.4	-2.5	1966-2006
5DF		3		-5	1966-2006
SMM	2.1	-1	-4.2	-6.7	1979-2012
7ST		-2.5		-5.2	1979-2012

Table A 2: Absolute changes for Dstart and Dstop in days per decade. Dstart_Klein and Dstop_Klein are taken from Klein et al. (2016) and cover the period 1970 whereas the periods for our station (pairs) are indicated in the table.

Acknowledgement

400 We thank Johannes Aschauer (SLF) for his valuable work in sussing out the metadata mess and contribution to the data set. We thank Francesco Issota (MeteoSwiss) for providing the sunshine duration (Sdur) data set and Wolfgang Schöner and Gernot Resch for valuable commentsdiscussions. This work is financed by the SNF. Our thanks go to Craig Smith and two anonymous referees whose valuable comments helped to improve the manuscript.

Data availability

405 The basis for all our analyses, the indicator data set, is available herefrom EnviDat: https://doi.org/10.16904/envidat.218 (2021).

Author contributions

References

410 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, Nat. Hazards Earth Syst. Sci., 8, 685–705, https://doi.org/10.5194/nhess-8-685-2008, 2008.

Brown, R. D. and Robinson, D. A.: Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty, The Cryosphere, 5, 219–229, https://doi.org/10.5194/tc-5-219-2011, 2011.

415 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, Int. J. Climatol., 41, https://doi.org/10.1002/joc.6863, 2021.

Buchmann, Moritz, Aschauer, Johannes, Begert, Michael, and Marty, Christoph: Snow climate indicators derived from parallel manuel snow measurements, https://doi.org/10.16904/envidat.218, 2021.

Foster, J. L.: The Significance of the Date of Snow Disappearance on the Arctic Tundra as a Possible Indicator of Climate
420 Change, Arct. Alp. Res., 21, 60, https://doi.org/10.2307/1551517, 1989.

Grünewald, T. and Lehning, M.: Are flat-field snow depth measurements representative? A comparison of selected index sites with areal snow depth measurements at the small catchment scale, Hydrol. Process., 29, 1717–1728, https://doi.org/10.1002/hyp.10295, 2015.

Haberkorn, A. (Ed.): European Snow Booklet, 363 pp., 2019.

425 Kirdyanov, A., Hughes, M., Vaganov, E., Schweingruber, F., and Silkin, P.: The importance of early summer temperature and date of snow melt for tree growth in the Siberian Subarctic, Trees, 17, 61–69, https://doi.org/10.1007/s00468-002-0209-z, 2003.

Klein, G., Vitasse, Y., Rixen, C., Marty, C., and Rebetez, M.: Shorter snow cover duration since 1970 in the Swiss Alps due to earlier snowmelt more than to later snow onset, Clim. Change, 139, 637–649, https://doi.org/10.1007/s10584-016-1806-y,
430 2016.

Livensperger, C., Steltzer, H., Darrouzet-Nardi, A., Sullivan, P. F., Wallenstein, M., and Weintraub, M. N.: Earlier snowmelt and warming lead to earlier but not necessarily more plant growth, AoB Plants, 8, plw021, https://doi.org/10.1093/aobpla/plw021, 2016.

Marcolini, G., Bellin, A., Disse, M., and Chiogna, G.: Variability in snow depth time series in the Adige catchment, J. Hydrol.
435 Reg. Stud., 13, 240–254, https://doi.org/10.1016/j.ejrh.2017.08.007, 2017.

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

Peng, S., Piao, S., Ciais, P., Friedlingstein, P., Zhou, L., and Wang, T.: Change in snow phenology and its potential feedback to temperature in the Northern Hemisphere over the last three decades, Environ. Res. Lett., 8, 014008,
440 https://doi.org/10.1088/1748-9326/8/1/014008, 2013.

Roland, J., Filazzola, A., and Matter, S. F.: Spatial variation in early-winter snow cover determines local dynamics in a network of alpine butterfly populations, Ecography, 44, 334–343, https://doi.org/10.1111/ecog.05407, 2021.

Scherrer, S. C. and Appenzeller, C.: Swiss Alpine snow pack variability: major patterns and links to local climate and large-scale flow, Clim. Res., 32, 187–199, https://doi.org/10.3354/cr032187, 2006.

445 Sen PK: Estimates of the regression coefficient based on Kendall's tau, J Am Stat Assoc, 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, J. Geophys. Res. 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 Math Sci, 53, 1950.

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

World Meteorological Organization, World Meteorological Organization, and Organisation Météorologique Mondiale: Guide to meteorological instruments and methods of observation, WMO, Geneva, 2018.

Zampieri, M., Scoccimarro, E., Gualdi, S., and Navarra, A.: Observed shift towards earlier spring discharge in the main Alpine
455 rivers, Sci. Total Environ., 503–504, 222–232, https://doi.org/10.1016/j.scitotenv.2014.06.036, 2015.