How does the homogenisation of snow measurements impact snow climatology in the Alps? (Hom4Snow)

Inaugural dissertation
of the Faculty of Science,

University of Bern

presented by

Moritz Buchmann

from Zurich

Supervisor of the doctoral thesis:

Prof. Dr. Stefan Brönnimann

Institute of Geography, University of Bern

Co-supervisor of the doctoral thesis:

Dr. Christoph Marty

WSL Institute for Snow and Avalanche Research, SLF Davos

source: https://doi.org/10.48549/3746 | downloaded: 26.4.2024

How does the homogenisation of snow measurements impact snow climatology in the Alps? (Hom4Snow)

Inaugural dissertation
of the Faculty of Science,

University of Bern

presented by

Moritz Buchmann

from Zurich

Supervisor of the doctoral thesis:

Prof. Dr. Stefan Brönnimann

Institute of Geography, University of Bern

Co-supervisor of the doctoral thesis:

Dr. Christoph Marty

WSL Institute for Snow and Avalanche Research, SLF Davos

Accepted by the Faculty of Science.

Bern, 17.8.2022

The Dean
Prof. Dr. Zoltan Balogh

This work is licensed under a CC-BY-NC 4.0 license. To view a copy of this license, please visit

https://creativecommons.org/licenses/by-nc/4.0.

Chapter I

Summary

How does the homogenisation of snow measurements impact snow climatology in the Alps?

The general principles of homogenisation are well known due to the numerous studies covering the homogenisation of temperature and precipitation. However, to date, there has been no attempt to tackle the homogenisation of the Swiss snow series. To tackle this problem, we must first establish which is the best variable to use when looking for breakpoints. This allows us to develop the toolboxes and methods to be able to then apply established homogenisation methods to snow time series.

During this thesis, the three steps of homogenisation (identify breakpoints in a time series, verify the breakpoints, and adjust the time series accordingly) are carried out in detail.

Before we could check the series for breakpoints, knowledge about which variable to use had to be acquired. We did this by using a unique purpose-built parallel snow data set. In the initial analysis of this data set, we used various snow climate indicators and found the number of days with snow on the ground (dHS1) to be more robust than the seasonal mean snow depth (Hsavg). This analysis was carried out regardless of existing breakpoints in the series, which has implications for future climatological studies, where the homogeneity of a network or single series cannot be adequately tested. For the first time, we quantified the local-scale variability of the various snow indicators. These snow indicators included Hsavg, dHS1, and return values for 50-year return periods of seasonal maximum snow depth, as well as snow onset (Dstart) and disappearance dates (Dstop). The latter three parameters have particular relevance for engineering and ecological studies, respectively. We found that seasonality is important, and most of the variability can be associated with the spring months.

For the breakpoint analysis, we selected Hsavg (as used by other existing snow studies) and dHS1 as the most robust of the snow variables. We used three well-established and state-of-the-art homogenisation methods for the break detection task. On the one hand, this enabled us to compare the methods when applied to a snow series. On the other hand, using the three independent methods allowed for a more robust approach, as for breakpoints to be interpreted valid, they had to be identified by at least two of the three methods. This is especially important as there is no ground truth where real-world observational are used. Incomplete or inaccurate metadata is quite common. Therefore, events recorded in the station history, as documented in the metadata can, but do not have to, lead to breakpoints. Combining the breakpoints identified

using either Hsavg and dHS1, we identified 45 valid breakpoints in 41 of the 184 investigated Swiss snow series. It is evident from the available metadata, that the majority of these breakpoints can be attributed to station relocations.

Having identified and verified the breakpoints in the Swiss snow network, the affected series can be adjusted. Here we used two established homogenisation methods (HOMER and Climatol) which allow the input of a manually adjusted break file, and a newly developed quantile matching method (interpQM). InterpQM yields daily homogenised values, whereas Climatol and HOMER deliver monthly homogenised values in our analysis framework. Running HOMER and Climatol with daily input data is computationally not feasible with the entire network. As the automation of these methods is one of their core advantages, the required pre-treatment (i.e. manual network building to reduce the number of possible stations to be able to deliver daily homogenised values) would remove that advantage. To answer the question; "how does the homogenisation of snow series impact snow climatology in the Alps?", trends are calculated for both the original homogeneous series and the subset of inhomogeneous and subsequently homogenised series. Here we found that trends for HSavg become more pronounced and the number of significant trends increased for the homogenised series. However, the variability in the series remained roughly the same. The only difference is that for stations between 1450 and 1650 m a.s.l., trends are no longer significant for all three homogenisation methods.

Dealing with snow depth and a diverse network (in terms of topography, climatic region, amount of snow, elevation, and micro-climate), elevation dependence is an ongoing issue in a mountainous study region, such as Switzerland. Network density, a crucial factor for the availability of suitable reference series, is also dependent on elevation as not many stations or available records exist below 500 and above 2000 m a.s.l., for a variety of reasons. Stations around and below roughly 500 m a.s.l. typically do not experience a continuous winter snow cover and were excluded from the analysis. The interesting elevation band of 2000 m a.s.l. and above is scarcely populated with manual snow stations due to the feasibility of manual support during the snowy winter months. The exception is for a few mountain resorts and cable car stations, however, as these places are normally only manned during the operational season of the resort, other issues can occur (i.e. no data in November).

One of the challenging issues encountered during the course of this thesis is the lack of ground truth via any reference data. This could be be in the context of analysing parallel measurements, breakpoints in a series, or the comparison between raw and homogenised data. Nevertheless, by quantifying the variability of snow climate indicators, using a robust approach for the detection of breakpoints, and comparing various homogenisation methods, the issue of uncertainty or variability is addressed, studied, and accounted for.

Acknowledgements

Special thanks go to:

... Amy Macfarlane, Janet Prevey, and John Coll for proofreading and scrutinising various sections of this manuscript

... Christoph Marty for his general support and supervision

... Stefan Brönnimann for his support and trust in the remote-supervision of this thesis

... Charles Fierz, Hansueli Rhyner and the past and current members of the winter sports and climate group

... Gernot Resch, Barbara Chimani, Christoph Marty, Johannes Aschauer, Michael Begert, and Wolfgang Schöner for the collaboration and discussions in the Hom4Snow project

... Gernot Resch for his contributions to manuscript no. 4

... Robert Kenner for the good company in our C148, for encouraging me to 'learn' ArcGIS myself, and the many off-topic discussions

... Urs Odermatt for providing the opportunities to broaden my skill set outside of SLF

... Lucie, Elisabeth, Amy, Gregor, Jürg T, Lukas, Kurt, Beni, Celia, Chris, Thomi, Frank, Matthias, Nicole, Dani, Stefan, Mark and Janet for their valuable contributions to the ClubMeteo series

... Jürg Schweizer for his support of the ClubMeteo series

... the front desk, maintenance, back office, and housekeeping teams at SLF for providing a functional and clean workplace and helpful answers to various questions and requests

This thesis was funded by the Swiss National Science Foundation SNF (grant no. 175920).

Contents

I Summary **I**

1 Introduction and structure **1**
1.1 Background and motivation 2
1.1.1 Inhomogeneities 3
1.1.2 Hom4Snow 4
1.1.3 Measurements and networks 5
1.1.4 Research questions and specific objectives 6

2 Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series **7**

3 Local-scale variability of seasonal mean and extreme values of in situ snow depth and snowfall measurements **23**

4 Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-)automatic homogenization methods **37**

5 Impact of homogenisation on extreme snow depth and decadal trends of snow days, mean and maximum snow depth: lessons from the Alps **59**

6 Conclusion and outlook **83**
6.1 Paper-focused summaries 84
6.2 Conclusion 85
6.3 Open issues and outlook 86

7 Literature **87**

A A quantile-based approach to improve homogenization of snow depth time series **97**

B Unique set of parallel Swiss snow series - a draft **115**

Chapter 1

Introduction and structure

Chapter 1 sets the scene, describes the project, presents a short data overview, and formulates the research questions. Chapters 2, 3, and 4 are the published articles. Chapter 5 is an article in preparation for submission. Chapter 6 summarises the work, draws conclusions, and provides an outlook. Chapter A is a co-authored paper relevant for the Hom4Snow project. Chapter B is a draft manuscript for a data paper covering the parallel data set.

1.1 Background and motivation

Snow is an integral part of the global climate system and hydrological cycle, and has many regional socio-economic implications and important physical properties (Armstrong and Brun, 2009). Its high albedo influences the global energy balance through the snow-albedo-feedback and decreases the local temperatures over snow-covered ground (e.g. Scherrer et al., 2012; Winter et al., 2017), whereas its thermal insulation enables plants and animals to survive under snow cover during the winter (e.g. Wipf et al., 2009). Dependent on a complex interplay between meteorological (temperature and precipitation) and topographical parameters, snow depth is more than a mere combination of the two (e.g. Gray and Male, 1981). The importance of snow depth is further reflected by it being one of the Essential Climate Variables (ECV) defined by the World Meteorological Organization (WMO) and its use as ground truth in model verification and forecasts (e.g. Brown et al., 2003; Essery et al., 2013).

For an Alpine country such as Switzerland, snow is of particular importance as it influences freshwater availability (e.g. Beniston et al., 2011), hydropower (Hänggi and Weingartner, 2012, e.g.), ecology (e.g. Jonas et al., 2008; Vorkauf et al., 2021), winter tourism (e.g. Beniston et al., 2018; Steger et al., 2012; Schmucki et al., 2017), and natural hazards such as avalanches and floods (e.g. Beniston, 2003; Beniston et al., 2011). In all these fields, there is a need for accurate and good quality data in order to understand the underlying processes and predict future states.

Long-term snow measurements are paramount for climate monitoring and climatological analyses (e.g. Aguilar et al., 2018; Venema et al., 2020). Snow climatology is well-studied (e.g. Laternser and Schneebeli, 2003; Marty, 2008; Scherrer et al., 2013; Schöner et al., 2019; Matiu et al., 2021; Brown et al., 2021), as are trends of other derived variables such as snow onset and disappearance dates (e.g. Klein et al., 2016). However, non-climatic shifts and their possible impacts have received less attention.

1.1.1 Inhomogeneities

As with every other meteorological variable, manual snow measurements are prone to inhomogeneities (shifts or breakpoints) introduced either naturally (climate change) or artificially by station relocation, observer change, change of the station environment (plant growth or urbanisation) or a combination of the above. Instrument change or automation are also potential causes for breakpoints. However, this thesis only focuses on manual measurements with no changes in instruments since the early 20th century. Knowledge about such inhomogeneities is key, because not every event is recorded in the station history, nor does every recorded event necessarily lead to a detectable breakpoint within the series. The existence of non-climatic inhomogeneities is especially important for climatological analyses such as trends or analysis of extreme values, where shifts or breaks within the series can have dramatic effects on the results.

To only focus on non-climatic shifts, relative breakpoint detection tests are used where the difference or ratio of two series (candidate and reference) is analysed for breakpoints, thus eliminating any climate signals (in contrast to absolute tests which focus on one series alone).

Homogenisation is the step-wise process of finding and correcting these breakpoints and is a well-established field in climate sciences with numerous contributions to temperature (e.g. Begert et al., 2005; Kuglitsch et al., 2012) and precipitation data sets (e.g. Kuya et al., 2021a; Scherrer et al., 2013; Guenzi et al., 2020; Coll et al., 2020). However, homogenisation has rarely been used when evaluating large snow data sets (Marcolini et al., 2017a, 2019; Schöner et al., 2019). Many homogenisation methods are available, most of them developed with temperature and precipitation in mind (WMO, 2017). However, as snow can be thought of as a (albeit complex) combination of temperature and precipitation, homogenisation should be feasible with the existing methods. First, a suitable variable is selected and then the series is checked for inhomogeneities or breakpoints. These identified breakpoints are verified and eventually adjusted (using suitable reference series), leading to homogenised series, which are not the same as homogeneous series (Peterson et al., 1998).

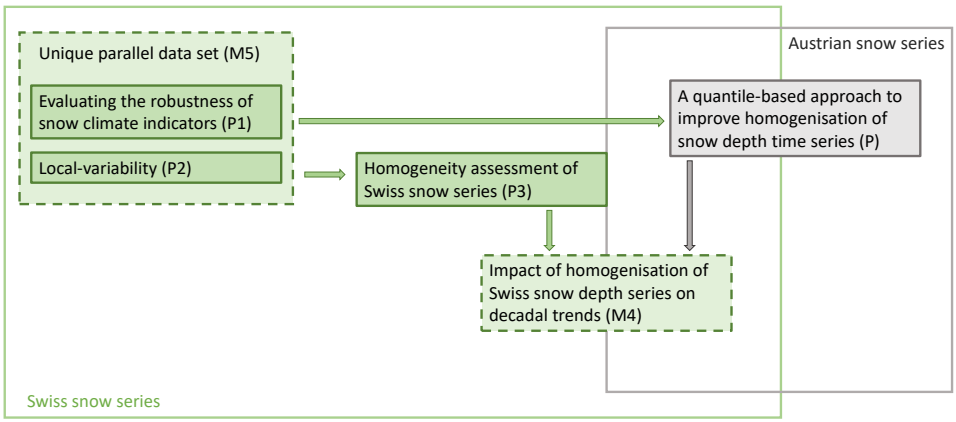


Figure 1.1: Shown here are the research articles of the Hom4Snow project (Austrian and Swiss part) and their relationship. First-authored articles belonging to this thesis are coloured green, numbered and marked as either published (Px) or as manuscript in preparation (Mx). Green and grey colours indicate Swiss and Austrian data sources respectively.

1.1.2 Hom4Snow

The Hom4Snow project (SNF grant Nr 175920) aims to investigate the impact of homogenisation of snow measurements on snow climatology (trends) in the Alps. Figure 1.1 shows the relationship of the various components of the Hom4Snow project. As an integral part of the project, this thesis contributes towards that aim by providing steps (a), (b), (c), (e), and (f) from the following list:

(a) Preparatory groundwork (Buchmann et al., 2020): compiling a set of parallel snow measurements and analysing the robustness of derived snow climate indicators given their suitability for breakpoint detection and homogenisation (first step of the homogenisation process).

(b) In-depth analyses (Buchmann et al., 2021) of local-scale variability of in-situ snow measurements using high-quality parallel series. The quantitative assessment of the natural variability associated with the snow measurements allows for a better evaluation of the eventually homogenised snow series.

(c) Proposing and applying a robust approach for the homogeneity assessment (breakpoint detection) of Swiss snow series (Buchmann et al., 2022a) (second step in the homogenisation process).

(d) InterpQM (Resch et al., 2022), a new method to homogenise snow depth time series using a quantile matching approach (third step).

(e) Impact assessment: Investigating the impacts of various homogenisation methods on trends and extreme values (Buchmann et al., 2022b) (fourth step).

(f) Furthermore, a detailed description of the compilation of the unique parallel data set - used in steps (a), (b) and (d) - is supplied as a draft manuscript in the appendix (Chapter B).

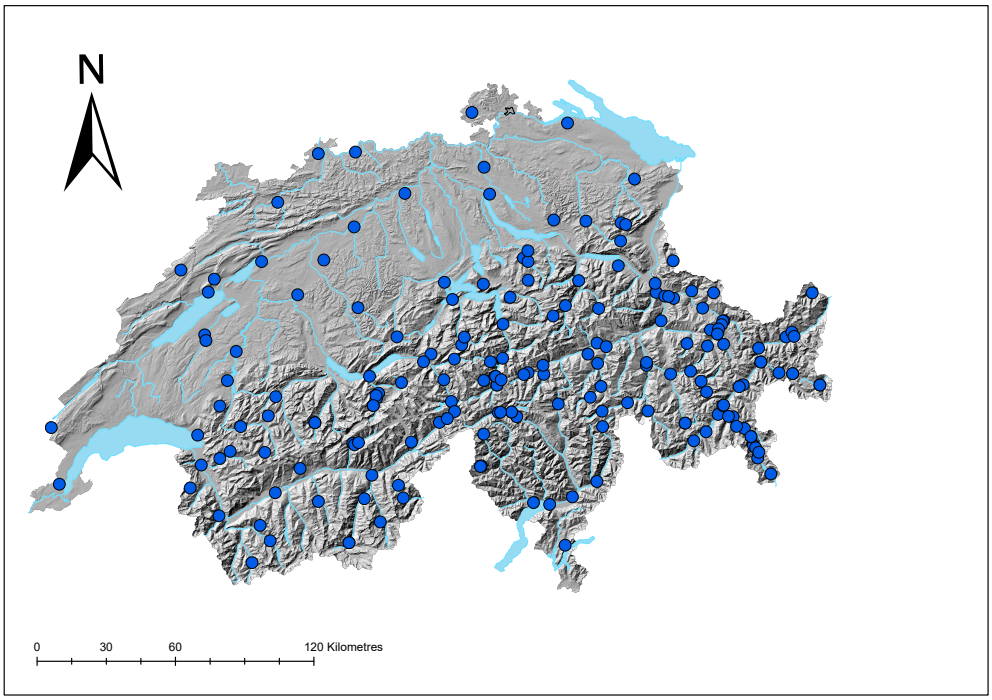


Figure 1.2: Manual snow measurements in Switzerland: The map shows the spatial distribution of the stations used and analysed in this thesis.

1.1.3 Measurements and networks

Snow depth is an essential climate variable that is monitored on a global scale by programmes such as GCW, GCOS and measured either manually (high quality, point measurement), automatic (point measurement) or remotely with various accuracy and spatial resolutions, with the manual measurements treated as ground truth and best-quality observations. Snow is measured as snow depth (HS) or height of new snow (HN). Until today, HS and HN are both, measured manually at around 200 locations in Switzerland. Figure 1.2 shows the spatial distribution of the stations used. The vast majority of manual snow stations in Switzerland are maintained by the Federal Office of Meteorology and Climatology (MeteoSwiss) or the WSL Institute for Snow and Avalanche Research (SLF). The various focuses of the two institutions are evident when looking at the spatial distribution of their networks. The SLF's priority is first and foremost the assessment of the avalanche situation (seasonal observations between November and April), in contrast to MeteoSwiss' focus on providing weather forecasts all year round. MeteoSwiss' stations are distributed over the entire country, covering valleys and the main plateau, whereas SLF-stations are predominantly situated in mountainous areas and ski resorts with seasonal observations. A detailed description of the two measurement networks used in this thesis can be found in Haberkorn (2019) as well as in Chapter B or the introductions in Buchmann et al. (2020) and Buchmann et al. (2021).

1.1.4 Research questions and specific objectives

We aim to answer the main question; *How does the homogenisation of snow measurements impact snow climatology in the alps?* This was broken down into the following research questions addressed in the following articles.

Paper 1: Robustness of snow climate indicators using parallel measurements

1. Which snow climate indicators are most robust for climate analyses?
2. Are there any elevation dependencies?
3. Study various error metrics including mean absolute deviations and relative changes

Paper 2: Local-scale variability of in-situ snow depth and snowfall measurements

1. Is the behaviour of the local bias a function of time series length; i.e. are relative percentage deviations dependent on the length of parallel series?
2. Is there any seasonal dependence of the local bias?
3. What is the impact of the local bias on return levels based on a commonly calculated return period of extreme events?

Paper 3: Homogeneity assessment of Swiss snow series

1. Which method or set-up works for breakpoint detection in Swiss snow depth time series?
2. Is there any elevation dependence affecting the capability of the methods for breakpoint detection?
3. How do results from the detection methods agree with available metadata?
4. How homogeneous (in terms of detected breakpoints) are the Swiss snow depth series investigated in our approach?
5. How does the use of different variables (average snow depth and days with snow cover) affect breakpoint detection?

Manuscript 4: Impact of homogenisation on trends (climatology)

1. How do the homogenised subset series compare across the three methods (Climatol, HOMER, and InterpQM)?
2. What is the impact on decadal trends of average snow depth?
3. How do the various homogenisation methods impact important metrics such as snow cover duration and maximum snow depths?
4. To what extent are maximum snow depths with a 50-year return period affected?

Manuscript 5: Data paper - Unique parallel snow data set

1. Description of the difficulties encountered and solutions applied during the preparation of the parallel data set
2. Provide a publicly available parallel snow data set for further analyses and use cases

Chapter 2

Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series

Status: Published in International Journal of Climatology https://doi.org/10.1002/joc.6863
Author contributions: The study was devised by CM, MBT, MBU, and SB. Data were compiled by CM, MBU with input from MBT. MBU performed the analysis, produced the figures and wrote the draft. MBU revised and wrote the final manuscript with inputs from CM, MBT, and SB.

This article investigates the robustness of popular snow climate indicators by analysing a purpose-built set of parallel snow measurements. Under the assumption that not all series are perfectly homogeneous, the aim is to see which of these indicators are more robust or less affected by inhomogeneities. Error metrics such as mean absolute deviations, root mean squared errors, and relative changes are used.

Received: 17 April 2020 | Revised: 8 September 2020 | Accepted: 29 September 2020 | Published on: 9 October 2020

DOI: 10.1002/joc.6863

RESEARCH ARTICLE

International Journal of Climatology RMetS

Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series

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

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

[2]Oeschger Centre for Climate Change Research, University of Bern, Switzerland

[3]Climate Division, Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

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

Correspondence
Moritz Buchmann, SLF Davos, Flüelastrasse 11, CH-7260 Davos, Switzerland,
Email: moritz.buchmann@slf.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: SNF:200021L_175920

Abstract

Snow on the ground is an important climate variable which is normally measured either as snow depth or height of new snow. Like any other meteorological variable, manually measured snow is prone to local influences, changes in the environment or procedure of the measurements. In order to investigate the robustness of snow measurement series towards such non-climatic changes, a unique set of parallel manual snow measurements over 25 years from 23 station pairs between 490 and 1800 m a.s.l. was compiled. A sensitivity analysis based on typical snow climate indicators (e.g., mean snow depth, sum of new snow) from these parallel time series was carried out to find the most robust snow climate indicators for climatological analyses. Results show that there are only small differences in the sensitivity of the various snow climate indicators with regards to local changes. However, the indicators number of days with snow on the ground as well as the maximum snow depth are least affected by local influences and changes at station level. Median values of all station pairs reveal relative differences of about 7% for the number of days with snow cover and 11–16% for all other indicators. However, in extreme cases, the deviations within a single station pair can reach 25–40%.

KEYWORDS

climate, climate indicator, parallel time series, snow, snow measurements

1 | INTRODUCTION

Snow has multiple implications for a wide range of areas like ecology, economy and society: ranging from plants, animals, habitats and cycles of life (Jonas *et al.*, 2008; Wipf *et al.*, 2009; Resano-Mayor *et al.*, 2019) to winter tourism, hydro power, fresh water availability, floods, avalanches and climate feedbacks (Marty, 2008; Scherrer *et al.*, 2012; Marcolini *et al.*, 2017b; Schmucki *et al.*, 2017). Modern climatological studies of past snow trends in the Alps have a relatively short history in science with first studies starting in the 1990s (Beniston *et al.*, 1994; Spreitzhofer, 1999). Since then, various studies have focused on a variety of aspects (such as trends, variations, or forecasts and modelling) in all the above-mentioned fields, although there are still questions regarding the quality and the representativeness of the actual measurements and corresponding climatological time series.

The longer a time series, the more likely it has experienced breaks due to changes of observer, location,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2020 The Authors *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

instrument or procedure. Such breaks can affect the quality of the data as mentioned by Aguilar et al. (2003) and Della-Marta and Wanner (2006). This is a fundamental issue in climate sciences which has been well-studied for temperature and precipitation, (e.g., Begert *et al*., 2005; Scherrer *et al*., 2013; Acquaotta *et al*., 2019; Guenzi *et al*., 2020) but very rarely addressed for snow (Marcolini *et al*., 2017a). It is widely accepted that "snow" measured in either height of new snow (HN) or snow depth (HS) can be spatially quite heterogeneous and the measurements are prone to local influences (e.g., Neumann *et al*., 2006; López-Moreno *et al*., 2015). Snow is not only a function of temperature and precipitation but also dependent on elevation, exposure to wind, solar irradiation, and to a large extent on its instability on a microstructural level (e.g., settlement, metamorphism). To address such issues, sensitivity analyses of parallel time series are paramount and have been conducted for temperature and precipitation (Acquaotta *et al*., 2016; Gubler *et al*., 2017; Hunziker *et al*., 2017). For snow only few studies exist (e.g., Accquaotta et al., 2015; Baronetti et al., 2019) analysing parallel snow series. However, data from two different measurement techniques were used (manual and automatic), large distances (up to 20 km) were allowed within their station pairs and no impact on indicator series was investigated.

The present study benefits from the fact that in Switzerland snow is monitored by two independent institutions (WSL Institute for Snow and Avalanche Research Davos SLF and Federal Office of Meteorology and Climatology MeteoSwiss) and that the measurement principle has not changed since the beginning. This circumstance allows the exploitation of a carefully constructed unique set of long-term, daily, manual data of independent, parallel snow measurements. The aim is first to investigate the sensitivity of indicators, derived from snow depth and new snow with regards to local changes (in either location or observer). Throughout this paper, the influence of such changes (environment, instructions, observer) is not accounted for because of lack of trustful metadata. Nevertheless, this is exactly the point of this study – to find the least sensitive indicators regardless of any changes that might have occurred during the analysed time period. This last point has practical implications, as normally neither observer metadata nor environmental changes of a station itself are completely documented.

The results of this analysis are used to assess which snow indicators are most robust for climate studies, that is, least sensitive to local changes. This has practical implications for any further homogenisation approaches as well as for climate services where recommendations about trend analyses of usually un-homogenised snow time series can be improved. This will be accomplished by introducing and analysing derived snow climate indicators, like mean snow depth, number of days with snowfall, etc. and assessing the stability by making use of the parallel long-term measurements.

This paper is organised as follows: Section 2 introduces the data set and outlines the statistical methods used for the analyses. Results are presented in Section 3, followed by a discussion in Section 4. Conclusions are drawn in Section 5.

2 | DATA AND METHODS

Switzerland consists of three major parts: Swiss Plateau in the north, Jura mountains in the west, and Prealps and Alps in the south. Topography and climate are consequently complex and diverse. Typical mean maximum annual snow depths range from 15 cm for the Swiss Plateau to 300 cm in the Alps, and 70–100 cm for inner Alpine valleys and pre-Alpine regions. Starting at an altitude of 1,200–1,500 m above sea level, precipitation during winter predominantly falls as snow, such that the area is often covered by a solid layer of snow for weeks, and even months at higher altitudes. Snowfall is relatively rare in the low-lying areas of western Switzerland (greater Geneva area) and northern Switzerland (greater Basel area) as well as in the lowland in the southern tip of Switzerland.

Daily operational, manual snow measurements usually entail at least two variables: height of new snow as a 24 h sum (HN) and snow depth (HS). HN has to be measured, as the difference in HS (retrieved, say in a 60 min interval) is usually not the amount of new snow accumulated in said interval due to settlement of the snow pack. In Switzerland, snow is measured since the late 19th century. Manual measurements, which are solely used in this study, are still conducted today, using basically the same instruments as in the beginning. Both MeteoSwiss and SLF maintain a network of manual snow observations, be it for slightly different objectives. For MeteoSwiss, as the Federal Office of Meteorology and Climatology, snow is one of many variables they are interested in and basically just one form of precipitation, in contrast to the SLF where snow and snowfall are important due to its main brief of avalanche forecasting. The station distribution of the two networks reflects that focus, as the MeteoSwiss-stations are located throughout Switzerland, whereas SLF-stations are distributed solely over the mountainous and alpine areas.

2.1 | Measurement procedures

Daily measurements are conducted each morning by reading off the value from a stake with centimetre scale (HS) and by taking three measurements with a ruler on

the snow measurement board (HN) both at 06:00 UTC (Haberkorn, 2019). Instructions vary slightly when it comes to reporting as SLF stations normally only measure between November and April (unless there is already or still snow present) in contrast to MeteoSwiss stations which measure all year round. Unlike MeteoSwiss, where HN smaller than 0.5 cm is recorded as zero, SLF allows to put in traces (fixed default value of 0.3 cm) for HN smaller than 0.5 cm. These traces have been subsequently set to zero in the data analysis in order to be consistent with MeteoSwiss procedures. Values are stored in two databases and have already been processed independently in terms of initial quality control and gap filling. But both data sets have been analysed again with a systematic manual quality control looking for gaps and implausibilities (see Section 2.2) prior to being used for the analyses in this study. However, according to MeteoSwiss and SLF experts, there is no publication available documenting the various QC methods.

2.2 | Parallel data set

Because of the various operational focuses, there are locations where both institutions operate (or have operated) an observer site within the same village at some period in time. To build the parallel station set, the candidates have to meet the following criteria: Roughly within the same village (±3 km), similar elevation (± 100 m), data from November to April, independent data (as we cannot exclude instances in which data were copied), and parallel measurement series for at least 20 consecutive years. Stations can, but do not have to be part of the same network.

For simplification, we define the hydrological year as follows:

$$hyear1980 := 1.10.1979\ to\ 30.9.1980 \tag{1}$$

and subsequently winter or snow season as the period from November to April.

Unfortunately, it is not just a question of comparing the coordinates from the two networks to build the station pairs. Stations can have two names and different coordinates but still be identical, due to the fact that the coordinates represent the meteorological station, rather than the actual location of the snow measurements, as they, for practical reasons often cannot be conducted too close to the measurement field of the automatic instruments. Additionally, in the past it was not deemed important to know the exact location of the snow measurements. Starting off with a rough list of possible station pairs, compiled on available metadata such as current elevation and location, each case had to be analysed in detail in order to make sure the pair was independent. A possible pair of two existing stations has then to be checked for independent data and overlap.

Sometimes a station pair only appears to be independent, but contains in fact exactly the same data. This could happen for the following reasons: In order to fill in missing values for one station, sometimes the data from the corresponding "partner" station was simply copied during one or multiple short periods of time. Or one observer reports to both networks separately. Moreover, the search for station pairs proved to be cumbersome as approximately one third of the MeteoSwiss stations was not digitally available and had gaps that needed careful treatment first.

Station metadata such as coordinates and observer names could sometimes help solving the conundrum. Unfortunately, these data are not always available and trustworthy, especially for past changes. Even if available, observer names such as "Swiss Border Force" are still vague (e.g., station Santa Maria 1970s). Additionally, each institution has its own data base and therefore different or no quality codes that inhibits a simple query to check whether some data were copied in the first place.

To address these issues, a simple quality control mechanism is introduced. Stations are treated as independent if more than 60% of the data are not equal (without counting zeros) for any given winter (empirical value, gained by visual analysis). The above criteria yield a preliminary set of more than 55 possible station pairs. By selecting only station pairs where its members have more than 80% of data available for each annual winter season the number of possible station pairs is further reduced. Considering the length of the overlap and the benefit of looking at the same period, a dataset of 23 station pairs (see supplementary material Table S1) results for the 25-year period between 1980 to 2004 (see map in Figure 1).

Data is subjected to quality control focusing on implausibilities (like for example, snow depth decrease of 50 cm within one day) and consistencies between HN and HS. All cases of possible implausibilities and inconsistencies were manually checked. It is important to mention that the number of such cases were rare, which is not surprising as most of the time series have already been checked by the data owners. Missing values were interpolated by manually fitting the evolution derived from the best correlated neighbouring station using median ratios (again, the occurrences were rare, as the 80% cut-off meant that only high-quality stations [almost complete series] were selected in the first place). After the interpolation process all analysed time series were complete. However, not all inconsistencies can be

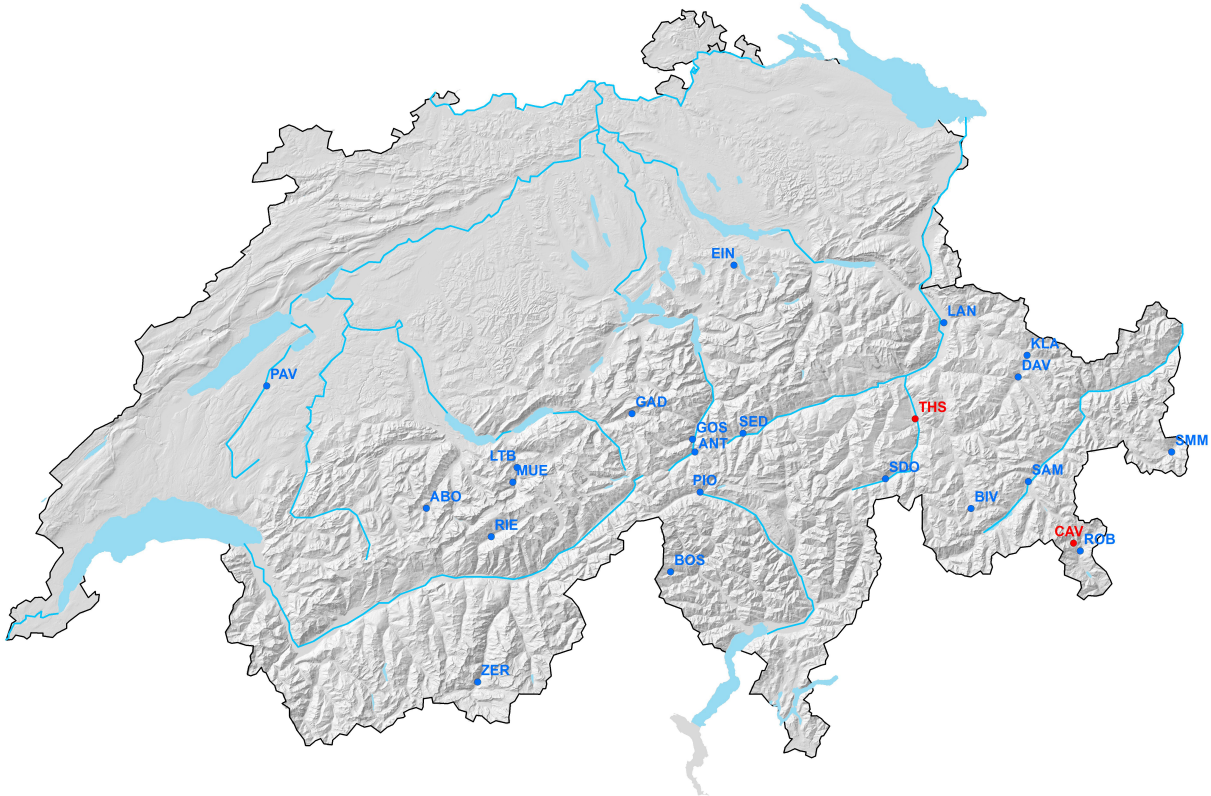


FIGURE 1 Map of Switzerland, showing the distribution of the station pairs used in this study. The red dots (THS and CAV) highlight station pairs that were excluded for the HN and trend analyses due to inadequate HN series

addressed as MeteoSwiss stations assign HN to the previous, rather than the actual day. This issue is historically partly addressed in the observer form, in the digitisation and/or in quality control process by simply moving the HN time series by one day. For these reasons this shift is unfortunately not always constant for all stations, time periods and sometimes not even an entire winter season. All possible combinations of no shift, two-day shift or shifts in the wrong direction can occur. Sometimes, even HS series have been affected. However, these cases could be easily detected and have been corrected accordingly. Fortunately, the HN shift does not affect the calculations, as the indicators are defined as annual values. For two stations the HN series had to be omitted, because HN was most of the time just calculated from the difference between today's and yesterday's snow depth (see red stations in map in Figure 1), resulting in 23 station pairs for HS- and 21 for HN-indicators. Hereafter, only the term "23 station pairs" is used for improved readability. Finally, all station pair time series were visually checked for possible remaining issues.

2.3 | Snow climate indicators

To be able to compare the station pairs and carry out the sensitivity analysis, snow climate indicators are introduced and defined as annual values from the daily HS and HN measurements similar to WMO's Global Climate Indicators; only with focus on snow and calculated for each station pair and hydrological year. Only years that have valid values for both stations are used to determine the snow climate indicators, which are:

Number of days with HS of at least 1 cm (dHS1) is widely used to establish whether the ground is snow-covered, an important factor for ecologists and climate scientists alike. To account for the fact that dHS1 might be sensitive to the observation time and method, dHS5 with a 5 cm threshold is introduced, as HS > =5 cm can still be regularly observed at all stations in the data set (see Section 4.3). The same applies to number of days with HN of at least 1 or 5 cm (dHN1 and dHN5) which are mainly of interest for tourism, road maintenance and climatology. The maximum sum of three consecutive

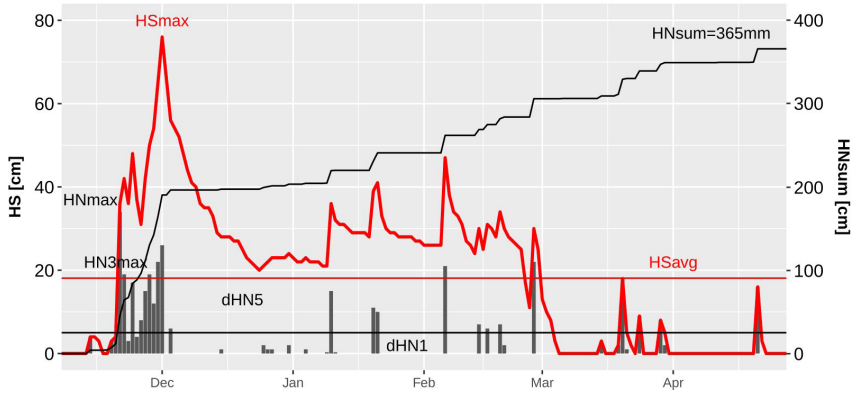


FIGURE 2 Winter 1997 at station Adelboden: The bold red line shows HS and the black bars indicate days with HN. All visible bars below the 5 cm threshold (black horizontal line) are counted as dHN1, all that exceed the 5 cm threshold as dHN5 (analogue for dHS1 and dHS5). The red horizontal line depicts HSavg. HSmax, HNmax, HN3max, and HNsum (shown as cumulative sum) are also indicated

TABLE 1 Summary of the snow climate indicators

Indicator	Description	Unit
HSavg	Mean HS	cm
HSmax	Max HS	cm
HNmax	Max HN	cm
HNsum	Sum HN	cm
HN3max	Max sum over three days	cm
dHS1	Number of days with HS > 0 cm	days
dHS5	Number of days with HS > =5 cm	days
dHN1	Number of days with HN > 0 cm	days
dHN5	Number of days with HN > =5 cm	days

snowfall days (HN3max) is of particular interest for avalanche forecasts, maximum snow depth (HSmax) has implications for national snow load codes and civil engineering, average snow depth (HSavg) is of interest for climate sciences, whereas the sum of new snow (HNsum) is more suited for hydrological or climatological analyses (refer to Figure 2 and Table 1).

2.4 | Statistical methods

2.4.1 | Correlations and absolute differences

Correlations (COR) between station pairs for all climate indicators are calculated using Spearman's rank coefficient as this method is less susceptible to outliers than Pearson's rho. Deviations within a station pair are expressed as relative percentage differences (RPD) between SLF (slf in the equations) and MeteoSwiss (mch) stations, because none of the two networks can be called a proper reference series.

$$RPD := \frac{mch - slf}{mean(mch, slf)} \tag{2}$$

RPD yield one annual value per station pair and indicator. Boxplots are used to compare the distribution of the RPD for all station pairs, separated for each snow climate indicator by using the mean RPD over all 25 years for each station pair.

To give a sense for the absolute scale of deviations, root mean squared errors (RMSE) are also calculated; always mch minus slf, analogue to RPD.

$$RMSE := \sqrt{mean\left((mch - slf)^2\right)} \tag{3}$$

2.4.2 | Median absolute deviations

In order to quantify the variability of the correlations and relative percentage deviations across all station pairs, median absolute deviations (MAD) are favoured over coefficients of variation due to their robustness with regards to outliers.

$$MAD := median(|X_i - median(X)|); with X \in \{COR, RPD, ..\} \tag{4}$$

2.4.3 | Relative changes

In order to assess the comparability of the long-term temporal evolution among station pairs changes are calculated using the Theil-Sen linear slope (Theil, 1950; Sen, 1968) and the non-parametric Mann-Kendal test (Mann, 1945; Kendall, 1975). Relative changes (RC) are defined as:

$$RC := \frac{\textit{fitted value at the end} - \textit{fitted value at the beginning}}{\textit{median of the fitted values}} \tag{5}$$

To investigate the differences within a station pair, the differences in the (above) relative changes (DRC) were calculated as:

$$DRC := RC(mch) - RC(slf) \tag{6}$$

3 | RESULTS

3.1 | COR, RPD, and RMSE

The median correlation values for each of the nine indicators range from 0.81 for HNmax to 0.95 for HSavg (Figure 3). However, median values are higher than 0.85 for all snow climate indicators except HNmax. HNmax also reveals by far the highest variation among all station pairs. In contrast, HSavg and HSmax, show by far the smallest variability. The outliers in Figure 3 consist of three different station pairs with no apparent connection.

Relative inter-pair deviations (Figure 4), expressed as median RPD for all station pairs range between 5% and 15%. The lowest values are clearly shown by dHS1 and dHS5, which together with HSmax also reveal small variation among the station pairs. On the other side, the largest variation result from dHN1 and dHN5. The outliers not shown in Figure 4 correspond to the lowest station pair in the data set. The visible outliers (HSmax, dHS1, and dHS5) relate to two snow-poor station pairs (PAV and ROB, see Table S1 for details).

In contrast to RPD, RMSE reveals the absolute deviation values (Figure 5). HNsum shows the largest RMSE with 24 cm and by far the largest variability. Median values for the other indicators range from 1.6 cm for HSmax to 3.5 cm for HN3max and from 2.2 days for dHN5 to 4.5 days for dHS5.

Because higher stations usually experience more snowfall, larger snow depths, and more snow days, possible elevation dependences were analysed for COR and RPD but no clear signal is found (see Figures S1–S4 and Table S2).

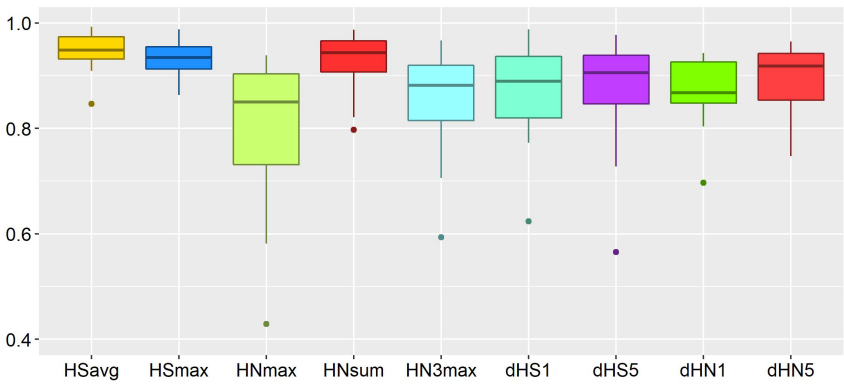


FIGURE 3 Spearman correlation for all station pairs grouped by snow climate indicators

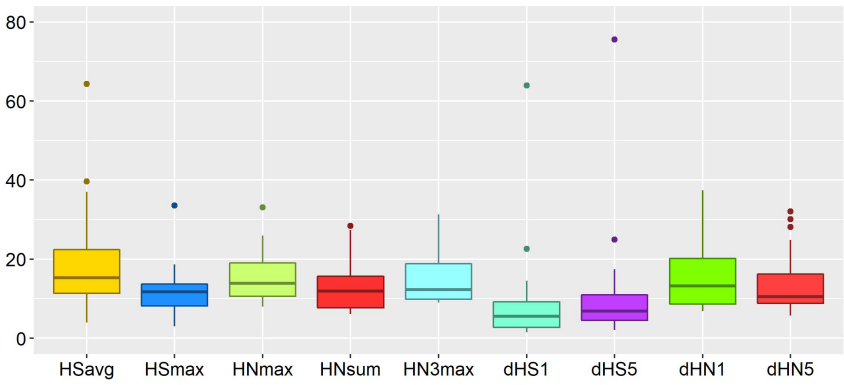


FIGURE 4 Comparison of mean RPD (in %) for each station pair and indicator. dHS1 and dHS5 both show median RPD smaller than 10%

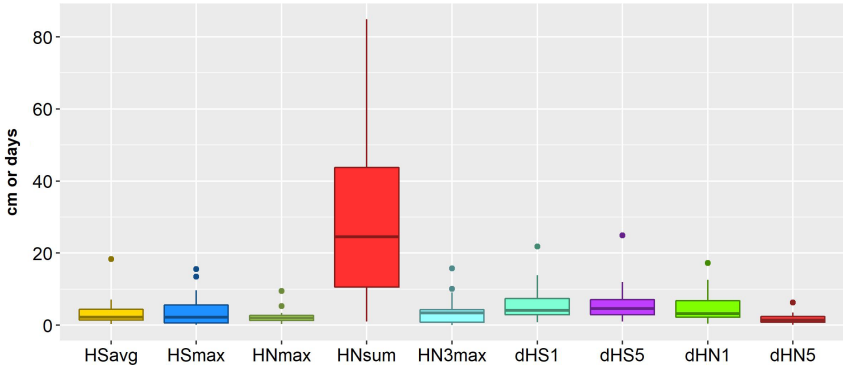


FIGURE 5 RMSE for the various snow climate indicators. Note that HNsum values are an order of magnitude larger than the rest and so is its RMSE

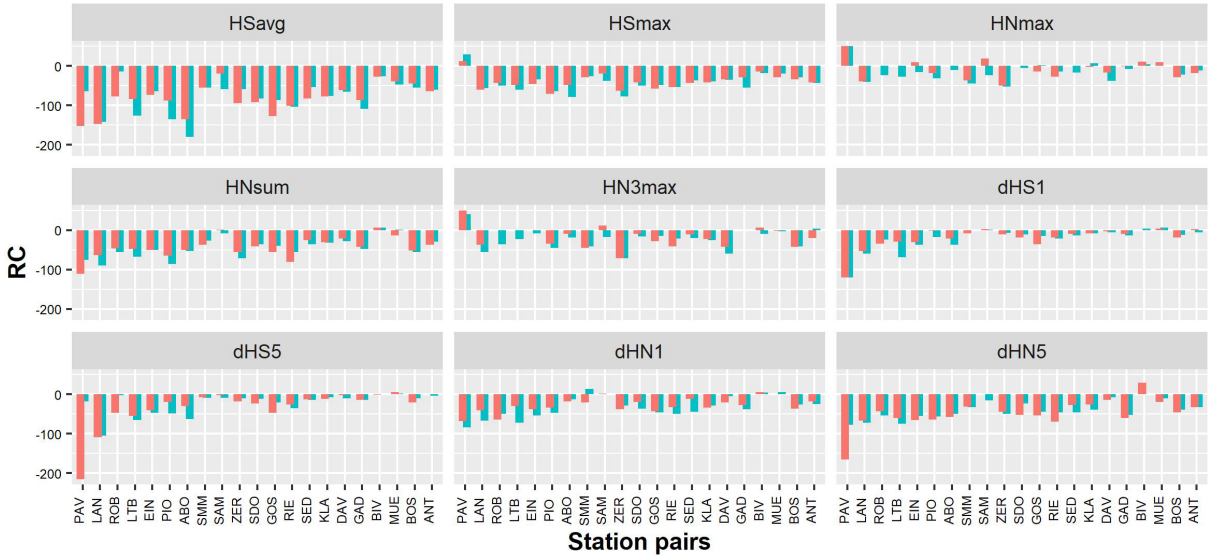


FIGURE 6 Relative changes (RC) for each station and indicator in %. Station pairs are ordered according to their mean HSavg from low (PAV) to high (ANT). Station names are defined in Table S1

3.2 | Relative changes

Figure 6 shows strength and direction of temporal RC for each station pair. Nearly all stations show negative RC for all indicators for the period under investigation. The sign of the changes agrees well within the station pairs; the only differences occur when one station has a value close to zero (see Figure 6 for details). The DRC (depicted in Figure 7) reveals median values for all station pairs below 10%. Furthermore, for a majority of stations and indicators, most values are below 20%, except for HNmax and dHN1. dHS1 and dHS5 show clearly the smallest values for the majority of stations, but several outliers which are associated with two snow-poor station pairs.

3.3 | MAD

Table 2 reveals the variability between the station pairs for COR, RPD, and DRC expressed as MAD in percentage values for all snow climate indicators. Larger values indicate a larger variability of the underlying metric and thus more variation across all station pairs (visible in Figures 3, 3, 4, and 7). MAD(COR) scores best for HSavg and HSmax with 4%. HNsum, dHS1, dHS5, dHN1, and dHN5 form the next group with values of 7%, followed by HNmax with 13%, illustrating the large spread in Figure 3. MAD(RPD) shows a more gradually distribution with values ranging from 4% to 11%. Again, HSmax, as well as dHS1 and dHS5 are associated with low values,

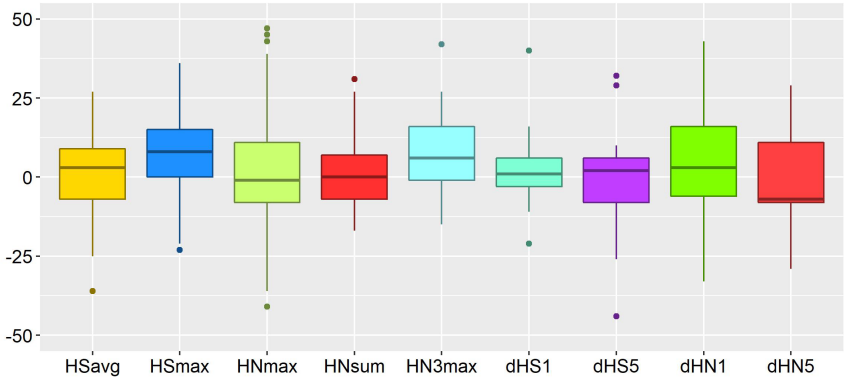


FIGURE 7 Differences in relative changes (DRC) for all snow climate indicators in %

TABLE 2 Comparison of the median absolute deviations for COR, RPD, and DRC for all snow climate indicators

Indicator	MAD COR [%]	MAD RPD [%]	MAD DRC [%]
HSavg	**3**	**7**	19
HSmax	**3**	**4**	**12**
HNmax	13	**7**	21
HNsum	**7**	8	**12**
HN3max	**7**	**4**	15
dHS1	**7**	**4**	**9**
dHS5	**7**	**4**	**9**
dHN1	**7**	8	18
dHN5	**7**	11	16

Note: Larger values indicate larger spreads (%) of the underlying metric. The two smallest values in each column are marked bold.

whereas here HNsum and dHN5 bring up the rear. MAD(DRC) values are generally higher than MAD(COR) and MAD(RPD) and like MAD(COR) separable into three groups: dHS1, dHS5 building the top with 9%, HSmax, HNsum, HN3max, and dHN5 in the middle with 12%–16% and dHN1, HSavg, and HNmax at the bottom with 19–21%.

4 | DISCUSSION

4.1 | Elevation dependences

The absence of any evident elevation dependences for COR, RPD, and DRC, determined with a simple linear regression (Figures S2-S4) and shown as coefficient of determination in Table S2, allows comparison and ranking of the snow climate indicators in order to find the most stable ones. However, due to the small sample size and unevenly distributed stations over the elevation range of our data set, a thorough analysis is not feasible. Having the largest RPD associated with the lowest station pair with the smallest absolute values is of no surprise and explained by the calculation of RPD itself.

The mere existence of elevation dependences for the mean indicator values (see Figure S1) does not come unexpectedly as one would expect more snow at higher elevations due to lower temperatures. However, the contrast with respect to elevation among the various snow climate indicators is interesting. HN-indicators are first and foremost precipitation dependent. Above a certain altitude as mentioned by Morán-Tejeda *et al.* (2013), the temperature is generally low enough for the occurrence of snow. HS-indicators on the other hand also require precipitation, but are much more dependent on temperature.

4.2 | COR, RPD, and RMSE

The COR values are similar to the ones retrieved by Acquaotta *et al.* (2015) for the comparison of two independent precipitation networks and generally quite high (median COR > 0.8). The overall weaker correlations for HNmax can be explained by the fact that HNmax is one single event per year and not necessarily recorded at the same date within a station pair. The small-scale nature of precipitation itself, the short lifespan of weak snowfall events, as well as local influences (shade, sun, exact measurement time) can also have a bearing.

When looking at RPD, the counting variables cover both ends of the spectrum: dHS1 and dHS5 display the smallest values, whereas dHN1 and dHN5 show high values and the largest variability. This can be explained by the overall greater absolute number of days with snow cover compared to days with snowfall. Small precipitation events coupled with varying time of observation and exposure of the actual measurement site have a greater impact on dHN1 which is visible in the larger spread of

dHN1 compared to dHN5. Overall, the variability in dHN1 and dHN5 can be attributed to small-scale variations in precipitation. The indicators HSavg, HNmax, HNsum, and HN3max show very similar values and distribution patterns, only HSmax has slightly smaller values than the others so a clear distinction is difficult. Other than HNmax, HSmax is not solely dependent on one heavy precipitation event, but also on the already existing snow pack. This is true for HSavg as well, however, HSavg might incorporate seasonal differences within a station pair due to varying exposure to sun light towards the end of the snow season, whereas HSmax does not. In terms of RMSE, the HN-derived indicators score slightly better than the HS-ones except HNsum. It is no surprise to see such high values for the RMSE of HNsum as HNsum itself is an order of magnitude greater than the rest of the indicators. However, the HN-indicators are generally smaller and therefore the small RMSE can be misleading as it is sensitive to the scale of the actual values. In general, HN-indicators consist of fewer and most of the time smaller values than HS-indicators. In conjunction with a measurable resolution of 1 cm (ruler, average of three measurements, round up to nearest cm) measurement uncertainties attributed to small values (mainly HN) may have larger absolute effects.

4.3 | Days with snow cover and days with snowfall

As an indicator for snow-covered ground dHS1 is widely used. Would a threshold of 5 cm be more stable when considering possible breaks in the long-term measurements? The reason for 5 cm rather than 10, 20, or 50 cm is that even at 400 m a.s.l. values of HS and HN greater than 5 cm are frequently observed, whereas a threshold of 20 or 50 cm would only cover very few events which would dramatically reduce the sample size and inhibit any analysis.

The indicator dHS1 is most likely skewed and certainly capped at the total number of days between November and April for snow rich stations and years (very few occasions in our data set). However, this only applies to stations at high elevations where there is either enough snow all season or the temperatures are cold enough to prevent the snow from melting. dHS5 does not necessarily solve this problem because the threshold of 5 cm is still regularly exceeded at higher elevation stations. However, this 5 cm margin addresses another issue in connection with small values of snowfall or snow depth; the sensitivity of the exact measurement time. Given the right conditions, a snowfall of 1 cm can easily melt or fall in 30 minutes and if the snowfall event takes place around the time of measurement it can be missed or recorded, solely dependent on the time the observer went out and conducted the measurements. Looking at the results, dHS1 and dHS5 are practically indistinct. Based on these results, there is no preference to either dHS1 or dHS5.

The same is true for dHN1 and dHN5; they both seem interchangeable when looking at COR, RPD, and RMSE. The only differences occur when looking at RC, where dHN5 (as well as dHS5) show stronger values than dHN1 (or dHS1) and RPD, where dHN1 shows a larger variability than dHN5, which is down to the temporal and spatial sensitivity of small precipitation events.

4.4 | Relative changes

RC only focus on a specific time window (1980–2004), which was chosen to maximise data availability. They serve mainly as additional information for inter-station-pair agreement and robustness. However, they generally agree with findings in Marty (2008) and Scherrer et al. (2013) of a decline in snow days in the Alps for the period after 1980. Within a station pair, the RC for various snow climate indicators are well in alignment with regards to the direction and strength, except for a few very weak RC that fluctuate around zero. Station pairs associated with large differences in RC for any indicator, visible as outliers in Figure 7, do not share any specific characteristics. For example, HSavg, which has the least similar RC per station pair, as reflected in the larger spread in Figure 7 and subsequently the largest MAD (Table 2). Low HSavg values normally imply less snow, which can particularly amplify the differences between stations at lower elevations where conditions can quickly change between freezing and melting, leading to short-lived snow appearances.

The stronger RC for dHS5 and dHN5 (compared to dHS1 and dHN1) highlight that a decrease in amount does not necessarily mean a decrease in frequency.

The overall smaller RC for HNmax can be explained by the large natural variability of the intensity of heavy precipitation events.

4.5 | Differences between Indicators

The memory effect of HS has an influence on persistence. Every HS at a specific day is to a certain amount dependent on the value of the day before and therefore more conservative, inertial, and stable in itself; whereas HN is a time series of (mostly) independent events. In contrast

to HSmax which is usually only dependent on accumulation, HSavg covers the entire season. Furthermore, HN-indicators such as HNmax, HN3max, dHN1 and dHN5 rather represent short and independent events that predominantly rely on precipitation which can vary on very small spatial scales; whereas HS-indicators are dependent on a combination of temperature and precipitation events therefore making them more inertial and apparently slightly more stable. However, HNsum and HN3max, which depend on a series of events as well showcase the most robust behaviour among the HN-indicators. Measurement uncertainties might have an impact as well, especially for small values (HN likely more affected than HS) as instruments and instructions do not permit for lower resolutions than 1 cm.

5 | CONCLUSIONS

When looking for the most robust indicators with regards to local changes, the following criteria have to be met: high COR, low RPD, small DRC, and low MAD for a majority of station pairs.

As the correlations for all indicators and a majority of station pairs is high (median COR > 0.8), only RPD, DRC and MAD are used to determine the most robust snow climate indicators. Low RPD are found in HSmax, dHS1, dHS5, and HNsum (closely followed by HN3max). Taking DRC into consideration has little effect as the values are similar for the snow climate indicators in question. The variability of the various RPD and DRC is quantifiable with the robust measure MAD (see Section 3.3 and Table 2). Smaller values indicate smaller variability and therefore imply a more robust indicator. Generally, DRC show larger MAD (Figure 3, 4, and 7 and Table 2). The most stable snow climate indicators with regards to local changes considering RPD, DRC, and MAD are: HSmax, dHS1, dHS5, and HNsum closely followed by HN3max. Surprisingly, HSavg, is according to our analyses, not among the most robust snow climate indicators and perhaps not the indicator to got to for future homogenisation efforts for snow, as it was used in Marcolini et al. (2017a).

Our analysis shows that median RMSE are about 5 cm for all height/depth related indicators and about 5 days for all time related indicators, except HNsum. Median relative percentage differences are about 7% for the number of days with snow cover and 11–16% for all other indicators. It is worth bearing in mind that in extreme cases the deviations within a station pair can reach 25–40%.

The differences between dHS1 and dHS5 respectively dHN1 and dHN5 are negligible except for the fact that both dHS5 and dHN5 show stronger changes than dHS1 and dHN1 and dHN1 has a larger variability than dHN5. A higher threshold of 5 cm (compared to 1 cm) does provide more stability on days with snowfall, but not on days with snow on the ground.

Nearly all stations show negative RC for all snow climate indicators during the period 1980 to 2004, suggesting that the climate signal during that period is stronger than the local inhomogeneities. But more importantly, the signs of strong RC agree at all station pairs for all the snow climate indicators.

All outliers in Figures 3–5 (correlation, relative differences and RMSE) can be attributed to the same five station pairs. However, the outliers in Figure 7 (difference in relative changes) show a more diverse pattern and consist of 10 different station pairs (including three from Figures 3–5).

ACKNOWLEDGEMENTS

We thank the two anonymous reviewers for their comments which helped to improve the manuscript.

ORCID

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

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

REFERENCES

Acquaotta, F., Colombo, N., Fratianni, S., Romeo, V. and Barbero, S. (2015) Preliminary results of a comparison study between two independent snow networks in north-western Italian Alps (Piemonte Region). In: Lollino, G., Manconi, A., Clague, J., Shan, W. and Chiarle, M. (Eds.) *Engineering Geology for Society and Territory - Volume 1*. Cham: Springer International Publishing, pp. 113–116.

Acquaotta, F., Fratianni, S., Aguilar, E. and Fortin, G. (2019) Influence of instrumentation on long temperature time series. *Climatic Change*, 156(3), 385–404. https://doi.org/10.1007/s10584-019-02545-z.

Acquaotta, F., Fratianni, S. and Venema, V. (2016) Assessment of parallel precipitation measurements networks in Piedmont, Italy: Assessment of parallel precipitation measurements networks in piedmont. *International Journal of Climatology*, 36 (12), 3963–3974. https://doi.org/10.1002/joc.4606.

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidance on metadata and homogenization, 54.

Baronetti, A., Fratianni, S., Acquaotta, F. and Fortin, G. (2019) A quality control approach to better characterize the spatial distribution of snow depth over New Brunswick, Canada. *International Journal of Climatology*, 39(14), 5470–5485. https://doi.org/10.1002/joc.6166.

Begert, M., Schlegel, T. and Kirchhofer, W. (2005) Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25(1), 65–80. https://doi.org/10.1002/joc.1118.

Beniston, M., Rebetez, M., Giorgi, F. and Marinucci, M.R. (1994) An analysis of regional climate change in Switzerland.

Theoretical and Applied Climatology, 49(3), 135–159. https://doi.org/10.1007/BF00865530.

Della-Marta, P.M. and Wanner, H. (2006) A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements. *Journal of Climate*, 19(17), 4179–4197. https://doi.org/10.1175/JCLI3855.1.

Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. and Rosas, G. (2017) The influence of station density on climate data homogenization. *International Journal of Climatology*, 37 (13), 4670–4683. https://doi.org/10.1002/joc.5114.

Guenzi, D., Acquaotta, F., Garzena, D., Baronetti, A. and Fratianni, S. (2020) An algorithm for daily temperature comparison: Co.Temp - comparing series of temperature. *Earth Science Informatics*, 13(1), 205–210. https://doi.org/10.1007/s12145-019-00414-y.

Haberkorn A (ed). 2019. European Snow Booklet.

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M. and Brönnimann, S. (2017) Identifying, attributing, and overcoming common data quality issues of manned station observations: Identifying, attributing, and overcoming common data quality issues. *International Journal of Climatology*, 37(11), 4131–4145. https://doi.org/10.1002/joc.5037.

Jonas, T., Rixen, C., Sturm, M. and Stoeckli, V. (2008) How alpine plant growth is linked to snow cover and climate variability. *Journal of Geophysical Research: Biogeosciences*, 113(G3), 2156–2202. https://doi.org/10.1029/2007JG000680.

Kendall, M.G. (1975) *Rank Correlation Methods*. London: Charles, Griffin.

López-Moreno, J.I., Revuelto, J., Fassnacht, S.R., Azorín-Molina, C., Vicente-Serrano, S.M., Morán-Tejeda, E. and Sexstone, G.A. (2015) Snowpack variability across various spatio-temporal resolutions: SNOWPACK VARIABILITY ACROSS VARIOUS SPATIO-TEMPORAL RESOLUTIONS. *Hydrological Processes*, 29(6), 1213–1224. https://doi.org/10.1002/hyp.10245.

Mann, H.B. (1945) Nonparametric tests against trend. *Econometrika*, 13, 245–259.

Marcolini, G., Bellin, A. and Chiogna, G. (2017a) 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. *International Journal of Climatology*, 37, 1267–1277. https://doi.org/10.1002/joc.4977.

Marcolini, G., Bellin, A., Disse, M. and Chiogna, G. (2017b) Variability in snow depth time series in the Adige catchment. *Journal of Hydrology: Regional Studies*, 13, 240–254. https://doi.org/10.1016/j.ejrh.2017.08.007.

Marty, C. (2008) Regime shift of snow days in Switzerland. *Geophysical Research Letters*, 35(L12501), 5. https://doi.org/10.1029/2008GL033998.

Morán-Tejeda, E., López-Moreno, J.I. and Beniston, M. (2013) The changing roles of temperature and precipitation on snowpack variability in Switzerland as a function of altitude. *Geophysical Research Letters*, 40(10), 2131–2136. https://doi.org/10.1002/grl.50463.

Neumann, N.N., Derksen, C., Smith, C. and Goodison, B. (2006) Characterizing local scale snow cover using point measurements during the winter season. *Atmosphere-Ocean*, 44(3), 257–269. https://doi.org/10.3137/ao.440304.

Resano-Mayor, J., Korner-Nievergelt, F., Vignali, S., Horrenberger, N., Barras, A.G., Braunisch, V., Pernollet, C.A. and Arlettaz, R. (2019) Snow cover phenology is the main driver of foraging habitat selection for a high-alpine passerine during breeding: implications for species persistence in the face of climate change. *Biodiversity and Conservation*, 28(10), 2669–2685. https://doi.org/10.1007/s10531-019-01786-9.

Scherrer, S.C., Ceppi, P., Croci-Maspoli, M. and Appenzeller, C. (2012) Snow-albedo feedback and Swiss spring temperature trends. *Theoretical and Applied Climatology*, 110(4), 509–516. https://doi.org/10.1007/s00704-012-0712-0.

Scherrer, S.C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R. and Appenzeller, C. (2013) Snow variability in the Swiss Alps 1864–2009. *International Journal of Climatology*, 33(15), 3162–3173. https://doi.org/10.1002/joc.3653.

Schmucki, E., Marty, C., Fierz, C., Weingartner, R. and Lehning, M. (2017) Impact of climate change in Switzerland on socioeconomic snow indices. *Theoretical and Applied Climatology*, 127 (3), 875–889. https://doi.org/10.1007/s00704-015-1676-7.

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

Spreitzhofer, G. (1999) Spatial, Temporal and Intensity Characteristics of Heavy Snowfall Events over Austria. *Theoretical and Applied Climatology*, 62(3–4), 209–219. https://doi.org/10.1007/s007040050085.

Theil, H. (1950) A rank-invariant method of linear and polynomial regression analysis. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series A*, 53, 386–392.

Wipf, S., Stoeckli, V. and Bebi, P. (2009) Winter climate change in alpine tundra: plant responses to changes in snow depth and snowmelt timing. *Climatic Change*, 94(1), 105–121. https://doi.org/10.1007/s10584-009-9546-x.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Buchmann M, Begert M, Brönnimann S, Marty C. Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series. *Int J Climatol*. 2021;41 (Suppl. 1):E2553–E2563. https://doi.org/10.1002/joc.6863

Supporting information

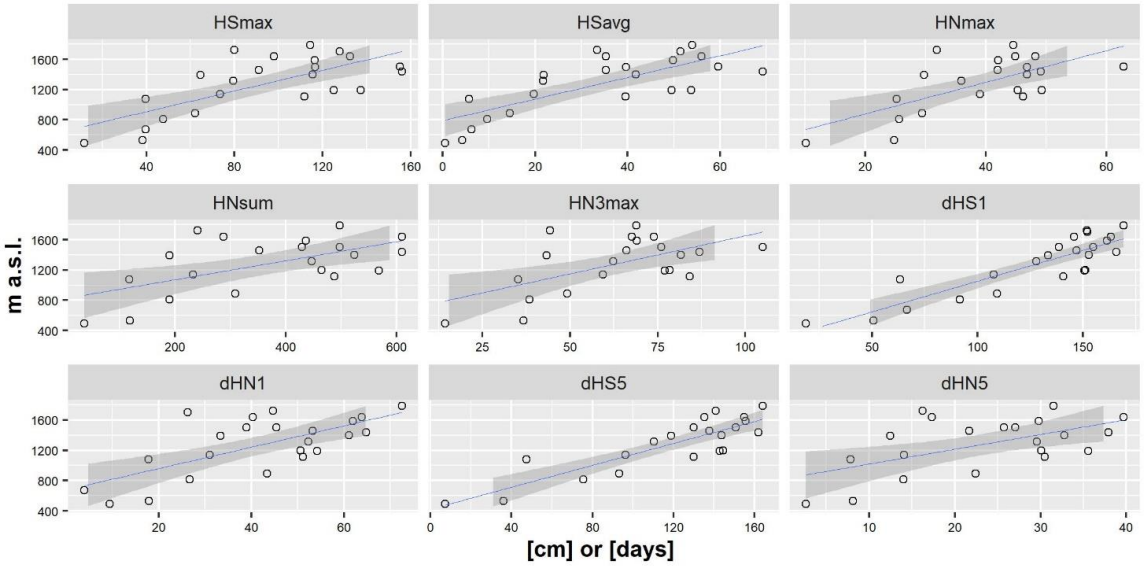


Figure S1: Elevation dependences of mean values for each indicator at station-pair level. R^2-values are in Table S2.

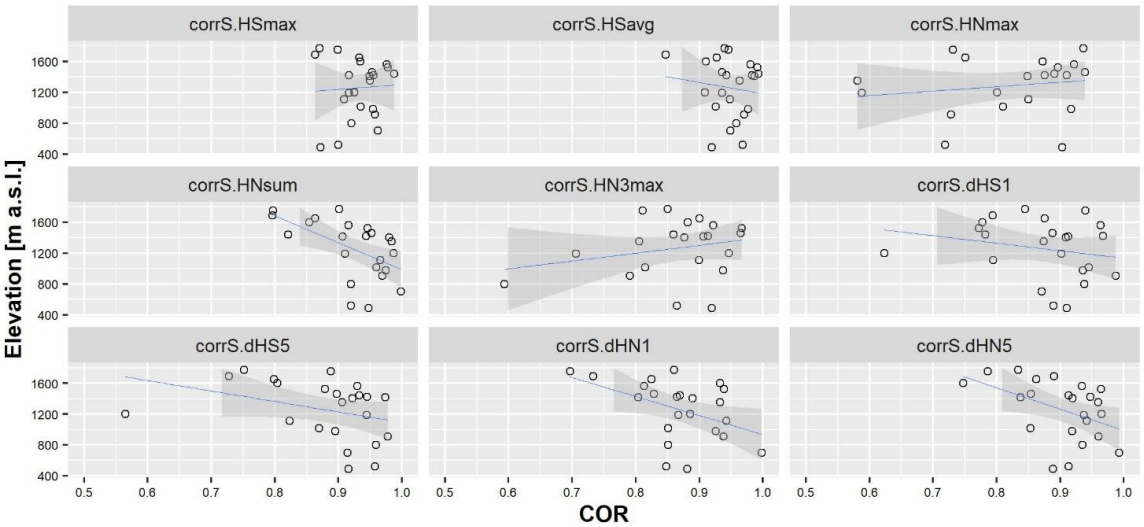


Figure S2: Elevation dependence of COR. R^2-values are in Table S2.

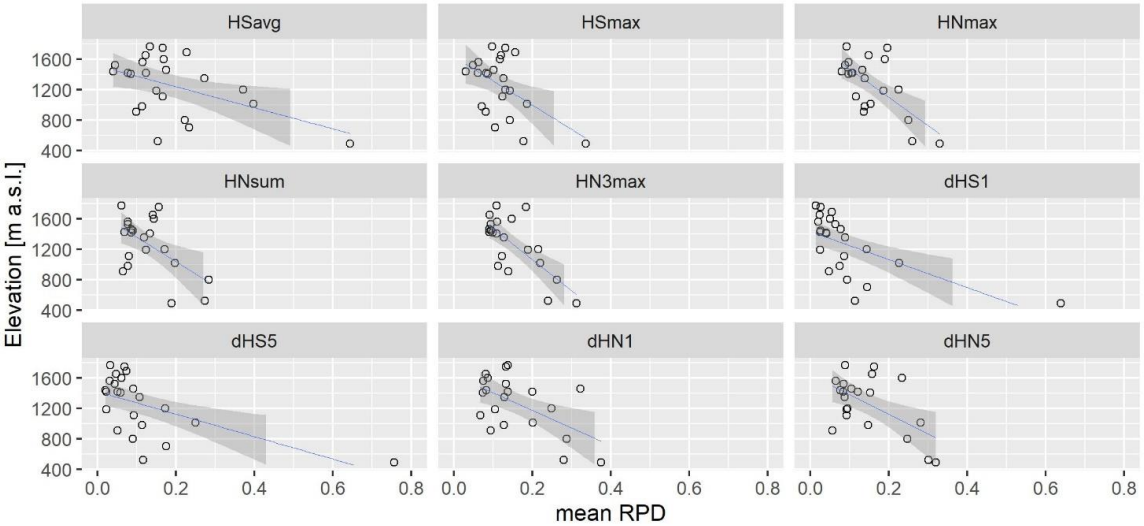


Figure S3: Elevation dependence of mean RPD. R^2-values are in Table S2.

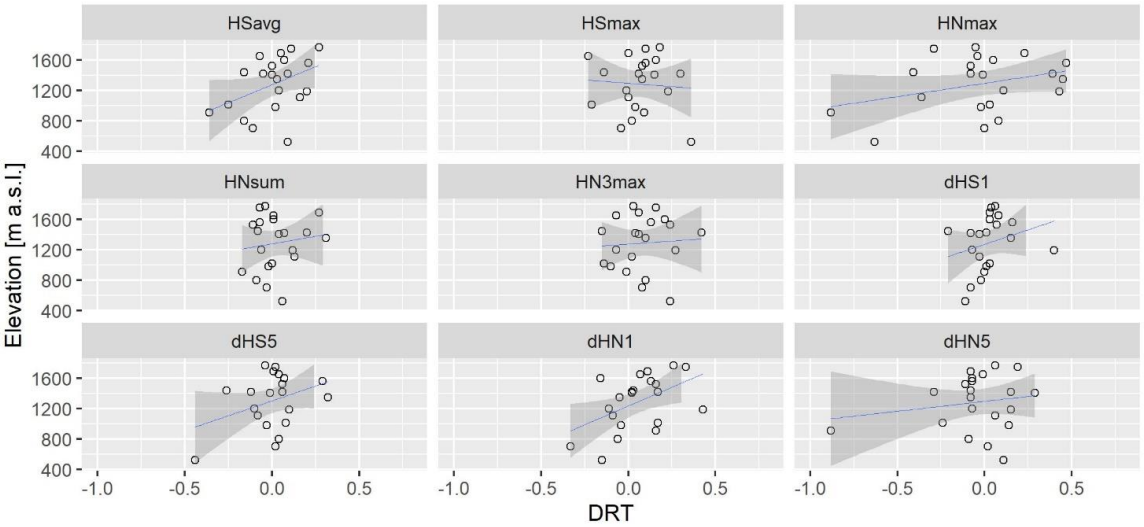


Figure S4: Elevation dependence of DRC. R^2-values are in Table S2.

Table S1: R^2-values for the linear regressions in Figures S1-S4 for all indicators.

Indicator	R^2, Fig S1	R^2 COR, Fig S2	R^2 RPD, Fig S3	R^2 DRC, Fig S4
HSavg	0.55	-0.03	0.21	-0.03
HSmax	0.5	-0.04	0.25	-0.02
HNmax	0.5	0.05	0.35	0.08
HN3max	0.39	-0.04	0.50	-0.05
HNsum	0.37	0.31	0.02	-0.02
dHS1	0.68	-0.03	0.36	0.03
dHS5	0.7	0.02	0.32	0.08
dHN1	0.48	0.11	0.10	0.05
dHN5	0.34	0.20	0.20	-0.05

Table S2: Station names and abbreviations used in this study.

Abbreviation	Location	Mean elevation [m a.s.l.]
ABO – 1AD	Adelboden	1315
ANT – 2AN	Andermatt	1440
BIV – 5BI	Bivio	1790
BOS – 6BG	Bosco/Gurin	1500
CAV – 7CA	Cavaglia	1700
DAV – 5DF	Davos	1570
EIN - SSE	Einsiedeln	900
GAD – 1GA	Gadmen	1190
GOS – 2GO	Göschenen	1110
KLA – 5KW	Klosters	1200
LAN – 5LQ	Landquart	530
LTB – 1LB	Lauterbrunnen	810
MUE – 1MR	Mürren	1640
PAV - PAY	Payerne	490
PIO – 6AM	Piotta	1140
ROB – 7PV	Poschiavo	1080
RIE—4WI	Lötschental	1500
SAM – 7SD	Samedan	1730
SDO – 5SP	Splügen	1460
SED – 5SE	Sedrun	1400
SMM – 7ST	Santa Maria	1390
THS – 5TH	Thusis	670
ZER – 4ZE	Zermatt	1640

Chapter 3

Local-scale variability of seasonal mean and extreme values of in situ snow depth and snowfall measurements

Status: Published in The Cryosphere https://doi.org/10.5194/tc-15-4625-2021
Data set can be downloaded from EnviDat https://www.doi.org/10.16904/envidat.218.
Author contributions: CM, MTB, and MBU designed the outline with input from SB. MBU performed the analyses and wrote the draft. The revised manuscript was written by MBU with inputs from all authors. CM supervised the work.

Snow observations and measurements can vary drastically over small distances, which can influence the accuracy of model verification or load factors for building codes. Here we present the first quantitative assessment of local-scale variations of commonly used snow variables. Local-scale variations in onset and disappearance dates as well as extreme values are studied using the parallel data set.

The Cryosphere, 15, 4625–4636, 2021
https://doi.org/10.5194/tc-15-4625-2021
© Author(s) 2021. This work is distributed under
the Creative Commons Attribution 4.0 License.

The Cryosphere Open Access EGU

Local-scale variability of seasonal mean and extreme values of in situ snow depth and snowfall measurements

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

[1]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2]Federal Office of Meteorology and Climatology (MeteoSwiss), Zurich, Switzerland
[3]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
[4]Institute of Geography, University of Bern, Bern, Switzerland

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

Received: 20 April 2021 – Discussion started: 30 April 2021
Revised: 12 July 2021 – Accepted: 8 September 2021 – Published: 30 September 2021

Abstract. Daily measurements of snow depth and snowfall 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 usually located in the same villages (or 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, and snow onset and disappearance dates, which are calculated for various seasons (December to February (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 variability. We found 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 variations of less than 5 % for all indicators. Local-scale variability ranges between less than 24 % (DJF) and less than 43 % (MA) for all indicators and 75 % of all station pairs. The highest percentage (90 %) of station pairs with variability of less than 15 % is observed for days with snow on the ground. The lowest percentage (30 %) of station pairs with variability of less than 15 % is observed for average snow depth. Median differences of snow disappearance dates are rather small (3 d) and similar to the ones found for snow onset dates (2 d). An analysis of potential sunshine duration could not explain the higher 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 d new snow sum, which are often used for avalanche prevention measures. The found return levels are within each other's 95 % confidence intervals for all (but three) station pairs, revealing no striking differences. The findings serve as an important basis for our understanding of variabilities of commonly used snow indicators and extremal indices. Knowledge about such variabilities in combination with break-detection methods is the groundwork in view of any homogenization efforts regarding snow time series.

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 snow cover are dependent on the local characteristics of the site: i.e. exposure to wind or solar radiation. Furthermore, nearby buildings or trees may have an impact on the measured snow quantities. In an ideal world this would not matter, as basic guidelines (e.g. World Meteorological Organization, 2018) recommend measuring snow in a flat, not-wind-exposed measurement field, which is at least at the same distance from a building or tree as

Published by Copernicus Publications on behalf of the European Geosciences Union.

the height of the obstacle. In reality, manual measurement locations sometimes do not fulfil these basic requirements for many reasons, such as the availability of suitable terrain and observers or easy access to the site. Due to this fact, the variability of the measured snow quantities on the 1 km scale (i.e. next to an open field) may be smaller than the variability on the 10 m scale (e.g. south or north of a house). In theory, this bias introduced by sometimes-not-ideal measuring locations that 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 invaluable in view of homogenization efforts of snow data series.

Moreover, climatological applications and studies usually focus either on the meteorological winter DJF (Scherrer and Appenzeller, 2006) or on the 6 months NDJFMA (Marcolini et al., 2017). However, winter in the Alps is not simply restricted to these 3 or 6 months. In contrast, for many ecological and hydrological applications, the melting spring snow cover is the main interest (Brown and Robinson, 2011; Livensperger et al., 2016; Zampieri et al., 2015) and for 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 (Vorkauf et al., 2021).

Using and extending the data set of parallel time series introduced by Buchmann et al. (2021a) in the available number of stations, months, and years enables the investigation of the impact of the above-mentioned bias introduced by sometimes-not-ideal measuring locations, hereafter referred to as "local bias" or variability. Investigating snow onset and disappearance dates, as well as extreme value analyses, we strive to answer the following questions.

1. Is the behaviour of the local bias a function of time series length (Sect. 4.1); i.e. are relative percentage deviations dependent on the length of 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 (Sect. 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 (Sect. 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 organized as follows: Sect. 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). HS measurements are taken from a fixed graduated stake, and HN measurements are taken from a board on top of the previous day's snow cover every morning at 06:00 UTC at least between November and April (for details refer to Haberkorn, 2019, and Buchmann et al., 2021a). 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). The set 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 it incorporates a total of 1338 station years, covering the time period from 1944 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 d 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) are an issue. Although we managed to compile complete metadata records, there is no guarantee that these are always precise and correct. Theoretically, the exact locations of the snow measurements are

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

Snow climate indicator	Abbreviation
Average snow depth	HSavg
Maximum snow depth	HSmax
New snow sum	HNsum
Maximum 3 d new snow sum	HN3max
Days with snow on the ground, HS > 0 cm	dHS1
Days with 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

known; however, until about two decades ago, only approximate coordinates were recorded. The main reason is the general 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 missing information.

Potential sunshine duration is obtained with the help of Swisstopo's digital elevation model DHM25, which has an accuracy in Alpine terrain of 5–8 m.

3 Methods

To be able to compare and quantify the differences of the various snow climate indicators, we use relative percentage differences (RPDs), 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.

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

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

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

To be able to capture all onset and disappearance dates, we defined the current hydrological year as the period from 1 September of the previous year to 31 August of the current year, as it is often used in snow hydrological modelling (e.g. Liston and Elder, 2006; Seibert and Vis, 2012).

There are various definitions for snow onset (Dstart) and disappearance dates (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 the first day (Dstart) and day after the last day (Dend) of the longest period with continuous snow cover. For the purposes in this study, we additionally allowed gaps of up to 3 consecutive days with no snow cover during the season. The chosen gap length allows the inclusion of full seasons in case they were fragmented in the middle of the winter by a maximum of 3 d 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.

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, 1968; Theil, 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.

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 d periods of the hydrological year.

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.

To analyse the impact of the local bias on potential extreme events based on the annual maximum 3 d 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 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 (RPDs) with time series length, we plotted the mean RPD (season NDJFMA) against the length of the underlying parallel time series for each indicator (Fig. 1). Here we found no clear pattern. RPDs are between 10 % and 50 % for all station pairs and indicators. The values are in line with findings from Buchmann et al. (2021a) 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 show RPDs smaller than 20 %. There is no difference between the shorter and the longer time series. This suggests that the lengths of the time series have no effect on the RPD values. This further implies that time series of different lengths can be compared and compiled into one data set 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 (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 signal of possible breaks from the noise.

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 6 months November to April (NDJFMA), 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 (Fig. 2 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 al. (2021a). Figure 2 (panel b) further reveals that the 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 (Fig. 2 panels c and d). Here we see that the largest median differences are not found in the same seasons for all indicators. Figure 2 (panel d) reveals that absD for HSavg and maximum snow depth (HSmax) seem to be largest in March and smallest in November, whereas for the 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 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 snowfall sums are observed in January and February, whereas November and April are months with usually the smallest snowfall sums. dHS1 shows the largest absD at the beginning and end of the snow season (first and last 2 months) likely due to larger differences in thermal conditions (ground and local air temperature) in these transition months. dHN1, in contrast, shows fewer 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 (Fig. 2 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 d 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 snowfall count indicator, dHN1, but on a much lower level. Additional information of absD allows us 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 applies to cases where the snow cover is quite low (a couple of centimetres). 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 also show 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 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 median winter seasons shorter than 60 d. We calculated the climatology of the daily snow depth differences for each station pair. We then constructed the difference series as the combination of the first (and last) 60 d for each station pair. The median of all these difference series is the thick black line in Fig. 3. These two periods stand for accumulation (first 60 d) and ablation (last 60 d). The 60 d 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 (Fig. 3). This suggests that the ablation period shows more variation

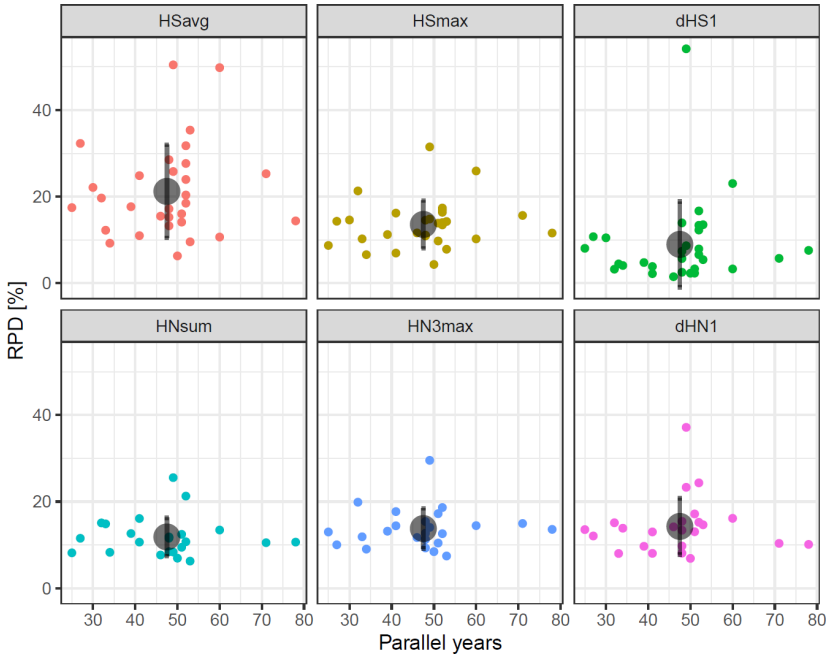


Figure 1. Relative percentage differences (RPDs) as a function of number of parallel years for six snow climate indicators. Black dots and error bars indicate mean and standard deviations. The variables are defined in Table 1. Snow depth (HS) and snowfall (HN) indicators are based on 30 and 24 station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

than the accumulation period, which corroborates the pattern found in the absD analysis.

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 (Dend) are analysed. Figure 4 (left) highlights the computed absolute differences for Dstart and Dend. We found that for 20 out of 30 station pairs, differences of Dend are larger than differences of Dstart. In contrast to the general impression that Dstart should be the same for parallel stations, our data show the differences of several days are not uncommon, which could be caused by the different thermal conditions of the ground of the measurement field. Figure 4 (right) depicts the mean inter-pair differences for Dstart and Dend. The median values over all station pairs are 2 d for snow onset and 3 d for snow disappearance dates. This corroborates the previous findings of 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 d, whereas Dend shows slightly larger differences (0 to 6 d). A possible explanation for the large differences associated with station 5KK is discussed in Sect. 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 variability within the station pairs. For this purpose, the temporal change in days for Dstart and Dend is calculated for each station pair (Fig. 5). The pairs are aligned from low (left) to high (right) elevation. Here we see that for a majority of station pairs, the direction of changes for Dstart and Dend is the same, and Dstart tends to be associated with positive (later) and Dend with negative (earlier) values.

Although inter-pair differences occur more pronounced and more frequently during the decline phase (spring) compared to the accumulation period (Fig. 3), the actual end of the snow season (snow disappearance date) does not show these larger variations (Fig. 4). Dend appears to be a rather stable indicator, varying for 50 % of the station pairs on average between 2 and 6 d. 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).

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

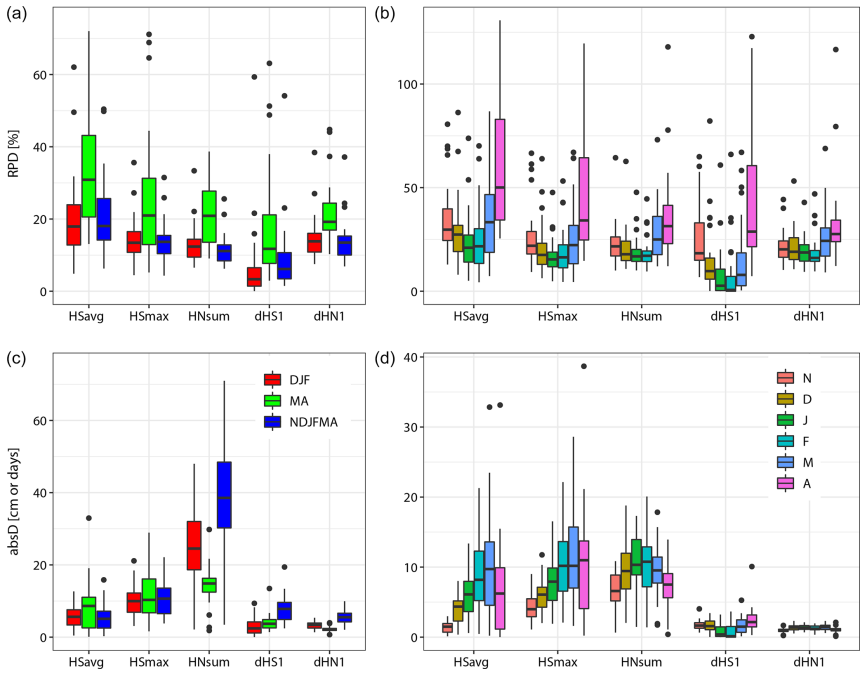


Figure 2. Shown are relative percentage deviations (RPD) and absolute differences (absD) for five snow climate indicators. Depicted are **(a, b)** three seasons from December to February (DJF, red), March–April (MA, green), and November to April (NDJFMA, blue) and **(c, d)** all months individually. Snow depth (HS) and snowfall (HN) indicators are based on 30 and 24 station pairs, respectively. The variables are defined in Table 1.

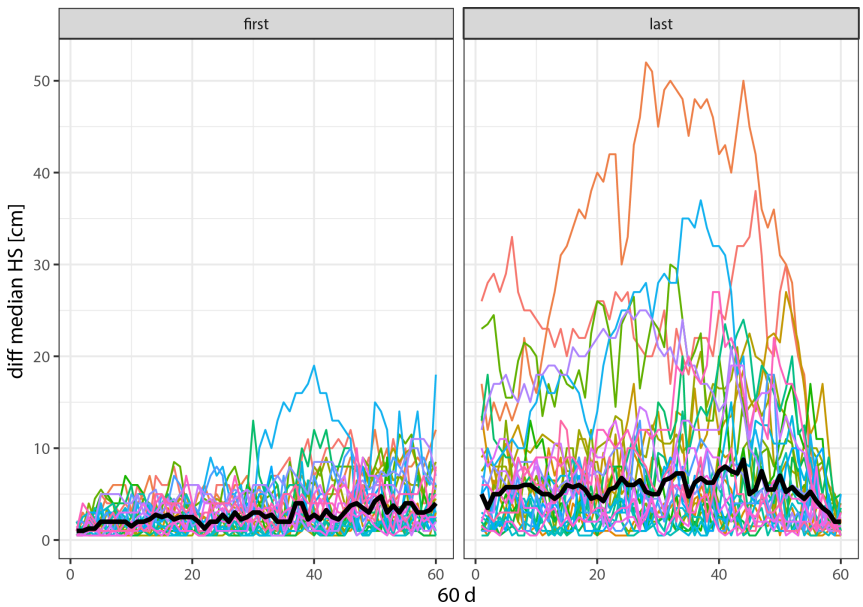


Figure 3. Median daily absolute differences in snow depth for each station pair for the first and last 60 d of each snow season. The bold line highlights the median of all station pairs.

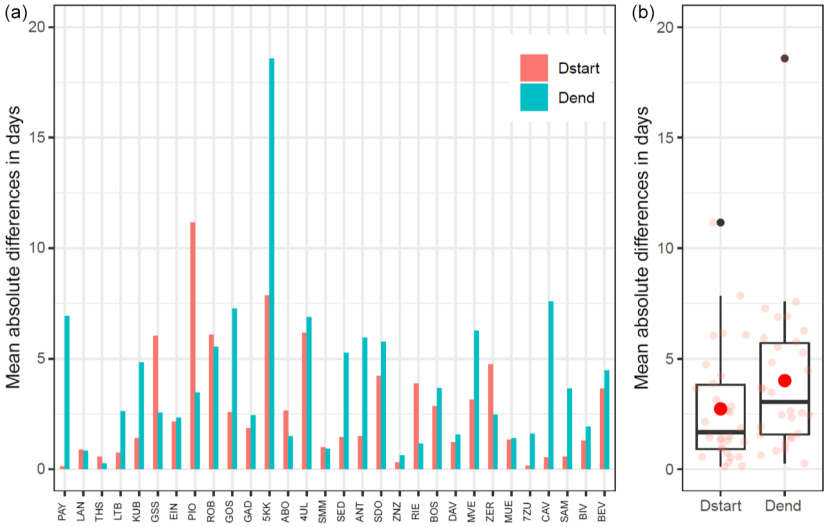


Figure 4. (a) Mean absolute differences in onset (Dstart) and disappearance (Dend) dates for all station pairs. The station pairs are ordered from left to right according to their elevation. **(b)** Variations of mean absolute differences of Dstart and Dend for all station pairs and available years. The light red dots mark the actual values, and the bold red dot indicates the mean.

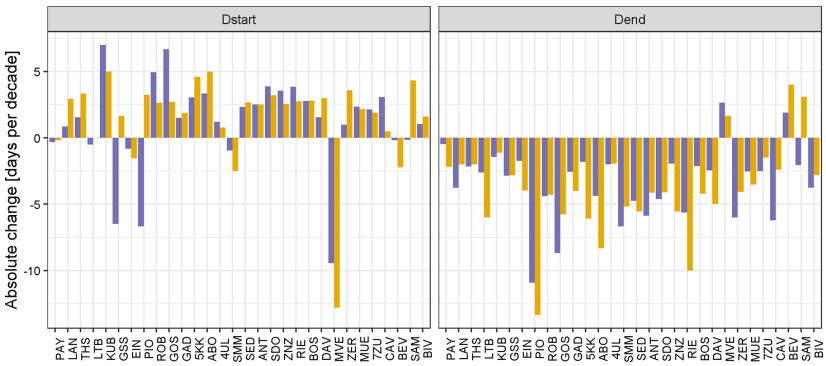


Figure 5. Absolute changes of snow onset (Dstart) and disappearance (Dend) dates calculated for all available parallel periods and expressed as days per decade. Station pairs are ordered according to elevation from left to right. Yellow and purple bars indicate individual stations within a pair.

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.

4.4 Influence of potential sunshine duration on snow disappearance dates

When looking for a possible explanation for the inter-pair differences in Dend with available local-scale variables, we compared them to differences in potential sunshine durations. Figure 6 displays the relationship of differences in Dend and differences in potential sunshine duration during March and April for a selection of station pairs. The largest difference in a station pair (64 h during MA) amounts to approximately 1 h per day. Mean and standard deviations of all unpaired combinations of this subset are −5 and 124 h. Here we found no relationship between potential sunshine duration and differences in Dend. As outlined in Sect. 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.

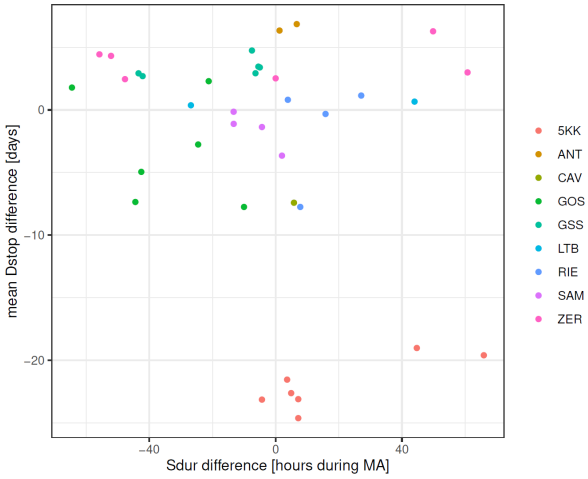


Figure 6. Relationship between differences in snow disappearance dates (Dend) and differences in potential sunshine durations (Sdur) for the months March and April (MA), calculated for a selection of station pairs with manually checked coordinates during different time periods.

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 that the return levels of the individual stations are usually within the 95 % confidence intervals, indicated by error bars in Fig. 7. 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.

The relative differences of 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 three station pairs these values are within each other's 95 % confidence intervals, which may be useful for applications where there is normally only a single time series available. 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.

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). Concerning metadata, at 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 and 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 5 m to the east of a high turbine house (with corresponding shading) at the power station. Both sites are affected by their surroundings, and corresponding 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.

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) 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, that remains impossible to prove.

Ideally, station metadata would include the exact coordinates of the snow stake and panoramic pictures of its surroundings. Also, regular visits to sites and observers would be beneficial. As our focus is solely on parallel measurements, the scalability to higher elevations is limited by the fact that there are no stations above 1800 m a.s.l. Therefore, any scalability attempts would involve a lot of speculation.

5 Conclusions

We presented the first assessment of seasonal local-scale variability of common extremal indices and snow climate indicators based on a unique data set of long-term parallel snow measurements.

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 % (days with

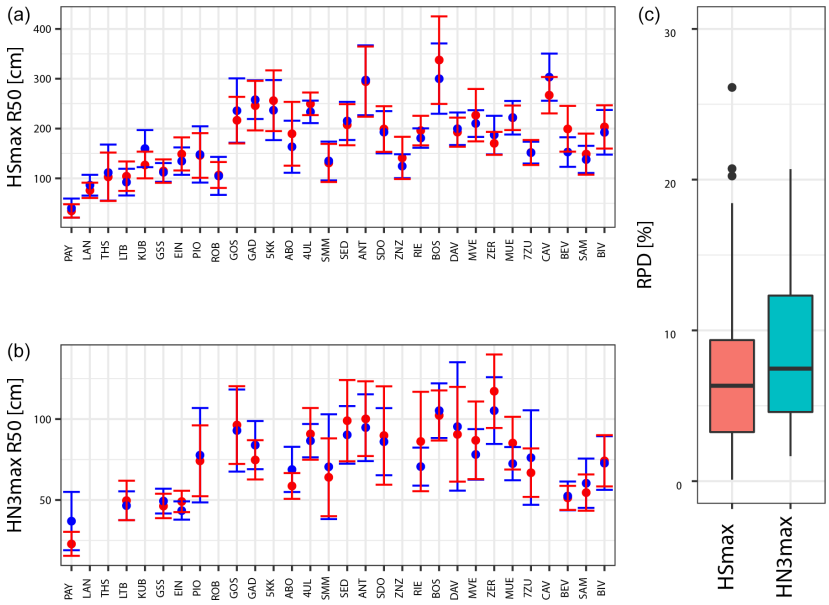


Figure 7. Return levels for seasonal maximum snow depth (HSmax) (top left) and maximum 3 d new snow sum (HN3max) (bottom left) for 50-year return periods. MeteoSwiss (SLF) stations are coloured red (blue). The error bars indicate the 95 % confidence intervals. Station pairs are ordered from left to right from low to high elevation. Relative percentage differences (RPDs) of return levels for HSmax and HN3max are depicted on the right.

snow on the ground) and 20 % (average snow depth) for all indicators, whereas average snow depth also demonstrated the largest spread of all indicators. 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 d) and the snow onset date (3 d) are small; nevertheless, there is considerable variation for some station pairs.

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 large effects and that local bias is probably amplified in MA (and November), 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 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, for the large majority of station pairs, are within each other's 95 % confidence intervals, corroborating that the knowledge of considering the uncertainty in return level estimation is indispensable.

Generally, local bias is often 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.

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

Appendix A

Table A1. 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 mDend are mean onset and disappearance dates, mHSavg indicates the mean average snow depth, and mdHS1 indicates the mean number of days with snow on the ground. Station indicators starting with a number refer to SLF stations, and station indicators without a number refer to MeteoSwiss stations.

Station pair		Start	Stop	Elevation [m a.s.l.]	Length [yr]	HN	mDstart	mDend	mHSavg [cm]	mdHS1 [d]
PAY	PAV	1970	2018	490	48		15 Jan	26 Jan	0.5	21
LAN	5LQ	1969	2020	520	51	no	3 Jan	6 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	5 Mar	9.8	94
KUB	5KU	1991	2020	815	29	no	7 Dec	15 Mar	15.2	109
GSS	EIN	1973	2020	895	47		18 Dec	9 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		3 Dec	6 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		3 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	1 May	69.8	164
SDO	5SP	1973	2020	1457	47		24 Nov	10 Apr	35.4	144
ZNZ	7ZN	1972	2020	1475	48		1 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		6 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

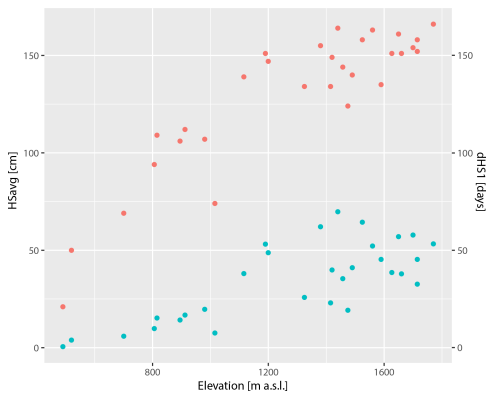


Figure A1. Shown in green are mean values of average snow depth (HSavg), and shown in red are mean values of days with snow on the ground (dHS1) for each station pair and NDJFMA. Station pairs are ordered according to their elevation.

Data availability. The basis for all our analyses, the indicator data set, is available from EnviDat: https://doi.org/10.16904/envidat.218 (Buchmann et al., 2021b).

Author contributions. CM, MiB, and MoB designed the outline with input from SB. MoB performed the analyses and wrote the draft. The revised manuscript was written by MoB with inputs from all authors. CM supervised the work.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. We thank Johannes Aschauer (SLF) for his valuable work in sorting 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 discussions. Our thanks go to Craig Smith and Kay Helfricht and one anonymous referee, whose valuable comments helped to improve the manuscript.

Financial support. This research has been supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant no. 175920).

Review statement. This paper was edited by Ketil Isaksen and reviewed by Craig Smith, Kay Helfricht, and one anonymous referee.

References

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.

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 E2553–E2563, https://doi.org/10.1002/joc.6863, 2021a.

Buchmann, M., Aschauer, J., Begert, M., and Marty, C.: Snow climate indicators derived from parallel manual snow measurements, EnviDat [data set], https://doi.org/10.16904/envidat.218, 2021b.

Foster, J. L.: The Significance of the Date of Snow Disappearance on the Arctic Tundra as a Possible Indicator of Climate Change, Arct. Alp. Res., 21, 60–70, 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.: European Snow Booklet – an Inventory of Snow Measurements in Europe, EnviDat, https://doi.org/10.16904/envidat.59, 2019.

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, 2016.

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

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. 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, 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.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, https://doi.org/10.5194/hess-16-3315-2012, 2012.

Sen, P. K.: 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.-Atmos., 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, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen Series A, 53, 386–392, 1950.

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 rivers, Sci. Total Environ., 503–504, 222–232, https://doi.org/10.1016/j.scitotenv.2014.06.036, 2015.

Chapter 4

Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-)automatic homogenization methods

Status: Published in The Cryosphere https://doi.org/10.5194/tc-16-2147-2022
The data set is available on EnviDat https://doi.org/10.16904/envidat.297.
Author contributions: MBU performed the analyses, produced the figures, and wrote the draft. The manuscript was written by MBU and JC with inputs from all authors. JA, CM, and MBT compiled the metadata. BC and MBT provided support for ACMANT and HOMER. The study was devised by MBU, SB, and CM with inputs from GR and WS and was supervised by CM.

The existence of inhomogeneities in climate series is a well-known problem and there are numerous methods designed to homogenise the series (identify breakpoints and adjust the series), however, there is limited experience in dealing with snow. A homogeneity assessment of 184 Swiss snow depth series is provided by introducing a new approach using three well-established methods and two variables to enhance confidence and robustness, which is especially useful for networks with incomplete or absent accompanying metadata. Using three state-of-the-art homogenisation toolboxes in a combined approach enables the robust identification of breakpoints in Swiss snow series. Having identified valid breakpoints in Swiss snow series paves the way for the adjustment step, which is presented in Chapter 5.

The Cryosphere, 16, 2147–2161, 2022
https://doi.org/10.5194/tc-16-2147-2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

The Cryosphere Open Access EGU

Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-)automatic homogenization methods

Moritz Buchmann[1,2,3], John Coll[4], Johannes Aschauer[1], Michael Begert[5], Stefan Brönnimann[2,3], Barbara Chimani[6], Gernot Resch[7], Wolfgang Schöner[7], 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
[3]Institute of Geography, University of Bern, Bern, Switzerland
[4]CGG, Crawley, United Kingdom
[5]Federal Office of Meteorology and Climatology (MeteoSwiss), Zurich Airport, Zurich, Switzerland
[6]Klimamonitoring und Kryosphäre, Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienna, Austria
[7]Department of Geography and Regional Science, University of Graz, Graz, Austria

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

Received: 23 February 2022 – Discussion started: 17 March 2022
Revised: 18 May 2022 – Accepted: 27 May 2022 – Published: 9 June 2022

Abstract. Knowledge concerning possible inhomogeneities in a data set is of key importance for any subsequent climatological analyses. Well-established relative homogenization methods developed for temperature and precipitation exist but have rarely been applied to snow-cover-related time series. We undertook a homogeneity assessment of Swiss monthly snow depth series by running and comparing the results from three well-established semi-automatic break point detection methods (ACMANT – Adapted Caussinus-Mestre Algorithm for Networks of Temperature series, Climatol – Climate Tools, and HOMER – HOMogenizaton softwarE in R). The multi-method approach allowed us to compare the different methods and to establish more robust results using a consensus of at least two change points in close proximity to each other. We investigated 184 series of various lengths between 1930 and 2021 and ranging from 200 to 2500 m a.s.l. and found 45 valid break points in 41 of the 184 series investigated, of which 71 % could be attributed to relocations or observer changes. Metadata are helpful but not sufficient for break point verification as more than 90 % of recorded events (relocation or observer change) did not lead to valid break points. Using a combined approach (two out of three methods) is highly beneficial as it increases the confidence in identified break points in contrast to any single method, with or without metadata.

1 Introduction

The quality of climate data time series analyses relies heavily on homogeneous input data, and such quality-controlled and homogenized climate data are needed to improve climate-related decision-making. Most decade- to century-scale meteorological time series are affected by inhomogeneities due to, for example, changes to instrumentation, changes to station location and observer practices, or changes in the local environment such as urbanization or plant growth (Tuomenvirta, 2001). Disentangling these break points from the underlying noise and variability in the data is challenging but crucial for improving confidence in any further analyses (e.g. Vertačnik et al., 2015). Accompanying metadata, if available, can be useful in helping to corroborate and verify breaks identified by statistical methods (Aguilar and Llanso, 2003). However, not every relocation or change in the station history necessarily leads to a break point in the data series in the first place.

Published by Copernicus Publications on behalf of the European Geosciences Union.

There is no "one-method" solution when it comes to the detection of break points but rather a collection of statistical tools. Homogeneity tests can be broadly divided into "absolute" and "relative" methods. The former are applied directly to individual candidate stations to identify statistically significant shifts in the section means (referred to as breaks or change points), while relative methods entail comparison of correlated neighbouring stations with a candidate station to test for homogeneity. If reference series exist, a relative (rather than an absolute) approach where candidate series are compared to reference series is considered state of the art (Venema et al., 2012) in contemporary climate sciences as it allows the practitioner to eliminate any erroneous climatological shifts (Della-Marta and Wanner, 2006). Another advantage of some relative methods is that they do not require the reference series to be homogeneous themselves (Szentimrey, 1999; Caussinus and Mestre, 2004). Given the frequent occurrence of inhomogeneities in many climate time series, considerable efforts have been made to address the issue. These efforts by the community have produced a number of relative homogenization methods and toolboxes with varying degrees of user interaction and ease of application to choose from. PRODIGE (French for miracle) (Caussinus and Mestre, 2004) proved to be among the best-performing methods evaluated in the COST (European Cooperation in Science and Technology) Action HOME alongside ACMANT (Adapted Caussinus-Mestre Algorithm for Networks of Temperature series) (Domonkos, 2011), Climatol (Climate Tools) (Guijarro, 2018), USHCN (US Historical Climatology Network) (Menne and Williams, 2009), and MASH (Multiple Analysis of Series for Homogenization) (Szentimrey, 1999).

Efforts towards efficient break detection have been made for many meteorological variables such as temperature (e.g. Kuglitsch et al., 2012; Begert et al., 2008), precipitation (e.g. Begert et al., 2005; Coll et al., 2020), and phenological series (Brugnara et al., 2020) using various methods and tools. In the case of snow time series, however, only a few studies exist: Marcolini et al. (2017) investigated the use of the SNHT (standard normal homogeneity test; Alexandersson, 1986; Alexandersson and Moberg, 1997) for detecting breaks in mean annual snow depth. Marcolini et al. (2019) compared the use of the SNHT and PRODIGE for the break detection and subsequent homogenization of mean seasonal snow depth. Schöner et al. (2019) focused on trend analysis of seasonal mean snow depth in the Swiss–Austrian domain, using PRODIGE to identify inhomogeneous series in the records analysed. In our approach we choose to use multiple reference series processed using several modern relative methods, thereby increasing confidence in the results. As information concerning potential inhomogeneities is crucial in providing an accurate and reliable snow time series record, an in-depth homogeneity assessment of Swiss snow depth series is necessary.

For our study we used ACMANT (Domonkos, 2011, 2020), Climatol (Guijarro, 2018), and the semi-automatic tool HOMER (HOMogenizaton softwarE in R; Caussinus and Mestre, 2004; Domonkos, 2011; Guijarro, 2018; Picard et al., 2011) as they were all used for break detection purposes in recent studies: Kuya et al. (2021a) used Climatol for precipitation and HOMER for temperature (Kuya et al., 2021b); Noone et al. (2016) used HOMER for precipitation; and Coll et al. (2020) compared break detection performance of various methods including ACMANT, Climatol, and HOMER. Climatol is based on the SNHT (Alexandersson and Moberg, 1997) and recommended by Coll et al. (2020) for detecting breaks in precipitation series. HOMER is the extension development of and direct successor to PRODIGE and therefore is one of the most used homogenization methods in climate sciences. In addition it provides the homogenization practitioner with more operational freedom (in terms of configuration possibilities) than Climatol or ACMANT. Applying the different methods to Swiss snow depth time series and experimenting with various configurations allow us to investigate the suitability of the different set-ups and to provide a homogeneity assessment of the manual Swiss snow observation network.

Benchmark analyses of various methods exist for temperature and precipitation at both monthly and daily resolutions (e.g. Venema et al., 2012; Killick et al., 2021); however, no clear favourite has emerged. In addition, the use of Climatol (based on the SNHT) and HOMER (based on PRODIGE) allows for a more direct comparison with work already undertaken to assess homogeneity and break detection for snow time series in the Alps (Marcolini et al., 2017, 2019; Schöner et al., 2019). Furthermore, this in-depth analysis of Swiss snow series allows for the identification of suspicious or erroneous data in the series analysed which would otherwise have escaped detection. The process of homogenization can be described in three steps: break detection, attribution (verification of break points), and correction. In this study, we focus on the first step and touch upon the second. To focus our study, our research questions are as follows:

1. Which method or set-up works for break point detection in (Swiss) snow depth time series?
2. Is there any elevation dependence affecting the capability of the methods for break point detection?
3. Are the detected break points consistent with changes inferred from available metadata?
4. How homogeneous (in terms of detected break points) are the Swiss snow depth series investigated in our approach?
5. Are break point results similar for different snow cover variables (average snow depth and days with snow cover)?

The paper is organized as follows: Sect. 2 introduces the data set, and Sect. 3 details the methods used for the analyses.

Results are presented in Sect. 4 and discussed in Sect. 5, and conclusions are drawn in Sect. 6.

2 Data

Our data consist of a newly compiled set of manually measured Swiss snow depth (HS) series obtained by the Federal Office of Meteorology and Climatology (MeteoSwiss) and the WSL Institute for Snow and Avalanche Research SLF. Manual snow measurements are conducted every morning between November and April at designated measurement fields with an observer reading off the snow depth from a graduated fixed stake; see Buchmann et al. (2021b) for more information. A favourable unique feature of using these manual snow depth measurements in Switzerland is that the instrument (graduated snow stake) and the general measurement procedure have not changed (Haberkorn, 2019), thereby eliminating one potential source of inhomogeneity in the records. We evaluated all the available series in the archives with data recorded between 1931 and 2021. Selection criteria were that time series have to be longer than 30 years and have at least 80 % complete data between November and April, and entirely missing single years were also allowed. Applying these criteria resulted in a data set of 184 station time series, distributed widely over Switzerland, from 200 to 2500 m a.s.l. Figure 1 shows the location and distribution of the stations used in this study (a list of all stations can be found in Table S1 in the Supplement). We use monthly (November–April) mean snow depth and monthly sums of days with snow on the ground for every hydrological year (November of the last year to April of the current year) as input data for the application of the break detection methods and set the remaining 6 months to zero. Days with snow on the ground are defined as days with snow depth equal to or greater than 1 cm. The final outputs used in the break point analysis were annual time series of winter mean snow depth (HSavg) and snow cover duration (dHS1).

Metadata

Since the instruments for manually measuring snow depth have not changed over our period of analysis, this allowed a clearer focus on other metadata components, such as coordinate changes as a proxy for station relocation and observer changes. Such information is generally available and was compiled from various sources (station records, archives, and operational databases). Unfortunately, for the MeteoSwiss network the exact locations of the snow measurements can differ from the coordinates of the associated meteorological station, and especially in the early days, MeteoSwiss did not record snow-specific coordinates. Moreover, the quality differs from station to station and is generally more vague for older periods (greater than 30 years) in the station records (Aschauer and Marty, 2020). Our metadata are not perfect (i.e. there is missing or incomplete information for some records); however, they are unlikely to be completely wrong and may offer some corroborative information for any breaks detected. Metadata are therefore used as additional verification for the identified break points where applicable/available with a tolerance of ±2 years.

3 Methods

3.1 Break detection algorithms

3.1.1 HOMER

HOMER is a collection of functions for break point detection and homogenization. The pairwise comparison (PRODIGE) is based on a penalized likelihood criterion (Caussinus and Lyazrhi, 1997) and is composed of optimal segmentation in conjunction with dynamic programming (Hawkins, 2001). Further methods are joint segmentation based on Picard et al. (2011) and ACMANT detection (Domonkos, 2011). We used pairwise comparison for our analyses because ACMANT detection depends on a seasonal cycle and Gubler et al. (2017) reported issues with HOMER's joint segmentation module. Coll et al. (2020) highlight that HOMER has its disadvantages when dealing with incomplete data, particularly when the missing data comprise contiguous blocks earlier or later in a series. For this reason the WMO Task Team on Homogenization recommends a missing data tolerance of 15 years for HOMER (WMO, 2017). However, as we are using mainly complete series and solely focus on pairwise detection, our analyses should not be affected. Network neighbourhoods in HOMER are constructed using station selection criteria, based on either distance or first-difference correlations. Here we used a minimum correlation threshold of 0.8 (empirical values) and a minimum number of five reference series (similarly to PRODIGE). In practice this means that if no reference series with correlations ≥ 0.8 are found for a particular candidate series, the five next best correlated stations are returned instead and, if more than five series exist with correlations ≥ 0.8, all are displayed.

HOMER is semi-automatic insofar as it provides the user with a set of graphical difference series whereby the candidate series are compared to the reference series in each of the sub-networks based on the selection criteria applied. For each candidate and reference series in the derived subnetworks, any breaks detected are displayed to help inform subsequent adjustment decisions by the user. Difference series in HOMER can be defined in two ways: either the candidate minus reference (Diff-mode) or the candidate divided by reference (Ratio-mode). As our data are skewed due to the seasonal nature of snow cover data and limited at zero (no negative snow depth), we used Ratio-mode instead of differences (details in Table 1). Break points are interpreted as valid if they occur in more than half of the first five valid

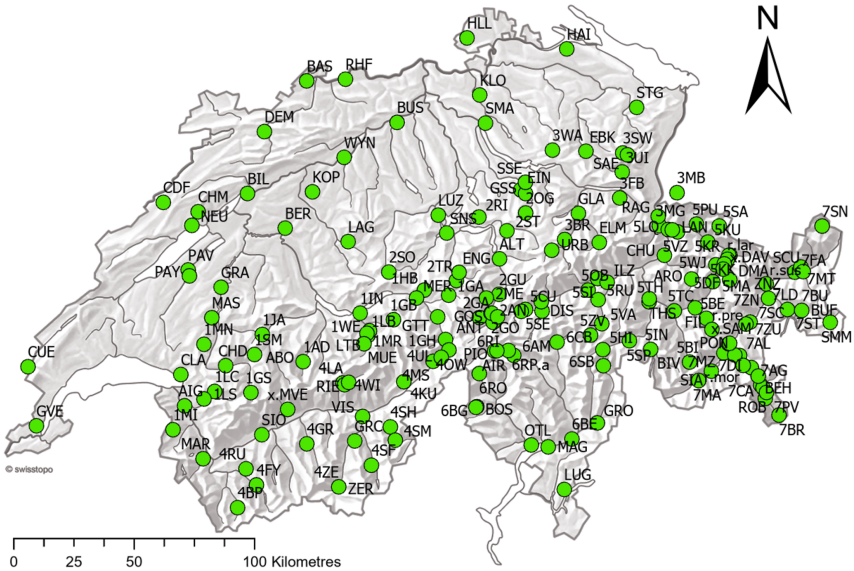


Figure 1. Map of Switzerland showing the distribution of the 184 stations used in this study.

Table 1. Settings for HOMER, ACMANT, and Climatol.

Method	Variable	Value	Comment
HOMER	Mode	Ratio, pairwise	
	Min correlation	0.8	
	Min reference series	5	
	Break detection	Annual	
	Break	3/5	Three out of five
	Tolerance for breaks	±2 years	
ACMANT	Mode	RR	Precipitation
	Min correlation	0.4	Fixed value
	Output	Default	
Climatol	`std`	2	Normalization: divide by the mean; preferred if data are skewed
	`wz`	0.1	Scale parameter for the vertical coordinate
	`wd`	0, 0, 100	Distance weight
	`dz.max`	22	Value from `expl-results`
	`snht1`	35	Determined by `expl-results` (left to default 25 for dHS1)
	`snht2`	45	Determined by `expl-results` (left to default 25 for dHS1)

(with standard deviation of the noise (sigma) smaller than 0.3 and similar length and geographical origin to candidate series) reference series (i.e. three out of five), within an uncertainty of ±2 years, meaning that breaks are accepted as valid if they occur within ±2 years of each other in at least three reference series. The ±2 years is adopted from Kuglitsch et al. (2012). Venema et al. (2020) pointed out that parallel analyses of statistically or geographically relevant station data are the best solution for identifying breaks in time series. Some of the investigated snow time series are parallel series (Buchmann et al., 2021b), and if a break point is suggested in such a parallel candidate–reference series, it is double counted. In cases where some of the first 5 reference series selected do not cover the whole time period of the candidate series, the next longest series with sigma smaller than 0.3 might be included in the analysis. Furthermore, break detection was set to "annual". Once the potential break points are adjusted based on operator interpretation and re-entered, HOMER is run again to calculate the break magnitude correction factors for each station.

3.1.2 ACMANT

ACMANT (Domonkos, 2011; Domonkos and Coll, 2017) is the most automatic of the homogenization methods used in this study, which means hardly any parameters can be changed by the user. Reference series are constructed via correlations between candidate and reference series with a fixed minimal correlation of 0.4. For this study we used ACMANTv4.3 (Domonkos, 2020) run in standard precipitation mode (see Table 1). Break points and the associated break magnitude corrections applied by ACMANT are retrieved from an automatically produced file. Automatic networking (AN) has been developed (Domonkos and Coll, 2019) as a preparatory operation for homogenizing data sets of more than 40 series (as is the case with the network here) with the ACMANT method. In AN, a specific network is constructed for each candidate series which provides optimal spatial comparison with the candidate series always in the centre of the network (Domonkos and Coll, 2017).

3.1.3 Climatol

In contrast to HOMER and ACMANT, Climatol uses composite reference series and detects breaks one by one with the standard normal homogenization test (SNHT; Alexandersson and Moberg, 1997) applied to anomaly series between candidate and reference series to detect break points. Anomalies are normalized through either the subtraction of or division by series averages, with the latter approach being recommended for skewed data series such as precipitation. Missing values are automatically infilled using values from neighbouring stations, thus allowing the method to compare series that do not share an intact data period prior to this adjustment. There are no pre-defined neighbourhoods as in HOMER, but a distance criterion is available to geographically confine the potential reference series (see Guijarro, 2018, and Coll et al., 2020, for more detail). Reference series are defined based on geographical proximity (Luna et al., 2012) using Euclidean distances. By default, the vertical coordinates in metres carry the same weight as the horizontal coordinates in kilometres. To account for the influence of elevation as a key control on snow, the scale parameter (`wz`) was adjusted so that elevation counts 100 times more, in practice meaning an elevation difference of 500 m is equivalent to a horizontal distance of 50 km in the approach used here. Climatol also provides an exploratory analysis of the data, which is necessary to properly set the parameters listed in Table 1. Once the parameters are set, Climatol can be run to produce an output file with suggested break points for each station. Break magnitude corrections are calculated as the change in mean before and after homogenization as Climatol does not automatically provide these values.

3.2 Evaluation of break detection algorithms

We define valid break points as break points detected at concordant times for any given series by at least two out of three independent methods within a tolerance of ± 2 years regardless of metadata. As an additional measure, we only accepted as valid those break points identified within 5 years of either the beginning or the end of a series to agree with procedures used in Kuya et al. (2021a). The break points detected were then compared to available metadata where applicable; this metadata information was the decisive factor used to attribute the exact year of break points. In the case of those series where no metadata information was available, either the common year detected by a majority of the methods or the first occurrence of a concordant break point within the defined tolerance threshold was used.

To evaluate the performance and set-up of each method, their suggested break magnitude corrections and their contributions to concordant break points (number of valid break points compared to total number of detected break points) were measured and compared. To test whether break point detection depends on elevation or the "amount of snow", detected breaks are compared to both station elevation and climatological (calculated for the entire period) snow values (i.e. mean HSavg). To assess how many valid break points can be explained by metadata and to see whether either station relocations or observer changes are more prone to cause detectable breaks, we compared the concordant break points to records in the station history.

Using this combined approach of integrating the break point information from the various methods alongside the information from metadata allows us to more confidently estimate the homogeneity of the Swiss snow network. The main focus is on breaks in HSavg; however, the opportunity to use dHS1 alongside HSavg as a complementary break point detection approach is discussed in Sect. 4.3. Series without any detected break points or break points detected by only one method are considered homogeneous; series with break points detected by at least two out of three methods (with or without metadata support based on our criteria) are considered inhomogeneous.

We are using HOMER 2.6, ACMANTv4.3, and Climatol 3.1.1; all analyses were run on a Windows system with R 4.1.1 (R Core Team, 2021).

4 Results

4.1 Comparison of method performances

4.1.1 Number of detected break points

To assess and be able to compare the three methods, we added up all the detected break points separately for each method and grouped them into time periods to investigate the temporal distribution of the break occurrences between

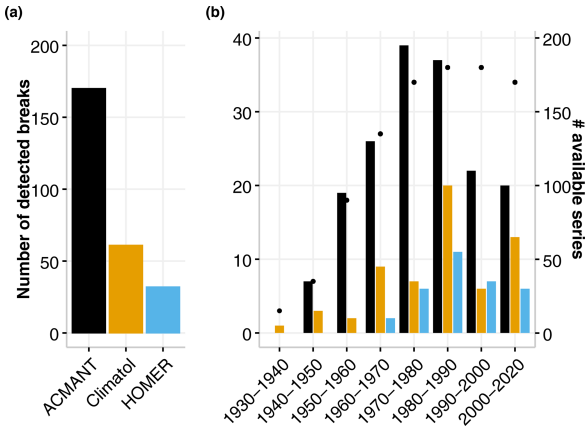


Figure 2. **(a)** Number of break points identified with ACMANT, Climatol, and HOMER for a total of 184 stations. **(b)** Number of break points and year of detection retrieved from ACMANT, Climatol, and HOMER for a total of 184 stations. The number of available series per decade is indicated with black dots and a corresponding right *y* axis. All results are for HSavg.

the methods. Here we found ACMANT returning the largest number of detected break points with 170 breaks in 98 of the 184 series. The total numbers of break detections for Climatol (61 in 54) and HOMER (32 in 30) are significantly smaller than the ACMANT break detections. Table 2 and Fig. 2a summarize the overall break detection frequencies between the three methods. Figure 2b shows the temporal distribution of the detected break points between the methods. Based on this summary ACMANT detects the maximum number of break points between 1970 and 1980, whereas for Climatol and HOMER the 1980s is the decade associated with most detections. This suggests that ACMANT seems to be more sensitive in detecting changes than Climatol and HOMER, and the time period associated with the maximum number of break detections is coincident with the period where the maximum number of station records is available.

4.1.2 Break magnitudes and detection capabilities

To investigate the sensitivity of the break detection capacity, the break magnitude corrections for all identified breaks are analysed across the methods. Figure 3a shows the number of detected breaks and their corresponding break magnitude correction categories. The majority of break magnitude corrections are distributed between 10 %–19 % (ACMANT and Climatol) and 20 %–29 % (HOMER). In contrast to ACMANT and Climatol, HOMER detects very few breaks with magnitude corrections below 10 %. This again indicates that ACMANT detects the most break points, including those of lower magnitude, whereas HOMER identifies fewer but higher-magnitude breaks. Climatol tends to mirror the pattern of ACMANT break point detection but with fewer detections overall. Figure 4 provides density distribution plots of the break magnitude corrections for the three methods separately. Here we found almost normal distributions for the Climatol and ACMANT break magnitude density distributions with a peak around 0 (suggesting no or very small corrections). The break magnitude distribution for ACMANT is bimodal with a local minimum at 0, reflecting more low-magnitude negative corrections than positive ones. The distribution of the break magnitude corrections for HOMER on the other hand is more uniform with no obvious peak. This reinforces that HOMER not only detects fewer break points than Climatol or ACMANT overall but also generally detects larger-magnitude breaks, a situation reflected in the broader distribution of break magnitude corrections.

4.1.3 Elevation and amount-of-snow dependencies

The availability (as well as quality) of suitable reference series for each candidate station is key for a proper break point detection. Snow distribution in Switzerland is highly elevation-dependent, and our stations range from 200 to 2500 m a.s.l. To test the hypothesis that lower stations might not have enough suitable reference series for proper break point detections, a possible elevation dependence is investigated. To explore a possible altitudinal or amount-of-snow influence on the break detection capability of the methods, the break magnitude corrections of the identified break points were compared to station elevation and climatological HSavg (mean over the entire period of the available records). This comparison of break detections with elevation found fewer break points below 500 and above 1800 m a.s.l. than between 1300 and 1700 m a.s.l., whereas in comparison to the climatological HSavg, no clear pattern is visible. For these comparisons all three methods exhibited similar break detection patterns. Figure 3b and c plot these relationships separately, while Fig. 5 summarizes the relationship between station elevation and climatological HSavg for every station where break points were detected.

4.2 Concordant break points in HSavg

As the location and magnitude of break points obtained from the individual methods differed, for the remainder of the analysis only concordant break points (identified across all three methods) are considered. In this approach only breaks identified by at least two of the methods and obtained within ±2 years are considered to be valid. The set of valid break points based on these criteria are shown in Table 3. This approach identified 31 break points in 30 series for HSavg, with only one series showing multiple break points detected across all three methods. The majority of valid break points were found with a combination of ACMANT and HOMER, whereas no break points in HSavg were found with only Climatol and HOMER. Figure 6 summarizes the method combinations which led to break points being assessed as valid

Table 2. Comparison of detected break points for each individual method. A valid break point means it was detected by at least two out of three methods. The complementary category uses break points from both dHS1 and HSavg (six break points are identical).

Method	Variable	Stations	Valid breaks	No. of breaks	Detected valid breaks	Efficiency (%)
ACMANT	HSavg	184	31	170	31	18
Climatol	HSavg	184	31	61	17	28
HOMER	HSavg	184	31	32	23	72
Valid	HSavg	184	31	31	31	100
ACMANT	dHS1	184	20	177	17	10
Climatol	dHS1	184	20	43	16	37
HOMER	dHS1	184	20	30	11	37
Valid	dHS1	184	20	20	20	100
Complementary	both	184	45	45	45	100

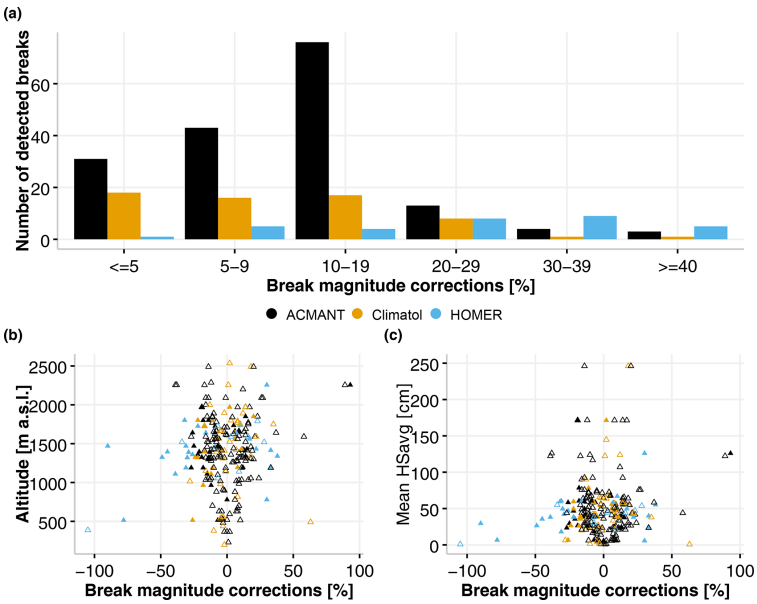


Figure 3. **(a)** Absolute break magnitude corrections and their distribution for ACMANT, HOMER, and Climatol. **(b)** Break magnitude corrections versus station altitude. **(c)** Break magnitude corrections against mean HSavg. Valid break points are bold.

based on the criteria, while Fig. 7 summarizes the stations with identified inhomogeneities based on these same criteria. Multiple detections for series based on these criteria being applied across all three methods indicate that 83 % of the 184 Swiss snow series analysed can be considered homogeneous. However, the individual contributions to valid break points based on these selection criteria is different for each method, and these are summarized in Table 2. ACMANT and Climatol contribute to the detection of 18 % and 28 % respectively of valid break points based on our selection criteria, whereas HOMER accounts for 72 % of the detections. To further support the break points with reference to the available metadata (either station relocation or observer changes), the break points obtained across the methods were compared to any recorded changes in the station history. This comparison with the available metadata found that 22 of the 31 detected break points were supported by metadata and that of these 19 were due to station relocation while 3 could be attributed to a change of observer. The remaining 11 break points had no metadata support in the station histories. Figure 7 (as well as Fig. S1 in the Supplement) summarizes the metadata information supporting the valid break points. This strongly suggests that station relocation is a likely explanation for the majority of the supported inhomogeneities detected in the snow series records analysed here.

Table 3. List of valid break points concordantly identified in 184 investigated Swiss snow time series. Stations are ordered according to break year, and series with break points identified in both HSavg and dHS1 are marked bold. Combination code refers to ACMANT (A), Climatol (C), and HOMER (H).

Code	ACMANT	Climatol	HOMER	Metadata	Metadata type	Combination	Variable	Break year	Altitude (m a.s.l.)
6RP.a	1955	1953			No information	AC	HSavg	1953	1800
x.MVE	1957	1957		1957	Relocation	AC	dHS1	1957	1590
1GH	1958	1960			No information	AC	HSavg	1958	1970
1GS	1966	1966	1965	1964	Relocation	ACH	HSavg	1964	1195
7ST	1967		1968	1968	Observer change	AH	HSavg	1968	1415
SNS	1968	1967		1968	Relocation	AC	dHS1	1968	439
1GB	1970	1970			No information	AC	HSavg	1970	1565
DIS	1971	1970			No information	AC	HSavg	1970	1190
GLA	1970	1970			No information	AC	HSavg	1970	515
STG	1970		1971		No information	AH	HSavg	1970	779
GRA	1970	1969		1971	Relocation	AC	dHS1	1971	634
LUZ	1971		1971	1971	Relocation	AH	dHS1	1971	456
5DF	1972	1972	1975		No information	AC	HSavg	1972	1560
GLA	1975		1975	1975	Relocation	AH	HSavg	1975	515
EBK	1974	1974		1976	Relocation	AC	dHS1	1976	629
6SB	1979		1981		Relocation	AH	HSavg	1979	1640
4SM	1979		1979	1980	Relocation	AH	HSavg	1980	1470
x.SAM	1980	1982	1979	1980	Relocation	AH	HSavg	1980	1726
GUE	1980	1980		1981	Relocation	AC	dHS1	1981	2287
URB	1980	1979		1981	Relocation	AC	dHS1	1981	1395
URB	1980	1980	1980	1981	Relocation	ACH	HSavg	1981	1395
BEH	1982		1982	1983	Observer change	AH	HSavg	1983	2256
ELM	1983	1984			No information	AC	HSavg	1983	965
ABO		1985	1985	1984	Relocation	CH	dHS1	1984	1325
ABO	1984	1981	1984	1984	Relocation	AH	HSavg	1984	1325
SED	1985	1982	1983	1985	Relocation	ACH	HSavg	1985	1432
1LC	1988		1987	1987	Relocation	AH	HSavg	1987	1360
1SM	1989	1990	1989	1988	Relocation	ACH	dHS1	1988	1390
1SM	1987		1988	1988	Relocation	AH	HSavg	1988	1390
2GO	1990	1991		1989	Relocation	AC	dHS1	1989	1110
2GO	1988	1988	1988	1989	Relocation	ACH	HSavg	1989	1110
2RI	1990	1990	1989		No information	ACH	dHS1	1989	1640
5ZV	1990		1989	1989	Observer change	AH	HSavg	1989	1735
ABO	1990		1990		No information	AH	dHS1	1990	1325
FIL	1990		1990	1990	Observer change	AH	dHS1	1990	1030
MUE	1989		1988	1990	Relocation	AH	HSavg	1990	1641
1MR	1991	1991			No information	AC	HSavg	1991	1650
MER		1989	1989	1992	Relocation	CH	dHS1	1992	595
5SP	1994		1993	1994	Relocation	AH	HSavg	1994	1457
3UI	1995		1995	1996	Relocation	AH	HSavg	1996	1340
2ME		1997	1996	1997	Relocation	CH	dHS1	1997	1320
2ME	1997	1997	1995	1997	Relocation	ACH	HSavg	1997	1320
5KK	1998		2000	1997	Relocation	AH	dHS1	1997	1190
5KK	2000		2000	1997	Relocation	AH	HSavg	1997	1190
7MZ	1997	1996		1998	Relocation	AC	HSavg	1998	1850
7AG	2000	1999			No information	AC	dHS1	1999	2090
ZNZ	2000		2001	2001	Relocation	AH	HSavg	2001	1470
SIA	2001	2004	2001	2001	Relocation	AH	HSavg	2001	1802
5SP	2003		2001	2004	Relocation	AH	dHS1	2004	1457
GSS	2008		2008	2009	Relocation	AH	dHS1	2009	897
LUZ	2012	2012			No information	AC	dHS1	2012	456

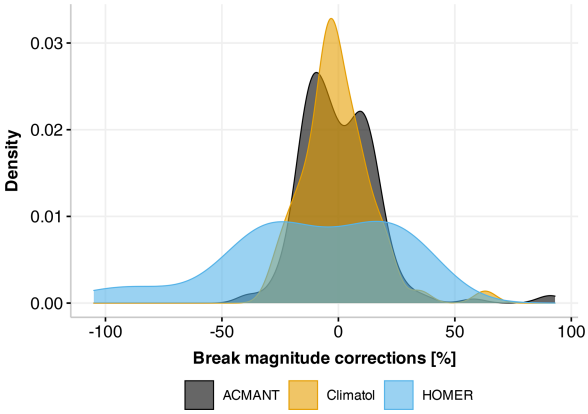


Figure 4. Density plots for break magnitude corrections for ACMANT, Climatol, and HOMER.

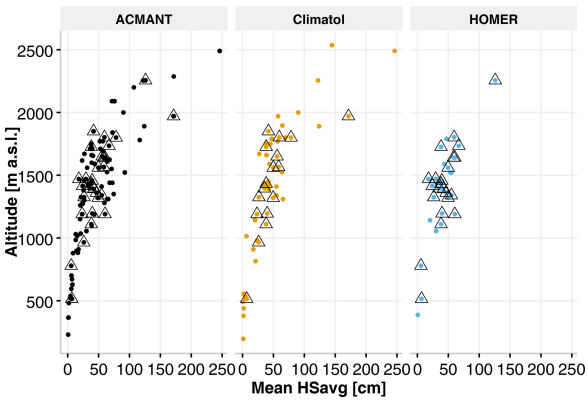


Figure 5. Break points in relation to station altitude and mean HSavg. Valid break points, identified by at least two methods, are marked with black triangles.

4.3 dHS1 as complementary information

The same procedures as outlined in Sect. 4.1 and 4.2 above were applied to another important and commonly used snow indicator: days with snow on the ground (dHS1). Here we found that for all methods except ACMANT, fewer break points are identified compared to the other measures (see Table 2). Here we found the majority of valid break points with the combination of ACMANT and Climatol (compared to AH for HSavg). From a total of 20 valid break points, ACMANT finds 17, followed by Climatol with 16 and HOMER with 11. This suggests that the dHS1 series appear to be more robust in terms of there being fewer breaks than for the HSavg series. Comparing the mean absolute break magnitude corrections for HSavg (14 %) with the ones for dHS1 (5 %) also reveals that the breaks identified in dHS1 are on average smaller than the ones found in HSavg. Table 2 (bottom) summarizes the number of identified breaks for the two indicators. Figure 8 compares the valid break points found for HSavg and dHS1. Six (out of 45) of the break points (13 %) agree, indicating that a complementary approach may be beneficial for detecting valid break points. Figure 7 provides an overview of the valid breaks and metadata information available for both HSavg and dHS1.

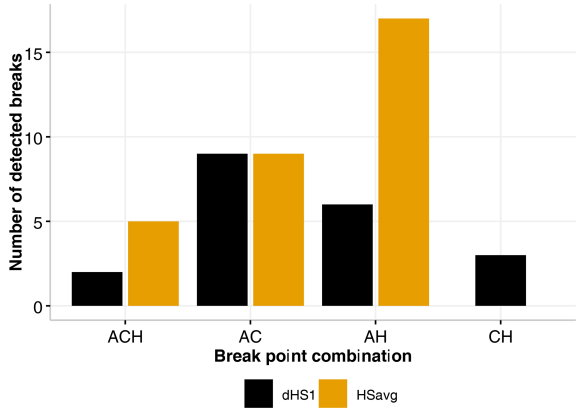


Figure 6. Method combinations for valid break points: ACMANT–Climatol–HOMER (ACH), ACMANT–Climatol (AC), ACMANT–HOMER (AH), and Climatol–HOMER (CH).

5 Discussions

5.1 Break detection

The main differences between the three methods arise from both the way reference series are constructed and how breaks are classified as valid. Reference series in ACMANT and HOMER are constructed based on correlation but with different correlation coefficient thresholds applied for the selection of reference series in the networks, whereas in Climatol, proximity (defined by Euclidean distances) is used. As these different configuration parameters allow for a large number of reference series combinations, as well as of ways breaks are classified, it is no surprise that the results differ between the methods. As the construction of reference series is an intrinsic part of the methods, a comparison with the same input network is not possible.

The detection capability of Climatol depends on the choice of wz (vertical scale parameter) and the SNHT thresholds as outlined by Guijarro (2018). However, the determination of appropriate SNHT thresholds is less straightforward, an issue which has been reported by Kuya et al. (2021a). Lowering the thresholds increased the number of identified break points. However, if the thresholds are too low, multiple breaks within the same years are detected, which implies that the method is too sensitive. Conversely, if the thresholds are too high, no breaks are detected at all. As suggested by Kuya

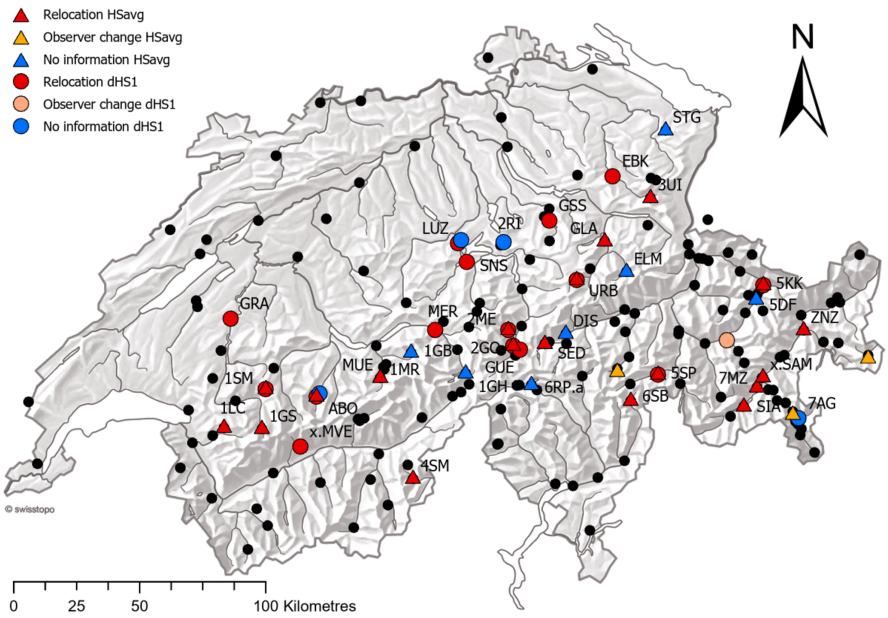


Figure 7. Map highlighting the location of series with identified valid break points and information from metadata where applicable. Series where no break points are detected are marked with black circles.

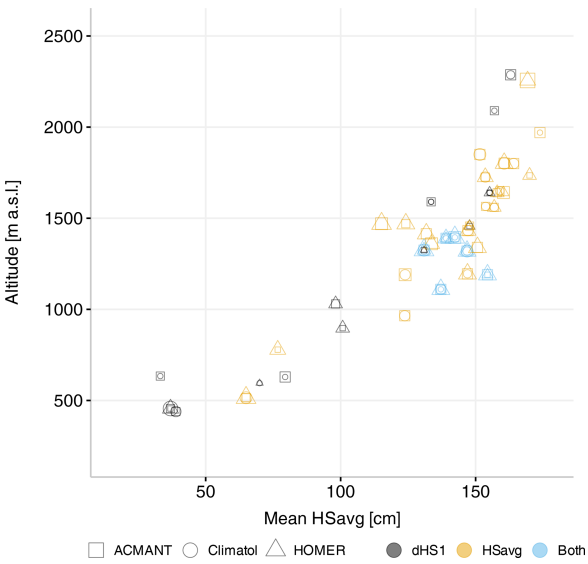


Figure 8. Comparison of valid break points found for HSavg and dHS1. Valid break points, detected by at least two methods, are coloured grey (dHS1), yellow (HSavg), or blue (both). The shape indicates which method detected the break, and the size is the corresponding break magnitude correction.

et al. (2021a), the optimal threshold choice is one which prevents multiple break detections within the same year. The weight of the vertical coordinate (wz) is especially important for snow, as elevation in conjunction with small-scale topographic variation is an important control on snow depth.

ACMANT on account of its automation (and hence no user interaction) is the most objective method, but consequently there are no settings to optimize based on the judgement/experience of a practitioner (Pérez-Zanón et al., 2015). However, there seems to be a sensitivity issue with ACMANT: if using 188 instead of 184 series, the break points for some stations shifted slightly (±2 years) due to different station combinations being used as reference series. This indicates that the break detection algorithm is less robust or more dependent on available stations, especially as the correlation threshold used (0.4) is low. For comparison reasons, we did not perform any pre-treatment outside the actual methods (such as pre-defined neighbourhood networks); however, this might be beneficial for ACMANT (Domonkos and Coll, 2019).

Break detection in HOMER relies on multiple reference series in conjunction with support decisions from an experienced user. While this makes it the most robust application, it is also the most subjective of the three methods used here.

While there are many benchmark data sets for other climate variables (e.g. Willett et al., 2014; Venema et al., 2012), to date there are no such benchmarks for snow, which is unfortunate as these could have helped to thoroughly benchmark the methods. As the method with the highest number of detected breaks, ACMANT seems to be the most sensitive of the methods investigated here (Pérez-Zanón et al., 2015; Fio-

ravanti et al., 2019), followed by Climatol and HOMER with smaller numbers of identified break points. Coll et al. (2020) found similar increments when analysing the break detection capabilities of ACMANT, Climatol, and HOMER for a large network of precipitation series in Ireland. Figure 5 shows no clear relationship between altitude, meanHSavg, and valid break points. Most valid break points are, however, detected at stations between 1000 and 1800 m a.s.l. A possible explanation for this is the large number of available stations in that particular elevation band, thereby ensuring that enough reference series are available. However, the network density for stations above 1800 m a.s.l. and below 1000 m a.s.l. is sparse by comparison, shown by Fig. 5. According to Gubler et al. (2017) the number of detected break points can be reduced by up to 50 % in sparse networks combined with a low signal-to-noise ratio (SNR).

Analysing the standard deviation of the ratio series (sigma, used as a proxy for noise) as well as the median correlations with HOMER reveals that for stations with altitudes below 800 and above 1800 m a.s.l., the range of the median sigma of the five reference series with the lowest sigma per candidate series is larger than for stations between 800 and 1800 m a.s.l. Figure S2 in the Supplement shows that relationship. A similar situation is clear when the median correlations of these subsets per candidate series are considered (Fig. S3). The higher noise (due to not enough suitable reference series being available) associated with these lower (and upper) stations may explain the different results between the networks at different densities and hence is a possible explanation for none of the methods detecting as many break points for stations in those elevation bands.

Break magnitude corrections

Break magnitude corrections and their corresponding density plots (Fig. 4) are similar to the ones found by Coll et al. (2020) working on Irish precipitation series. Differences in break magnitude corrections are also affected by different station neighbourhoods and arising from this the different subsets of stations used for the homogenization process. With a default internal correlation threshold of 0.4, ACMANT's neighbour selections cannot be the same as those of HOMER based on the 0.8 correlation threshold used.

5.2 Choice of method

According to our analyses, HOMER performs better than ACMANT or Climatol when the number of detected valid break points is compared to the number of break points detected overall. However, this comes at a cost based on the time input of the user and also requires extensive expert knowledge. In terms of ease of use based on automation, ACMANT would be the method of choice. However, sifting through the large number of break points detected involves a lot of post hoc processing and also requires in-depth knowledge about the network. Climatol provides the second-best efficiency (in terms of the ratio of valid breaks to total identified break points). However with 17 valid break points detected, 14 would have gone undetected without a combined approach; see Table 2. As outlined in Sect. 5.1 and reported by Kuya et al. (2021a), the results from Climatol heavily depend on the initial set-up. Coll et al. (2020) recommended Climatol as the method to use for precipitation because the identified break points appeared likely to be more realistic than the ones found with either ACMANT or HOMER in the analysis of their network. However, our analysis shows that this might not be the same when applied to snow series in a more topographically complex region. We found rather that a combination of the three methods works best and that HOMER performs significantly better than Climatol or ACMANT in the context of our network setting. This application of multiple methods was also used and recommended by Kuglitsch et al. (2012) and Toreti et al. (2012) for Swiss temperature series, as well as by Marcolini et al. (2019) for Austrian snow depth series. A recent study by Brugnara et al. (2020) applying homogenization methods to the phenological network in Switzerland also recommended this application of multiple methods.

Choice of variable – HSavg versus dHS1

Results from Buchmann et al. (2021b, a) concerning the stability, robustness, and variability in the snow variables HSavg and dHS1 investigated here suggest that HSavg is on average less stable and shows more variation than dHS1. The larger variability associated with HSavg when compared to dHS1 would be expected to lead to larger breaks and thus result in an increased likelihood of break point detection in HSavg. However, associated with these larger variations, there is also increased noise across the series, which has the effect of generally reducing the detection capability for lower-magnitude break points altogether. The variable dHS1, on the other hand, is more stable and shows less variability across the series, suggesting less noise for this variable and therefore break points with lower amplitudes overall. However, only looking at an average variability (or standard deviation) does not necessarily improve our understanding as the temporal evolution of variation of any single station is more important but is also a property affected by inhomogeneities. Analysing the mean absolute break magnitude corrections for the six break points identified in both HSavg and dHS1 reveals that the amplitudes of those detected for dHS1 are significantly smaller than the ones for HSavg. Break magnitude corrections retrieved from HOMER using breaks detected in dHS1 and inserted into HSavg do not differ from the ones obtained through breaks detected purely in HSavg. Based on this finding, break points detected in dHS1 may be used to calculate corrections for HSavg. These results tend to corroborate the hypothesis that the three methods detect smaller break points with dHS1 than with HSavg. For stations below

1000 m a.s.l. the majority of valid break points are detected in dHS1; hence these results tend to support the benefits of a complementary (dHS1 and HSavg) approach. As only six break points are identified with both variables, a complementary approach seems beneficial. Combining the results from both HSavg and dHS1 returns 45 (Table 2) valid break points in 184 Swiss snow time series.

5.3 Homogeneity of Swiss snow series

Our analysis shows the need for a combined use of some of the available methods in order to retrieve a set of break points for Switzerland where we can have higher confidence in their validity. In the majority of cases the methods agree well, with HOMER and Climatol returning the highest proportion of break points deemed valid based on the criteria we applied. Kuglitsch et al. (2012) could explain most of their break points for Swiss temperature series based on a combination of reasons, whereas our results tend to indicate that station relocation is the most likely source of inhomogeneities for the snow series analysed (Coll et al., 2020; Kuya et al., 2021a). However, due to the incomplete nature of the metadata available to support our study, we are unable to investigate this in more detail. For example, the entry "observer change" can mean change of observer only but may unfortunately also imply a combination of relocation and observer change. In spite of having identified relocation as the main explanation for the majority of break points, this is not consistently the case, since from a total of 519 recorded location changes in the station histories across the network, only 45 relocated stations (9 %) produced a valid break point across all three methods. The majority of our identified valid break points are located between 1000 and 1800 m a.s.l. at stations which normally experience continuous snow cover from November to April of the following year. From Fig. 7, the main geographical locations of series with breaks are the northern Prealps, Bernese Alps, and Engadine. The lack of inhomogeneity detections for series south of the Alps (Ticino) can largely be explained by a lack of suitable reference series available. The southern parts of Switzerland are dominated by low valleys surrounded by high mountains. This steep gradient in conjunction with small-scale varying climatic conditions seems to be a reason for not having enough reference series and subsequently inhibits any suitable break detection due to a lack of appropriate reference series.

Comparison with Italy, Austria, and the Alps

Similar break detection investigations have been performed for Italy, Austria, and the Austrian–Swiss domain: Marcolini et al. (2017) applied the SNHT to 106 closely adjacent snow time series for the same region of Italy, and in that work they reported 20 % of the series to be inhomogeneous. In terms of altitudinal and temporal extent, the data set used in that work is similar to ours; however, the 184 stations comprising our network are not from such a climatically coherent region.

Marcolini et al. (2019) investigated only 25 series between 200 and 1600 m a.s.l., whereas we analysed 184 series between 200 and 2500 m a.s.l. but only found valid break points in series between 500 (400 dHS1) and 2300 m a.s.l. (and the majority of these between 1000 and 1800 m a.s.l.). Moreover, Marcolini et al. (2019) found 11 (5 with the SNHT and PRODIGE) breaks in 25 series. They also reported an agreement of 45 % between the SNHT and PRODIGE in identifying the same break points. We only found 20 % (25 % for HSavg and 16 % for dHS1) agreement between Climatol and HOMER. In other work, Schöner et al. (2019) performed a break detection analysis on 96 Swiss snow series between 1961 and 2012 using PRODIGE and reported 25 series (26 %) as inhomogeneous. We found similar numbers: 17 % as inhomogeneous time series when only HSavg breaks are considered and 25 % with the complementary approach (HSavg and dHS1).

Using the complementary approach (HSavg and dHS1), we found similar numbers of potential break points to those that have been reported in previous studies (Marcolini et al., 2017, 2019; Schöner et al., 2019); this is in spite of the time periods and data sets used in our study not being identical. However, if we only compare the break points found in HSavg, our results show fewer break points than have been identified by the previous studies. Possible explanations for the differences are the combined use of three break detection methods in our work (rather than reliance on one method), as well as our fairly strict criteria for defining valid break points.

5.4 Further issues

The attribution of each valid break point to a single hydrological year is not straightforward, even when combined with the information from the available metadata, as not all series we defined as having valid break points had metadata available or, when metadata were available, they were not necessarily complete or entirely correct.

Furthermore, as the number and location of true break points in our data set remain unknown, statements about the probability of detection and false alarm rates (Brugnara et al., 2019) are not possible.

Stations below 500 m a.s.l. regularly experience winter months with little snow and only a few snow days; thus for a number of years in these records the recorded monthly means can be near zero. This lack of consistent snow records and the large variability (noise) associated with lower stations make it virtually impossible to have sufficient suitable reference series.

The next step to obtain homogenized snow series is to find the best method to use and further validate and subsequently correct the identified break points.

6 Conclusions

We present the first in-depth break point detection comparison of three homogenization methods for snow depth time series. In addition, we present the first detailed homogeneity assessment of Swiss snow depth series using state-of-the-art homogenization methods. Our analyses suggest that using ratios, working with monthly input data and annual detection, and combining the results from three methods (ACMANT, Climatol, HOMER) and two variables (HSavg and dHS1) offer a promising configuration to more accurately identify inhomogeneities in (Swiss) snow depth series.

Treating break points as valid when identified by at least two out of three methods and the application of strict criteria increase the robustness of, as well as the confidence in, the results compared to the use of a single method. For all the methods applied, expert knowledge about the network in question is indispensable. If, however, the practitioner is limited to only one method, based on the data and analysis here the method of choice would be HOMER. However, the approach combining multiple methods introduced here for application to snow depth series is more rigorous as it provides more confidence in the results. Concerning the total set of valid break points, ACMANT and Climatol appear to overestimate, whereas HOMER underestimates, the number of valid break points, for both HSavg and dHS1 based on the criteria used.

We identified 45 valid break points in 41 of 184 (22 %) series investigated using a complementary approach of HSavg and dHS1, of these 71 % could be explained by metadata, and based on the metadata, 88 % of the identified breaks could be attributed to station relocation. At low elevations, we identified a lack of suitable reference series due to many stations having inconsistent snow lie and being associated with large year-to-year variations, possibly masking the signal within the noise. However, further work is required, especially in view of the growing research effort in relation to attribution (break point verification) and homogenization (correction) efforts.

Data availability. Daily manual snow depth measurements in Switzerland are subject to copyright but can be obtained on request directly from the two sources: MeteoSwiss and SLF (data@slf.ch). The basis for our analysis, the monthly input values, is available on request through EnviDat at https://doi.org/10.16904/envidat.297 (Buchmann et al., 2022), as is the corresponding metadata.

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/tc-16-2147-2022-supplement.

Author contributions. MoB performed the analyses, produced the figures, and wrote the draft. The manuscript was written by MoB and JC with inputs from all authors. JA, CM, and MiB compiled the metadata. BC and MiB provided support for ACMANT and HOMER. The study was devised by MoB, SB, and CM with inputs from GR and WS and was supervised by CM.

Competing interests. The contact author has declared that neither they nor their co-authors have any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. Thanks are due to Elinah Khasandi Kuya, Peter Domonkos, and Jose Guijarro for tips regarding HOMER, ACMANT, and Climatol. Thanks are also due to Ross Brown and the two anonymous reviewers, whose comments helped to improve the manuscript.

Financial support. This research has been supported by the Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung (grant no. 175920).

Review statement. This paper was edited by Chris Derksen and reviewed by Ross Brown and two anonymous referees.

References

Aguilar, E. and Llanso, P.: Guidelines on climate metadata and homogenization, World Meteorological Organization, WCDMP-No. 53, https://library.wmo.int/doc_num.php?explnum_id=10751 (last access: 8 June 2022), 2003.

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 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, Int. J. Climatol., 17, 25–34, https://doi.org/10.1002/(sici)1097-0088(199701)17:1<25::aid-joc103>3.0.co;2-j, 1997.

Aschauer, J. and Marty, C.: Providing Data Provision for a Sensitivity Analysis of Snow Time Series, resreport, WSL Institute for Snow and Avalanche Research SLF, research Report for GCOS Switzerland, https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/Internationale-Zusammenarbeit/GCOS/doc/Final_report_Poviding_Data_Provision_for_a_Sensitivity_Analysis_of_Snow_Time_Series.pdf (last access: 8 June 2022), 2020.

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

Begert, M., Zenklusen, E., Häberli, C., Appenzeller, C., and Klok, L.: An automated procedure to detect discontinuities; performance assessment and application to a large

European climate data set, Meteorol. Z., 17, 663–672, https://doi.org/10.1127/0941-2948/2008/0314, 2008.

Brugnara, Y., Good, E., Squintu, A. A., van der Schrier, G., and Brönnimann, S.: The EUSTACE global land station daily air temperature dataset, Geosci. Data J., 6, 189–204, https://doi.org/10.1002/gdj3.81, 2019.

Brugnara, Y., Auchmann, R., Rutishauser, T., Gehrig, R., Pietragalla, B., Begert, M., Sigg, C., Knechtl, V., Konzelmann, T., Calpini, B., and Brönnimann, S.: Homogeneity assessment of phenological records from the Swiss Phenology Network, Int. J. Biometeorol., 64, 71–81, https://doi.org/10.1007/s00484-019-01794-y, 2020.

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, Int. J. Climatol., 41, E2553–E2563, https://doi.org/10.1002/joc.6863, 2021b.

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

Caussinus, H. and Lyazrhi, F.: Choosing a Linear Model with a Random Number of Change-Points and Outliers, Ann. I. Stat. Math., 49, 761–775, https://doi.org/10.1023/a:1003230713770, 1997.

Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, J. Roy. Stat. Soc. C-App., 53, 405–425, https://doi.org/10.1111/j.1467-9876.2004.05155.x, 2004.

Coll, J., Domonkos, P., Guijarro, J., Curley, M., Rustemeier, E., Aguilar, E., Walsh, S., and Sweeney, J.: Application of homogenization methods for Ireland's monthly precipitation records: Comparison of break detection results, Int. J. Climatol., 40, 6169–6188, https://doi.org/10.1002/joc.6575, 2020.

Della-Marta, P. M. and Wanner, H.: A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements, J. Climate, 19, 4179–4197, https://doi.org/10.1175/JCLI3855.1, 2006.

Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), International Journal of Geosciences, 2, 293–309, https://doi.org/10.4236/ijg.2011.23032, 2011.

Domonkos, P.: ACMANTv4: Scientific content and operation of the software, GitHub [code], https://github.com/dpeterfree/ACMANT (last access: 8 June 2022), 2020.

Domonkos, P. and Coll, J.: Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests, Int. J. Climatol., 37, 1910–1921, https://doi.org/10.1002/joc.4822, 2017.

Domonkos, P. and Coll, J.: Impact of missing data on the efficiency of homogenisation: experiments with ACMANTv3, Theor. Appl. Climatol., 136, 287–299, https://doi.org/10.1007/s00704-018-2488-3, 2019.

Fioravanti, G., Piervitali, E., and Desiato, F.: A new homogenized daily data set for temperature variability assessment in Italy, Int. J. Climatol., 39, 5635–5654, https://doi.org/10.1002/joc.6177, 2019.

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, Int. J. Climatol., 37, 4670–4683, 2017.

Guijarro, J. A.: Homogenization of climatic series with Climatol, Climatol manual, https://www.climatol.eu/homog_climatol-en.pdf (last access: 8 June 2022), 2018.

Haberkorn, A.: European Snow Booklet – an Inventory of Snow Measurements in Europe, EnviDat [data set], https://doi.org/10.16904/envidat.59, 2019.

Hawkins, D. M.: Fitting multiple change-point models to data, Comput. Stat. Data An., 37, 323–341, https://doi.org/10.1016/S0167-9473(00)00068-2, 2001.

Killick, R. E., Jolliffe, I. T., and Willett, K. M.: Benchmarking the performance of homogenisation algorithms on synthetic daily temperature data, Int. J. Climatol., 42, 3968–3986, https://doi.org/10.1002/joc.7462, 2021.

Kuglitsch, F. G., Auchmann, R., Bleisch, R., Brönnimann, S., Martius, O., and Stewart, M.: Break detection of annual Swiss temperature series, J. Geophys. Res., 117, D13105, https://doi.org/10.1029/2012JD017729, 2012.

Kuya, E. K., Gjelten, H. M., and Tveito, O. E.: Homogenization of Norwegian monthly precipitation series for the period 1961–2018, Tech. rep., Norwegian Meteorological Institute, mETreport, https://www.met.no/publikasjoner/met-report/_/attachment/download/829d6a8e-1ec8-417d-8231-0268bddac4bb:7eb30ca105a6cf1d470453976bb8d729a64b25a4/METreport04-21-Homogenizationofprecipitation1961-2018.pdf (last access: 8 June 2022), 2021a.

Kuya, E. K., Gjelten, H. M., and Tveito, O. E.: Homogenization of Norway's mean monthly temperature series, Tech. rep., Norwegian Meteorological Institute, mETreport, https://www.met.no/publikasjoner/met-report/_/attachment/download/94481ff9-6366-4569-84f5-881d8e671245:e83e7420a5810019bd2241684070d31d74a35c81/METreport03-2020.pdf (last access: 8 June 2022), 2021b.

Luna, M. Y., Guijarro, J. A., and López, J. A.: A monthly precipitation database for Spain (1851–2008): reconstruction, homogeneity and trends, Adv. Sci. Res., 8, 1–4, https://doi.org/10.5194/asr-8-1-2012, 2012.

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, Int. J. Climatol., 39, 4514–4530, https://doi.org/10.1002/joc.6095, 2019.

Menne, M. J. and Williams, C. N.: Homogenization of Temperature Series via Pairwise Comparisons, J. Climate, 22, 1700–1717, https://doi.org/10.1175/2008jcli2263.1, 2009.

Noone, S., Murphy, C., Coll, J., Matthews, T., Mullan, D., Wilby, R. L., and Walsh, S.: Homogenization and analysis of an expanded long-term monthly rainfall network for the Island of Ireland (1850–2010), Int. J. Climatol., 36, 2837–2853, https://doi.org/10.1002/joc.4522, 2016.

Pérez-Zanón, N., Sigró, J., Domonkos, P., and Ashcroft, L.: Comparison of HOMER and ACMANT homogenization methods us-

ing a central Pyrenees temperature dataset, Adv. Sci. Res., 12, 111–119, https://doi.org/10.5194/asr-12-111-2015, 2015.

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.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (last access: 8 June 2022), 2021.

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.

Szentimrey, T.: Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, World Meteorological Organization, WCDMP-No. 41, https://library.wmo.int/index.php?lvl=notice_display&id=11624 (last access: 8 June 2022), 1999.

Toreti, A., Kuglitsch, F. G., Xoplaki, E., and Luterbacher, J.: A Novel Approach for the Detection of Inhomogeneities Affecting Climate Time Series, J. Appl. Meteorol. Clim., 51, 317–326, https://doi.org/10.1175/JAMC-D-10-05033.1, 2012.

Tuomenvirta, H.: Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data, Int. J. Climatol., 21, 495–506, https://doi.org/10.1002/joc.616, 2001.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89–115, https://doi.org/10.5194/cp-8-89-2012, 2012.

Venema, V. K. C., 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 (last access: 8 June 2022), 2020.

Vertačnik, G., Dolinar, M., Bertalanič, R., Klančar, M., Dvoršek, D., and Nadbath, M.: Ensemble homogenization of Slovenian monthly air temperature series, Int. J. Climatol., 35, 4015–4026, https://doi.org/10.1002/joc.4265, 2015.

Willett, K., Williams, C., Jolliffe, I. T., Lund, R., Alexander, L. V., Brönnimann, S., Vincent, L. A., Easterbrook, S., Venema, V. K. C., Berry, D., Warren, R. E., Lopardo, G., Auchmann, R., Aguilar, E., Menne, M. J., Gallagher, C., Hausfather, Z., Thorarinsdottir, T., and Thorne, P. W.: A framework for benchmarking of homogenisation algorithm performance on the global scale, Geosci. Instrum. Method. Data Syst., 3, 187–200, https://doi.org/10.5194/gi-3-187-2014, 2014.

WMO: Web site of the task team on homogenization, World Meteorological Organization, http://www.climatol.eu/tt-hom/ (last access: 8 June 2022), 2017.

Supplement of The Cryosphere, 16, 2147–2161, 2022
https://doi.org/10.5194/tc-16-2147-2022-supplement
© Author(s) 2022. CC BY 4.0 License.

The Cryosphere Open Access EGU

Supplement of

Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-)automatic homogenization methods

Moritz Buchmann et al.

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

The copyright of individual parts of the supplement might differ from the article licence.

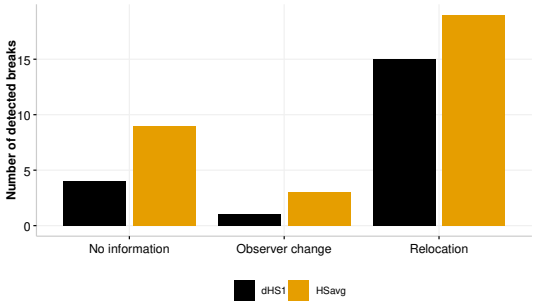


Figure S 1. Metadata type associated with valid break points.

1

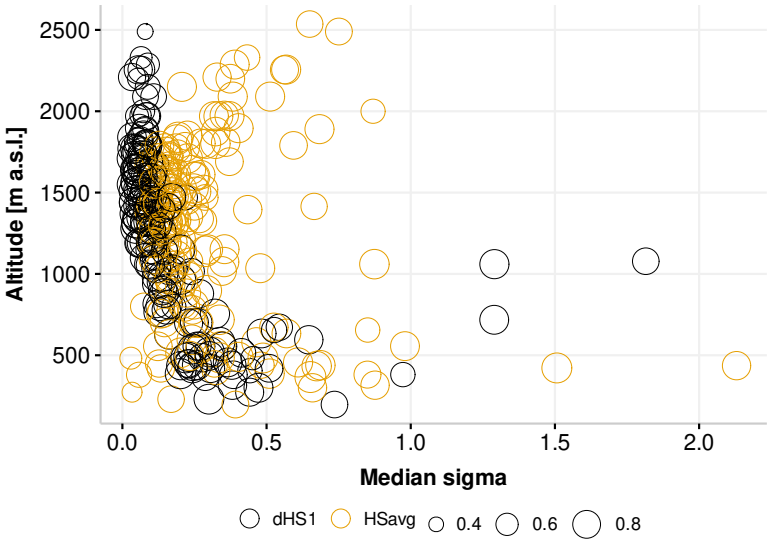


Figure S 2. For both variables HSavg and dHS1, the median sigma of the subset of 5 reference series with lowest sigma for each candidate series is plotted against the altitude of the candidate series. The size of the dots corresponds to the median correlation of these subsets. All correlation and sigma values are from HOMER.

2

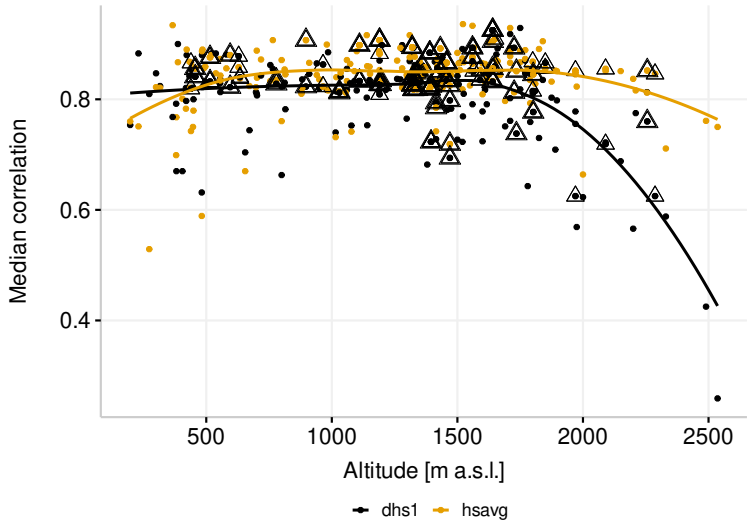


Figure S 3. For both variables HSavg and dHS1, the median correlation of the subset of 5 reference series with lowest sigma for each candidate series is plotted against the altitude of the candidate series. The coloured lines are fitted means for both dHS1 (red) and HSavg (green). Valid break points are highlighted with triangles (break identified by HOMER with thick triangles). All correlation and sigma values are from HOMER.

3

Name	Altitude [m a.s.l.]	mHSavg [cm]	mdHS1 [days]	Name	Altitude [m a.s.l.]	mHSavg [cm]	mdHS1 [days]
MAG	197	1	23	ABO	1325	26	131
6BE	230	1	16	5CU	1330	16	113
LUG	273	1	14	3UI	1340	55	151
RHF	300	1	25	4UL	1345	67	156
BAS	316	1	23	3SW	1350	74	157
OTL	366	1	20	1LC	1360	47	134
GRO	380	2	21	4OW	1370	81	161
AIG	381	1	17	5BE	1380	25	131
BUS	387	1	22	1SM	1390	42	139
CLA	405	0	13	URB	1395	36	142
HLL	419	1	35	4MS	1410	55	148
GVE	420	1	18	7ST	1415	24	132
WYN	422	2	38	SMM	1415	24	135
BIL	433	1	29	5OB	1420	34	144
KLO	436	1	32	5SE	1420	42	149
DEM	439	1	31	r.sus	1432	28	135
SNS	439	2	39	SED	1432	38	147
PAV	447	0	16	2AN	1440	72	164
ALT	449	1	34	ANT	1440	68	162
LUZ	456	2	37	4WI	1450	42	136
SIO	482	2	32	5SP	1457	40	148
KOP	483	1	38	5IN	1460	27	142
NEU	485	1	27	4SM	1470	30	115
PAY	490	1	27	ZNZ	1470	18	124
RAG	496	4	51	7ZN	1475	22	127
GLA	515	7	65	4FY	1500	68	161
5LQ	520	5	53	RIE	1500	40	139
LAN	538	5	54	BOS	1505	63	153
BER	555	2	42	5SA	1510	69	162
CHU	555	4	55	1JA	1520	54	146
SMA	556	2	44	1MN	1520	55	151
1IN	580	3	54	r.lar	1522	92	167
MER	595	6	70	6BG	1525	64	156
EBK	629	8	79	2GA	1550	71	162
GRA	634	1	33	4GR	1560	31	134
VIS	655	3	47	5DF	1560	50	157
THS	672	7	71	x.DAV	1560	51	160
ILZ	698	11	84	1GB	1565	58	154
5TH	701	6	68	x.MVE	1590	44	134
HAI	702	4	70	4ZE	1600	39	145
MAS	718	2	52	ZER	1600	36	147
LAG	755	4	60	3MB	1610	47	155
3WA	765	11	90	5HI	1610	64	162
STG	779	6	77	GRC	1617	23	129
1LB	800	10	95	r.wol	1625	68	159
7BR	800	1	31	2RI	1640	57	155
LTB	812	10	91	6SB	1640	61	158

4

5KU	815	21	119	MUE	1641	58	160
5TC	879	9	79	1MR	1650	56	159
SSE	882	17	112	5MA	1655	41	149
GSS	897	14	101	7SC	1660	39	152
EIN	910	18	112	4BP	1670	27	137
5PU	940	19	101	7CA	1690	57	150
2GU	950	13	93	7FA	1710	40	151
ELM	965	26	124	7LD	1710	53	166
6AM	980	24	118	7ZU	1710	36	147
CHD	985	13	102	DMA	1710	48	158
7PV	1015	6	70	x.SAM	1726	38	154
CDF	1018	15	105	5ZV	1735	67	170
FIL	1030	13	98	7SD	1750	38	151
ENG	1035	21	124	7SN	1750	45	157
GTT	1055	30	124	5BI	1770	52	163
2OG	1060	38	141	BIV	1770	57	167
MAR	1060	2	53	2TR	1780	117	172
CHM	1073	18	109	r.pre	1789	50	160
ROB	1078	9	78	4SF	1790	47	161
5VZ	1090	19	102	6RI	1800	69	158
GOS	1099	39	139	6RP.a	1800	78	164
2GO	1110	38	137	SIA	1802	59	161
AIR	1139	31	126	7MA	1810	76	165
PIO	1141	20	107	1HB	1825	107	168
2SO	1150	20	117	ARO	1840	72	168
CUE	1152	25	117	7MZ	1850	42	152
5KR	1188	43	140	PON	1875	38	152
1GA	1190	55	147	6RO	1890	124	167
3MG	1190	27	127	r.mor	1896	64	164
5KK	1190	60	154	1GH	1970	171	174
DIS	1190	23	124	7BU	1970	58	167
1GS	1195	40	147	BUF	1970	57	163
5RU	1200	18	120	4LA	1975	85	163
6CB	1215	41	137	4SH	2000	90	170
5VA	1260	20	125	7AG	2090	72	157
5SI	1273	24	123	7DI	2090	76	168
1WE	1280	24	135	7MT	2150	64	165
2ST	1280	59	154	4RU	2200	107	165
SCU	1298	21	118	4KU	2210	88	168
1LS	1300	27	120	7BP	2255	122	169
3BR	1310	66	155	BEH	2256	126	169
3FB	1310	64	156	GUE	2287	172	163
1MI	1320	49	146	7AL	2330	83	174
2ME	1320	50	147	SAE	2490	246	164
1AD	1325	21	129	5WJ	2536	145	178

Table S 1: List of all stations used in this study, along with altitude values for mean HSavg and mean dHS1.

5

Chapter 5

Impact of homogenisation on extreme snow depth and decadal trends of snow days, mean and maximum snow depth: lessons from the Alps

Status: Manuscript in preparation for submission
Data set will be made available on EnviDat
Author contributions: Buchmann and Resch contributed equally to this manuscript by performing the analyses, producing the figures, and writing the draft manuscript. The study was designed by MBU, CM, and WS with inputs from GR, BC, MBT and SB. CM supervised the work.

By making use of the robust breakpoints identified in the Swiss snow network, the impacts of multiple homogenisation methods (HOMER, Climatol, and interpQM) are compared using trends and return levels for extreme values.

Impact of homogenization on extreme snow depths and decadal trends of snow days, mean and maximum snow depths: lessons from the Alps

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

Abstract.

We investigated the impact of homogenization on trends and extreme values of commonly used snow variables such as snow days and seasonal average and maximum snow depth. For the first time, three homogenization methods, Climatol, HOMER, and interpQM, run with the same break points and input data, are applied to the large network of Swiss snow depth time series. This multiple-method approach allows to investigate and quantify the various impacts of these methods. Here we found that for trends of seasonal average snow depth, the homogenization methods agree quite well, whereas for seasonal maximum snow depth and the corresponding extreme values, differences across the methods are visible, with interpQM performing slightly better than Climatol or HOMER. The main cause for these differences is the way the reference series are constructed. We found that a carefully constructed sub-network of reference series outweighs the benefits gained from using a (semi-) automatic homogenization method in terms of quality and robustness of the results. Using snow depth in combination with a complex terrain, the combined used of multiple selection criteria is superior to only using correlations or distances respectively.

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

1 Introduction

Long-term measurements are necessary for climate monitoring and climatological analyses. However, the longer a time series, the more likely for it to incorporate inhomogeneities due to station relocation, instrument or observer change, change of station environment, or a combination of the above. These unknown inhomogeneities can have significant influences on trends (Begert et al., 2005) and extreme values (Kuglitsch et al., 2009). To address that problem, the time series can be homogenized by first identifying the break points or inhomogeneities and then adjusting the time series accordingly in the next step.

Widespread used metrics to describe the snowpack include snow days and maximum and average snow depths. Snow days are especially important for ecology (e.g. Stone et al., 2002; Jonas et al., 2008) and climatology (e.g. Scherrer et al., 2004; Marty, 2008), whereas the average and maximum snow depths are more applicable to climatology and engineering applications. Trends and extreme value analyses of snow variables (e.g. Scherrer et al., 2013; Matiu et al., 2021) are relevant for multiple applications such as 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), and definitions of snow loads and limits for building-codes (e.g. 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., 2022; Resch et al., 2022), it is still not clear which of the many proposed homogenization approaches work best and what the impacts from these various methods are on trends and extreme values. So far, Swiss snow depth time series have not been homogenized and no impact assessment has been carried out.

We are using the set of break points, recently identified by Buchmann et al. (2022) for manual Swiss snow series, as input for three modern homogenization methods; Climatol(Guijarro, 2018), HOMER (Caussinus and Mestre, 2004; Domonkos, 2011; Picard et al., 2011; Guijarro, 2018), and interpQM (Resch et al., 2022). These methods are then applied to the large network of Swiss snow depth time series. This enables us to calculate and compare the impact of homogenization on trends of seasonal mean and maximum snow depths, and days with snow on the ground as well as extreme values of maximum snow depths. To focus our study, our research questions are the following:

1. How do the homogenized series compare across the three methods used (Climatol, HOMER, and InterpQM)?
2. What is the impact of homogenization on decadal trends of average snow depth?
3. How do the three homogenization methods influence important metrics such as snow cover duration and maximum snow depths?
4. To what extent are maximum snow depths with a 50-year return period affected by the various homogenization 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.

2 Data

Daily manual snow depth measurements (HS) from 184 Swiss stations are used in this study. Seasonal and monthly average (HSavg), maximum snow depths (HSmax), and the number of snow days (dHS5) are calculated from daily values. Figure 1 shows the station distribution. The stations, maintained by either the Federal Office of Meteorology and Climatology (MeteoSwiss) or the WSL Institute for Snow and Avalanche Research (SLF) provide data within the period from 1931 to 2021 and span from 200 to 2500 m a.s.l. Only stations with complete data between November and April and at least 30 years of data are considered. For a detailed description of the data set, please see Buchmann et al. (2022).

2.1 Break points

We use 45 valid break points (found in 40 series), obtained with a combined use of ACMANT (Domonkos, 2011), Climatol(Guijarro, 2018), and HOMER (Caussinus and Mestre, 2004; Domonkos, 2011; Picard et al., 2011; Guijarro, 2018), where break points are treated as valid if they are at least identified by 2-out-of-3 methods, see Buchmann et al. (2022) for details. Break points are detected with seasonal values. Available metadata was used to verify break points where applicable. The analysis described in Buchmann et al. (2022) is re-run with the available qualified Austrian data series (three stations) 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)). Figure 1 shows all 184 series, the inhomogeneous ones are coloured pink. Two series with insufficient data between 1961 and 2021 were removed from the subset of 42 series reported in Buchmann et al. (2022).

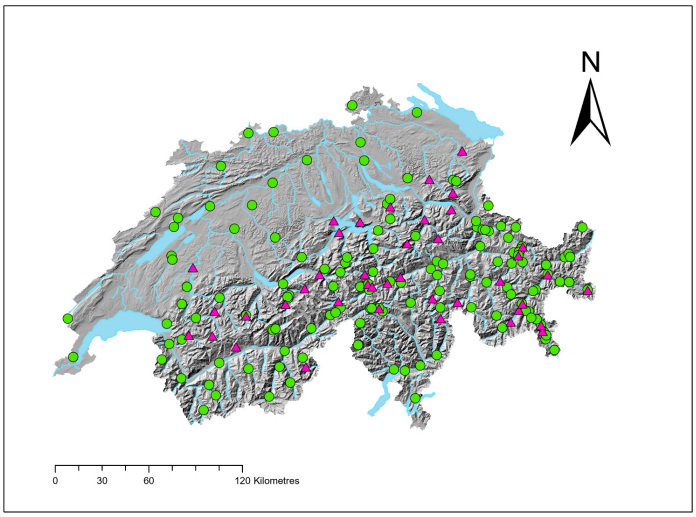


Figure 1. Map of Switzerland showing all 184 stations used in this study. The 42 inhomogeneous stations (identified in this study) with valid break points are highlighted in pink. Green circles are series considered homogeneous. Two series are subsequently removed from the analysis presented here due to insufficient data within the period 1961 to 2021.

3 Methods

3.1 Homogenization methods

Climatol (Guijarro, 2018) is an automatic homogenization function based on SNHT (Alexandersson, 1986) and uses a step-wise detection of homogeneous sub-periods. Reference series are constructed using euclidean distances. Climatol allows the use of a manually adjusted break file which in our case is provided by Buchmann et al. (2022). In Climatol, series are adjusted backwards from the last homogeneous sub-period, by adjusting every detected homogeneous sub-period with an annual factor (see (Guijarro, 2018) for details).

HOMER (Caussinus and Mestre, 2004; Domonkos, 2011; Picard et al., 2011; Guijarro, 2018) is a toolbox which provides many functions for break detection and adjustments like pairwise comparison (PRODIGE), joint-segmentation (Picard et al., 2011) and ACMANT-detection (Domonkos, 2011). Like Climatol, HOMER can be run with a separately compiled break file. In HOMER, series are adjusted with a single annual factor for the entire period before a break point. The adjustment factor is retrieved using variance analysis ANOVA (Caussinus and Mestre, 2004; Mestre et al., 2013) based on the reference sub-networks, which are built in our case by correlations >0.8.

InterpQM (Resch et al., 2022) is an extension of INTERP Vincent et al. (2002), using quantile matching to improve the adjustments. It yields daily homogenized data, which then allows the analysis of subsequently derived snow indicators. For this, the daily measurements of candidate and reference series are split into several interquantile ranges (IQR), which are then compared. An adjustment factor is calculated for each IQR and then linearly interpolated between neighbouring IQR's to avoid artificial jumps. Reference series can either be selected manually or calculated using a weighted mean of a network of selected stations (< 100 km horizontal and < 300 m vertical distance, > 0.7 correlation, no detected breaks), which can be manually refined and optimized by making use of local knowledge and experience. The distribution of the weighting between the stations of a network can either be exponential or linear. For reducing the strong impact of single highly correlated stations on the result, linear distribution of the weighting was chosen. Identified breaks from a break file were used. Snow days ($HS > 0$) are neither created or deleted with the homogenization.

3.2 Snow days, trends, and extreme values

The use of homogenization techniques that adjust daily values allows the analysis of the impact on derived parameters which require daily data for their calculation. Snow days with different thresholds (1 - 50 cm) are commonly used to describe the snowpack (e.g. Abegg et al., 2020; Schmucki et al., 2017). They are defined as the number of days within a time period with a snow depth greater or equal to a specified threshold. Since interpQM does not change the total number of snow days ($HS > 0$), the thresholds of 5, 30 and 50 cm (dHS5, dHS30, dHS50) were used and summed up for each hydrological year between November and April.

Trends are detected using the 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).

Theil-Sen slopes (Theil, 1950; Sen, 1968) are used to estimate the strength of the trends. Decadal trends are expressed as change in [cm/decade]. To compare the homogenized subsets of 40 stations, we focus on the period from 1961 to 2021 since most stations have data within this period. Trends for all available decades are provided in the supplement (Figure A3).

To study the impact of homogenization on extreme snow depths, return levels for seasonal maximum snow depth (HSmax) (Marty and Blanchet, 2012) are calculated for a fixed return period of 50 years based on original and homogenized data (R50HSmax). Calculations are done with the R package extRemes (Gilleland and Katz, 2016) run in standard settings (GEV, estimation method MLE, and 95% confidence intervals).

3.3 Experiment and procedure

We use the subset of 40 stations with identified breaks as input and homogenize this subset with interpQM using daily input values. Additionally, HOMER and Climatol are once run with monthly HSavg and once with monthly HSmax as input for the same subset. Decadal trends are calculated for seasonal HSavg, aggregated either from monthly HSavg (HOMER and Climatol) or daily HS (interpQM). Furthermore, trends for snow days are compared for original and interpQM. Overall maximum HSmax (maxHsmax) are compared for homogenized and original values. Return levels for seasonal HSmax with a 50-year return period (R50HSmax) are retrieved from either daily homogenized HS aggregated into seasonal HSmax (interpQM) or monthly homogenized HSmax (HOMER and Climatol). Trends and R50HSmax for the subset 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), HOMER v2.6, InterpQM (Resch et al., 2022), and R 4.2 (R Core Team, 2022).

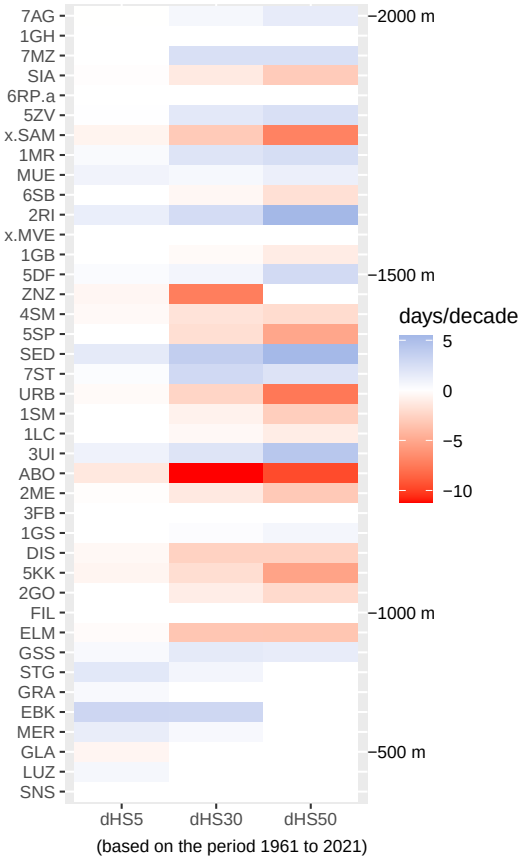


Figure 2. Difference of trends between Original and interpQM for dHS5, dHS30 and dHS50.

4 Results

In the following section, the impact 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.

4.1 Trends of snow days

The seasonal number of snow days was investigated using thresholds of 5, 30 and 50 cm (dHS5, dHS30, dHS50) for the original and homogenized data. This analysis is applied to all homogenized stations and a subset of stations with elavation

>1500m, which is referred to as "high-elevation stations" in this chapter. For the sake of readability, the high-elevation stations are excluded from Table 1. With the exception of one station at 1850m which has no trend (7MZ), all are negative in the original data for dHS5 (mean -4.9 days/decade), where 77% are significant. For dHS30, three stations (1MR (1650m), 3UI (1340m), 7MZ) have a small positive trend between 0.3 - 1.7 days/decade, five stations have no trend, whereas the other 80% (32 stations) have a negative trend (mean: -5.2 days/decade), which is significant for 56% of them. Larger snow depths are less common for lower elevations. Therefore, if calculated for all stations, the proportion with a negative (68%) and significant negative (59%) trend is smaller for dHS50 than for the other thresholds. This elevation-dependent pattern is also shown in Appendix 2, where the difference of the original minus homogenized snow days is shown.

The proportion of stations with a negative trend declines for high-elevation stations for dHS5 from 98% (all stations) to 93%, but increases for both dHS30 (from 80% to 86%) and dHS50 (from 68% to 93%). On the contrary, the proportion of stations with a significant negative trend declines from 59% (all stations) to 50% of high-elevation stations for dHS5, from 56% to 33% for dHS30 and from 59% to 31% for dHS50.

As shown in Figure 2, the performed homogenization with interpQM had the strongest impact on dHS50, with changes between -10 to +5 days/decade. In contrast, the changes for dHS5 were minor, ranging from -1 to +3 days/decade. InterpQM weakened the dHS5-trends for 35% of all stations, strengthened it for 30% and did not change it for 35%. For dHS30, 38% of all stations had weaker trends after the adjustments, 40% had a stronger one and for 22% it did not change. For dHS50, the trend was weakened for 30%, strengthened for 38% and did not change for 32% of all stations. The mean negative trend over all stations for dHS5 was lowered from -4.8 to -5 days/decade, weakened for dHS30 from -5.2 to -4.9 and from -4.9 to -4.3 for dHS30. For the high elevation stations, the dHS5-trend was lowered from -3.3 to -3.4 days/decade, for dHS30 from -4.3 to -4.7 and for dHS50 from -5.4 to -5.8.

While there were no positive trends before the homogenization for dHS5, a small percentage of stations for dHS30 (8% of all, 14% of high-elevation stations) and dHS50 (5%, 7%) was positive whereas in the homogenized data set, they all turned negative. The performed adjustments did not change the proportion of stations with a negative trend for dHS5 (98% if all stations are considered and 93% high-elevation stations). For dSH30 and dHS50 the impact is larger. For dHS30, it went up from 80% of all stations having a negative trend to 88% and from 68% to 78% for dHS50. For the high-elevation stations, the percentage of negative stations did not change but all are now negative for dHS30 and dHS50. The number of stations with significant negative trends increased for almost all thresholds and altitude levels: If all stations are considered, from 77% to 82% for dHS5, 56% to 57% for dHS30 but lowered from 59% to 52% for dHS50. For the high-elevation stations from 69% to 77% for dHS5, 33% to 43% for dHS30 and from 31% to 50% for dHS50.

4.2 Trends of mean and maximum snow depth

To assess the impact of homogenized snow depth series on trends of HSavg and HSmax, decadal trends are calculated for each homogenization method and the original data respectively. Figure 3 shows the trends for Hsavg in panel (a) and Hsmax in panel (b). Trends are expressed as cm/decade for the period from 1961 to 2021 for each method and station of the inhomogeneous subset, black dots indicate significant trends. For Hsavg, we found an overall similar pattern across the methods. Figure A2

Table 1. Statistics for snow days (dHS) for the period 1961 to 2021 on a seasonal basis with thresholds of 5, 30 and 50 cm.

	dHS5		**dHS30**		**dHS50**	
	Original	interpQM	Original	interpQM	Original	interpQM
Median trend [days/decade]	-4.4	-4.5	-4.8	-4.7	-3.6	-3.9
Mean trend [days/decade]	-4.8	-5	-5.2	-4.9	-4.9	-4.3
Positive [%]	0	0	8	0	5	0
Negative [%]	98	98	80	88	68	78
Significant Positive* [%]	0	0	0	0	0	0
Significant Negative* [%]	77	82	56	57	59	52
Positive to Negative [%]	-	0	-	8	-	5
Negative to Positive [%]	-	0	-	0	-	0
Significant to not significant [%]	-	2	-	8	-	10
Not significant to significant [%]	-	8	-	12	-	15

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

shows the trends as differences between original and homogenized values for Climatol, HOMER, and interpQM and both HSavg in panel (a) and HSmax in panel (b). Two original series (7MZ and 3UI) show positive trends, whereas HOMER displays positive trends for 4SM and GLA. No trends are positive with interpQM or Climatol. Except GLA (HOMER), none of the positive trends are significant. Table 2 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. Mean trends of HSavg for Climatol and HOMER appear to be weaker than for Original and interpQM.

Figure 3 panel (a) further reveals that the homogenized 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, 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 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 right side of Table 2 for details. The number of significant trends is about 20% lower than for Hsavg. The order of the methods is maintained with interpQM still showing the largest and HOMER the lowest number of significant trends. The most striking difference between the pattern in panel (a) and (b) of Fig. 3 is the area with no significant trends. For Hsavg, this is located between 1500 and 1600 m a.s.l., whereas for Hsmax it concerns stations below 1000 m a.s.l. The exception is interpQM which shows significant trends for stations below 1000 m a.s.l. There seems to be no particular altitudinal pattern other than that trends below 1000 m a.s.l. are weak across the methods and increase in strength from 1200 to 1400 m a.s.l. This suggests that in contrast to Hsavg, trends for Hsmax seem to be more sensitive to the underlying homogenization methods.

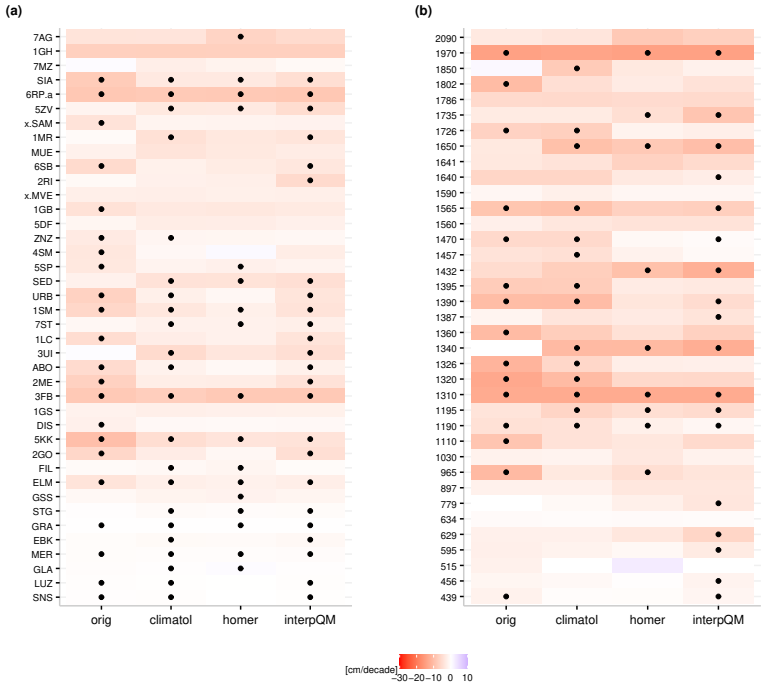


Figure 3. Comparison of trends calculated with original and homogenized data (Climatol, HOMER, and interpQM) for the period 1961-2021. Panel (a) shows trends for HSavg and panel (b) trends for HSmax. Stations are ordered according to elevation (low-high, bottom-top) with an corresponding altitudinal scale in panel (b). Black dots indicate statistical significance with p-values below 0.05.

Table 2. 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.

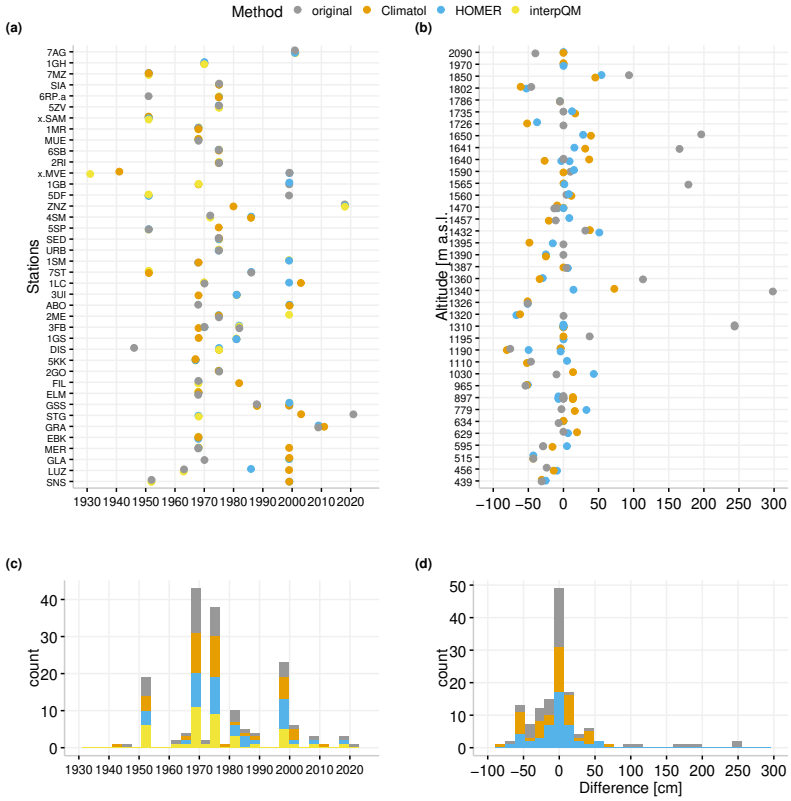


Figure 4. Maximum values of HSmax recorded for each station and method. Panel (a) shows the year for which the overall maximum snow depth is recorded. Panel (b) displays the differences between original and homogenized values. Panel (c) and (d) are the corresponding histograms.

4.3 Influence on maximum snow depth

To investigate a possible influence of the homogenization procedure on absolute maximum snow depths recorded at each station, the year with the overall maximum snow depth, as well as the difference between original and homogenized maxHSmax are plotted for each station and homogenization method. Figure 4 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 across the methods observed with Climatol (Panel (d) of Figure 4). This again suggests that, in contrast to the trends of HSavg, differences across the methods are more visible for HSmax. Furthermore, panel (c) of Figure 4 clearly highlights the four snow-rich winters of 1951, 1968, 1975, and 1999.

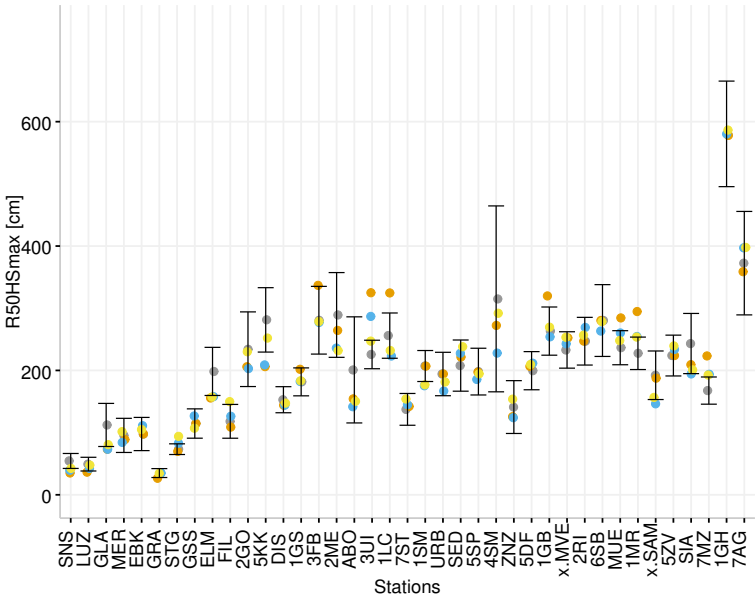


Figure 5. HSmax with 50-year return periods and 95% confidence intervals for both original (grey) and data homogenized with Climatol (orange), HOMER (blue), and interpQM (yellow). The black whiskers represent the 95% confidence interval for the original values. Stations are ordered left to right according to elevation.

Return levels for 50-year return periods of maximum snow depth (R50HSmax) are calculated from homogenized data and compared to values obtained from the original data including the 95% confidence intervals. Figure 5 shows the original values in black with their associated 95% confidence intervals and the homogenized values in colour. Here we found a similar pattern for a majority of stations across the methods. However, between six (interpQM) and 13 (Climatol) stations have R50HSmax outside the 95% confidence intervals of the original values. For Climatol, seven are above and six below, HOMER shows four above and seven below, whereas for interpQM the numbers are three above and three below, see Table 3 for details. This again suggests that the differences between the homogenization methods are more pronounced for R50HSmax than for trends of HSavg, with interpQM performing slightly better than Climatol or HOMER.

Table 3. Statistics for R50HSmax: Number and percentage of stations that are outside the origininal's 95% confidence intervals for each homogenization 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 analysis of dHS5, which shows almost no differences between the original and the homogenized series, confirms the stability of this metric as reported in Buchmann et al. (2021b). The altitudinal pattern for dHS30 and dHS50 can be explained by the fact that stations below 1000 m a.s.l. hardly ever experience dHS30 and dHS50 and therefore small changes can have an influence. Those high-elevation stations that show large differences between trends before and after the homogenization in Fig. 2 (SIA, x.Sam, 6SB, ZNZ, 4SM, and 5SP) are all located in the Engadin, the Southern Alps, or at the border between these two areas. Especially the Engadin, a high-elevated, inner alpine valley with its dry and cold climate, is normally 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 data) are negative, corresponding to findings from previous snow studies (e.g. Laternser and Schneebeli, 2003; Marty, 2008; Scherrer et al., 2013; Matiu et al., 2021). Marcolini et al. (2019) report an increase in series showing significant trends for HSavg after homogenization (40 to 44%). We can see the same effect for interpQM, but not for Climatol (no change) and HOMER, which shows a decrease of significant trends after homogenization (Table 2). The same increase can be observed for trends in 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.

For the majority of stations, R50HSmax from homogenized data are still within the 95% confidence intervals of the original values. However, depending on the homogenization method, between 3 to 7 (see Table 3) stations show R50HSmax exceeding the original values beyond the 95% confidence intervals, with potential implications for engineering applications and building codes. Values well above the 95% intervals are predominantly from Climatol. The reference networks in Climatol are built with the help of Euclidean distances between candidate and reference series, with an optional scale-factor for the vertical component. We set this threshold to $wz = 100$, as in that configuration, the station pair Davos (1570 m a.s.l.) and Weissfluhjoch (2535 m a.s.l.) cannot be selected. However, it maybe the case that this threshold is just not low enough to inhibit 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 station.

The observed differences across the methods can be explained by the various intrinsic means of handling the construction of the sub-networks of reference series and the adjustments. HOMER adjusts the entire period before an identified break point with a single factor, whereas Climatol uses multiple factors dependent on the reference series constructed using homogenized sub-periods. InterpQM on the other hand, uses multiple adjustment factors based on quantile matching. The range of the applied adjustments for interpQM is shown in Appendix A1.

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

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

6 Conclusions

This study presents the results of the first indepth comparison of different homogenization methods applied to a large network of snow depth series. It analyses their impact on extreme snow depths as well as on decadal trends of snow days with different thresholds, mean and maximum snow depths. The results confirm the importance of homogenization (e.g. Auer et al., 2007; Venema et al., 2020), especially for drawing conclusions from long-term time series due to the impact on derived indices. Regarding trends of HSavg and dHS5, the overall picture does not change when homogenized data is used instead of the original. However, the picture becomes more pronounced with interpQM showing the strongest effect with only negative trends for HSavg and a slight increase in the number of significant ones. However, differences across the homogenization methods are more visible when maximum values are considered: trends of HSmax, overall maximum snow depths and extreme values. In terms of snow days, the adjustments had the strongest impact on the larger thresholds.

A separation of break point detection and adjustment using the robust combined detection approach described in Buchmann et al. (2022) in conjunction with the adjustment procedure from Resch et al. (2022) looks promising as the ability to conveniently intervene with the building of the sub-networks is key. The possibility to use a combination of multiple selection criteria such as correlations, horizontal, and vertical distances as well as manual interference seems beneficial to only using one when dealing with snow in conjunction with a complex topography as present in the Alps.

As the break detection step is best done separately, there is no incentive or obvious benefit in using only automatic homogenization methods like HOMER and Climatol. To obtain reasonable results, a degree of user interference is needed for all

the methods, so the most efficient way would be to choose interpQM, which provides the most flexibility and control, with the additional benefit of daily output values. On the other hand, pre-defining sub-networks for HOMER and Climatol using thresholds for correlation, horizontal and vertical distances should also be possible, albeit laborious. Furthermore, the added value of daily output resolution of interpQM in combination with a fast and modern code which is not reliable on outdated libraries suggests a promising and future-proof approach.

Code and data availability. Data and R code are available on EnviDat envidat.ch/someLink

Author contributions. TEXT

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 and evaluation, R 4.2 (R Core Team, 2022) with the tidyverse package (Wickham et al., 2019) was used. We would also like to thank all snow observers. Without their continuous effort, this study would not have been possible.

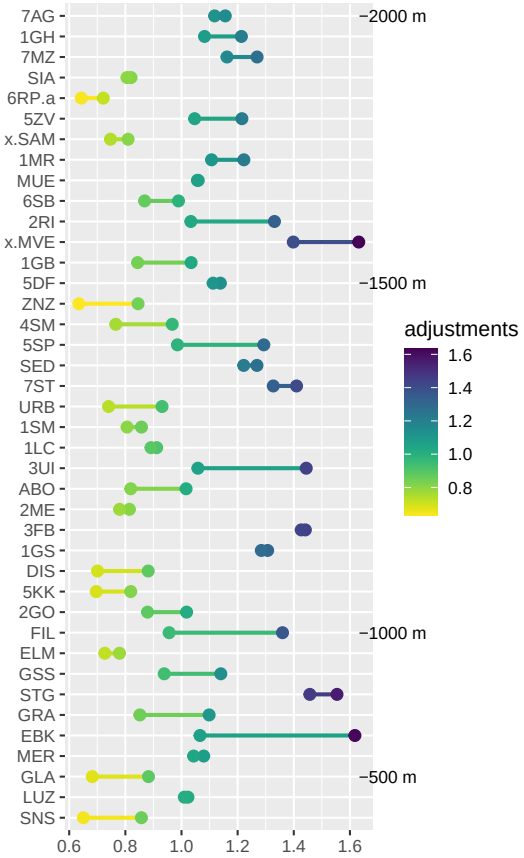


Figure A1. Adjustment factors for the performed homogenization with interpQM.

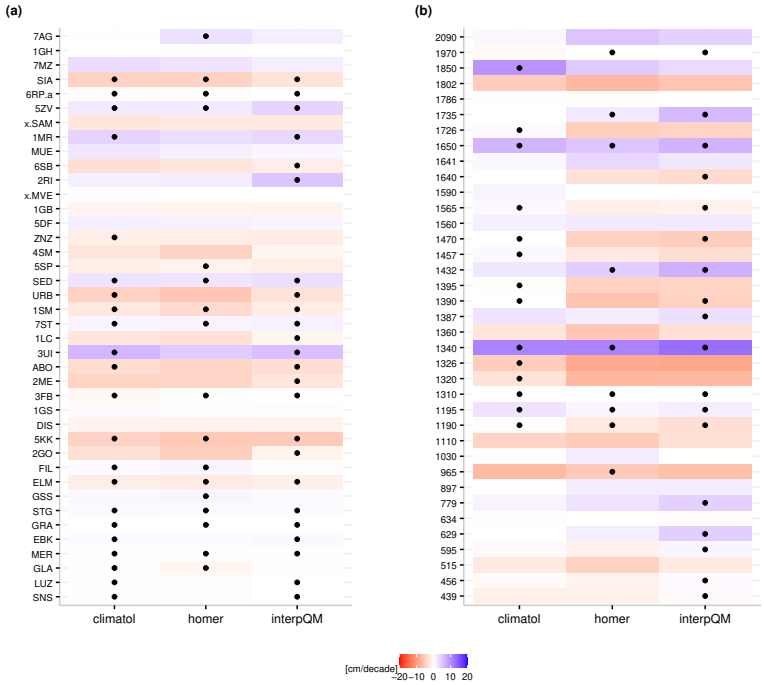


Figure A2. Comparison of differences of trends calculated with Climatol, HOMER, and interpQM for the period 1961-2021.. Differences are expressed as original minus homogenized. Panel (a) shows the differences for HSavg and panel (b) the differences for HSmax. Stations are ordered according to elevation (low-high, bottom-top) with an corresponding altitudinal scale in panel (b). Black dots indicate statistical significance with p-values below 0.05.

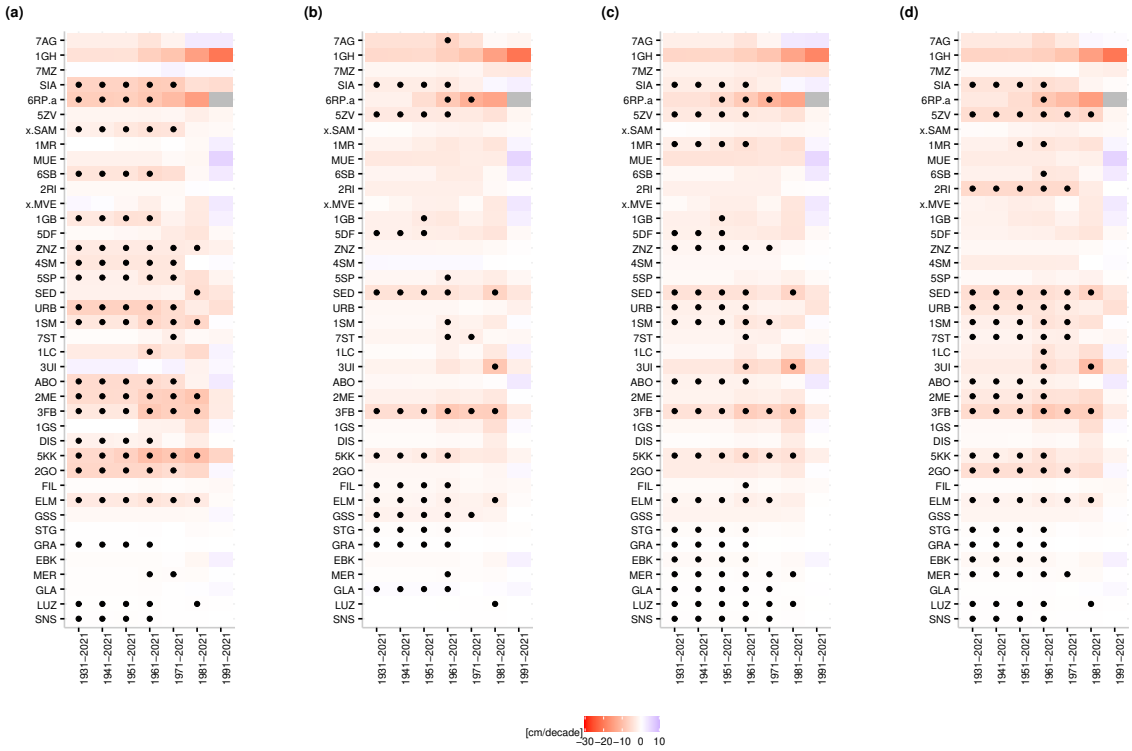


Figure A3. Trends for HSavg: Shown are all methods and all decades. Original (a), HOMER (b), Climatol (c), and interpQM (d).

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., 1986.

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.

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.

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.

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.

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.

Guijarro, J. A.: Homogenization of climatic series with Climatol, 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.

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.

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

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.

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.

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, 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.

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.

22

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.

Chapter 6

Conclusion and outlook

This thesis consists of three published articles, one manuscript in preparation for submission, and a further manuscript as a draft. In the following sections, each article is briefly summarised and the overall conclusions are drawn. The chapter ends with an outlook.

6.1 Paper-focused summaries

P1 Inhomogeneities can occur undetected in long-term snow measurement time series with unknown effects on derived snow climate indicators and climatology. This article investigates the robustness of popular snow climate indicators by analysing a purpose-built set of parallel snow measurements. Under the assumption that not all series are perfectly homogeneous, the aim is to see which of these indicators are more robust or less affected by inhomogeneities. Error metrics such as mean absolute deviations, root mean squared errors, and relative changes are used.

P2 Snow observations and measurements can vary drastically over small distances, which can influence the accuracy of model verification or load factors for building codes. Here we present the first quantitative assessment of local-scale variations of commonly used snow variables. Local-scale variations in onset and disappearance dates as well as extreme values are studied using the parallel data set.

P3 The existence of inhomogeneities in climate series is a well-known problem and there are numerous methods designed to homogenise the series (identify breakpoints and adjust the series), however, there is limited experience in dealing with snow. A homogeneity assessment of 184 Swiss snow depth series is provided by introducing a new approach using three well-established methods and two variables to enhance confidence and robustness, which is especially useful for networks with incomplete or absent accompanying metadata. Having identified 45 valid breakpoints in 41 Swiss snow series via this approach, this paved the way for the adjustment step presented in M4.

M4 Homogenisation is important, but its effects on (Swiss) snow climatology are not well known. Here, an impact assessment is provided by making use of the valid breaks found in P3 and applied corrections to monthly or seasonal data. In spite of recent efforts towards the homogenisation of snow data, it is still not clear which methods work best. Using the results from P3 as input for HOMER and Climatol in combination with a newly developed quantile matching adjustment approach (Resch et al., 2022) allows us to thoroughly study and compare the impacts of these homogenisation methods on trends of the seasonal mean snow depth, the maximum snow depth, and the number of snow days. For HSavg, trends are similar across the methods, however, compared to the original values, the trends are more pronounced and the number of significant trends increased slightly. The results for trends of HSmax are less clear and differ across the methods.

M5 Parallel data sets are important for various verification and analysing purposes. However, there was no set of parallel manual Swiss snow observations. By combining the independent snow series from two institutions (MeteoSwiss and SLF) a unique parallel snow data set is constructed, described, and presented for further analyses and use cases.

6.2 Conclusion

The main aim of the Hom4Snow project was to analyse the impact of homogenisation of snow measurements on trends. To answer that question, initial extensive groundwork was needed to attempt the first homogenisation of the available Swiss snow depth series. This was also preceded by a thorough evaluation of potential snow climate indicators, an in-depth analysis of the natural local-scale variations associated with these indicators and the first homogeneity assessment of Swiss snow series using a robust multi-method approach.

We showed that the homogenisation of snow depth series is possible, however, the quality and availability of suitable reference series is crucial. Moreover, we now have a better understanding of which snow climate indicators are better for climatological analyses when homogenisation is not feasible. The first quantitative assessment of local-scale variations helped us to deal with the variability of snow measurements for multiple applications, alongside the insight knowledge about the quality and availability of various Swiss snow depth series.

The use of a multi-method approach to assess the homogeneity of the Swiss snow depth network is beneficial compared to a single method as the results are more robust. The separation of the breakpoint detection and adjustment step allows the user to maintain control over all the variables and is not method-dependent. The ability to manually adjust the selection of reference series (sub networks) can be important for complex topographies in conjunction with elevation-dependent variables.

Comparing three homogenisation methods with various intrinsic ways of building reference series and different output resolutions, we showed that the homogenisation of snow depth series is generally possible. Results from the various homogenisation approaches do not differ dramatically for trends of HSavg and snow days. However, differences are visible for Hsmax, especially for the overall maximum (maxHSmax) and extreme values (R50HSmax), with interpQM performing slightly better than Climatol or HOMER.

The benefit of using multiple homogenisation methods is that the results are more robust. This can help identify stations that are potentially not feasible to homogenise in the first place. However, using multiple methods is cumbersome; so which one should we choose?

As a certain user engagement is required for all these methods (manual or automatic), we suggest using the one which also provides additional benefits in terms of simplicity, control, or degrees of freedom. The separation of the breakpoint detection and adjustment steps (P3) in conjunction with the adjustment procedures from Resch et al. (2022) looks promising (M4) as the ability to conveniently intervene with the building of the sub-networks is a key feature. Furthermore, the possibility in interpQM to use a combination of multiple selection criteria such as correlations, horizontal, and vertical distances, as well as manual interference, seems beneficial to only using one such criterion (Climatol or HOMER). Especially when dealing with a complex variable, such as snow depth, in conjunction with a complex topography as present in the Alps.

6.3 Open issues and outlook

In spite of the thorough analyses presented here, some issues remain open. So far, the methods are constrained from deleting or creating existing snow days that need to be changed to investigate the impacts on the onset and disappearance dates of homogenised series. Moreover, including the automatic measurement series (hardly used for climatological analyses so far) would increase the network size considerably but possibly introduce new breakpoints. And last but not least, instead of snow depth HS, the amount of new snow HN (snowfall) needs to be looked at in terms of inhomogeneities and potential homogenisation. Furthermore, constructing snow depth series from homogenised temperature and precipitation series using a model and comparing the results to directly-homogenised snow series would be interesting.

Despite these open issues, the synthesis of knowledge presented here already allows us to initiate the next important step; that is, to operationalise the Swiss snow depth homogenisation process, i.e. to devise an operational workflow compatible with the existing methods for homogenisation of other meteorological variables used at MeteoSwiss or SLF.

Chapter 7

Literature

Bibliography

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.

Acquaotta, F., Colombo, N., Fratianni, S., Romeo, V., and Barbero, S.: Preliminary Results of a Comparison Study Between Two Independent Snow Networks in North-Western Italian Alps (Piemonte Region), in: Engineering Geology for Society and Territory - Volume 1, edited by Lollino, G., Manconi, A., Clague, J., Shan, W., and Chiarle, M., pp. 113–116, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-09300-0_22, 2015.

Acquaotta, F., Fratianni, S., and Venema, V.: Assessment of parallel precipitation measurements networks in Piedmont, Italy, Int. J. Climatol., 36, 3963–3974, https://doi.org/10.1002/joc.4606, 2016.

Acquaotta, F., Fratianni, S., Aguilar, E., and Fortin, G.: Influence of instrumentation on long temperature time series, Climatic Change, 156, 385–404, https://doi.org/1007/s10584-019-02545-z, 2019.

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J.: Guidance on metadata and homogenization, World Meteorological Organization WMO, URL https://library.wmo.int/doc_num.php?explnum_id=10751, last access: 9 June 2022, 2018.

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/1002/(sici)1097-0088(199701)17:1⟨25::aid-joc103⟩3.0.co;2-j, 1997.

Armstrong and Brun: Snow and Climate, Cambridge University Press, 2009.

Aschauer, J. and Marty, C.: Providing Data Provision for a Sensitivity Analysis of Snow Time Series, resreport, URL https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/Internationale-Zusammenarbeit/GCOS/doc/Final_report_Poviding_Data_Provision_for_a_Sensitivity_Analysis_of_Snow_Time_Series.pdf, research Report for GCOS Switzerland, 2020.

Aschauer, J. and Marty, C.: Evaluating methods for reconstructing large gaps in historic snow depth time series, Geoscientific Instrumentation, Methods and Data Systems, 10, 297–312, https://doi.org/5194/gi-10-297-2021, 2021.

Baronetti, A., Fratianni, S., Acquaotta, F., and Fortin, G.: A quality control approach to better characterize the spatial distribution of snow depth over New Brunswick, Canada, International Journal of Climatology, 39, 5470–5485, https://doi.org/10.1002/joc.6166, 2019.

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.

Begert, M., Zenklusen, E., Häberli, C., Appenzeller, C., and Klok, L.: An automated procedure to detect discontinuities; performance assessment and application to a large European climate data set, metz, 17, 663–672, https://doi.org/1127/0941-2948/2008/0314, 2008.

Beniston, M.: Climatic Change in Mountain Regions: A Review of Possible Impacts, Climatic Change, 59, 5–31, https://doi.org/1023/a:1024458411589, 2003.

Beniston, M., Rebetez, M., Giorgi, F., and Marinucci, M. R.: An analysis of regional climate change in Switzerland, Theor Appl Climatol, 49, 135–159, https://doi.org/1007/BF00865530, 1994.

Beniston, M., Stoffel, M., and Hill, M.: Impacts of climatic change on water and natural hazards in the Alps: can current water governance cope with future challenges? Examples from the European "ACQWA" project, Environmental Science & Policy, 14, 734–743, https://doi.org/1016/j.envsci.2010.12.009, 2011.

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

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. 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/5194/tc-5-219-2011, 2011.

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.

Brown, R. D., Smith, C., Derksen, C., and Mudryk, L.: Canadian In Situ Snow Cover Trends for 1955–2017 Including an Assessment of the Impact of Automation, Atmosphere-Ocean, 59, 77–92, https://doi.org/10.1080/07055900.2021.1911781, 2021.

Brugnara, Y., Good, E., Squintu, A. A., van der Schrier, G., and Brönnimann, S.: The EUSTACE global land station daily air temperature dataset, Geoscience Data Journal, 6, 189–204, https://doi.org/10.1002/gdj3.81, 2019.

Brugnara, Y., Auchmann, R., Rutishauser, T., Gehrig, R., Pietragalla, B., Begert, M., Sigg, C., Knechtl, V., Konzelmann, T., Calpini, B., and Brönnimann, S.: Homogeneity assessment of phenological records from the Swiss Phenology Network, 64, 71–81, https://doi.org/10.1007/s00484-019-01794-y, 2020.

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

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, 2020.

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, 2021.

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, 2022a.

Buchmann, M., Resch, G., Begert, M., Brönnimann, S., Chimani, B., Schöner, W., and Marty, C.: Impact of homogenisation on trends, In preparation, 2022b.

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 Lyazrhi, F.: Choosing a Linear Model with a Random Number of Change-Points and Outliers, 49, 761–775, https://doi.org/10.1023/a:1003230713770, 1997.

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.

Coll, J., Domonkos, P., Guijarro, J., Curley, M., Rustemeier, E., Aguilar, E., Walsh, S., and Sweeney, J.: Application of homogenization methods for Ireland's monthly precipitation records: Comparison of break detection results, International Journal of Climatology, 40, 6169–6188, https://doi.org/1002/joc.6575, 2020.

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.

Della-Marta, P. M. and Wanner, H.: A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements, J. Climate, 19, 4179–4197, https://doi.org/10.1175/JCLI3855.1, 2006.

Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), International Journal of Geosciences, 02, 293–309, https://doi.org/4236/ijg.2011.23032, 2011.

Domonkos, P.: ACMANTv4: Scientific content and operation of the software, 2020.

Domonkos, P. and Coll, J.: Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests, International Journal of Climatology, 37, 1910–1921, https://doi.org/1002/joc.4822, 2017.

Domonkos, P. and Coll, J.: Impact of missing data on the efficiency of homogenisation: experiments with ACMANTv3, Theoretical and Applied Climatology, 136, 287–299, https://doi.org/1007/s00704-018-2488-3, 2019.

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.

Fioravanti, G., Piervitali, E., and Desiato, F.: A new homogenized daily data set for temperature variability assessment in Italy, International Journal of Climatology, 39, 5635–5654, https://doi.org/1002/joc.6177, 2019.

Foster, J. L.: The Significance of the Date of Snow Disappearance on the Arctic Tundra as a Possible Indicator of Climate Change, Arctic and Alpine Research, 21, 60, https://doi.org/2307/1551517, 1989.

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.

Gray, D. M. and Male, D. M., eds.: Handbook of Snow: principles, processes, management use, The Blackburn Press, Caldwell, NJ, USA, 1981.

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, Hydrological Processes, 29, 1717–1728, https://doi.org/1002/hyp.10295, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10295, 2015.

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.

Guenzi, D., Acquaotta, F., Garzena, D., Baronetti, A., and Fratianni, S.: An algorithm for daily temperature comparison: Co.Temp - comparing series of temperature, Earth Sci Inform, 13, 205–210, https://doi.org/1007/s12145-019-00414-y, 2020.

Guijarro, J. A.: Homogenization of climatic series with Climatol, https://doi.org/13140/RG.2.2.27020.41604, 2018.

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

Haberkorn, A.: European Snow Booklet – an Inventory of Snow Measurements in Europe, https://doi.org/16904/envidat.59, 2019.

Hawkins, D. M.: Fitting multiple change-point models to data, 37, 323–341, https://doi.org/10.1016/s0167-9473(00)00068-2, 2001.

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., and Brönnimann, S.: Identifying, attributing, and overcoming common data quality issues of manned station observations: IDENTIFYING, ATTRIBUTING, AND OVERCOMING COMMON DATA QUALITY ISSUES, Int. J. Climatol., 37, 4131–4145, https://doi.org/10.1002/joc.5037, 2017.

Hänggi, P. and Weingartner, R.: Variations in Discharge Volumes for Hydropower Generation in Switzerland, Water Resour Manage, 26, 1231–1252, https://doi.org/10.1007/s11269-011-9956-1, 2012.

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/1029/2007JG000680, 2008.

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

Killick, R. E., Jolliffe, I. T., and Willett, K. M.: Benchmarking the performance of homogenisation algorithms on synthetic daily temperature data, International Journal of Climatology, n/a, https://doi.org/10.1002/joc.7462.

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.

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, Climatic Change, 139, 637–649, https://doi.org/10.1007/s10584-016-1806-y, 2016.

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.

Kuglitsch, F. G., Auchmann, R., Bleisch, R., Brönnimann, S., Martius, O., and Stewart, M.: Break detection of annual Swiss temperature series, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017729, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012JD017729, 2012.

Kuya, E. K., Gjelten, H. M., and Tveito, O. E.: Homogenization of Norwegian monthly precipitation series for the period 1961-2018, Tech. rep., Norwegian Meteorological Institute, URL https://www.met.no/publikasjoner/met-report/_/attachment/download/829d6a8e-1ec8-417d-8231-0268bddac4bb:7eb30ca105a6cf1d470453976bb8d729a64b25a4/METreport%2004-21-Homogenization%20of%20precipitation%201961-2018.pdf, mETreport, 2021a.

Kuya, E. K., Gjelten, H. M., and Tveito, O. E.: Homogenization of Norway's mean monthly temperature series, Tech. rep., Norwegian Meteorological Institute, URL https://www.met.no/publikasjoner/met-report/_/attachment/download/94481ff9-6366-4569-84f5-881d8e671245:e83e7420a5810019bd2241684070d31d74a35c81/METreport03-2020.pdf, mETreport, 2021b.

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.

Liston, G. E. and Elder, K.: A Meteorological Distribution System for High-Resolution Terrestrial Modeling (MicroMet), Journal of Hydrometeorology, 7, 217–234, https://doi.org/1175/JHM486.1, 2006.

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/1093/aobpla/plw021, 2016.

Luna, M. Y., Guijarro, J. A., and López, J. A.: A monthly precipitation database for Spain (1851–2008): reconstruction, homogeneity and trends, Advances in Science and Research, 8, 1–4, https://doi.org/5194/asr-8-1-2012, 2012.

López-Moreno, J. I., Revuelto, J., Fassnacht, S. R., Azorín-Molina, C., Vicente-Serrano, S. M., Morán-Tejeda, E., and Sexstone, G. A.: Snowpack variability across various spatio-temporal resolutions, Hydrological Processes, 29, 1213–1224, https://doi.org/1002/hyp.10245, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hyp.10245, 2015.

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

Marcolini, G., Bellin, A., and Chiogna, G.: Performance of the Standard Normal Homogeneity Test for the homogenization of mean seasonal snow depth time series, Int. J. Climatol, 37, 1267–1277, https://doi.org/1002/joc.4977, 2017a.

Marcolini, G., Bellin, A., Disse, M., and Chiogna, G.: Variability in snow depth time series in the Adige catchment, Journal of Hydrology: Regional Studies, 13, 240–254, https://doi.org/10.1016/j.ejrh.2017.08.007, 2017b.

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/1002/joc.6095, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.6095, 2019.

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.

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/5194/tc-15-1343-2021, 2021.

Menne, M. J. and Williams, C. N.: Homogenization of Temperature Series via Pairwise Comparisons, Journal of Climate, 22, 1700–1717, https://doi.org/1175/2008jcli2263.1, 2009.

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, 2013.

Morán-Tejeda, E., López-Moreno, J. I., and Beniston, M.: The changing roles of temperature and precipitation on snowpack variability in Switzerland as a function of altitude, Geophysical Research Letters, 40, 2131–2136, https://doi.org/10.1002/grl.50463, 2013.

Neumann, N. N., Derksen, C., Smith, C., and Goodison, B.: Characterizing local scale snow cover using point measurements during the winter season, Atmosphere-Ocean, 44, 257–269, https://doi.org/3137/ao.440304, 2006.

Noone, S., Murphy, C., Coll, J., Matthews, T., Mullan, D., Wilby, R. L., and Walsh, S.: Homogenization and analysis of an expanded long-term monthly rainfall network for the Island of Ireland (1850–2010), International Journal of Climatology, 36, 2837–2853, https://doi.org/10.1002/joc.4522, 2016.

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, Environmental Research Letters, 8, 014 008, https://doi.org/10.1088/1748-9326/8/1/014008, 2013.

Pérez-Zanón, N., Sigró, J., Domonkos, P., and Ashcroft, L.: Comparison of HOMER and ACMANT homogenization methods using a central Pyrenees temperature dataset, Advances in Science and Research, 12, 111–119, https://doi.org/10.5194/asr-12-111-2015, 2015.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and Parker,

D.: Homogeneity adjustments of in situ atmospheric climate data: a review, International Journal of Climatology, 18, 1493–1517, https://doi.org/10.1002/(SICI)1097-0088(19981115)18:13⟨1493::AID-JOC329⟩3.0.CO;2-T, 1998.

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.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/, 2021.

Resano-Mayor, J., Korner-Nievergelt, F., Vignali, S., Horrenberger, N., Barras, A. G., Braunisch, V., Pernollet, C. A., and Arlettaz, R.: Snow cover phenology is the main driver of foraging habitat selection for a high-alpine passerine during breeding: implications for species persistence in the face of climate change, Biodivers Conserv, 28, 2669–2685, https://doi.org/.1007/s10531-019-01786-9, 2019.

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.

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/1111/ecog.05407, 2021.

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

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., Ceppi, P., Croci-Maspoli, M., and Appenzeller, C.: Snow-albedo feedback and Swiss spring temperature trends, Theor Appl Climatol, 110, 509–516, https://doi.org/10.1007/s00704-012-0712-0, 2012.

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, Theor Appl Climatol, 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/1002/joc.5902, 2019.

Seibert, J. and Vis, M. J. P.: Teaching hydrological modeling with a user-friendly catchment-runoff-model software package, Hydrol. Earth Syst. Sci., 16, 3315–3325, https://doi.org/5194/hess-16-3315-2012, 2012.

Sen, P.: Estimates of the regression coefficient based on Kendall's tau, J. Am. Stat. Assoc., 63, 1379–1389, 1968.

Spreitzhofer, G.: Spatial, Temporal and Intensity Characteristics of Heavy Snowfall Events over Austria, Theoretical and Applied Climatology, 62, 209–219, https://doi.org/1007/s007040050085, 1999.

Steger, C., Kotlarski, S., Jonas, T., and Schär, C.: Alpine snow cover in a changing climate: a regional climate model perspective, Climate Dynamics, 41, 735–754, https://doi.org/10.1007/s00382-012-1545-3, 2012.

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.

Szentimrey, T.: Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, 1999.

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

Toreti, A., Kuglitsch, F. G., Xoplaki, E., and Luterbacher, J.: A Novel Approach for the Detection of Inhomogeneities Affecting Climate Time Series, Journal of Applied Meteorology and Climatology, 51, 317–326, https://doi.org/1175/JAMC-D-10-05033.1, 2012.

Tuomenvirta, H.: Homogeneity adjustments of temperature and precipitation series—Finnish and Nordic data, International Journal of Climatology, 21, 495–506, https://doi.org/10.1002/joc.616, 2001.

Venema, V., Trewin, B., and Wang, X.: Guidelines on Homogenization 2020 edition, World Meteorological Organization Issue WMO-No. 1245, URL https://library.wmo.int/?lvl=notice_display&id=21756, last access: 9 June 2022, 2020.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Climate of the Past, 8, 89–115, https://doi.org/10.5194/cp-8-89-2012, publisher: Copernicus GmbH, 2012.

Vertačnik, G., Dolinar, M., Bertalanič, R., Klančar, M., Dvoršek, D., and Nadbath, M.: Ensemble homogenization of Slovenian monthly air temperature series, International Journal of Climatology, 35, 4015–4026, https://doi.org/10.1002/joc.4265, 2015.

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

Vincent, L. A., Milewska, E. J., Wang, X. L., and Hartwell, M. M.: Uncertainty in homogenized daily temperatures and derived indices of extremes illustrated using parallel observations in Canada, International Journal of Climatology, 38, 692–707, https://doi.org/1002/joc.5203, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5203, 2018.

Vorkauf, M., Marty, C., Kahmen, A., and Hiltbrunner, E.: Past and future snowmelt trends in the Swiss Alps: the role of temperature and snowpack, Climatic Change, 165, 44, https://doi.org/1007/s10584-021-03027-x, 2021.

Willett, K., Williams, C., Jolliffe, I. T., Lund, R., Alexander, L. V., Brönnimann, S., Vincent, L. A., Easterbrook, S., Venema, V. K. C., Berry, D., Warren, R. E., Lopardo, G., Auchmann, R., Aguilar, E., Menne, M. J., Gallagher, C., Hausfather, Z., Thorarinsdottir, T., and Thorne, P. W.: A framework for benchmarking of homogenisation algorithm performance on the global scale, Geoscientific Instrumentation, Methods and Data Systems, 3, 187–200, https://doi.org/10.5194/gi-3-187-2014, 2014.

Winter, K. J. P. M., Kotlarski, S., Scherrer, S. C., and Schär, C.: The Alpine snow-albedo feedback in regional climate models, Clim Dyn, 48, 1109–1124, https://doi.org/10.1007/s00382-016-3130-7, 2017.

Wipf, S., Stoeckli, V., and Bebi, P.: Winter climate change in alpine tundra: plant responses to changes in snow depth and snowmelt timing, Climatic Change, 94, 105–121, https://doi.org/10.1007/s10584-009-9546-x, 2009.

WMO: World Meteorological Organization: Web site of the task team on homogenization, URL http://www.climatol.eu/tt-hom/, last access: 9 June 2022, 2017.

WMO, W. M. O.: Guide to Instruments and Methods of Observation, World Meteorological Organization WMO, URL https://library.wmo.int/index.php?id=12407&lvl=notice_display, last access: 9 June 2022, 2018.

Zampieri, M., Scoccimarro, E., Gualdi, S., and Navarra, A.: Observed shift towards earlier spring discharge in the main Alpine rivers, Science of The Total Environment, 503-504, 222–232, https://doi.org/1016/j.scitotenv.2014.06.036, 2015.

Appendix A

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

Status: Published in International Journal of Climatology https://doi.org/10.1002/joc.7742
Author contributions: The study concept was defined by GR, WS and CM. GR, RK, MBT, MBU were involved in the data curation. Used scripts were written by GR, formal analysis was conducted by GR with input for methodology from WS, CM, MBT, BC and JA. The visualization was done by GR with input from CM and WS. The original draft was written by GR with assistance from RK. The paper review was performed by GR, RK, CM, WS, BC, MBU and JA. Funding acquisition was performed by CM and WS.
This paper is included here as it is an integral part of the Hom4Snow project (see Fig. 1.1), serving in particular as the basis for manuscript no. 4 (Chapter 5).

A new quantile based adjustment method (interpQM) is described and assessed making use of the parallel Swiss data set. InterpQM is among the methods used to homogenise the Swiss snow series and the subsequent impact assessment described in Chapter 5.

Received: 18 November 2021 | Revised: 7 April 2022 | Accepted: 26 May 2022

DOI: 10.1002/joc.7742

RESEARCH ARTICLE

International Journal of Climatology RMetS

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

Gernot Resch[1] | **Roland Koch**[2] | **Christoph Marty**[3] | **Barbara Chimani**[2] | **Michael Begert**[4] | **Moritz Buchmann**[3,5,6] | **Johannes Aschauer**[3] | **Wolfgang Schöner**[1]

[1]University of Graz, Department of Geography and Regional Science, Graz, Austria

[2]Zentralanstalt für Meteorologie und Geodynamik ZAMG, Vienna, Austria

[3]WSL Institute for Snow and Avalanche Research SLF, Davos, Switzerland

[4]Federal Office of Meteorology and Climatology MeteoSwiss, Zürich, Switzerland

[5]Institute of Geography, University of Bern, Bern, Switzerland

[6]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

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

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

Abstract

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

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2022 The Authors. *International Journal of Climatology* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

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

1 | INTRODUCTION

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

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

The two most frequently observed variables, snow depth (HS, the total height of the snow cover from the base to the snow surface) and depth of snowfall (HN), are measured at about 7000 stations worldwide (Armstrong et al., 2009). In Austria, snow is measured at about 160 stations by the Zentralanstalt für Meteorologie und Geodynamik (ZAMG) and at about 690 stations by the Hydrographischer Dienst (HD). The snow records of the HD are usually much longer and date back to 1895. The reason for this difference is, that a large part of the Austrian observation data collected by ZAMG was destroyed during the Second World War. While the density of the measurement network for weather stations in inhabited areas is quite good, the station network for snow measurements is patchy, especially in mountainous regions (Hock et al., 2019). Although measurement techniques have changed over time and automatic stations measuring snow depth by ultrasound or laser are becoming more common (Matiu et al., 2021), the method of snow observation has not changed for most stations in Austria and is still done manually at 07:00 CET (Haberkorn et al., 2019). This is a major advantage for the analysis of long-term snow trends, as changes in the measurement technique of climate variables often introduce systematic errors.

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

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

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

2 | DATA AND METHODS

2.1 | Swiss parallel dataset

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

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

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

2.2 | Austrian dataset

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

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

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

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

Gaps with a length of up to 6 years were filled using the gradient-plus-inverse-distance-squared method GIDS if the necessary conditions were fulfilled (Nalder and Wein, 1998). Unfortunately, gaps between 1940 and 1946 often could not be filled due to the low number of reference stations. GIDS is a combination of multiple linear regression (MLR) and inverse distance weighting (IDW). In regression (1), X is the snow depth (to be predicted), Lon (longitude), Lat (latitude) and h (station height) are the predictors. The resulting coefficients b1, b2 and b3 were used to adjust the snow depth of each reference station to the geographical location of the candidate station.

$$X = b_0 + b_1 Lon + b_2 Lat + b_3 h \tag{1}$$

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

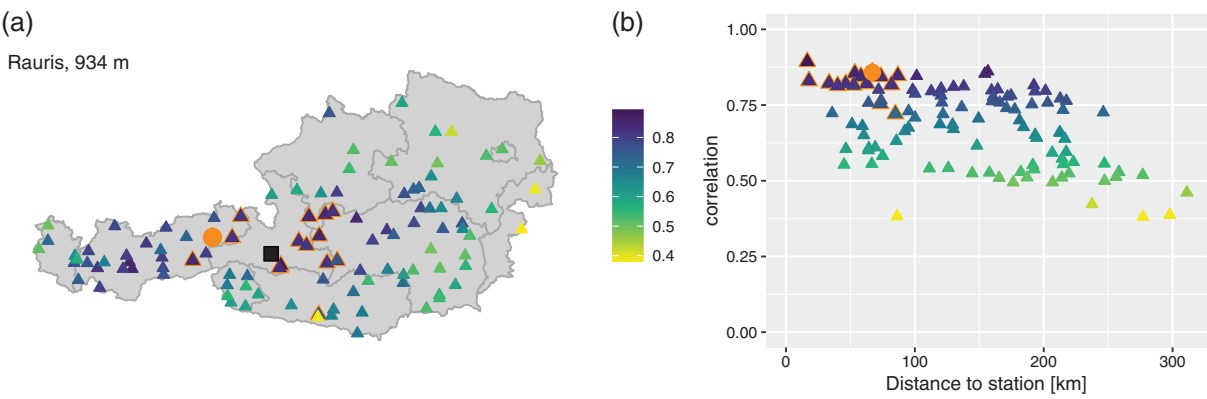


FIGURE 1 Correlations between the Austrian candidate station Rauris (black square) and all other stations in the dataset on a map (a) and against horizontal distance (b). The orange circle indicates the selected reference station, the orange borders show that the necessary conditions for selection as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

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

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

2.3 | Homogenization of snow depth series

2.3.1 | Reference series

To detect breaks in the time series of a candidate station, the observations are compared with either one or several other stations, so called reference stations, which should be subject to the same climate as the candidate time series. Reference stations for snow depth homogenization face several challenges: As shown in Figure 1 using the example of the Austrian station Rauris (934 m), the correlation based on daily values with neighbouring stations does not change linearly with distance. Due to complex orographic situations and related influences such as exposure to shortwave radiation, wind speed and direction, amount of snowfall and temperature, closest stations do not necessarily show the highest correlation. Thus, it can be quite low for stations at a smaller distance, but above an acceptable threshold for stations >200 km away. The amount of snow changes with the vertical distance as well, which is a good reason to exclude stations exceeding a certain elevation distance as it has a strong influence on the adjustment factor.

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

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

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

2.3.2 | Break detection

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

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

Snow depth shows a strong interannual variability, especially compared with long-term trends (Marke et al., 2015). This is caused by its sensitivity to different atmospheric conditions, location and altitude. To overcome these high-frequency variations and to provide robust trends at larger temporal and spatial scales, previous studies of (at least to our knowledge) break-detection in snow time series were based on seasonal data (NDJFMA, DJF). Although there are several different methods for detecting systematic shifts (breaks) in climate data series, to our knowledge only two are known today to be used for detecting breaks in snow height (HS) series: SNHT (Alexandersson, 1986; Alexandersson and Moberg, 1997; Moberg and Alexandersson, 1997) and PRODIGE (Caussinus and Mestre, 2004), which is implemented in ZAMG's homogenization package HOMOP. A comparison of the two methods showed good overall performance for both, with equally credible results. Differences occurred mainly for suspicious fractures at very low snow depths, which were detected by HOMOP but not by SNHT (Marcolini et al., 2019). Due to the assumption of homoscedasticity in the homogeneous time series, both tests are not able to detect inhomogeneities in trends, where inhomogeneities in the trend often show up as a string of single jumps.

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

2.3.3 | Adjustment methods

INTERP

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

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

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

InterpQM

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

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

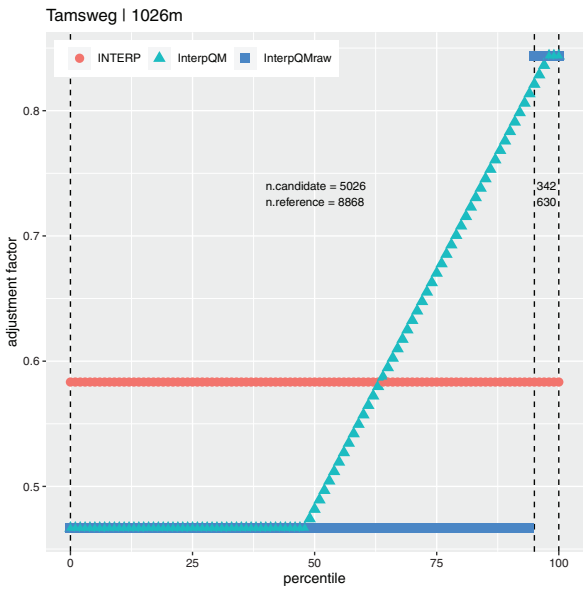


FIGURE 2 Calculated adjustment factors for the second break of the Austrian station Tamsweg based on interquantile ranges (IQR). Red circles: INTERP, blue squares: Adjustment factors of the IQRs (InterpQMraw), turquoise triangles: Interpolated adjustment factors (InterpQM), vertical dashed lines: Separation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

2.3.4 | Evaluation of adjustment methods

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

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

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

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

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

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

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

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

2.3.5 | Impact on time series

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

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

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

3 | RESULTS

3.1 | Break detection

Inhomogeneities in the Swiss dataset of parallel observations were detected applying the SNHT on the quotients of HSmean (NDJFMA) between the candidate and the respective parallel station series: About 66% of the station pairs showed breaks. An overview is presented in Table 2. Between one and two breaks were detected with PRODIGE and confirmed with the SNHT and metadata in six of the 86 Austrian stations (Table 4, Marcolini et al., 2019): Bad Gastein (1092 m), Galtür (1577 m), Oed (400 m), Rauris (934 m), St. Leonhard im Pitztal (1335 m) and Tamsweg (1026 m). A possible explanation for the large difference in the detected number of breaks between SNOWPAT and the parallel dataset are significantly higher correlations between the time series of the station pairs for the latter. This indicates a lower noise of the quotient series, which otherwise would mask smaller breaks. For additional prove, break detection was also performed for the Swiss HN pairs, indicating that for HS and HN different breaks were identified for almost every station pair. This finding could be explained by different influences on HS and HN observations (e.g., ablation processes for HS), different measurement devices and observation and procedures. If compared with variables like temperature (Venema et al., 2012), a lower number of breaks was detected in the analysed snow depth time series.

3.2 | Evaluation of InterpQM

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

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

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

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

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

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

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

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

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

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

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

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

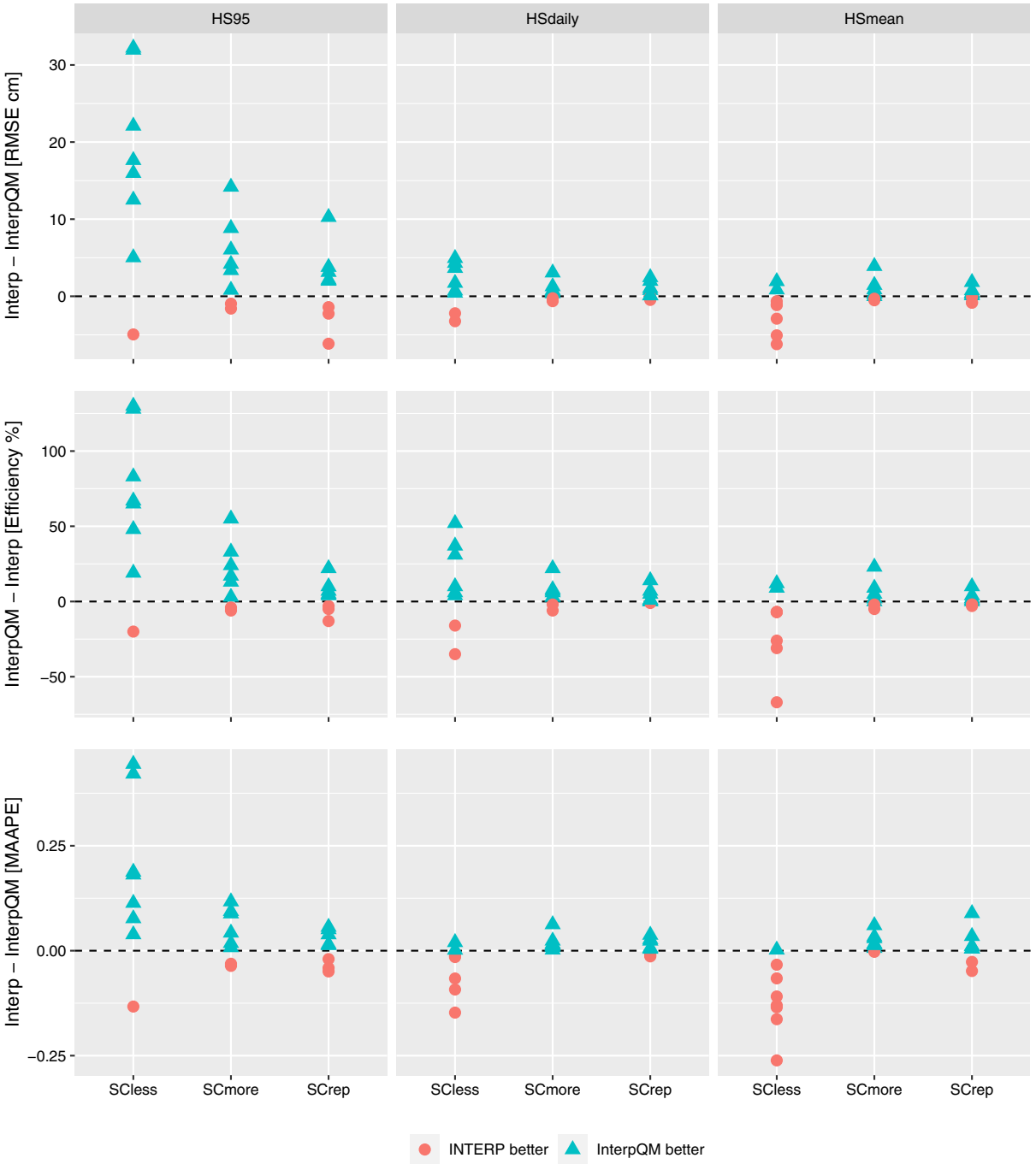


FIGURE 3 Comparison of the error measures (RMSE, efficiency and MAAPE) of the homogenization performed for the three Swiss inhomogeneity scenarios SCless, SCmore and SCrep. Triangles and dots show the difference between the two methods for each station, with triangles showing better performance of InterpQM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

stations of the SCless scenario for RMSE and Efficiency and seven for MAAPE, but in contrast lower RMSE and Efficiency in six (5) and lower MAAPE in seven (6) stations with InterpQM for SCmore (SCrep). Since the MAAPE can be roughly considered as a percentage error, it can be summarized that InterpQM reduced the relative

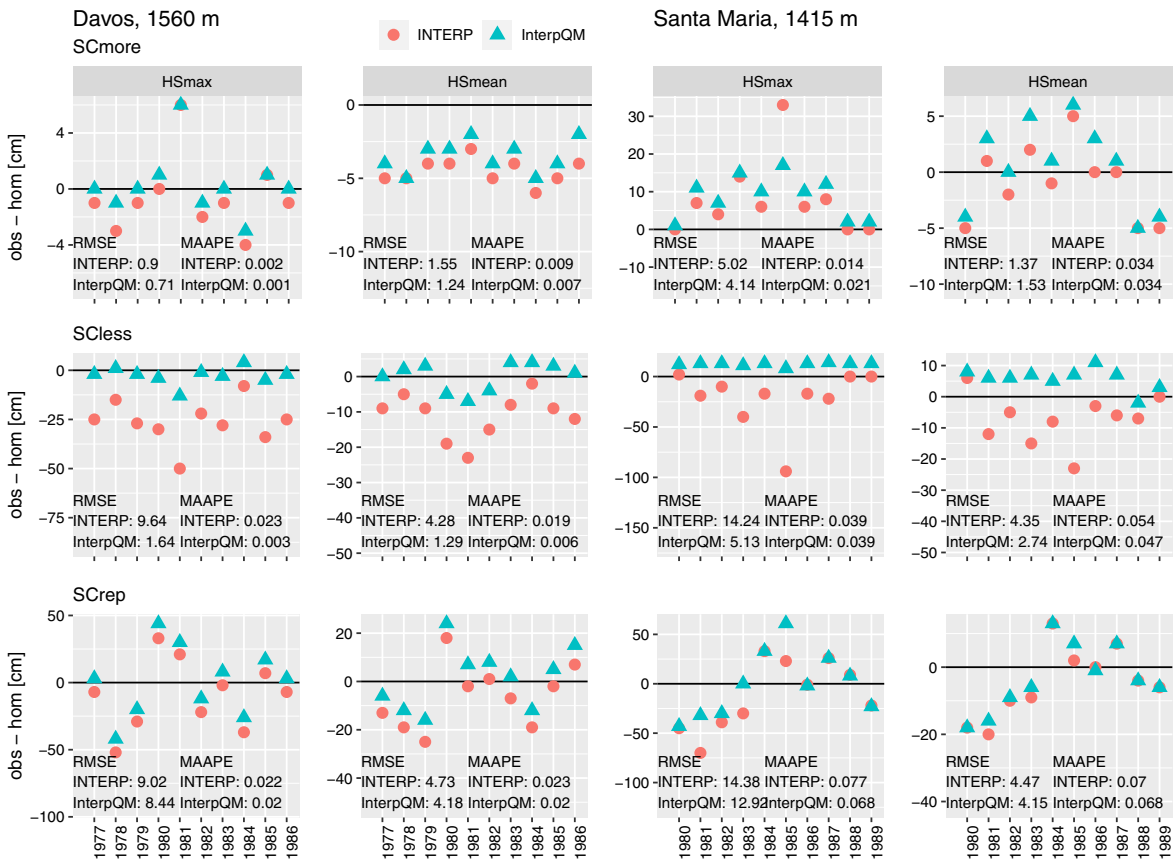


FIGURE 4 Difference between observed and homogenized data for HSmax (maximum seasonal snow depth) and HSmean (mean seasonal snow depth) of the evaluation data set for the Swiss stations Davos and Santa Maria. [Colour figure can be viewed at wileyonlinelibrary.com]

errors of HS95 in SCless between 7% and 19%, between 1% and 4% in SCmore and between 2% and 3% in SCrep better than INTERP. On the contrary, errors of HSmean were 3–26% larger using InterpQM in SCless, but smaller for SCmore (between 1% and 3%) and for SCrep (3–9%). For HSdaily errors were also between 1% and8% larger with InterpQM for SCless but 1–4% smaller for SCmore and 1–3% for SCrep.

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

3.3 | Homogenization of Austrian snow depth series

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

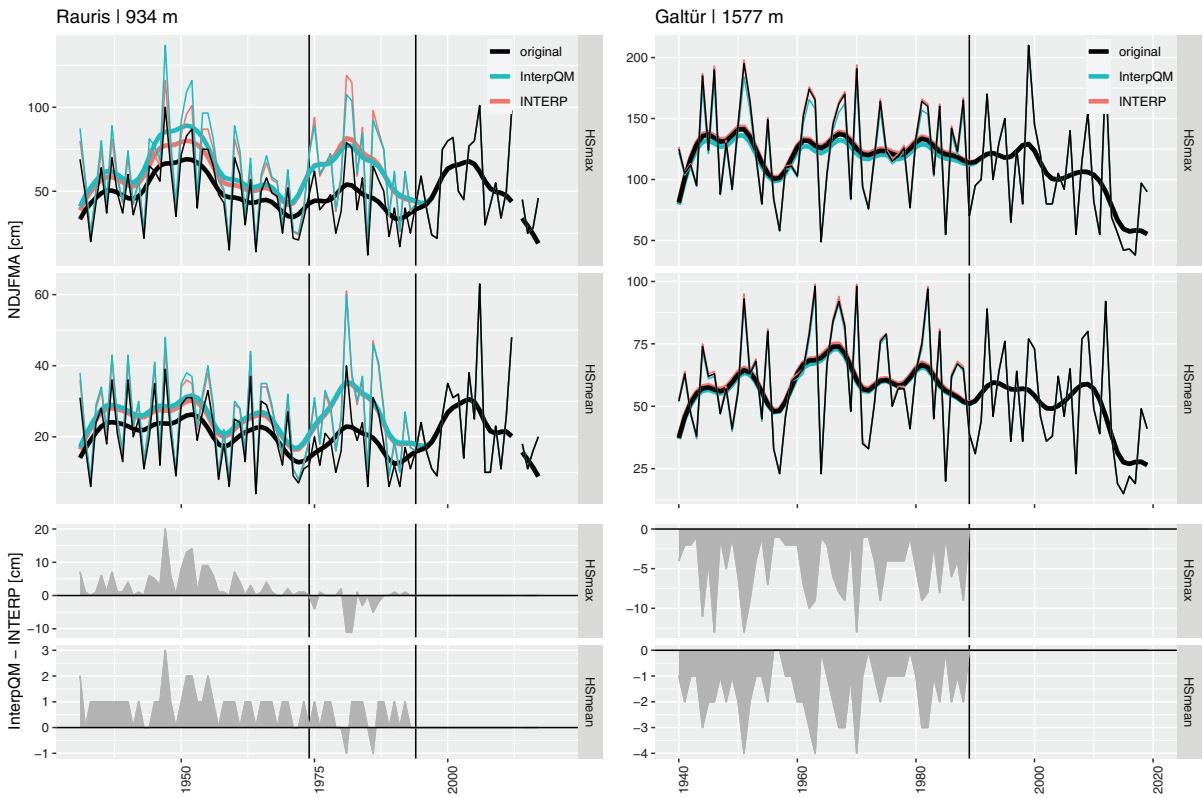


FIGURE 5 Comparison of the original time series and the homogenized versions with INTERP and InterpQM for the Austrian stations Rauris and Galtür for seasonal maximum and mean snow depth. The thick lines are Gaussian filtered with an 11y window. The vertical black lines indicate the breaks in Rauris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

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

3.4 | Impact of homogenization

The Wilcoxon and Kruskal–Wallis Test revealed the different impact of the two adjustment methods to the homogenized series: While the adjustments did not cause a significant change for dHS1 for any of the stations, they had a significant effect for HSdaily of all stations with InterpQM, but only for three with INTERP (Galtür, Rauris and Tamsweg). HSmean was only significant for Bad Gastein and St. Leonhard with InterpQM and Bad Gastein, St. Leonhard, Oed and Tamsweg with INTERP. Figure 6 shows long-term trends of snow depth as the mean change per decade and its significance, calculated for the entire time series for HSmax and HSmean. In the original time series, all stations showed a negative trend for HSmax, only Galtür (−5.7 cm/decade) and Tamsweg (−3.5) were significant. When homogenized with INTERP/InterpQM, the negative trend of Bad Gastein was increased by 2.8/2.4 cm/decade, of Oed by 0.9/2.9 and of Rauris by 0.6/1.3. INTERP weakened the trend of St. Leonhard by 0.7 cm/decade while InterpQM slightly intensified it by 0.1. Both methods reduced it in Tamsweg (1/3.4), where the adjustments by InterpQM were so strong that the trend (−0.1 cm/decade) was almost removed. The trends of Bad Gastein and Oed were also significantly changed by both methods, while only InterpQM removed both the negative trend and its significance of Tamsweg.

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

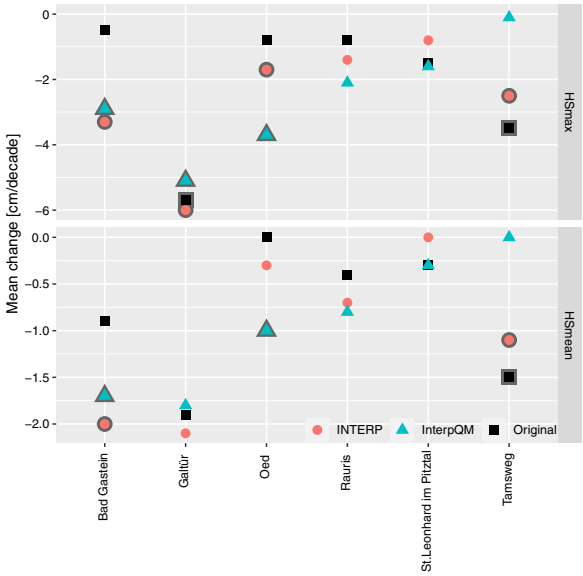


FIGURE 6 Summary of trend and significance tests performed for HSmean and HSmax for the adjusted Austrian stations. Colours and symbols denote the original (black), INTERP (red circle) or the InterpQM version (light blue triangle); outlined symbols indicate a significant (p value < 0.05) trend shown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

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

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

4 | CONCLUSIONS

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

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

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

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

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

AUTHOR CONTRIBUTIONS

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

ACKNOWLEDGEMENTS

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

ORCID

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

REFERENCES

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

Alexandersson, H. and Moberg, A. (1997) Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology*, 17(1), 25–34. https://doi.org/10.1002/(SICI)1097-0088(199701)17:1%3C25::AID-JOC103%3E3.0.CO;2-J.

Armstrong, R., Steffen, K., Monteduro, M., & Sessa, R. (2009). Assessment of the status of the development of the standards for the terrestrial essential climate variables: Snow. In: Global terrestrial observing system, Vol. 10. http://www.fao.org/gtos/doc/ecvs/t12/t12.pdf. Accessed 23 September 2021.

Armstrong, R.L. and Brun, E. (2008) *Snow and Climate: Physical Processes, Surface Energy Exchange and Modeling*. Cambridge, UK: Cambridge University Press.

Aschauer, J., & Marty, C. (2020). Providing data provision for a sensitivity analysis of snow time series. https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/

Internationale-Zusammenarbeit/GCOS/doc/Final_report_Poviding_Data_Provision_for_a_Sensitivity_Analysis_of_Snow_Time_Series.pdf. Accessed 23 September 2021.

Aschauer, J. and Marty, C. (2021) Evaluating methods for reconstructing large gaps in historic snow depth time series. *Geoscientific Instrumentation, Methods and Data Systems*, 10(2), 297–312. https://doi.org/10.5194/gi-10-297-2021.

Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J.-M., Begert, M., et al. (2007) HISTALP: historical instrumental climatological surface time series of the greater alpine region. *International Journal of Climatology*, 27(1), 17–46. https://doi.org/10.1002/joc.1377.

Ballotta, L., Fusai, G., Kyriakou, I., Papapostolou, N.C. and Pouliasis, P. K. (2020) Risk management of climate impact for tourism operators: an empirical analysis on ski resorts. *Tourism Management*, 77, 104011. https://doi.org/10.1016/j.tourman.2019.104011.

Buchmann, M., Begert, M., Brönnimann, S. and Marty, C. (2021) Evaluating the robustness of snow climate indicators using a unique set of parallel snow measurement series. *International Journal of Climatology*, 41(S1), E2553–E2563. https://doi.org/10.1002/joc.6863.

Buchmann, M., Coll, J., Aschauer, J., Begert, M., Brönnimann, S., Chimani, B., Resch, G., Schöner, W. and Marty, C. (2022) Homogeneity assessment of Swiss snow depth series: comparison of break detection capabilities of (semi-) automatic homogenisation methods. *The Cryosphere Discuss.* https://doi.org/10.5194/tc-2022-48.

Caussinus, H. and Mestre, O. (2004) Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(3), 405–425. https://doi.org/10.1111/j.1467-9876.2004.05155.x.

Chimani, B., Venema, V., Lexer, A., Andre, K., Auer, I. and Nemec, J. (2018) Inter-comparison of methods to homogenize daily relative humidity. *International Journal of Climatology*, 38(7), 3106–3122. https://doi.org/10.1002/joc.5488.

Domonkos, P. (2015) Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology*, 122(1–2), 303–314. https://doi.org/10.1007/s00704-014-1298-5.

Domonkos, P., Venema, V., Mestre, O., & France, M. (2012). Efficiencies of homogenization methods: our present knowledge and its limitation. Proceedings of the 7th Seminar and Quality Control in Climatological Databases, 19–32.

Goodison, B., Louie, P.Y.T. and Yang, D. (1998) WMO solid precipitation measurement intercomparison. *Final Report*, 67, 318.

Gubler, S., Hunziker, S., Begert, M., Croci-Maspoli, M., Konzelmann, T., Brönnimann, S., Schwierz, C., Oria, C. and Rosas, G. (2017) The influence of station density on climate data homogenization. *International Journal of Climatology*, 37(13), 4670–4683. https://doi.org/10.1002/joc.5114.

Gultepe, I. (2015) In: Dmowska, R. (Ed.) *Chapter Three - Mountain weather: Observation and modeling*, Vol. 56. Elsevier, pp. 229–312. https://doi.org/10.1016/bs.agph.2015.01.001.

Haberkorn, A., Helmert, J., Leppänen, L., López-Moreno, J.I. and Pirazzini, R. (2019) *European Snow Booklet - an Inventory of Snow Measurements in Europe*. https://doi.org/10.16904/envidat.59.

Hock, R., Rasul, G., Adler, C., Cáceres, B., Gruber, S., Hirabayashi, Y., Jackson, M., Kääb, A., Kang, S., Kutuzov, S., Milner, A., Molau, U., Morin, S., Orlove, B., & Steltzer, H. (2019). High mountain areas. In: IPCC special report on the ocean and cryosphere in a changing climate. https://www.ipcc.ch/srocc/

Johnston, A.N., Bruggeman, J.E., Beers, A.T., Beever, E.A., Christophersen, R.G. and Ransom, J.I. (2019) Ecological consequences of anomalies in atmospheric moisture and snowpack. *Ecology*, 100(4), 1–12. https://doi.org/10.1002/ecy.2638.

Kendall, M. (1975) In: Griffin, C. (Ed.) *Rank correlation methods*, 4th edition. San Francisco, CA: Charles Griffin.

Kim, S. and Kim, H. (2016) A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003.

Lamprecht, A., Semenchuk, P.R., Steinbauer, K., Winkler, M. and Pauli, H. (2018) Climate change leads to accelerated transformation of high-elevation vegetation in the Central Alps. *New Phytologist*, 220(2), 447–459. https://doi.org/10.1111/nph.15290.

Lehning, M., Löwe, H., Ryser, M. and Raderschall, N. (2008) Inhomogeneous precipitation distribution and snow transport in steep terrain. *Water Resources Research*, 44(7), 1–19. https://doi.org/10.1029/2007WR006545.

Lettenmaier, D.P., Wood, E.F. and Wallis, J.R. (1994) Hydro-climatological trends in the continental United States, 1948-88. *Journal of Climate*, 7, 586–607. https://doi.org/10.1175/1520-0442(1994)007%3C0586:HCTITC%3E2.0.CO;2.

Mann, H.B. (1945) Nonparametric tests against trend. *Econometrica*, 13(3), 245–259.

Marcolini, G., Bellin, A. and Chiogna, G. (2017) Performance of the standard Normal homogeneity test for the homogenization of mean seasonal snow depth time series. *International Journal of Climatology*, 37, 1267–1277. https://doi.org/10.1002/joc.4977.

Marcolini, G., Koch, R., Chimani, B., Schöner, W., Bellin, A., Disse, M. and Chiogna, G. (2019) Evaluation of homogenization methods for seasonal snow depth data in the Austrian Alps, 1930–2010. *International Journal of Climatology*, 39(11), 4514–4530. https://doi.org/10.1002/joc.6095.

Marke, T., Strasser, U., Hanzer, F., Stötter, J., Wilcke, R.A.I. and Gobiet, A. (2015) Scenarios of future snow conditions in Styria (Austrian Alps). *Journal of Hydrometeorology*, 16(1), 261–277. https://doi.org/10.1175/JHM-D-14-0035.1.

Matiu, M., Crespi, A., Bertoldi, G., Carmagnola, C.M., Marty, C., Morin, S., Schöner, W., Cat Berro, D., Chiogna, G., de Gregorio, L., Kotlarski, S., Majone, B., Resch, G., Terzago, S., Valt, M., Beozzo, W., Cianfarra, P., Gouttevin, I., Marcolini, G., et al. (2021) Observed snow depth trends in the European Alps: 1971 to 2019. *The Cryosphere*, 15(3), 1343–1382. https://doi.org/10.5194/tc-15-1343-2021.

Mestre, O., Gruber, C., Prieur, C., Caussinus, H. and Jourdain, S. (2011). SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50(11), 2343–2358. https://doi.org/10.1175/2011jamc2641.1.

Moberg, A. and Alexandersson, H. (1997) Homogenization of Swedish temperature data. Part II: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology*, 17(1), 35–54. https://doi.org/10.1002/(sici)1097-0088(199701)17:1<35::aid-joc104>3.3.co;2-6.

Nalder, I.A. and Wein, R.W. (1998) Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest.

Agricultural and Forest Meteorology, 92(4), 211–225. https://doi.org/10.1016/S0168-1923(98)00102-6.

Nemec, J., Gruber, C., Chimani, B. and Auer, I. (2013) Trends in extreme temperature indices in Austria based on a new homogenized dataset. *International Journal of Climatology*, 33(6), 1538–1550. https://doi.org/10.1002/joc.3532.

Olefs, M., & Koch, R. (2020). SNOWPAT: long-term homogenized snow observations in Austria. https://data.ccca.ac.at/dataset/snowpad_2019-v01/resource/36f3a5ac-fde2-492f-8248-adf79b26ee8e. Accessed 24 September 2021.

Olefs, M., Koch, R., Schöner, W. and Marke, T. (2020) Changes in snow depth, snow cover duration, and potential snowmaking conditions in Austria, 1961–2020: a model based approach. *Atmosphere*, 11(12), 1330. https://doi.org/10.3390/atmos11121330.

Prein, A.F. and Gobiet, A. (2017) Impacts of uncertainties in European gridded precipitation observations on regional climate analysis. *International Journal of Climatology*, 37(1), 305–327. https://doi.org/10.1002/joc.4706.

Pröbstl-Haider, U., Lund-Durlacher, D., Olefs, M. and Prettenthaler, F. (2021) Tourismus und Klimawandel. In: Pröbstl-Haider, U., Lund-Durlacher, D., Olefs, M. and Prettenthaler, F. (Eds.) *Tourismus und Klimawandel*. Berlin Heidelberg: Springer. https://doi.org/10.1007/978-3-662-61522-5.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T. and Norberg, J. (2020) Patterns and trends of northern hemisphere snow mass from 1980 to 2018. *Nature*, 581(7808), 294–298. https://doi.org/10.1038/s41586-020-2258-0.

R Core Team. (2021) *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Scherrer, S.C., Wüthrich, C., Croci-Maspoli, M., Weingartner, R. and Appenzeller, C. (2013) Snow variability in the Swiss Alps 1864-2009. *International Journal of Climatology*, 33(15), 3162–3173. https://doi.org/10.1002/joc.3653.

Schöner, W., Koch, R., Matulla, C., Marty, C. and Tilg, A.-M. (2019) Spatiotemporal patterns of snow depth within the Swiss-Austrian Alps for the past half century (1961 to 2012) and linkages to climate change. *International Journal of Climatology*, 39(3), 1589–1603. https://doi.org/10.1002/joc.5902.

Sen, P.K. (1968) Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, 63(324), 1379–1389. https://doi.org/10.2307/2285891.

Sevruk, B. (1985) Correction of precipitation measurements: Swiss experience. *Zürcher Geogr. Schriften*, 23, 187–196.

Statistik_Austria. (2019). Tourismus in Österreich 2018: Ergebnisse der Beherbergungsstatistik. https://www.statistik.at/web_de/services/publikationen/13/index.html?includePage=detailedView§ion-Name=Tourismus&pubId=569

Steinbauer, M.J., Grytnes, J.-A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., Rixen, C., Winkler, M., Barni, M.B.-D.E., Bjorkman, A.D., Breiner, F.T., Burg, S., Czortek, P., Dawes, M. A., Delimat, A., Dullinger, S., Erschbamer, B., Felde, V.A., Fernández-Arberas, O., et al. (2018) Mountain summits is linked to warming. *Nature*, 556, 231–234.

Szentimrey, T. (1999). Multiple Analysis of Series for Homogenization (MASH). Proceedings of the second seminar for homogenization of surface climatological data, Budapest, Hungary. WMO, 41, 27–46.

Toreti, A., Kuglitsch, F.G., Xoplaki, E., Della-Marta, P.M., Aguilar, E., Prohom, M. and Luterbacher, J. (2011) A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series. *International Journal of Climatology*, 31(4), 630–632. https://doi.org/10.1002/joc.2088.

Vaughan, D., Comiso, J., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., Ren, J., Rignot, E., Solomina, O., Konrad, S. and Zhang, T. (2013) Observations: Cryosphere. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, UK: Cambridge University Press. https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter04_FINAL.pdf.

Venema, V., Trewin, B., Wang X.L., Tamás, S., Lakatos, M., Aguilar, E., Auer, I., Guijarro, J. A., Menne, M., Oria, C., Louamba, W. S. R. L., Argiriou, A., Hechler, P., Vertačnik, G., & Yosef, Y. (2020). Guidelines on Homogenization 2020 edition (Issue WMO-No. 1245). https://library.wmo.int/?lvl=notice_display&id=21756

Venema, V.K.C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., et al. (2012) Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1), 89–115. https://doi.org/10.5194/cp-8-89-2012.

Vincent, L.A., Zhang, X., Bonsal, B.R. and Hogg, W.D. (2002) Homogenization of daily temperatures over Canada. *Journal of Climate*, 15(11), 1322–1334. https://doi.org/10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2.

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., et al. (2019) Welcome to the Tidyverse. *The Journal of Open Source Software*, 4(43), 1–6. https://doi.org/10.21105/joss.01686.

SUPPORTING INFORMATION

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

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

Appendix B

Unique set of parallel Swiss snow series - a draft

Status: Initial draft, work in progress, not yet complete or ready for submission
Author contributions: This current draft was written by MBU without input from either co-author.

This is a draft manuscript which describes the unique parallel snow data set used for analyses in the articles presented in Chapters 2 and 3 as well as in Chapter A. Although the parallel data set is briefly introduced in Chapter 2, the immense effort needed to obtain the final data set went unaccounted for. As the data set still offers potential use for various applications, the motivation to describe and publish it in full detail emerged.

Unique data set of parallel, manual, quality-controlled snow measurement time series for Switzerland

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

[1]WSL-Institute for Snow and Avalanche Research SLF, Davos, Switzerland
[2]Oeschger Centre for Climate Change Research, Bern, Switzerland
[3]Institute of Geography, University of Bern, Bern, Switzerland
[4]Federal Office of Meteorology and Climatology (MeteoSwiss), Zurich Airport, Switzerland

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

Abstract. TEXT

1 Introduction

Parallel measurements are important (Venema et al., 2020) as they improve the quality of the variable under investigation. They are most commonly compiled from manual and automatic measurements for instrument verification purposes. Further applications include, but are not limited to, validation of new instruments, sensors, or measurement procedures, gap filling, break point detection and homogeneity assessments, and benchmark tests.

Many parallel data sets exist for temperature and precipitation (e.g. Gubler et al., 2017; Hunziker et al., 2017; Vincent et al., 2018; Acquaotta et al., 2016, 2019). For snow, only one parallel data set exists so far (Acquaotta et al., 2015). However, their data set is confined to one specific climatic region (Piemont), includes only six pairs with an average temporal extend of 13 years and consist of a combination of automatic and manual measurements. Here we present a unique set of 30 series of parallel, manual and quality checked snow measurements for Switzerland. The parallel series are between 24 and 77 years long and cover elevations from 490 to 1770 m a.s.l. Their spatial extend covers many different climatic regions. The set was purpose-built for the Hom4Snow project (Buchmann et al., 2020, 2021, 2022a, b; Resch et al., 2022).

It allows for the first time to investigate the variability of, and local influences on, snow measurements by using parallel pairs of high-quality measurements within close proximity. Such a high-quality, long-term, parallel data set is invaluable for any climatological analysis. It provides in-depth knowledge about the Swiss snow climate in general and local variations in particular. Moreover, it enables further investigations and applications and various other use cases beyond the scope of the initial project; such as model verification, gap filling validation (Aschauer, 2021) or others. Furthermore, the work is not complete, there are possibly more stations still to be digitised, quality-checked and added. Elaborate why: NIME; snow observations only on paper, never used for anything

This article is structured as follows: Section 2 describes the data and the two sources, followed by an overview of the methods used in this study in Section 3. Section 4 covers the encountered issues. Results are shown and discussed in Section 5 and conclusions are drawn in Section 6.

2 Data description and sources

Snow depth (HS) and height of new snow (HN) are measured manually every day from at least November until April at roughly 200 locations in Switzerland. Snow depth is read off a stake and the height of new snow is measured on a board on top of yesterday's snow surface. These two instruments have not changed in principle since the early 1900 (see Haberkorn (2019); Buchmann et al. (2020) for details). Accompanying metadata, in the form of year of relocation or observer change, is generally available. However, the quality and availability differs from station to station and period in time. A detailed account is provided in section 5.3.

2.1 Manual snow measurements in Switzerland - a brief history

CM, MBT, MBU, SB

In Switzerland, the Federal Office for Meteorology and Climatology (MeteoSwiss) as well as the WSL-Institute for Snow and Avalanche Research (SLF) both maintain a large network of manually measured snow data. Additionally, private entities such as hydropower companies or smaller meteorological outlets sometimes provide snow data as well. We included the station Sihlsee (SSE) in our data set, manned by Etzelkraftwerke and provided by Meteodat. In the past, the Ration Railways (RhB) operated a large network of snow measurements in the canton of Grison along their railway lines and train stations. With the increased automation of the railways in the 1990, these snow time series are either transferred to the SLF or ceased to exist.

2.1.1 A brief history

MeteoSwiss Anetz, Enet, Observer network, (SwissMetNet)

SLF Enet, Observer network, (IMIS)

NiME text

Yeti text

RhB text

Power stations text

Ski resorts text

Road maintenance text

2.2 Unique set-up for parallel measurements

The different interests and focuses of two institutions both with a vast manual snow observation network, enables the construction of the parallel snow data set. MeteoSwiss, with its primarily role as a weather forecast provider, maintains a huge network of diverse meteorological stations distributed all over Switzerland, with snow being just one of the many observed variables. For the SLF however, snow measurements are the main focus. Although the SLF's area of expertise of rather diverse, its main focus and most popular product is the Swiss Avalanche Bulletin. For that, daily, high-quality snow measurements from the Alpine or mountainous area are key. The contrast of the two main objectives of the involved institutions manifests itself in the distribution of their snow observational networks. Whereas the MeteoSwiss-stations are distributed all over Switzerland, including the lower valleys and the Mittelland, the SLF-stations are solely located in mountainous terrain such as pre-Alps, Alps et cetera. However, there is an overlap with the MeteoSwiss network, which leads to the unique parallel snow data set.

3 Methods

3.1 Parallel set

For the purpose of this study, valid parallel station pairs are defined as follows:

- Only manual measurements are considered
- Distance between stations smaller or equal to 4 km horizontally and 150 m vertically
- At least snow depth (HS) present
- At least 20 years of consecutive data
- Data between November and April with no missing months
- Independent data

The evolution of our parallel data set is shown as a flowchart in Fig. 1. First, all possible pairs are constructed based on available measurements and geographical coordinates In a second step, the possible pairs are checked for data availability: Are the data digitally available or only on paper? After a quick plausibility and quality check, mainly focused on missing values and outliers), stations were either excluded or kept. To decide whether the effort of digitisation was worthwhile for a potential station, the paper-records were check on legibility and completeness (Nov-Apr and at least 20 years).

We encountered several complications and pitfalls during the process of building the parallel station pairs which are described in detail in section 4.

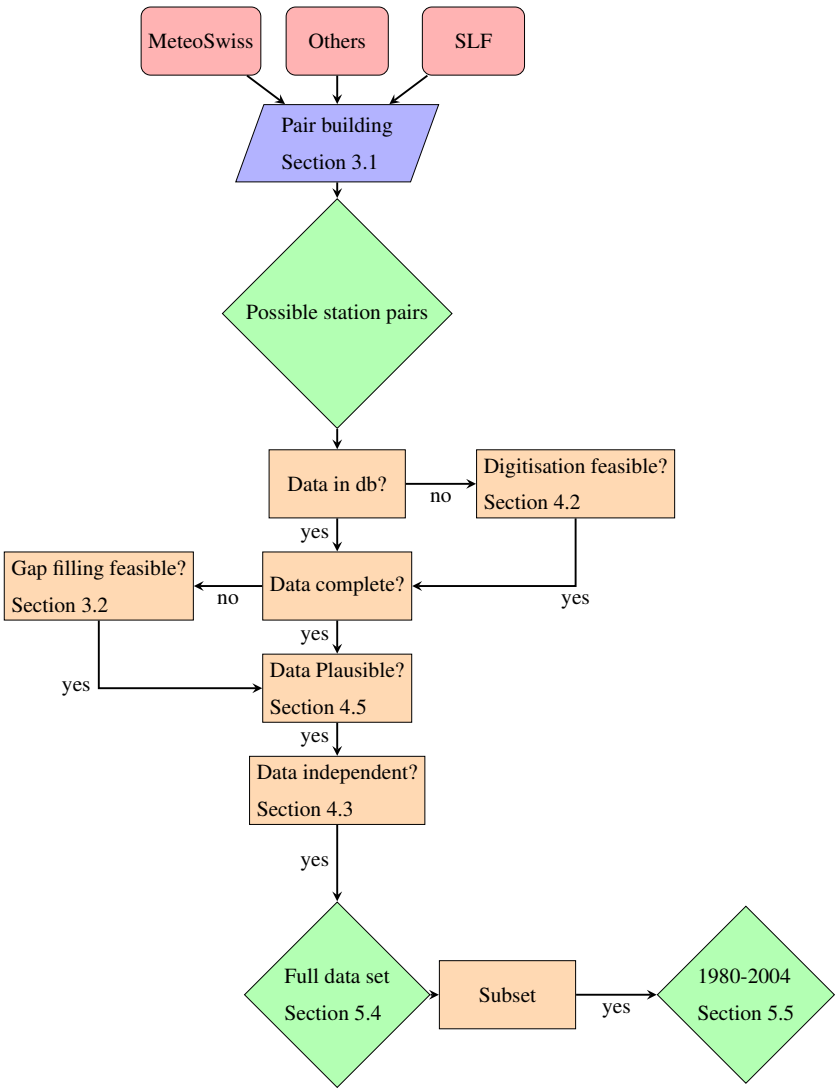


Figure 1. Flowchart illustrating the evolution of the parallel data set.

3.2 Gap filling

CM or JA

Gaps up to X days between November and April have been filled manually to increase the number of complete series and thus possible candidate series for our parallel data set. Individual missing days were filled using a set of nearby and highly correlated reference series. Occasionally, series were filled using the original paper records from the archives.

3.3 Snow climate indicators

Snow climate indicators (Table 1) derived from daily snow depth (HS) and height of new snow (HN) values, defined by Buchmann et al. (2020), and useful for climatological analyses, are incorporated in the contemporaneous subset (see Sect. 5.5).

Table 1. Definition of snow climate indicators.

Snow climate indicator	Definition	Abbreviation
Average snow depth		HSavg
Maximum snow depth		HSmax
Days with snow on the ground	HS $>$ 0 cm	dHS1
Days with snow on the ground	HS $\geq$ 5 cm	dHS5
Sum of new snow		HNsum
3-day consecutive sum of new snow		HN3max
Days with snowfall	HN $>$ 0 cm	dHN1
Days with snowfall	HN $\geq$ 5 cm	dHN5
Snow onset date		Dstart
Snow disappearance date		Dend

4 Issues

4.1 Unclear coordinates

Due to missing or incomplete metadata, the exact location of the snow measurements remained unclear for certain stations and it was therefore not straightforward to decide if two stations of a potential pair were close enough to be considered a genuine pair. Further location-related issues are how to treat relocations that exceeded the horizontal or vertical distance thresholds, the uncertainty associated with historic coordinates (pointing to the snow measurements or the official meteorological site?).

4.2 Digitisation

Include picture: Station Sihlsee (Fig. 2), mention digitisation process and people involved (SLF front desk) Change of header over time, possible impact on quality.. Only analogue station data above a certain quality and with less than X missing values were considered worthy of digitisation. X station years needed to be digitised and the work was carried out by various interns at SLF.

4.3 Independence

CM, JA, MBU

Independence is more than just using different coordinates or names. Especially, when the exact nature of the coordinates is unknown and the resolution is not more refined than 50 metres. The following examples highlight the possible issues associated with independence:

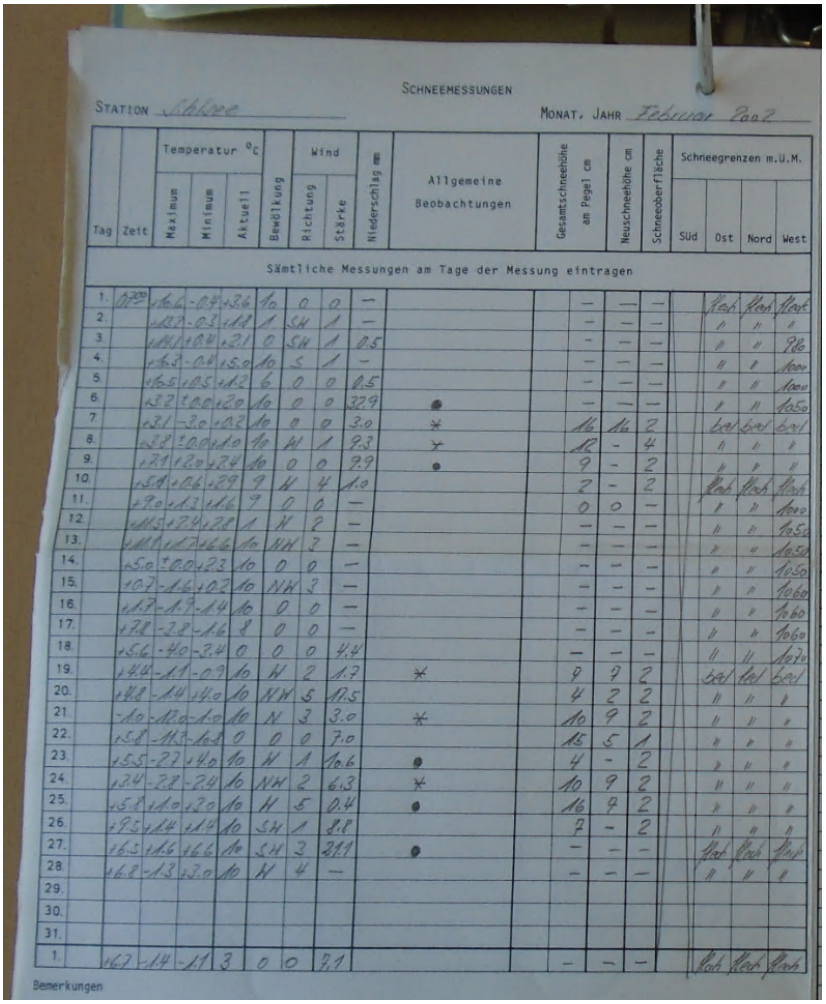
Abandoned stations, merged with "parallel" series: After the manual snow measurements stopped at DAV in 2006, data from 5DF are used by simply copying the values.

Same station, but different provider-based quality-control: WFJ/5WJ, i.e. not the same data, provide figure or plot. Here, due to different provider-based QC approaches, the processed data are not the same, even-though they originated from the same single source.

Different name, but the same data: DIS/5DI

BEH/7BP Pass location,

GRH/1GH Pass location,



SCHNEEMESSUNGEN

STATION Säkse — MONAT, JAHR Februar 2002

Tag	Zeit	Temperatur °C Maximum	Temperatur °C Minimum	Temperatur °C Aktuell	Bewölkung	Wind Richtung	Wind Stärke	Niederschlag mm	Allgemeine Beobachtungen	Gesamtschneehöhe am Pegel cm	Neuschneehöhe cm	Schneeoberfläche	Schneegrenzen m.ü.M. Süd	Ost	Nord	West
Sämtliche Messungen am Tage der Messung eintragen																
1.	07:30	+6.6	-0.4	+2.6	10	0	0	—		—	—	—	Flach	Flach	Flach	
2.		+0.2	-0.3	+0.8	1	SW	1	—		—	—	—	"	"	"	
3.		+0.4	+0.4	+2.1	0	SW	1	0.5		-	—	—	"	"	980	
4.		+6.3	-0.6	+5.0	10	S	1	—		—	—	—	"	"	1000	
5.		+6.5	+0.5	+1.2	6	0	0	0.5		-	—	—	"	"	1000	
6.		+3.7	+0.0	+2.0	10	0	0	32.9	●	—	—	—	"	"	1050	
7.		+3.1	-2.0	+0.3	10	0	0	3.0	*	16	16	2	bal	bal	bal	
8.		+2.8	+0.0	+1.0	10	W	1	9.3	*	18	-	4	"	"	"	
9.		+2.1	+2.0	+2.4	10	0	0	9.9	●	9	-	2	"	"	"	
10.		+5.4	+0.6	+2.9	9	W	4	1.0		2	-	2	Flach	Flach	Flach	
11.		+7.0	+1.3	+1.6	9	0	0	—		0	0	—	"	"	1000	
12.		+11.5	+2.4	+2.8	1	W	2	—		—	—	—	"	"	1050	
13.		+11.1	+1.2	+6.6	10	NW	2	—		—	—	—	"	"	1050	
14.		+5.0	+0.0	+2.3	10	0	0	—		—	—	—	"	"	1050	
15.		+0.7	-1.6	+0.2	10	NW	3	—		—	—	—	"	"	1060	
16.		+1.7	-1.9	-1.4	10	0	0	—		—	—	—	"	"	1060	
17.		+2.8	-2.8	-1.6	8	0	0	—		—	—	—	"	"	1060	
18.		+5.6	-4.0	-2.4	0	0	0	4.4		—	—	—	"	"	1070	
19.		+4.4	-1.1	-0.9	10	W	2	1.7	*	9	9	2	bal	bal	bal	
20.		+1.8	-1.4	+4.0	10	NW	5	11.5		4	2	2	"	"	"	
21.		-1.0	-12.0	-1.0	10	N	3	3.0	*	10	9	2	"	"	"	
22.		+5.8	-11.7	-1.2	0	0	0	7.0		15	5	1	"	"	"	
23.		+5.5	-2.2	+4.0	10	W	1	10.6	●	4	-	2	"	"	"	
24.		+2.4	-2.8	-2.4	10	NW	2	6.3	*	10	9	2	"	"	"	
25.		+5.8	+1.0	+3.0	10	W	5	0.4	●	16	9	2	"	"	"	
26.		+7.5	+1.4	+1.4	10	SW	1	8.8		7	-	2	"	"	"	
27.		+6.5	+1.6	+6.6	10	SW	3	21.1	●	—	—	—	Flach	Flach	Flach	
28.		+6.8	-1.3	+3.0	10	W	4	—		—	—	—	"	"	"	
29.																
30.																
31.																
1.		+6.7	-1.4	-1.1	3	0	0	7.1		—	—	—	Flach	Flach	Flach	

Bemerkungen

Figure 2. Example of a station record (Sihlsee, SSE) that had to be digitised.

8

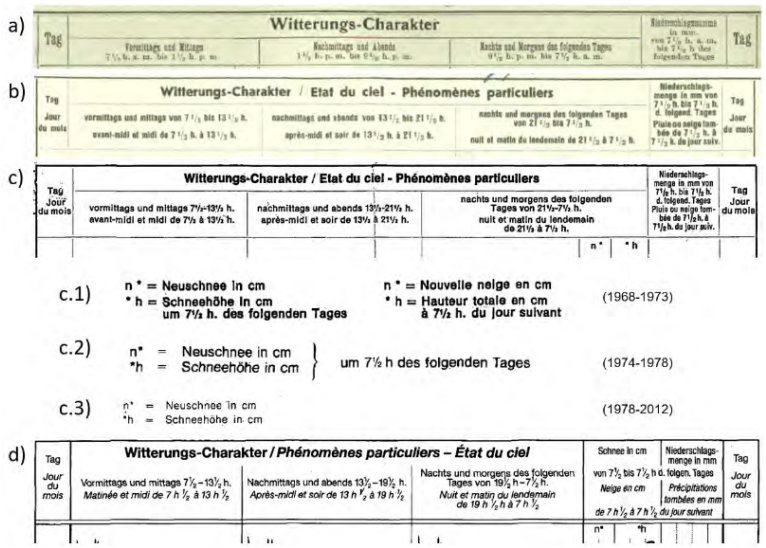


Figure 2: *Observer sheet headers and footnotes and their change over time for the MeteoSwiss Station Bernina. a) 1917-1958: no columns for snow in the header b) 1959-1967: no columns for snow in the header c) 1968-2001: dedicated columns for HN and HS in the header c1)- c3) different definitions of n* (HN) and *h (HS) in the footnote of the observer sheets e) 2002-2012: snow treated as precipitation in the header.*

Figure 3. Figure 2 from Aschauer and Marty (2020) highlighting the various layouts of the observer sheets.

4.4 Day shift

The ideas to attribute HN to different dates proved challenging. At the SLF, HN-measurements at 6:00 UTC are attributed to the current day, whereas MeteoSwiss, at least for some time, attributed the 6:00 UTC HN-measurements to the previous day. The definition on the MeteoSwiss-station-sheets changed over time (see 3) and it is likely that confusion arose at some point either during the input of the data on the sheet by the observer, the digitisation or the quality control step. The use of various QC codes over time and the lack of a consistent system makes it unfeasible to see, if these issues had been rectified or (re-)introduced during the digitisation or QC process. Therefore, the possible shifts can also be greater than one day and do not have to be in the same direction.

4.5 Plausibility

The applied plausibility checks include the search for erroneous values due to possible HN/HS mix-ups or digitisation errors such as missing decimal points. Another important test here is to check whether ΔHS between two days with snowfall always matches the amount of HN reported for that day. If true, this indicates, that either HS was constructed from the HN-measurements, or vice versa, neglecting the settling of the snow-pack altogether, and suggesting that this series needs to be treated with care or excluded from the subsequent analyses.

5 Results and Discussion

Combining the geographical requirements, obtained by metadata, with quality control mechanisms led to the quality-checked data set of 30 parallel snow series.

5.1 Quality control; plausibility and independence

JA?

Summarise the results from the QC-routines: how many stations affected? How many excluded? How many additionally digitised? etc

State of the observer network in terms of quality Mention QC instructions at MeteoSwiss

5.2 Gap filling

CM or JA

How many gaps were filled? How many series had gaps to be filled?

X gaps shorter than Y days in Z series were filled using expert knowledge and data from nearby stations.

5.3 Metadata

The metadata consists of annual values for observer changes and relocations (expressed as change of coordinates). There is a certain degree of uncertainty associated with these metadata, as for some time, the exact location of the snow measurements was not of great importance, and coordinates were only recorded as approximate values (50 m scale upwards) or taken from the corresponding meteorological site instead. Moreover, the term "observer change" normally refers to a change of observer at the same location, but can also imply a combination of observer change and relocation.

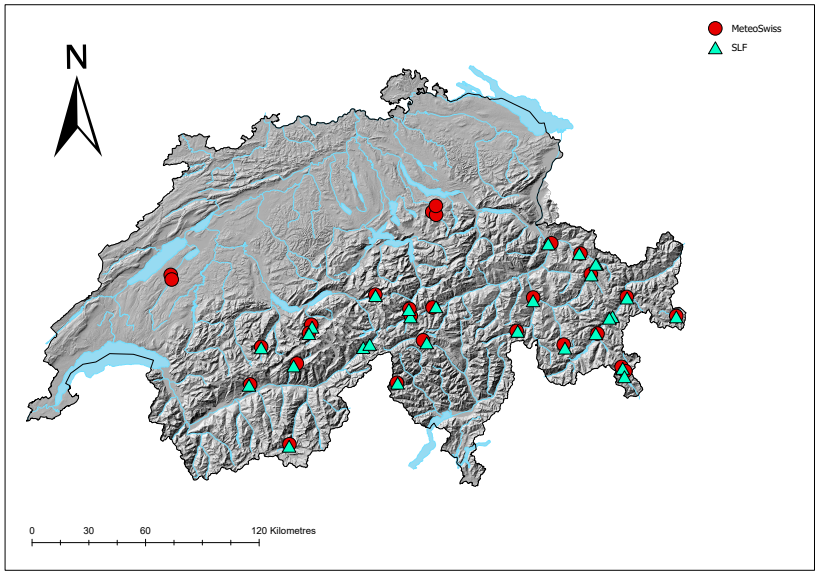


Figure 4. Map showing the spatial extent of the full parallel data set. It consists of 30 station pairs. MeteoSwiss stations are marked with a red circle, whereas SLF stations are highlighted with a tourmaline green triangle. Station pairs can either consist of MeteoSwiss/SLF, SLF/SLF, or MEteoSwiss/MeteoSwiss.

5.4 Complete parallel data set

The complete parallel set consists of 30 station pairs with various lengths between 1943 and 2020. See Table 2 for details.

5.5 Contemporaneous parallel data set

To be able to analyse various snow climate indicators within the same time window, a contemporaneous subset was constructed. An optimisation of number of station pairs and available period resulted in a subset of 23 station pairs and covering 25 years of parallel measurements from 1980 until 2004 (see Table 3). This subset was used by Buchmann et al. (2020) to analyse the robustness of snow climate indicators.

Table 2. Complete parallel data set.

Station pairs		From	To	m a.s.l.	Years	HN	mDstart	mDend	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 3. Contemporaneous parallel subset of 23 station pairs with complete data between 1980 and 2004.

Station pair	Village	m a.s.l.
ABO 1AD	Adelboden	1315
ANT 2AN	Andermatt	1440
BIV 5BI	Bivio	1790
BOS 6BG	Bosco/Gurin	1500
CAV 7CA	Cavaglia	1700
DAV 5DF	Davos	1570
EIN SSE	Einsiedeln	900
GAD 1GA	Gadmen	1190
GOS 2GO	Göschenen	1110
KLA 5KW	Klosters	1200
LAN 5LQ	Landquart	530
LTB 1LB	Lauterbrunnen	810
MUE 1MR	Mürren	1640
PAV PAY	Payerne	490
PIO 6AM	Piotta	1140
ROB 7PV	Poschiavo	1080
RIE 4WI	Lötschental	1500
SAM 7SD	Samedan	1730
SDO 5SP	Splügen	1460
SED 5SE	Sedrun	1400
SMM 7ST	Santa Maria	1390
THS 5TH	Thusis	670
ZER 4ZE	Zermatt	1640

6 Conclusions

We present a unique parallel and quality-controlled snow measurement data set for Switzerland. The spatial extend of this parallel data set is shown in Figure 5.

So far, this data set was used to: (1) evaluate the robustness of various snow climate indicators in Buchmann et al. (2020), (2) investigate and quantify the local-scale variability associated with commonly-used snow variables as well as extreme values and on-set and disappearance dates in Buchmann et al. (2021), (3) Aschauer (2021) used the parallel data set to evaluate multiple gap-filling methods for snow depth time series, (4) and Resch et al. (2022) made use of the parallel data set to test the newly-developed quantile matching homogenization method interpQM. However, apart from these already investigated topics, there are many more use cases still unexplored: e.g. applications focusing on snowfall or model verification studies.

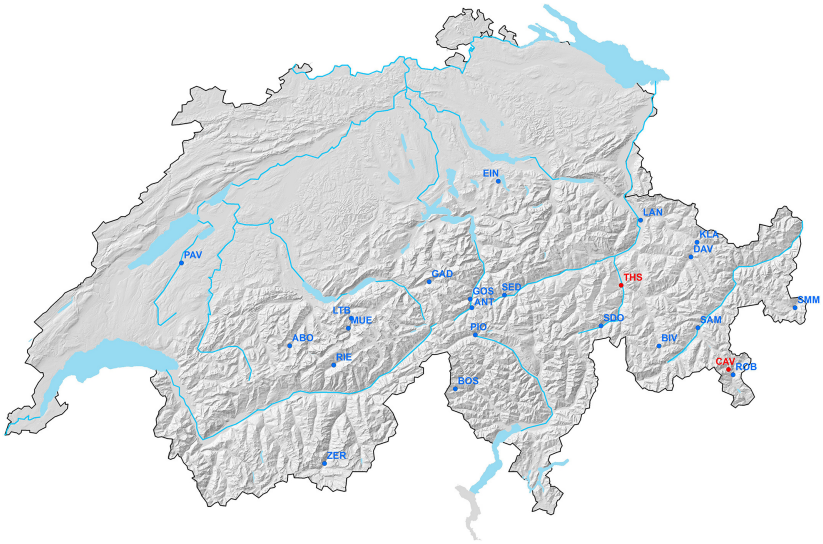


Figure 5. Map showing the spatial extent of the contemporaneous parallel data set. It consists of 23 station pairs. The two red pairs in the do not have snowfall values.

As the construction of said parallel data set has not proven to be straightforward, the detailed steps described here may help facilitate similar efforts in the future.

7 Data availability

Data are subject to copyrights but are available on request through EnviDat. doi ..

Appendix A

A1

Author contributions. MBU initial analyses, JA metadata analysis, CM gap filling, MBT metadata research. Manuscript was written by MBU.

Acknowledgements. MeteoDat for providing Sihlsee data, MeteoSwiss and SLF for the rest.

References

Acquaotta, F., Colombo, N., Fratianni, S., Romeo, V., and Barbero, S.: Preliminary Results of a Comparison Study Between Two Independent Snow Networks in North-Western Italian Alps (Piemonte Region), in: Engineering Geology for Society and Territory - Volume 1, edited by Lollino, G., Manconi, A., Clague, J., Shan, W., and Chiarle, M., pp. 113–116, Springer International Publishing, Cham, https://doi.org/https://doi.org/10.1007/978-3-319-09300-0_22, 2015.

Acquaotta, F., Fratianni, S., and Venema, V.: Assessment of parallel precipitation measurements networks in Piedmont, Italy, Int. J. Climatol., 36, 3963–3974, https://doi.org/https://doi.org/10.1002/joc.4606, 2016.

Acquaotta, F., Fratianni, S., Aguilar, E., and Fortin, G.: Influence of instrumentation on long temperature time series, Climatic Change, 156, 385–404, https://doi.org/https://doi.org/1007/s10584-019-02545-z, 2019.

Aschauer, J.: source code: Evaluating methods for reconstructing large gaps in historic snow depth time series,, https://doi.org/https://doi.org/5281/zenodo.5547996, 2021.

Aschauer, J. and Marty, C.: Providing Data Provision for a Sensitivity Analysis of Snow Time Series, resreport, https://www.meteoschweiz.admin.ch/content/dam/meteoswiss/en/Forschung-und-Zusammenarbeit/Internationale-Zusammenarbeit/GCOS/doc/Final_report_Poviding_Data_Provision_for_a_Sensitivity_Analysis_of_Snow_Time_Series.pdf, research Report for GCOS Switzerland, 2020.

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/https://doi.org/10.1002/joc.6863, 2020.

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/https://doi.org/10.5194/tc-15-4625-2021, 2021.

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/https://doi.org/10.5194/tc-16-2147-2022, 2022a.

Buchmann, M., Resch, G., Begert, M., Brönnimann, S., Chimani, B., Schöner, W., and Marty, C.: Impact of homogenisation on trends, In preparation, 2022b.

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/https://doi.org/10.1002/joc.5114, 2017.

Haberkorn, A.: European Snow Booklet – an Inventory of Snow Measurements in Europe, https://doi.org/https://doi.org/16904/envidat.59, 2019.

Hunziker, S., Gubler, S., Calle, J., Moreno, I., Andrade, M., Velarde, F., Ticona, L., Carrasco, G., Castellón, Y., Oria, C., Croci-Maspoli, M., Konzelmann, T., Rohrer, M., and Brönnimann, S.: Identifying, attributing, and overcoming common data quality issues of manned station observations: IDENTIFYING, ATTRIBUTING, AND OVERCOMING COMMON DATA QUALITY ISSUES, Int. J. Climatol., 37, 4131–4145, https://doi.org/https://doi.org/10.1002/joc.5037, 2017.

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/https://doi.org/10.1002/joc.7742, 2022.

Venema, V., Trewin, B., and Wang, X.: Guidelines on Homogenization 2020 edition, World Meteorological Organization Issue WMO-No. 1245, https://library.wmo.int/?lvl=notice_display&id=21756, last access: 9 June 2022, 2020.

Vincent, L. A., Milewska, E. J., Wang, X. L., and Hartwell, M. M.: Uncertainty in homogenized daily temperatures and derived indices of extremes illustrated using parallel observations in Canada, International Journal of Climatology, 38, 692–707, https://doi.org/https://doi.org/1002/joc.5203, _eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.5203, 2018.